NEURAL MECHANISMS OF VISUAL AWARENESS AND THEIR MODULATION BY SOCIAL THREAT

Spas Vladimirov Getov

Wellcome Trust Centre for Neuroimaging
Institute of Neurology
University College London

Prepared under the supervision of
Professor Geraint Rees and Professor Vincent Walsh

Dissertation submitted for the degree of
DOCTOR OF PHILOSOPHY, UNIVERSITY COLLEGE LONDON

February 2020
Declaration

I, Spas Vladimirov Getov, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
Contributions and publications

The structure-from-motion fMRI data used in Chapter 3 were collected by Megumi Fukuda (see Section 3.2.1). Some of the resting-state fMRI data used in Chapter 4 were collected by Maren Urner (see Section 4.2.1). All other data used in these chapters and the experimental design as well as all associated analysis and interpretation were my own. Some of the scripts for analysis of EMG and heart rate used in Chapters 7 and 8 were written with significant input from Joel Winston. However, I performed the experimental design, data collection, analysis and interpretation.

The following research tools were provided by colleagues from other laboratories:

- Face stimuli for binocular rivalry and continuous flash suppression experiments were obtained from Alex Todorov’s lab ([http://www.princeton.edu/~atodorov/](http://www.princeton.edu/~atodorov/)).
- Face stimuli used for the functional face localiser in Chapter 7 were obtained from Michael Tarr’s lab ([https://www.cmu.edu/dietrich/psychology/tarrlab/](https://www.cmu.edu/dietrich/psychology/tarrlab/)).
- Masks delineating sub-regions of the superior parietal lobule, as used in Chapters 3, 4 and 6 were kindly provided by Ryota Kanai and Rogier Mars.

The work presented in Chapter 6 has been published in the following peer-reviewed papers:


The work presented in Chapter 7 has been submitted for publication in NeuroImage.

The work presented in Chapter 8 has been submitted for publication in the Journal of Neuroscience.
Acknowledgements

The working environment at the Wellcome Trust Centre for Human Neuroimaging (FIL), where I was based for the duration of my three-year Clinical Research Training Fellowship, was outstandingly stimulating and supportive. I shared an office (which we affectionately named “the aquarium”) with a diverse group of PhD students and post-doctoral researchers. In particular I am grateful to Peter Smittenaar, James Lu and Eleanor Loh for keeping me sane, dragging me out to Gourmet Burger Kitchen often enough to avoid developing iron deficiency, and for being crazy enough to convince me to attempt and even complete the Three Peaks challenge with zero preparation (Eleanor had the good sense not to join us on that one). It was a great privilege and luxury to be at the birthplace of SPM and its creators were literally across the corridor. Guillaume Flandin, amongst others, was always patient and generous with his time on the many occasions when I got stuck. Finally, the excellent atmosphere of the FIL is largely determined by the outstanding core staff who were instrumental in countless ways from managing my grant, to keeping my MATLAB software up to date, to finding a way to connect the neuronavigation console to the stimulus PC, to helping me run the MRI scanner or eye tracker, to organising lively and exciting social events: in particular, but not exclusively, I am grateful to – Peter Aston, David Bradbury, Marcia Bennett, Eric Featherstone, Alphonso Reid, Jan Glensman, Elaine Williams, Ric Davis, Chris Fremantle, and Tom Simpson.

I would like to thank all the members of the inimitable Awareness Group, all of whom helped me both directly in my work and more generally in developing as a researcher and in enjoying my time in Queen Square. Particularly, I am grateful to Ryota Kanai for teaching me about VBM and FreeSurfer and for several exciting and frenetic brainstorming sessions; to Sarah Gregory, Adeel Razi and Maren Urner for teaching me just enough to dabble in some stochastic DCM; and to Jin Ekanayake for, well… for being Jin.

Bahador Bahrami, from my first day in the lab, was like a big brother, providing me with a huge amount of help and guidance, as well as encouragement and confidence in my abilities. His contributions ranged from the large-scale (advice on careers or overall theoretical thrust of my work) to the countless necessary details (resisting defeat by MATLAB or forcing the MRI scanner to work) and I really cannot thank him enough.

Joel Winston was a good friend and brilliant teacher with outstanding technical knowledge and a great eye for detail, spending many hours guiding me in analysis of fMRI and psychophysiological data. It was also during the days whileed away running experiments in
the FIL basement that we somehow convinced each other to become clinical neurophysiologists, a decision I am very glad to have made.

I had many exciting discussions with Megumi Fukuda and we shared the tribulations and pitfalls of running behavioural and scanning experiments to pursue our closely aligned interests. I have a sneaking suspicion that she taught me more than I taught her!

Vincent Walsh, my second supervisor, was readily on hand for the very infrequent times I asked to meet. These were memorable and highly instructive conversations to provide much-needed perspective on my work and development.

Geraint Rees has been a brilliant main supervisor, and is an outstanding individual, who leads by example. His support and confidence in me have been unwavering, and yet he has always given me ample room to develop independence and self-motivation as a researcher. Geraint’s diverse research interests, his open-mindedness, and his high profile, meant that the research lab was always an eclectic and exciting place, changing size and shape around me as I made my way through my fellowship. During our regular group interactions, serious and satisfyingly (often intimidatingly) nerdy science was balanced with healthy doses of gossip, obscure YouTube clips, and wacky lab outings (some of which were even disguised as conference trips).

I must of course thank the Wellcome Trust, a truly unique and inspiring organisation. I feel very privileged that they chose to fund this essential step in my professional development, and that they continue to fund the excellent centre where I was based.

Early in the (now really quite prolonged) period I have spent writing this thesis, I was unexpectedly compelled to gain a totally new perspective on the workings of the human brain: that of a patient. In 2015 I was diagnosed with a benign cerebellar tumour, which was causing significant mass effect and required prompt surgical excision. By this point I was working at King’s College Hospital as a clinical neurophysiology registrar. I was given initial assessment and advice, and was shown great compassion and professionalism, by the neurosurgical team there, most notably Mr Richard Selway and Mr Daniel Walsh. Jin Ekanayake, who had been a close friend, colleague and occasional collaborator at the Functional Imaging Lab, also came selflessly to my rescue at this time. He provided both friendly support and neurosurgical expertise, putting in sustained efforts to convey me rapidly towards treatment (as it turned out successful) by the outstanding Mr Neil Kitchen and his excellent team at the National Hospital for Neurology and Neurosurgery, literally across the road from where the work within this thesis was carried out. This was not an easy time but the outcome has thankfully been favourable and the circumstances brought...
me in contact with some truly exceptional individuals, and gave me a patient’s-eye-view of
the many brilliant qualities of our National Health Service. These people and experiences
have given me new inspiration and impetus to develop into the best clinician and academic
that I can, in the hope of making my own modest contribution to the benefit of other
patients.

Many friends and relatives supported and encouraged me during the preparation of this
thesis. In particular, my mother-in-law Aviva Petrie spent a number of hours with me
discussing, and providing expert advice on, statistical methods.

My parents and brother have of course shaped me into the person I am. I have always been
able to count on their love and support. Amongst a great deal else, they instilled in me the
importance of consideration for others and a respect for knowledge and scholarly
endeavour. These are values I continue to believe in very strongly, and they have been key
motivators in bringing this thesis to its eventual completion.

Last, and in many ways most of all, I owe a huge thanks to my wife Nina and my daughters
Sofia and Clara. In their different ways, they have all supported me through many moments
of seeming disaster and continued to believe in me when I was ready to give up. They have
put up with me shutting myself away or running off to a library on countless evenings and
weekends. I owe Nina hundreds of hours of childcare and quality time together; I owe Sofia
a million outings on her bike to buy ice cream; and I owe Clara many cuddles and quite a
few trips to go and see ‘the dinosaurs’. Despite these debts I have accrued they have never
stopped encouraging me and looking after me, and I dedicate this piece of work to the three
of them.
Abstract

The human brain can extract an enormous wealth of visual information from our surroundings. However, only a fraction of this immense data set ever becomes available to the observer’s awareness. How and why certain information is selected for awareness are questions under active investigation. Following two introductory chapters, this thesis contains six inter-related experimental chapters, through which I will explore two key outstanding questions in this field, using bistable perceptual paradigms to study conscious and non-conscious visual processing in healthy human volunteers.

The first major theme focuses on adding new insight into the brain regions and networks that facilitate transfer between non-conscious and conscious modes of visual processing. In Chapters 3 and 4 I will use fMRI, both in task-related and resting-state conditions, to delineate areas, and their interactions (in terms of effective connectivity), which are relevant for transition between different conscious perceptual experiences. In Chapter 5 I will focus on one specific region in the proposed perceptual transition-related network (the frontal eye field) and explore its causal role in access to awareness using repetitive TMS.

The second key question explored in this thesis is how social cues in the visual environment influence non-conscious visual processing as well as transition to conscious vision. In Chapter 6 I will study behavioural effects of non-conscious social cues from faces, and the relationship of these effects to focal brain anatomy. Based on finding slower perceptuomotor performance when non-conscious faces contain threatening cues in Chapter 6, I hypothesise that a defensive freezing response is engaged in such situations. The final two experimental chapters will explore the correlates of putative human freezing in the context of non-conscious social threat: using fMRI and psychophysiological measurements to study effects on perceptual transition in Chapter 7, and relating TMS-induced motor-evoked potentials and concurrent psychophysiological measurements to non-conscious perceptuomotor performance in Chapter 8.

Taken together, the presented findings shed new light on the network of brain regions involved in transition between non-conscious and conscious modes of visual processing. In addition, they uncover novel mechanisms through which socially relevant visual cues shape our awareness of the visual world, with particular emphasis on the engagement of defensive responses by socially threatening stimuli. The concluding chapter discusses the implications of these findings and explores relevant avenues for future work.
Impact statement

An understanding of the mechanisms that underpin human consciousness remains one of the greatest challenges in science. The phenomena of consciousness and the associated scientific study are extremely broad in scope but this thesis focuses on addressing two key questions specifically in the field of visual awareness. An observer only ever becomes conscious of a small proportion of visual information processed in the brain and the nature of this selection process is an area of active investigation. In this thesis I review relevant theoretical background and experimental findings as well as relevant research methodology in two introductory chapters, and then explore two key outstanding questions in this field in six inter-related experimental chapters describing experiments performed by recruiting healthy human volunteers. I seek to further understanding of neurobiological mechanisms of visual awareness and threat perception using behavioural testing, structural and functional MRI, and transcranial magnetic stimulation methods.

The first key objective of this thesis is to add new insight into the brain regions and networks that facilitate transfer between non-conscious and conscious modes of visual processing in the human brain. Influential theories and conceptual frameworks posit that evaluating the differences between these types of brain processing is a key step towards understanding the brain mechanisms that underpin conscious experience. The experimental work presented in this part of the thesis, and the associated discussions, provide new knowledge and perspective to bolster our understanding of the Neural Correlates of Consciousness, and to advance this important branch of the scientific study of consciousness. Moreover, refining knowledge of the mechanisms for switching between non-conscious and conscious processing modes in the brain holds promise for new approaches to understanding and ultimately treating disorders of consciousness in clinical populations. Epilepsy, where sudden and unexpected changes in consciousness represent a significant burden of disease, is just one important example. Finally, an improved understanding of human consciousness would have wide-reaching societal impacts far beyond the fields of experimental and clinical neuroscience, ultimately allowing us to understand how our experiences are shaped by the workings of our brains.

The second main theme is to explore how social visual cues (specifically, facial traits such as trustworthiness) influence non-conscious and conscious visual processing, as well as the balance between these processing modes. There is particular focus on the role of threatening social cues in promoting defensive behaviours and responses. Social signals play an extremely important part in human relations, and there are even established theories arguing that a key function of human consciousness is to support complex social
interaction. While defensive responses (for example, freezing or flight) are important for survival in the animal kingdom, their importance in humans is likely more in navigating complex social landscapes (for example, in the workplace or school playground) as well as forming part of the mechanisms by which psychological disturbances such as panic disorders can develop. This strand of the presented work thus has broad implications both for the social sciences and for application in psychological medicine.
## Contents

1. **General introduction** ......................................................... 27  
   1.1 Overview ............................................................................. 27  
   1.2 Cortical visual processing systems ..................................... 29  
      1.2.1 Functional organisation of visual processing ................. 29  
      1.2.2 Neural systems for human face processing ................. 32  
   1.3 Neuroscience of visual consciousness ................................. 34  
      1.3.1 Definitions and brief history of the scientific study of consciousness .......... 34  
      1.3.2 The neural correlates of consciousness ............................ 37  
      1.3.3 Experimental dissociation of sensory stimulation and conscious perception: visual bistability ........................................ 39  
      1.3.4 Non-conscious and conscious visual processing during bistable perception ... 43  
      1.3.5 Perceptual transitions and associated neural activity .......... 46  
   1.4 Social visual perception .................................................... 47  
      1.4.1 Social face perception, including evolutionary and historical perspectives ... 48  
      1.4.2 A 2-D model space for social face evaluation ................ 50  
      1.4.3 Neural mechanisms for social face processing ............... 52  
      1.4.4 Non-conscious processing of social visual cues ............... 53  
      1.4.5 Social modulation of vision and visual awareness ............ 54  
   1.5 Defensive responses to threat ............................................ 55  
      1.5.1 Conceptual models of defence responses ........................ 55  
      1.5.2 Freezing responses and their neurobiology .................. 57  
      1.5.3 Neural correlates of human freezing responses ............... 59  
   1.6 Individual variability in visual perception ............................ 60  
   1.7 Summary and approach to present thesis ............................. 63  

2. **General Methods** .................................................................. 67  
   2.1 Introduction ........................................................................... 67  
   2.2 MRI and fMRI ...................................................................... 68  
      2.2.1 Fundamental principles of MRI and fMRI ....................... 69  
         2.2.1.1 Behaviour of atomic nuclei within magnetic fields .......... 69  
         2.2.1.2 Generating a signal: magnetic resonance and relaxation .... 72  
         2.2.1.3 Spatial localisation within the MR signal .................. 72  
         2.2.1.4 MR image acquisition: contrast and echoes ................. 73  
         2.2.1.5 Biochemical and neural basis of the BOLD signal ........... 75
2.2.2 fMRI preprocessing .................................................................................. 78
  2.2.2.1 Bias correction .................................................................................... 79
  2.2.2.2 Field maps to correct for B field distortions ..................................... 79
  2.2.2.3 Realignment and unwarping .............................................................. 79
  2.2.2.4 Co-registration to a T1 structural image ............................................. 80
  2.2.2.5 Segmentation ..................................................................................... 80
  2.2.2.6 Spatial normalisation ........................................................................ 80
  2.2.2.7 Spatial smoothing ............................................................................. 81
2.2.3 fMRI analysis and statistics ..................................................................... 81
  2.2.3.1 The General Linear Model ................................................................. 82
  2.2.3.2 Statistical inference ......................................................................... 84
2.2.4 Resting-state fMRI analysis ..................................................................... 85
2.2.5 Dynamic causal modelling ....................................................................... 88
2.2.6 Voxel-based morphometry ....................................................................... 89
  2.2.6.1 Image segmentation ........................................................................ 90
  2.2.6.2 DARTEL registration ...................................................................... 90
  2.2.6.3 Smoothing ........................................................................................ 91
  2.2.6.4 Statistical analysis .......................................................................... 91
2.2.7 Surface-based MRI analyses using FreeSurfer ......................................... 92
  2.2.7.1 Constructing an accurate cortical surface model ......................... 93
  2.2.7.2 Cortical thickness estimation ........................................................... 95
  2.2.7.3 Brain segmentation using FreeSurfer .............................................. 95
2.3 Transcranial magnetic stimulation ................................................................ 97
  2.3.1 Electromagnetic induction in the brain ................................................. 97
  2.3.2 TMS hardware .................................................................................... 99
2.3.3 Stimulation protocols ............................................................................. 100
  2.3.4 TMS safety ........................................................................................ 102
2.4 Other general experimental methods .......................................................... 103
  2.4.1 Participant inclusion criteria and consent ............................................. 104
  2.4.2 Procedures for behavioural experiments .............................................. 104
  2.4.3 Face stimuli for BR and CFS experiments .......................................... 104
  2.4.4 Procedures for CFS and BR visual tasks .............................................. 105
    2.4.4.1 CFS experimental paradigm ....................................................... 106
    2.4.4.2 Binocular rivalry paradigm: static face and grating .................... 107
    2.4.4.3 Binocular rivalry paradigm: two rotating gratings ...................... 107
  2.4.5 Procedures for SFM visual task ............................................................ 108
  2.4.6 Procedures for visual tasks inside the MRI scanner ............................ 109
  2.4.7 MRI and fMRI acquisition protocols .................................................. 109
  2.4.8 fMRI preprocessing methods ............................................................... 110
  2.4.9 Eye tracking ....................................................................................... 110
3. **Stimulus-invariant brain activity underlying perceptual switches and relationship to resting-state networks** ........................................ 111

3.1 Introduction ................................................................................................................. 111

3.2 Materials and methods ................................................................................................. 115

3.2.1 Contribution acknowledgement and reuse of data ................................................. 115

3.2.2 Participants ................................................................................................................ 115

3.2.3 Display apparatus inside MRI scanner ...................................................................... 116

3.2.4 Binocular rivalry: stimuli and experimental procedures ......................................... 116

3.2.5 Structure-from-motion: stimuli and experimental procedures ................................. 118

3.2.6 fMRI data acquisition and pre-processing ............................................................... 118

3.2.7 fMRI data analysis .................................................................................................... 119

3.3 Results .......................................................................................................................... 122

3.3.1 Binocular rivalry experiment results ........................................................................ 122

3.3.2 Structure-from-motion experiment results ............................................................... 125

3.3.3 Stimulus-invariant transition-related BOLD signal patterns ................................... 125

3.3.4 Relationship of fMRI findings to SPL subregions and associated resting-state networks ................................................................. 126

3.3.5 Results from small-volume-corrected analyses centred at other predetermined regions of interest ................................................................................... 128

3.4 Discussion .................................................................................................................... 131

3.4.1 Neural correlates of endogenous perceptual transitions during BR and SFM ................. 131

3.4.2 Stimulus-invariant brain activity related to perceptual transitions ............................ 132

3.4.3 Possible roles of regions within the stimulus-invariant network in access to awareness ................................................................................................. 133

3.4.4 SPL subregions, associated resting state networks and perceptual transition .......... 135

3.4.5 Is frontoparietal BOLD signal cause or consequence of perceptual transitions? ........ 137

3.4.6 Summary .................................................................................................................. 138

4. **Effective resting-state connectivity and individual dynamics of bistable perception** ......................................................................................... 141

4.1 Introduction ................................................................................................................... 141

4.2 Materials and methods ................................................................................................. 143

4.2.1 Contribution acknowledgement ................................................................................ 143

4.2.2 Participants ................................................................................................................ 144

4.2.3 Behavioural experiment: stimuli, display apparatus and experimental procedures .... 144

4.2.3.1 Binocular rivalry experimental procedure ............................................................ 144

4.2.3.2 Structure-from-motion experimental procedure ................................................... 146

4.2.4 Resting-state fMRI: experimental procedures .......................................................... 146
5. Functional role of right frontal eye field in the dynamics of bistable perception .................................................. 165

5.1 Introduction ........................................................................ 165

5.2 Methods ........................................................................... 169

5.2.1 Participants .................................................................. 169

5.2.2 Experimental sequence .................................................. 169

5.2.3 SFM behavioural paradigm ............................................. 170

5.2.4 Eye tracking .................................................................. 171

5.2.5 MRI preprocessing and TMS target localisation ................. 171

5.2.6 TMS procedure ............................................................. 172

5.2.7 Ratings for participant experience of TMS ......................... 173

5.3 Results ............................................................................. 174

5.3.1 Individual variability in mean percept duration ................. 174

5.3.2 Group-level effects of TMS on percept duration ................. 174

5.3.3 Participant ratings of the experience of TMS ...................... 176

5.3.4 Effects of visual fixation, pupil size and saccades ................ 176
6. **Evaluation of social face traits outside of awareness** .......... 187

6.1 Introduction ................................................................................................. 187

6.2 Behavioural materials and methods ............................................................... 189
  6.2.1 Participants ............................................................................................ 189
  6.2.2 Stimuli and Display Apparatus ............................................................... 189
  6.2.3 Behavioural Procedures ......................................................................... 190

6.3 Structural MRI methodology ....................................................................... 191
  6.3.1 MRI Data acquisition ............................................................................ 191
  6.3.2 VBM analysis of MRI data .................................................................... 191
  6.3.3 Surface-based analysis of MRI data ....................................................... 192

6.4 Results ......................................................................................................... 193
  6.4.1 Behavioural experiment results ............................................................... 194
    6.4.1.1 Main effects of dominance and trustworthiness on t2e ..................... 194
    6.4.1.2 Individual variability of dominance and trustworthiness effects .......... 196
    6.4.1.3 Error rates in the behavioural task .................................................... 196
    6.4.1.4 Testing for outliers in the behavioural results ..................................... 197
  6.4.2 MRI experiment results .......................................................................... 199
    6.4.2.1 Correlations between GM volume and behavioural measures: whole-brain VBM analysis ........................................................................................................... 199
    6.4.2.2 Correlations between GM volume and behavioural measures: small-volume-corrected VBM analysis ........................................................................ 202
    6.4.2.3 Linearity/outliers in correlations between GM volume and behaviour .......... 203
    6.4.2.4 Relationship between amygdala volume and behavioural measures ........ 204
    6.4.2.5 Correlation between behaviour and surface-based measures of cortical thickness, surface area and volume .............................................................. 205
    6.4.2.6 Testing whether the brain structural correlates of dominance and untrustworthiness evaluation are dissociable ......................................................... 208
    6.4.2.7 Brain structural correlates of individual differences in t2e and experimental task error ................................................................................................................... 209

6.5 Discussion ..................................................................................................... 210
  6.5.1 Dominant and untrustworthy faces emerge into awareness more slowly than neutral faces ............................................................................................................. 211
  6.5.2 Structural correlates of dominance-related slowing .................................... 212
  6.5.3 Structural correlates of untrustworthiness-related slowing ......................... 213
7.  **Modulation of conscious perception in binocular rivalry by social face traits** ................................................................. 219

7.1 Chapter introduction ................................................................................. 219

7.2 Behavioural experiment .......................................................................... 222

7.2.1 Introduction .............................................................................................. 222

7.2.2 Materials and methods ......................................................................... 222

7.2.2.1 Participants .......................................................................................... 222

7.2.2.2 Stimuli and display apparatus .............................................................. 223

7.2.2.3 Binocular rivalry experimental procedures ......................................... 223

7.2.2.4 Eye tracking experimental procedures ............................................... 224

7.2.2.5 Self-rated personality questionnaires ............................................... 224

7.2.3 Results ..................................................................................................... 224

7.2.3.1 Binocular rivalry results ................................................................... 224

7.2.3.2 Eye tracking results .......................................................................... 227

7.2.3.3 Self-rated personality questionnaire results .................................. 230

7.2.4 Discussion ............................................................................................... 231

7.2.4.1 Visibility durations for individual percepts ..................................... 232

7.2.4.2 Visual fixation and pupil effects of face dominance ....................... 233

7.2.4.3 Lack of correlation between individual binocular rivalry dynamics or eye gaze fixation and personality traits ... 233

7.3 fMRI experiment ....................................................................................... 234

7.3.1 Introduction ........................................................................................... 234

7.3.2 Materials and methods ......................................................................... 234

7.3.2.1 Participants ........................................................................................ 235

7.3.2.2 Stimuli, display apparatus and behavioural procedures ............................. 235

7.3.2.3 Functional face localiser .................................................................... 236

7.3.2.4 fMRI data acquisition and pre-processing ....................................... 236

7.3.2.5 fMRI data analysis ......................................................................... 236

7.3.2.6 Psychophysiological analysis of heart rate data ................................ 238

7.3.3 Results .................................................................................................... 239

7.3.3.1 Behavioural results ........................................................................ 239

7.3.3.2 fMRI results .................................................................................... 241

7.3.3.3 Transition-evoked changes in heart rate ........................................... 250

7.3.4 Discussion ............................................................................................... 252

7.3.4.1 Behavioural and eye tracking measures inside the scanner ............. 253

7.3.4.2 Neural correlates of endogenous perceptual transitions ................ 253
Physiological evidence for freezing responses to non-conscious social threat .............................................. 259

8.1 Introduction ......................................................................................................................... 259

8.2 Materials and methods ....................................................................................................... 262
  8.2.1 Participants ....................................................................................................................... 262
  8.2.2 Experimental procedures ............................................................................................... 263
    8.2.2.1 Overview of experimental procedure .................................................................... 263
    8.2.2.2 Stimuli, display apparatus and procedure for CFS .................................................. 264
    8.2.2.3 EMG and ECG recording ......................................................................................... 264
    8.2.2.4 Single-pulse TMS .................................................................................................... 265
    8.2.2.5 Eye tracking ............................................................................................................. 265
  8.2.3 Data analysis .................................................................................................................... 265
    8.2.3.1 MEP analysis ........................................................................................................... 265
    8.2.3.2 ECG and pupil size analysis .................................................................................... 266

8.3 Results .............................................................................................................................. 267
  8.3.1 Main experiment results ............................................................................................... 267
    8.3.1.1 Effects of non-conscious face threat on behavioural and motor physiological measures ......................................................................................... 267
    8.3.1.2 Effects of non-conscious face threat on autonomic indices ................................... 269
  8.3.2 Auxiliary experiment results .......................................................................................... 272
  8.3.3 Effects of visual fixation and eye dominance ................................................................. 274

8.4 Discussion .......................................................................................................................... 275
  8.4.1 Slowing of t2e for non-conscious threatening faces ....................................................... 276
  8.4.2 Effect of non-conscious threat on CSE ......................................................................... 277
  8.4.3 TMS-MEP as a measure of CSE and motor freezing ....................................................... 278
  8.4.4 Non-conscious threat association with bradycardia ....................................................... 279
  8.4.5 Threat-invariant pupillary dilatation .............................................................................. 280
  8.4.6 Proposed neurobiological basis of the effects on CSE and autonomic indices ................. 280
  8.4.7 Summary ....................................................................................................................... 281
9. General discussion ................................................................. 283

9.1 Overview ........................................................................... 283

9.2 Neural mechanisms of perceptual transition ......................... 284

9.2.1 Discussion and context of findings from Chapters 3-5 .......... 284

9.2.2 Proposing extended functional networks contributing to perceptual transition ........................................................................ 286

9.2.3 Significance of transition-related frontoparietal BOLD signals ................................................................. 288

9.3 Social modulation of visual awareness ...................................... 292

9.3.1 Effects of non-conscious social face traits on behavioural responses ......................................................................................... 293

9.3.2 Structural and functional neural correlates of social visual cues: non-conscious processing and access to awareness ................................................................. 295

9.4 Defensive freezing in humans .................................................. 296

9.4.1 Placing findings in the context of evolutionary vigilance, dual competition and defence cascade frameworks ......................................................................................... 296

9.4.2 Proposed role of perceptual freezing ...................................... 299

9.4.3 Freezing responses elicited by non-conscious images ............ 301

9.4.4 Possible mechanisms for non-conscious freezing ................. 301

9.5 Future research directions ...................................................... 303

9.5.1 Future work on neural mechanisms of perceptual transition ......................................................................................... 303

9.5.2 Future work on the social modulation of visual awareness ......................................................................................... 305

9.5.3 Future work on mechanisms of human defensive freezing ......................................................................................... 307

9.5.4 Future clinical relevance of work presented in this thesis .... 309

9.6 Concluding remarks .............................................................. 310

Appendix 1: List of Abbreviations ................................................. 313

Appendix 2: Details of apparatus and software ............................... 317

Appendix 3: References ............................................................... 319
List of Figures

Figure 1.1: Simplified hierarchy of cortical visual areas in the macaque monkey .......... 31
Figure 1.2: Model of the distributed neural systems for face perception ......................... 33
Figure 1.3: The two major dimensions of consciousness .............................................. 35
Figure 1.4: Bistable visual stimuli ................................................................................ 42
Figure 1.5: Clusters of frontal and parietal BOLD activation associated with changes in the contents of visual awareness ............................................................. 47
Figure 1.6: Correlations between trait judgments and emotional ratings attributed to naturalistic emotionally neutral faces .............................................................. 49
Figure 1.7: Creation of a two-dimensional trustworthiness-by-dominance model space for evaluation of social face traits ..................................................................... 51
Figure 1.8: Defence response cascade model ................................................................ 57

Figure 2.1: Hydrogen atoms align either parallel or anti-parallel to a static magnetic field ....................................................................................................................... 70
Figure 2.2: The effects of a static magnetic field and an additional superimposed field gradient on protons and their spins ................................................................. 71
Figure 2.3: How average neural activity is translated into fMRI signal ............................ 77
Figure 2.4: Summary of SPM preprocessing steps for fMRI ......................................... 78
Figure 2.5: Overview of SPM preprocessing and analysis ............................................... 82
Figure 2.6: Accounting for frequency drift in fMRI time series ...................................... 83
Figure 2.7: Resting state networks displayed on the surface of inflated template brains .... 87
Figure 2.8: Ambiguities inherent in interpretation of VBM results .................................. 92
Figure 2.9: Contrasting cortical surface modelling with standard deformable models and FreeSurfer ........................................................................................................... 94
Figure 2.10: Intensity histograms from different tissue classes and brain structures ....... 96
Figure 2.11: Basic principles of TMS ............................................................................. 98
Figure 2.12: Topographical representations of electrical field strength induced below circular and figure-of-eight TMS coils .............................................................. 100
Figure 2.13: Social face model with face traits varying along orthogonal dimensions of dominance and trustworthiness ................................................................. 105
Figure 2.14: CFS visual task paradigm ........................................................................... 106
Figure 2.15: Binocular rivalry paradigm (static faces and gratings) ................................. 107
Figure 2.16: Binocular rivalry paradigm (two rotating gratings) ...................................... 108
Figure 2.17: Structure-from-motion visual paradigm ..................................................... 109

Figure 3.1: Stimuli and experimental paradigms employed in BR and SFM experiments ............................................................................................................................. 117
Figure 3.2: Five subregions of SPL, proposed on the basis of diffusion-weighted tractography results ................................................................................................. 121
Figure 3.3: Brain regions with increased BOLD signal for the contrast endogenous transitions > replayed transitions in the BR and SFM experiments .................................. 124
Figure 3.4: Overlap between SPL subregions and increased BOLD signal for the contrast endogenous transitions > replayed transitions in BR masked with SFM and in SFM masked with BR .............................................................. 127

Figure 3.5: Overlap between anterior SPL subregion, regions with associated resting state functional connectivity and stimulus-invariant perceptual-transition-related BOLD activity ............................................................ 128
Figure 3.6: Overlap between SPL subregions defined by Kanai et al. (2010, 2011) and the anterior SPL subregion from the parcellation of Mars et al. (2011) .................................................. 136

Figure 4.1: Stimuli used in BR and SFM behavioural experiments .................................. 145
Figure 4.2: Graphical representation of VOI signal extraction procedure using SPM8 .......................................................... 148
Figure 4.3: Behavioural experiment results ................................................................. 152
Figure 4.4: Percept duration histograms from BR and SFM behavioural experiments .............................................................................. 153
Figure 4.5: DCM analysis and results ............................................................................. 155

Figure 5.1: Experimental SFM stimulus ........................................................................ 170
Figure 5.2: Processing sequence for placing TMS targets onto individual structural MRI scans ...................................................................................................................... 171
Figure 5.3: Spherical markers used for neuronavigation with the ANT Visor 2 system ......................................................................................................................... 173
Figure 5.4: Inter- and intra-individual variability in mean percept durations ................ 174
Figure 5.5: Group-level behavioural results .................................................................... 175
Figure 5.6: Individual eye position heat maps ............................................................... 178
Figure 5.7: Group level eye tracking results .................................................................... 179

Figure 6.1: Face stimuli from the model of Oosterhof and Todorov (2008) .................. 189
Figure 6.2: Schematic representation of CFS behavioural paradigm .............................. 190
Figure 6.3: Frequency distribution of t2e values .......................................................... 194
Figure 6.4: Effects of social face traits on t2e ............................................................... 195
Figure 6.5: Frequency distributions of t2e response times ........................................... 197
Figure 6.6: Location of fast-t2e individuals on dominance-related slowing and untrustworthiness-related slowing distributions ................................................................. 198
Figure 6.7: Structural brain correlates of individual differences in non-conscious dominance-related slowing ..................................................................................... 199
Figure 6.8: Cytoarchitectonically-defined regions near the locus where GM volume is correlated with dominance-related slowing ................................................................. 200
Figure 6.9: Structural brain correlates of individual differences in non-conscious untrustworthiness-related slowing ................................................................................ 201
Figure 6.10: Spatial overlap between the right pTPJ result of the present study and the previously described TPJp subregion (Mars et al., 2011) .................................................. 202
Figure 6.11: Correlations between focal GM volume and dominance-related slowing or untrustworthiness-related slowing.......................................................... 204
Figure 6.12: Regions where cortical thickness or cortical surface area was correlated with individual dominance-related slowing ............................................ 206
Figure 6.13: Regions where cortical thickness was correlated with individual untrustworthiness-related slowing .......................................................... 206
Figure 6.14: Regions where cortical volume was correlated with individual untrustworthiness-related slowing ................................................. 207

Figure 7.1: Stimuli and paradigm for behavioural experiment ........................................ 223
Figure 7.2: Behavioural experiment results .................................................................. 225
Figure 7.3: Percept visibility duration histograms ......................................................... 227
Figure 7.4: Eye position and pupillary area data from eye tracking measurements ........... 229
Figure 7.5: Individual variability in behavioural performance for binocular rivalry ......... 231
Figure 7.6: Replay behavioural condition employed in the fMRI experiment ............... 236
Figure 7.7: fMRI experiment behavioural results .......................................................... 240
Figure 7.8: Correlations between individual binocular rivalry switch rate and individual effect of social dominance on switch rate ........................................ 241
Figure 7.9: Brain regions with increased BOLD signal for the contrast rivalry transitions > replay transitions ........................................ 244
Figure 7.10: Brain regions with increased BOLD signal for the contrast replay transitions > rivalry transitions .............................................................. 245
Figure 7.11: Brain regions with increased BOLD signal for the interaction term: dominant transitions>neutral transitions (replay) > dominant transitions>neutral transitions (rivalry) ............................................................. 248
Figure 7.12: Brain regions with increased BOLD signal for the interaction term: dominant transitions>neutral transitions (rivalry) > dominant transitions>neutral transitions (replay) ............................................................. 249
Figure 7.13: Perceptual transition-evoked heart rate .................................................... 251

Figure 8.1: Behavioural paradigm and stimuli ............................................................... 264
Figure 8.2: Group and individual effects of face threat on behavioural and physiological measures .................................................................................. 268
Figure 8.3: Group and individual effects of face threat on autonomic measures .......... 270
Figure 8.4: Group-level effects of face threat and TMS latency on behavioural, physiological and autonomic measures ................................................... 271
Figure 8.5: Time-varying effects of face threat on HR and pupil size in the period preceding t2e ......................................................................................... 272
Figure 8.6: Eye position heat map plots ....................................................................... 274

Figure 9.1: Theoretical frameworks for the neural mechanisms underlying perceptual transitions .................................................................................. 290
Figure 9.2: Speculative frameworks for the neural mechanisms (brain regions and their possible interactions) involved in defensive freezing responses to non-conscious socially threatening stimuli ................................................. 302
List of Tables

Table 3.1: Results of previous studies that motivate ROIs to be used for small volume correction ................................................................. 121
Table 3.2: Stimulus-specific perceptual-transition-related activation; whole brain analyses ................................................................. 123
Table 3.3: Brain regions activated during endogenous (versus replayed) perceptual transitions in a stimulus-invariant fashion; whole brain analyses ......................... 126
Table 3.4: Brain regions activated during endogenous (versus replayed) perceptual transitions in a stimulus-invariant fashion; small volume corrected results ........ 130

Table 4.1: Anatomical location of brain regions to be used in DCM analysis .......... 148
Table 4.2: Statistical results for correlations between individual mean BR durations or mean SFM durations and individual Bayesian parameter average values ............. 156

Table 6.1: Coordinates of loci used for small volume correction analysis .......... 192
Table 6.2: VBM analysis, whole brain statistics ......................................................... 200
Table 6.3: Coordinates and statistical results for small volume correction analyses .... 203

Table 7.1: Loci used for small volume correction analysis ........................................... 238
Table 7.2: Brain regions reflecting the main effect of transition type (rivalry or replay) .......................................................................................... 243
Table 7.3: Brain regions reflecting the main effect of social face trait (neutral or dominant) ........................................................................... 246
Table 7.4: Brain regions reflecting an interaction between the main effects of transition type (rivalry or replay) and social face trait (neutral or dominant) ................. 247
Chapter 1

General introduction

“Men ought to know that from the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear, and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant.”

Attributed to Hippocrates, fifth century BC

1.1 Overview

For each of us, consciousness represents a private and highly individual set of experiences of our internal and external world. Almost invariably, regardless of social or educational background, we find this phenomenon to be a fascinating enigma. After several centuries of studying consciousness scientifically, there is fairly broad consensus that it is generated in the brain; however, delineating the neurobiological underpinnings of conscious experience has proven very challenging (see Section 1.3). One useful approach to gain leverage on this problem has been to focus on a specific subset of information from the environment (such as a particular image) and contrast the brain’s processing of this information in situations when the observer is aware of it with processing in situations when they are not. While this approach can only encompass a very narrow subset of our conscious experience (such as the conscious awareness of a particular image), this very narrowness makes this a realistic and possible route to developing an understanding of consciousness through empirical experimental methods. Work in this area has most frequently been undertaken in the visual domain, where the broad and easily modifiable range of stimuli as well as the substantial
existing knowledge regarding the neurobiology and organisation of visual processing provide favourable conditions for experimentation and theoretical development (see Section 1.2). Experimental paradigms of bistable perception instantiate perceptually ambiguous conditions, enabling us to study spontaneous changes that occur against a background of unvarying sensory stimulation (see Section 1.4; Blake and Logothetis, 2002; Kim and Blake, 2005). These experimental approaches, their application to visual perception and awareness, and their relevance to understanding of the underlying neuronal mechanisms, form the foundations of this thesis.

We only ever become consciously aware of a small fraction of the vast and diverse amalgamation of visual information that our brains produce from the neural signals originating in our retinas. However, a subset of visual percepts does gain access to our conscious experience. How this information is selected and prioritised in the brain has been a central question in visual awareness research (Crick and Koch, 1990a). It appears that the contents of such visual conscious perception can be influenced by various factors associated with the source of visual information itself, with the context in which it is acquired, or with the dispositions of the observer (e.g. Gray et al., 2009; Kanai and Rees, 2011; Sterzer et al., 2009b; Yang et al., 2007). The neurobiological mechanisms of access to awareness remain incompletely understood and many of the factors that modulate these mechanisms also require further exploration. The work in this thesis will address some key outstanding questions in this field, focusing on two main themes through six chapters (Chapters 3-8) that detail inter-related experimental studies. In Chapters 3-5 I will explore the neural mechanisms that underpin transition from non-conscious to conscious visual processing modes in the human brain. Then, in Chapters 6-8, I will study the effects of social visual cues contained in faces (when perceived non-consciously) on behaviour, specifically investigating relationships with brain structure and with the engagement of defensive responses to threat. The particular questions addressed in each chapter, and the experimental methodologies used, are briefly outlined in Section 1.7.

The remainder of the present chapter attempts to broadly review several areas of existing knowledge that are of particular relevance to the experimental work presented in this thesis. These include our understanding of the brain’s systems for visual processing, the neural correlates of visual consciousness, the nature and utility of bistable visual paradigms, social perception and social modulation of vision and awareness, the neurobiology of defensive responses to threat, and individual variability in perception. Chapter 2 contains a summary of the general principles and relevant analytical methods pertaining to the main techniques used in the experimental chapters (namely, fMRI and structural MRI as well as TMS). Chapter 2 also contains a section describing general methodological approaches used in multiple experiments in this thesis, including participant recruitment, construction
of visual paradigms and fMRI preprocessing. The six experimental chapters (Chapter 3-8) contain the specific rationale and hypotheses, more specific methods such as recording of motor-evoked potentials or psychophysiological measurement of heart rate and pupillary changes, results, and finally interpretation and discussion relating to the experiments described therein. Following these experimental chapters, I will attempt a synthesis of all the findings presented in the thesis, discuss how they relate to, and extend, existing knowledge, and outline some key outstanding questions and directions for future work in the final chapter (Chapter 9).

1.2 Cortical visual processing systems

In humans (as in primates), vision is the most highly developed sensory modality; more than 20% of the human brain is at least partly devoted to visual processing (Kandel et al., 2000; Price, 2013). Our visual systems extract huge amounts of information from the environment. This visual information, encoded in neural signals, is subjected to an enormously varied, complex and distributed range of processing operations to facilitate its use in many distinct yet interrelated ways, in order to shape perception and ultimately guide decisions and behaviour. While visual processing occurs through a combination of cortical and subcortical mechanisms, here I will provide a brief review specifically of cortical visual processing systems since these are of particular relevance to the work in this thesis.

1.2.1 Functional organisation of visual processing

The extensive network of cortical regions that contribute to visual processing is very well characterised anatomically up to quite advanced hierarchical levels, especially in primates (Figure 1.1). The functional organisation of the human visual cortex has also become increasingly well understood, with especially large contributions made by work using fMRI in the last couple of decades (e.g. Grill-Spector and Malach, 2004). In both primates and humans, light is converted to spatiotemporal patterns of neuronal activity by the retina and thus transmitted to the thalamus, primary visual cortex and beyond. In primary visual cortex (V1), neurons are organised retinotopically: adjacent cortical regions receive information encoding adjacent portions of the visual field (Engel et al., 1997), although cortical representation of the fovea is greatly expanded. Moreover, at a smaller spatial scale, there is an intricate organisation such that neurons with similar orientation selectivity are arranged together in columns, information from the two eyes is segregated,
and colour information is processed in separate ‘blobs’ of tissue (Hubel and Wiesel, 1962; see also Kandel et al., 2000). A large number of functionally specialised processing areas are found beyond (or at higher hierarchical levels than) such early (or striate) visual cortex. A highly influential model, based on work in primates, organises these extrastriate cortical areas into two parallel processing streams (Mishkin et al., 1983; Figure 1.1). A dorsal stream, passing through area MT and travelling into parietal regions, was proposed to act as a ‘where’ pathway, concerned with visuospatial analyses. Conversely, a ventral stream, originating in V4 and travelling into inferior temporal cortex was conceptualised as a ‘what’ pathway involved in object recognition. This framework remains largely valid, although subsequent accounts have focused on the objectives achieved by the two streams (rather than the type of information they process) and have redefined them as ‘processing for action’ and ‘processing for perception’ streams (Goodale and Milner, 1992). Apart from being fundamental to understanding the organisation of visual processing, these models are also of particular relevance to a core theme of this thesis since they underpin the proposal that only the ventral stream supports awareness. This is backed up by empirical evidence (Fang and He, 2005; Milner, 2012), although other work suggests that neural activity in both processing streams is tightly coupled to awareness (Hesselmann and Malach, 2011).

The described theoretical models suggest that the dorsal and ventral processing streams are functionally segregated. However, topological analyses of primate visual system organisation suggest that they eventually converge at the DLPFC and superior temporal polysensory areas (Young, 1992).
A ubiquitous characteristic of visual processing, which is fulfilled in the retina, and is conserved all the way along the system, into the dorsal and ventral extrastriate streams, is parallel processing of different stimulus components (such as colour, orientation and motion) so that there is both functional and anatomical segregation. This segregation occurs at multiple spatial scales and at multiple hierarchical levels. Given this principle, it is remarkable that our visual experience is so unified: when we are thrown a red apple we are oblivious of the largely separate computations of colour, form and motion that ensue. A second fundamental principle, which is also highly relevant, concerns the rich interconnectivity between different functionally specialised areas and different hierarchical levels, including through feed-forward and feedback connections. This means that despite
the functional segregation and hierarchical organisation described, the system is in fact highly integrated. For example, it turns out that retinotopic organisation occurs not only in early cortical areas that are hierarchically close to the retina but also at much higher levels, for example in FFA (Grill-Spector and Malach, 2004). Another striking example of integration is that eye-specific neural signals in LGN are modulated in accordance with changes in conscious visual experience (Haynes et al., 2005). Such findings suggest that regions found very early (LGN) or much later (FFA) in the visual processing hierarchy have access to computations derived from processes that occur at hierarchically distant sites; this is likely made possible by extensive feed-forward and feedback projections in the system. In general, the principles of segregation and interconnectivity in the visual system permeate much of the work contained in this thesis (as well as the wider associated literature).

1.2.2 Neural systems for human face processing

Faces are probably the most complex stimuli dealt with by the human visual system. They contain multiple types of information and besides their structural and configural intricacy they are also dynamic and highly socially informative (see Section 1.4). It is therefore no surprise that studies of face perception and associated neural processing have received a great deal of attention in visual and social neuroscience. For a considerable length of time, the most influential cognitive model of face perception has been that of Bruce and Young (1986), which emphasised distinct processing systems to deal with different kinds of information relevant to faces. A key distinction was made between processes supporting identity recognition and those supporting recognition of facial expressions and speech. It is noteworthy that these face perception systems bear some analogy to the ventral and dorsal streams of visual processing and conform to the described general principle of functional segregation in the visual system.

The advent of fMRI (see Section 2.2) allowed brain areas involved in human face processing to be determined with considerable anatomical specificity, leading to landmark descriptions of regions where neural signal was triggered specifically by face images. These include the fusiform gyrus (subsequently named FFA; Kanwisher et al. 1997) as well as STS (Puce et al. 1998) and inferior occipital cortex (Puce et al., 1996), amongst others. Such findings have proven very robust to replication and, alongside physiological recordings in non-human primates, have informed the development of a more neurobiologically based framework for human face processing (Haxby et al., 2000). Bruce and Young’s model has thus been extended into a proposed core system for face perception made up of OFA (facilitating visual analysis of facial features), FFA (facilitating face recognition, or analysis
of invariant face information), and STS (facilitating processing of the dynamic aspects of faces such as expression and eye gaze). The model of Haxby and colleagues (2000) also describes interactions between this core system and an extended system that includes regions involved primarily in non-visual processing, but which make important contributions to face perception, such as the amygdala (emotion evaluation) and anterior temporal cortex (retrieval of biographical information; Figure 1.2).

More recent studies of face perception have for the most part adopted the framework of Haxby and colleagues (2000). However, despite the conceptual utility of the segregated modules proposed for the core system, it is increasingly apparent that there is considerable crossover in the roles of regions subserving dynamic and invariant face processing and that they are far from completely dissociable. For example, there is evidence for the involvement of STS in representing facial identity (Winston et al. 2004) as well as involvement of the FFA in representing facial expression (Fox et al. 2009). The remit of STS, in particular, is increasingly understood to be much broader than the originally proposed role in representing the dynamic aspects of faces. STS is also important for the processing of multimodal information regarding social relevance of faces, especially in terms of inferring the goals and intentions of others (whether this is conveyed by dynamic or static signals;
SV Getov, Chapter 1. General introduction

Allison et al., 2000; Gallagher and Frith, 2003; Hein and Knight, 2008). A detailed critique of the traditionally proposed separable face processing modules has been undertaken, along with proposals of alternative models (Calder and Young 2005).

Our detailed understanding of the anatomical and functional organisation of the cortical visual system and face processing network thus makes vision in general (and faces more specifically) highly suitable experimental modalities and stimuli for studies of human awareness (Section 1.3) and social perception (Section 1.4), as well as for studying behavioural and physiological responses to social threat (see Section 1.5).

1.3 Neuroscience of visual consciousness

1.3.1 Definitions and brief history of the scientific study of consciousness

Human consciousness has been called ‘one of the most enigmatic features of the universe’ (Koch, 2013) and the pursuit of its understanding is said to be ‘one of the most challenging goals of contemporary science’ (Schneider & Velmans, 2007). While consciousness is completely central to the experience of waking life for all healthy humans, agreement about its universal definition remains limited (perhaps because popular definitions come from so many different perspectives and refer to different aspects of what is likely a group of phenomena and processes rather than a single entity). Very generally, consciousness is understood to be the vivid, internal experience of our surroundings as well as of our thoughts and states. A common philosophical definition refers to “what it is like to be something”. There are many other important aspects of consciousness, some of which have become generally accepted as core features. These include its unitary and serial nature (we can only be conscious of one thing at a time), the fact that it is a first-hand experience that belongs to one’s self (Graziano, 2013), the existence of its contents for short periods (up to a few seconds), and finally the fact that this phenomenon is selective and with limited capacity with respect to the represented information even though this information can be sourced flexibly from any of the sensory or psychological processes in the brain (Zeman, 2001). Detailed analyses of the physiological and psychological properties associated with consciousness have been undertaken (e.g. Seth et al., 2005). The aforementioned definitions and characterisations relate to phenomenological aspects of consciousness and each particular irreducible experience (e.g. experience of the colour red) is known as a quale. Seeking an understanding of qualia is the Hard Problem of consciousness, which aims to tackle how feeling can arise from physical actions in the brain (Chalmers, 2007), and has thus far
proven largely intractable (Crick and Koch, 2003). Nevertheless, ways to study the neurobiology of phenomenological consciousness without tackling the Hard Problem directly have been developed (Section 1.3.2).

It is important to clarify that the phenomenological aspects of consciousness (which represent a major focus in this thesis) are a distinct (although related) entity compared to the state or level of consciousness that translates to a particular level of wakefulness or arousal. These two aspects of consciousness can be represented as two major and orthogonal dimensions (Laureys, 2005). Generally, high levels of arousal (being awake) are associated with phenomenological consciousness whereas low levels (sleep, coma) are not, although there are important exceptions to this in both healthy and pathological states. This relationship between level of arousal and the richness of conscious experience can be represented in a two-dimensional model space (Figure 1.3).

Figure 1.3: The two major dimensions of consciousness
The figure shows level of consciousness and content of consciousness on two orthogonal axes. In normal physiological states (turquoise colour), there is a roughly linear relationship between these dimensions. Dream activity during REM sleep is an exception to this; here there may be rich awareness and yet reduced wakefulness. Pharmacologically of pathologically induced coma (red), where both wakefulness and awareness are severely reduced, also generally conforms to this linear relationship. In purple are a group of conditions including sleepwalking, seizures and vegetative states where there is dissociation between the two dimensions (individuals are awake but not aware; figure adapted from Laureys, 2005).
Although the quote at the beginning of this Chapter, attributed to Hippocrates, is a clear example of deep reflection on the nature of mental processes from as long ago as the fifth century BC, it is claimed that specific conceptions regarding consciousness came much later (Wilkes, 1984). Modern views of consciousness and associated scientific study probably date back to the work of Western philosopher-scientists of the early 17th century (see Frith and Rees, 2007 for a more detailed account). René Descartes (1596–1650) defined thought in terms of reflexive consciousness or self-awareness. He also famously formulated dualism (a clear distinction between body and mind) and related non-conscious and conscious processes to this, postulating that the brain was sufficient for the former but a mind was required for the latter. Cartesian dualism is not a popular view in contemporary science, although the Hard Problem of consciousness (see above) continues to focus on bridging a conceptual gap between body and mind. In an opposing wave of thought to dualism, materialism, Pierre Gassendi (1592–1665) and others argued that matter was the fundamental substance in nature and formed the basis of all phenomena, including consciousness. Such principles have endured in the scientific study of consciousness until the present day. Philosophical ideas continued to develop during the 18th century, for example through the work of Immanuel Kant (1724–1804) who argued that phenomenal consciousness needed to reflect the experience of a conscious self situated in an objective world structured with respect to space, time and causality. The 19th century saw the development of psychophysics and experimental psychology, sciences for measurement and empirical study of mental phenomena, most notably driven by work from Gustave Fechner (1801–1887). Among a number of important physiological discoveries, Hermann von Helmholtz (1821–1894) first developed the idea that the brain processes sensory information in an inferential manner, and that this processing is non-conscious. In the early 20th century the study of consciousness was largely eclipsed within psychology by the rise of behaviourism, particularly its more radical forms as promoted famously by Burrhus Skinner (1904–1990), who aimed to eliminate mentalistic concepts from the study of psychology. The strong influence of behaviourism eventually weakened in the 1960s with the development of cognitive psychology and mathematical representations of cognitive processes. In the second half of the 20th century, the innovations of priming experimental techniques (Marcel, 1983) as well as studies of patients with brain lesions (Warrington and Weiskrantz, 1968) were important sources of new evidence for subliminal perception. Development of the scientific study of consciousness has flourished and accelerated markedly since the final decade of the 20th century, largely facilitated by dramatic technological developments that have permitted non-invasive imaging of brain function through techniques such as PET and fMRI (see Chapter 2), but also aided by increased understanding of the broad scope of non-conscious processes (Zeman, 2008).
In order to study consciousness empirically, an operational definition has been of central importance, and this is now widely agreed to be the “accurate voluntary report” of the human experimental participant (Baars, 2007). Although generally accepted, this approach to measurement carries a number of disadvantages, which will not be discussed here (but see Overgaard and Sandberg, 2012; Seth et al., 2005). Theoretical models of consciousness aim to relate these measures of phenomenal experience to the neural properties of consciousness. Such models seek to be generally applicable and therefore differ from studies of the neural correlates of consciousness (see Section 1.3.2), where the aim is to identify more specific correlations between a certain conscious experience and the associated neuronal mechanisms. The diversity of the properties of consciousness has resulted in a variety of established theoretical models (reviewed in Seth, 2007). Some of these models, such as the global workspace theory (Baars, 1997) and related neuronal network model (Dehaene et al., 2003), as well as the information integration theory (Tononi, 2008), have proposed clear associated neuroanatomical substrates and have been applicable to empirical findings.

1.3.2 The neural correlates of consciousness

Since the early 1990s, a key focus for the neuroscientific study of consciousness has been on the objective of identifying and understanding the neural correlates of consciousness, defined as ‘the minimal neuronal mechanisms jointly sufficient for any one specific conscious percept’ (Crick and Koch, 1990b). This approach seeks to discover which components or types of brain structure or function are essential for the production of conscious experience, although it is recognised that the NCC for each individual conscious experience (or quale) will be unique and therefore different to those for any other. Nevertheless, the premise is that by increments we would eventually be able to entirely explain the relationship between phenomenological mental states (such as the feeling of pain) and physical brain states. Study of the NCC is argued to be a necessary step towards development of a successful theoretical account of consciousness (Mormann and Koch, 2007; see Section 1.3.1).

A common experimental approach to studying the NCC is to contrast brain activity when an external stimulus (or internal state) is experienced with brain activity when it is not (see also Section 1.3.3). If all other variables are kept constant or adequately controlled for and the only difference between conditions is the presence of conscious experience, then the corresponding difference in neural signal should represent the neural correlates of consciousness. Performing such a contrast means that consciousness can be treated as a variable and therefore studied empirically (Baars, 2007). Once the NCC for particular
stimuli or states are identified, an important next step would be to determine what is common across the NCC for different conscious experiences and thereby to gradually delineate what distinguishes patterns of neural activity that support consciousness form those that do not. Bistable visual paradigms have been important in facilitating such experimental approaches (see Section 1.3.3). Coupled with functional neuroimaging, and in particular fMRI, they form a highly useful set of methodologies for exploring differences in brain activity during non-conscious and conscious visual processing as well as exploring activity patterns associated with gating of initially non-consciously processed information towards conscious appraisal. Broadly speaking, evidence from such experimental approaches suggests that the NCC for visual stimuli include coordinated activity in ventral (extrastriate) visual cortical areas and frontoparietal regions (Haynes and Rees, 2006; Kleinschmidt et al., 2012; Rees, 2007; Sterzer et al., 2009b; see Section 1.3.4 for a more detailed summary). There is a growing body of work that has used other functional neurophysiological measurement approaches (in particular EEG and MEG) in conjunction with visual psychophysics to gather (often complimentary) evidence about NCC and the mechanisms of awareness. This work will not be reviewed here (there are existing reviews including Koivisto and Revonsuo, 2010; Kornmeier and Bach, 2012), although relevant studies will be referenced throughout the thesis.

Another approach to the study of NCC is to focus on the wakefulness dimension of consciousness (Section 1.3.1) and evaluate what changes occur in the brain when consciousness is globally diminished (such as in non-REM sleep, anaesthesia, coma and epileptic seizures; Tononi and Koch, 2008). These different causes of reduced wakefulness (often with corresponding reduction in phenomenological consciousness) seem to have at least partially overlapping mechanisms that often relate to changes in the activity of thalamocortical networks. Pathologically both diffuse changes in cortical function and localised damage to various nuclei in the thalamus, midbrain and pons (especially if bilateral) can lead to loss of consciousness (Mormann and Koch, 2007). Additional insights into the neural correlates of consciousness have been gained from studies of so-called ‘split brain’ patients (following therapeutic callosotomy), which generally show that in such cases both hemispheres are capable of independent conscious experience (Gazzaniga, 1995). In contrast to the correlational approach of functional imaging, lesion studies enable causative roles of specific elements of the NCC to be evaluated. Such studies can be performed in a reversible and anatomically directed manner using TMS (Boly et al., 2013; Vuilleumier and Driver, 2007). Alternatively, following recent developments in optogenetics, much more precise techniques can be directed at modifying specific neuronal populations to understand mechanisms for changes in arousal in non-human animals (Adamantidis et al., 2007). Despite considerable progress, our understanding of the NCC remains incomplete and these multimodal approaches will undoubtedly be important in advancing it further.
Another important perspective when seeking to understand the NCC is to consider which aspects of neural processing, and indeed human behaviour, absolutely require consciousness. One compelling thought experiment that has been described is to consider how a ‘Haitian zombie’, capable of normal human unconscious processes but lacking only consciousness, differs from conscious beings (Koch & Crick 2001). An important line of scientific enquiry, which will not be examined in detail here, seeks to generate and test hypotheses regarding the *functions* of consciousness (Crick and Koch, 2003; Dehaene and Naccache, 2001; Frith, 2010; Graziano and Kastner, 2011; Posner and Rothbart, 1998).

### 1.3.3 Experimental dissociation of sensory stimulation and conscious perception: visual bistability

Psychophysics is the scientific discipline that deals with the relationship between physical stimuli and the sensations or perceptions resultant from these stimuli (see also Section 1.3.1). Over the past 25 years or so, a group of psychophysical experimental paradigms have contributed importantly to studies of the NCC by providing a powerful means of dissociating sensory stimulation and conscious perception, allowing for the exploration of distinctions and boundaries between conscious and non-conscious visual processes empirically (review by Kim and Blake, 2005). So-called bistable stimuli produce spontaneous changes in visual awareness in the context of invariant sensory stimulation, allowing dissociation of these two aspects of visual processing and exploration of the associated neural mechanisms.

The most widely recognised class of bistable visual stimuli is perceptual illusions, such as the Necker cube (Necker, 1832; Figure 1.4A) or Rubin’s face-vase illusion (Rubin, 1915; Figure 1.4B). The 12 lines of the Necker cube can be perceived as one of two depth configurations, while Rubin’s face-vase illusion can be perceived either as a vase or as two faces looking at each other. When an observer looks at these illusory figures for some time, perception is found to spontaneously alternate between one possibility and the other.

In binocular rivalry (Blake and Logothetis, 2002), a different image is shown to each eye, and the observer’s conscious experience alternates between perception of one image or the other while sensory stimulation remains unchanging (Figure 1.4 C-D). The existence of binocular rivalry has been known for centuries, but it was not until the 19th and 20th centuries that this phenomenon was studied systematically and in detail. Most notably, detailed observations and several laws that describe relations between physical stimuli and phenomenological experience in binocular rivalry were outlined by Levelt (1968) and have
been reviewed more recently (Brascamp et al., 2015). It is established that perceptual changes during binocular rivalry (and other forms of bistable perception) are influenced by both bottom-up mechanisms (including neural fatigue and competition between the neural representations of two different percepts) and top-down mechanisms typified by effects of learning, volition and expectancy (Long and Toppino, 2004; Wang et al., 2013). These complex and distributed systems are reflected by neural activity changes (that mirror perceptual changes) seen at multiple sites and hierarchical levels in the brain (e.g. in early visual cortical areas, in regions involved in category-level representations relevant to the experimental stimulus, and in frontal and parietal association cortices; Tong et al., 2006; Sterzer et al., 2009b). Binocular rivalry has become an indispensable tool for studying visual consciousness and whole books have been compiled on the phenomenon (Alais and Blake, 2005) and its relationship to awareness (Miller, 2013). The relative ease with which successful binocular rivalry can be achieved using a wide range of stimulus categories and levels of complexity has ensured the enduring popularity of this technique (including for studying mechanisms of conscious access, Section 1.3.5; or influences of social cues on visual awareness, Section 1.5). I have made use of this paradigm for three experiments within this thesis (Chapters 3, 4, and 7).

Another relevant experimental paradigm is flash suppression (Wolfe, 1984), where an image shown to one eye can be hidden from conscious perception by briefly flashing another image into the opposite eye. This technique has the advantage of providing external control over the timing of the change in perception. Single neuron recordings in humans and primates have provided convergent findings, namely that cells in medial temporal regions cease to be active when the test image becomes perceptually suppressed (despite the fact that sensory stimulation from the image is continuous; Kreiman et al., 2002).

More recently, it was discovered that combining binocular rivalry and flash suppression could enable visual stimuli to be rendered perceptually invisible for prolonged periods. In the technique of continuous flash suppression (Tsuchiya and Koch, 2005) an image of interest is presented to one eye and a sequence of rapidly flickering patterns (typically randomly generated ‘Mondrian’ masks) is concurrently shown to the opposite eye. This configuration usually suppresses the image of interest from awareness for several seconds. CFS has therefore proven extremely useful for studying non-conscious visual processes (e.g. Stein et al., 2012; Yang et al., 2007). Gradually increasing the strength of the suppressed stimulus (for example through a graded increase in contrast) results in its eventual emergence from suppression and the time to emergence, or time to break CFS (t2e; b-CFS), has been used as a probe for the strength of non-conscious processing associated with particular stimuli (Jiang et al., 2007; Stewart et al., 2012). There have been
some criticisms regarding the depth and uniformity of suppression from awareness provided by CFS under different conditions (Stein et al., 2011a; Yang et al., 2014; see also Section 9.3.1). While there are clearly important caveats that should be borne in mind when using this technique and interpreting associated findings, it is clear that CFS provides much deeper perceptual suppression than other forms of bistability (Tsuchiya et al., 2006) and a number of experiments have shown that, as far as a criterion of reportability is concerned, observers are no better than chance at determining what is hidden by a CFS mask when the paradigm is correctly executed (e.g. Fang and He, 2005; Sterzer et al., 2009a). The relative ease of implementing CFS (both for experimenter and participant) and the relatively reliable and deep perceptual suppression achieved have made this technique a suitable choice for some of the experiments within this thesis (Chapters 6 and 8).
Figure 1.4: Bistable visual stimuli

Illusory figures like the Necker cube (A) and Rubin’s face-vase illusion (B) are viewed binocularly and there is spontaneous alternation between two possible perceptual interpretations. In structure-from-motion, three-dimensional rotating spheres (which spontaneously change direction of rotation) can be perceived when viewing a collection of moving dots (C). In binocular rivalry a different stimulus is shown to each eye. There is considerable flexibility in the types of stimuli that can be used. Panel (D) shows visual gratings (which are especially suited to studying early visual cortical areas) placed in a configuration where stimuli are separated on a screen so that each is visible to only one eye (stimuli can then be made to appear superimposed by the use of mirrors or prism glasses, which redirect light from the images). Panel (E) shows face and house stimuli suitable for studying category-specific processing regions in extrastriate cortex and a configuration where stimuli are shown superimposed in different colours (face in green and house in red) and each is made visible only to one eye by the use of red-green glasses. Panel (F) shows an example of flash suppression, where a flash stimulus shown to one eye abolishes perception of another stimulus (in this case a face) shown to the opposite eye. Panel (G) is an example of continuous flash suppression, where a dynamic flashing (and in this case colourful) pattern shown to one eye suppresses awareness of the image shown to the other eye for an extended period (figures adapted from Blake and Logothetis, 2002; Tong et al., 1998; Kreiman et al., 2002; Tsuchiya and Koch, 2005).
While binocular rivalry, flash suppression and CFS provide excellent opportunities to study a wide range of both simple and complex static images, other bistable paradigms are better suited to images that are moving or have stereoscopic properties. Structure-from-motion stimuli (Wallach and O’Connell, 1953) fall into this category, and are also extensively used in this thesis (Chapters 3, 4 and 5). Commonly, sinusoidally moving dots are placed in a two-dimensional spatial configuration so that they can be perceived as rotating three-dimensional cylinders or spheres with an ambiguous surface order (or direction of rotation). Analogous to the binocular rivalry findings, changes in perceptual experience are tracked by neuronal activity in hierarchically advanced motion-sensitive visual areas (e.g. area MT; Bradley et al., 1998). The choice of bistable paradigm can thus permit different parts or hierarchical levels of the visual processing systems to be targeted for study (e.g. CFS is more suited for probing object recognition systems and SFM better for studying motion-selective mechanisms; see Section 1.2.1).

A number of other techniques permit dissociation between physical stimulation and perceptual experience, usually through allowing hitherto visible images to be erased from awareness by various means (reviewed in Kim and Blake, 2005). These techniques, including masking, motion-induced blindness, the attentional blink, and change blindness, have all been used extensively to explore the neural correlates of consciousness but will not be discussed further at present and their use does not directly form part of the work in this thesis.

1.3.4 Non-conscious and conscious visual processing during bistable perception

As discussed above (Section 1.3.2), contrasting conscious and non-conscious visual processes is a central approach to delineating and understanding the NCC. A crucial step in this pursuit is to explore and characterise the nature of non-conscious processes (Rees, 2007). It is now widely accepted that the majority of the brain’s processing occurs without our awareness (Dehaene and Naccache, 2001; Velmans, 1991). A proposed subdivision of non-conscious processes is into those that remain non-conscious due to limited bottom-up stimulus strength (termed subliminal) and those that remain non-conscious due to withdrawal of top-down attention (termed pre-conscious; Dehaene et al., 2006). This is an important distinction with implications for the associated neural processes. Throughout this thesis I will use the broader term ‘non-conscious’ since my experiments will not systematically manipulate bottom-up versus top-down effects.
Evidence from experiments using psychophysical manipulations to suppress and mask visual stimuli from awareness (see Section 1.3.3) suggests that a variety of complex evaluations can be performed in the brain without conscious appraisal. These include processing of organisational visual structures and scenes that require high-order perceptual computations and integration (e.g. Jiang et al., 2007; Mudrik et al., 2011), social evaluation (Stewart et al., 2012), and even semantic processing or arithmetic (Sklar et al., 2012). In terms of relevant neuroimaging findings, early visual areas are seen to be active even when the observer is not aware of any stimulus (for example during visual saccades in darkness; Sylvester et al., 2005). Further up the visual processing hierarchy, both ventral stream and dorsal stream areas can be activated by perceptually invisible objects, albeit at lower strength than for a visible condition (Fang and He, 2005; Moutoussis and Zeki, 2002). Likewise, subcortical structures, namely the amygdala, can be reliably activated by perceptually invisible fearful face stimuli (Morris et al., 1998; Williams et al., 2004; see also Section 1.4.4). Sophisticated analysis methods for fMRI data, which make use of machine learning algorithms to examine fine-grained and spatially distributed patterns of neural activation, have revealed a great deal of extra evidence for non-conscious neural processing where it has not been detectable with standard fMRI analyses (Haynes and Rees, 2006). Such MVPA approaches have shown that even for perceptually invisible stimuli, orientation can be decoded from V1 activity (Haynes and Rees, 2005a); category-specific neural processing can be detected in extrastriate cortex (Sterzer et al., 2008); and neural activity related to decisions can be detected in prefrontal and parietal regions (Soon et al., 2008). The boundaries of non-conscious processing are therefore continually being extended, implicating a highly distributed group of brain regions including subcortical structures and multiple parts of the dorsal and ventral visual pathways (Rees, 2007), although the limits and relevant modulatory influences for such processes remain incompletely characterised. As outlined above, non-conscious processing can be importantly influenced both by top-down factors, such as attention (e.g. Bahrami et al., 2007) or prior knowledge (e.g. Anderson et al., 2011); and bottom-up influences such as stimulus orientation (e.g. Stein et al., 2012) or stimulus emotional content (e.g. Yang et al., 2007; see also Section 1.4.4).

Having discussed neural processing of images that are not experienced consciously, I will now briefly consider situations where neural activity is correlated with the contents of awareness. Again, the approach of dissociating perceptual experience from sensory stimulation (see Section 1.3.3) has been employed extensively to this end. A large number of relevant findings have been based on non-bistable stimuli that sometimes reach awareness and at other times remain non-conscious (for example, stimuli at perceptual threshold, among others; reviewed in Rees, 2007). Here I will briefly summarise relevant findings specifically from studies of bistable perception. Reliable reports of ongoing
perceptual experience during binocular rivalry have been achieved in macaque monkeys undergoing intracranial EEG recording; such perceptual report correlates with activity in higher hierarchical areas along the ventral visual stream (e.g. area IT). In contrast, activity in early visual areas (V1) is only weakly modulated by changes in perceptual experience and is instead more closely reflective of the sensory conditions (Logothetis, 1998). In human studies, where binocular rivalry has been used extensively in conjunction with fMRI, perception-related changes in neural activity at many levels of the visual processing hierarchy have been observed (Blake and Logothetis, 2002; Sterzer et al., 2009b). In particular, a series of findings have contrasted with the conclusions drawn from primate electrophysiology, suggesting that changes in perceptual experience are observed (Haynes and Rees, 2005a; Lee et al., 2005; Polonsky et al., 2000) and can even be induced (Pearson et al., 2007) in association with neural activity changes in early visual areas. On the other hand, it has been argued that activity in these earlier areas is modulated by attention rather than correlating directly with conscious perception (Tononi and Koch, 2008). In hierarchically more advanced regions, such as FFA, responses during binocular rivalry clearly and unarguably reflect perceptual experience (Tong et al., 1998). The overall pattern of findings has led to proposals that binocular rivalry (and other forms of bistability) result from ongoing interaction between regions at multiple hierarchical levels of visual processing (Sterzer et al., 2009b; Tong et al., 2006). This view is supported by work based on attractor-network models, showing that distributed dynamic brain activity during bistable perception can be described as moving between three energy landscape minima, which relate to underlying anatomical features: a visual-area-dominant state, a frontal-area-dominant state, and an intermediate state (Braun and Mattia, 2010; Watanabe et al., 2014).

As is apparent from the findings reviewed in this section, no clear anatomical subdivision has been found between brain areas that support visual phenomenological awareness and others that do not (Rees, 2007). There may instead be important differences in the nature of activity related to conscious and non-conscious processing. Often perceptually visible stimuli give rise to stronger activation in functionally specialised visual cortex than their perceptually suppressed counterparts, although this is not always the case (Jiang and He, 2006; Rees et al., 2000). Moreover, the fine-grained patterns of neural activity detectable by MVPA can reveal additional more subtle differences in activity supporting conscious and non-conscious processes that are not detectable by univariate analyses (Haynes and Rees, 2006). There is also evidence that timing and integration of processing signals is important. For example, in monkeys late activity in V1 (likely the result of feedback from extrastriate areas) is correlated with awareness (Supèr et al., 2001). In another example, awareness of motion is impaired if feedback signals from V5 to V1 are disrupted with TMS (Pascual-Leone and Walsh, 2001).
1.3.5 Perceptual transitions and associated neural activity

A related area of focus, that is complimentary to understanding the nature of conscious and non-conscious visual processing, concerns the transition between these two processing modes, or in other words, the neural mechanisms that govern access of visual information to conscious awareness. In this case, instead of focusing on sustained periods of perception experienced during bistable visual paradigms (when conscious or non-conscious processing can be examined; Section 1.3.4), analysis can be directed at the time points when there is a transition between different perceptual experiences. A number of fMRI studies have shown that BOLD signals time-locked to such perceptual transitions are observed in a group of predominantly right-hemisphere brain regions, especially in inferior frontal and superior parietal cortex (Kleinschmidt et al., 1998; Lumer et al., 1998; Sterzer et al., 2002; Figure 1.5). Spontaneous changes in the contents of awareness in various other non-bistable yet ambiguous perceptual paradigms are associated with similar activity patterns (Rees, 2007).

Subsequent work has gone further to demonstrate that perceptual transition-related activity in high-order regions has temporal precedence over transition-related activation in visual areas. Such evidence exists for right inferior frontal gyrus (measured using fMRI; Sterzer and Kleinschmidt, 2007), and for right inferior parietal cortex (measured using EEG; Britz et al., 2009). These findings have been interpreted as indirect demonstrations that frontoparietal regions mediate active reinterpretation of perceptual experience (Rees, 2007; Sterzer et al., 2009b). However, whether frontoparietal activation indeed represents the cause of perceptual transitions remains under considerable debate. For some time it has been suggested that the observed changes in frontoparietal regions may be the result of feed-forward signals linked to the resolution of perceptual ambiguity in visual areas through reciprocal inhibition (e.g. see Alais et al., 2010). More recent findings argue that transition-related frontoparietal activity may in fact be the consequence of neural responses to transitions (e.g. Knapen et al., 2011), perhaps representing functions of the ventral attention network with which these patterns of activation largely overlap (Corbetta and Shulman, 2002; see Sections 3.4.5 and 9.2.3 for further discussion). More direct support for a causal role, particularly for parietal regions, comes from studies on patients with visual neglect (Vuilleumier and Driver, 2007), and studies using repetitive TMS to show that disrupting function of right SPL has an effect on the rate of switching between percepts in binocular rivalry or SFM paradigms (Carmel et al., 2010; Kanai et al., 2010, 2011; Zaretskaya et al., 2010; Section 3.1). Some work has also been done with TMS to explore the role of frontal regions (de Graaf et al., 2011) as well as the interaction between frontal and parietal areas (Vernet et al., 2015). These issues will be discussed further in Chapter 5, where I will explore the causal importance of frontal regions in perceptual transitions, specifically focusing on FEF.
Figure 1.5: Clusters of frontal and parietal BOLD activation associated with changes in the contents of visual awareness

Each black dot represents a single focus of BOLD signal time locked to perceptual transitions during bistable perception. Findings are pooled from three studies (Kleinschmidt et al., 1998; Lumer et al., 1998; Sterzer et al., 2002). The majority of activations are found in superior parietal and inferior frontal regions (circled; figure adapted from Rees, 2007).

1.4 Social visual perception

One of the great strengths and specialisms of the human visual system is in the processing of social cues. As with other facets of visual perception, social vision ultimately allows us to select appropriate behaviours, which in this context are based on visual cues about others’ mental or emotional states (Nakayama, 2011). These functions are extremely important in human societies, where competition is balanced by cooperation: any individual’s relation to any other individual may be highly relevant to success. The foundations of such theories are based in Darwin’s work where he argued that emotions prompt actions that are beneficial to the organism and its survival (Darwin, 1872). Two categories of visual social information have received the overwhelming majority of scientific attention: point light demonstrations that convey body movement, and human faces (Nakayama, 2011). Of all visual social signals, faces are undoubtedly the most multidimensional and socially informative (Adams et al., 2011; see Section 1.4.1). The systematic description of different face attributes has especially benefitted from the development of mathematical, analytic and image processing techniques, resulting in the creation of novel metrical spaces to represent faces (e.g. Valentine, 1991; see also Section 1.4.2). A large proportion of the work within this thesis makes use of faces as experimental stimuli; the rest of this section will therefore focus on several relevant aspects of face perception.
1.4.1 Social face perception, including evolutionary and historical perspectives

Faces carry information about their owner’s character and intentions; their societal importance has been consistent across history as well as across diverse cultures. Socially relevant visual cues from faces include dynamic signals such as eye gaze and emotional expressions as well as more static cues based on facial features, or traits (Adams et al., 2011). We draw inferences about others’ personalities and dispositions rapidly and effortlessly from both invariant and dynamic facial information. This has led to proposals that humans perceive compound social cues facilitated by early integration of processing of multiple social features from faces. In addition, our own stable or transient social states also guide social perception, meaning that individual factors relevant to the observer must be carefully considered (Adams et al., 2011; see also Section 1.6). There has been extensive study of the neural systems for emotional processing and recognition, particularly relating to dynamic facial cues (comprehensive review by Adolphs, 2002). At present (and for the experimental work in this thesis) I will focus on non-dynamic visual attributes (traits) of faces, which have received less attention in the existing literature.

We evaluate faces using many trait-based measures; if given no constraints regarding the descriptors used, groups of human observers use hundreds of different terms; examples of the more common ones include ‘sociable’, ‘mean’, ‘weird’, or ‘confident’ (Oosterhof and Todorov, 2008). Other examples of facial traits, which have been extensively studied due to their proven evolutionary and societal importance, include attractiveness (Thornhill and Gangestad, 1999), dominance (Adams et al., 2011) and trustworthiness (Todorov et al., 2005). Although the relatively invariant nature of facial traits sets them apart from the constantly dynamic emotional facial expressions, evaluation of these aspects of faces is not independent. For example, judgments of trustworthiness are influenced by dynamic information such as viewing angle (Sutherland et al., 2017). Moreover, there is support for commonality in the evaluation strategies and mechanisms applied to these two types of social cue. Judgments of face traits and emotional expressions are largely inter-correlated so that, for example, faces rated as dominant are also rated as angry (Engell et al., 2010; Said et al., 2011; Figure 1.6). Principal components analysis also shows that emotion and trait judgements have very similar dimensional structures (Said et al. 2011). Complementary evidence for common processing mechanisms across different emotional and social cues comes from adaptation studies demonstrating that emotionally neutral faces were judged as more trustworthy after adaptation to angry faces and less trustworthy after adaptation to happy faces (Engell et al., 2010).
Contemporary studies of social face traits exist in an interesting and at times troubling historical context. During the late 18th century, as well as much of the 19th century, the concept of physiognomy (that inner human character could be inferred from external facial appearance) became highly popular. This popularity was in large part founded on Johann Caspar Lavater’s *Essays on Physiognomy* (1789-1793) where he claimed that physiognomy reflects all that a person does and feels. By many, physiognomy was considered a science, but it also found a broad appeal within religious thought (Hartley, 2005), and was even applied to forensic science most famously by Cesare Lombroso who developed theories that criminals carried physical stigmata of their criminality (especially in relation to skull and jaw dimensions). It has been argued that the ideas of physiognomy were naturally replaced in the second half of the 19th century by a new science of character and behaviour founded on Darwin’s work on emotional expression (Darwin, 1872), and then recast as part of Galton’s theories on hereditary transmission, which were influenced by the work of Darwin (Hartley, 2005). Indeed, in the present day, the premise of physiognomy appears highly fanciful and unscientific. Nevertheless, these ideas seem to, at least in part, reflect the very rapid inferences we automatically make about others’ character and intentions form their facial appearance (Willis and Todorov, 2006). Indeed, such inferences can include perceptions of criminality (Flowe, 2012).

![Figure 1.6: Correlations between trait judgments and emotional ratings attributed to naturalistic emotionally neutral faces](image)

Figure 1.6: Correlations between trait judgments and emotional ratings attributed to naturalistic emotionally neutral faces

Figure shows correlations between trait judgments attributed to 66 images of naturalistic emotionally neutral faces and emotional ratings attributed to the same faces by an independent group of raters. Both sets of ratings were performed on 9-point scales; since faces were emotionally neutral, ratings of ‘subtle’ emotional content were asked for. The figure shows high inter-correlation between emotion and trait judgments with 65% of all correlations statistically significant (asterisks) and a number of correlations being very strong. The colour bar indicates strength of correlation (*r*), with red representing a strong positive correlation and blue representing a strong negative correlation (figure adapted from Said et al., 2011).
1.4.2 A 2-D model space for social face evaluation

Although we rate faces on multiple social dimensions (see Section 1.4.1), these turn out to be highly inter-correlated (Oosterhof and Todorov, 2008). Therefore, social face evaluation can be characterised by models that reduce these multidimensional ratings to principal axes. The most established model of this nature describes social judgments along dimensions of warmth and competence (Fiske et al., 2007). A more recent model by Oosterhof and Todorov (2008) similarly reduces social trait judgments to two orthogonal dimensions of dominance and trustworthiness as motivated by principal components analysis of a large data set of unconstrained descriptions of real-life face images (Figure 1.7A). The novelty and applicability of the approach of Oosterhof and Todorov came in the subsequent steps of their analysis when they used a modified version of the FaceGen Modeller software package to create computer-generated faces that could be manipulated parametrically along their two dimensions, varying in standard deviations from a mean (neutral) face (Figure 1.7B-C). Finally, the faces generated by the programme were validated using an independent sample of participants, showing that ratings given to the computer-generated faces were highly correlated with ratings predicted by the model (Figure 1.7D). At the extremes of the dominance or trustworthiness dimensions, faces were also rated as having emotional expressions (e.g. dominant faces were rated as angry; this accords with independent findings; see Section 1.4.1 and Figure 1.6). However, this was not the case for more moderate variations in dominance and trustworthiness (up to ± 3 s.d., as shown graphically in Figure 1.7C); such faces were reliably rated as emotionally neutral. The main diagonal in the two-dimensional model was shown to represent threat, such that faces that were both dominant and untrustworthy were rated as threatening (Oosterhof and Todorov, 2008). Dominance ratings (both from this model and from real-life faces) are highly correlated with threat ratings, while trustworthiness ratings are highly correlated with judgments about attractiveness and intelligence.

Using the model of Oosterhof and Todorov face images are created in an unbiased data-driven and parameterised fashion, and can form a flexible yet structured framework for studying the impact of social traits on visual processing and awareness. I therefore used stimuli constructed using this model in several experiments within this thesis (Chapters 3, 6, 7 and 8).
Figure 1.7: Creation of a two-dimensional trustworthiness-by-dominance model space for evaluation of social face traits.

(A) Solution of PCA of trait judgments for 66 naturalistic emotionally neutral faces. The first PC approximates to trustworthiness and the second PC approximates to dominance. (B) Computer-generated face model constructed using FaceGen Modeller software, where each specific face is represented as a function of the average face by calculating differences on all vertices in a surface mesh. Clockwise from top left: a surface mesh is superimposed on an average face; linear changes in vertex positions are represented in 50 shape dimensions; frontal view of average face with texture; frontal view of average face with surface mesh and texture. (C) Examples of the computer-generated faces varying on orthogonal dimensions of trustworthiness and dominance. Face modulation along each trait is represented in s.d. units on both axes. (D) Validation showing that trustworthiness and dominance judgments of faces generated by the model are correlated with the model ratings (figures adapted from Oosterhof & Todorov, 2008).
1.4.3 Neural mechanisms for social face processing

The neural mechanisms supporting social perception and evaluation are complex and distributed; some of the key brain regions implicated consistently across large and diverse bodies of work include the amygdala (e.g. Adolphs, 2010), the medial prefrontal cortex (Amodio and Frith, 2006), the superior temporal sulcus (Allison et al., 2000; Hein and Knight, 2008), and the temporoparietal junction (Saxe and Kanwisher, 2003).

Many studies have explored the neural processing of socially relevant information contained within faces using fMRI. These have included work on the processing of emotional facial expressions (Blair et al., 1999; Phillips et al., 2004; reviews by Vuilleumier and Pourtois, 2007 and Adolphs, 2002), and eye gaze perception (Grosbras et al., 2005). In addition, important work has highlighted gender differences in the neural processing of emotional facial expressions (Hofer et al., 2006). In the domain of social face traits, dominance is known to be a key dimension of social evaluation in fields of social psychology and anthropology (Adams et al., 2011); data-driven approaches also suggest it represents a principal axis of social face evaluation (Oosterhof and Todorov, 2008; see Section 1.4.2). A single study has explored the neural processing of social dominance as conveyed by head posture (rather than facial traits) using fMRI and EEG, showing associated activity occurs relatively late and is found in the fusiform, superior temporal and lingual gyri (Chiao et al., 2008). Other work has focused on the brain mechanisms relating to social dominance hierarchies in human society (Chiao, 2010; Kumaran et al., 2012). Little is known about the neural processing of social dominance information as specifically derived from facial traits. There is a fairly substantial literature focusing on neural processing of other face traits including trustworthiness (Engell et al., 2007; Said et al., 2009; Winston et al., 2002) and attractiveness (Aharon et al., 2001; Chatterjee et al., 2009; Iaria et al., 2008; Winston et al., 2007; see Getov and Winston, 2015 for a review). Trait-specific patterns of BOLD signal change have been described, with trustworthiness evaluation more likely to engage higher order social perception regions such as medial prefrontal cortex (Todorov et al., 2008) or STS (Winston et al., 2002) and attractiveness evaluation more likely to involve subcortical structures implicated in reward processes (Getov and Winston, 2015). On the other hand, a meta-analysis has shown considerable overlap in the brain regions implicated in neural processing of attractiveness and trustworthiness, as well as variable involvement dependent on valence of the face stimulus (Mende-Siedlecki et al., 2012). The amygdala appears to be important for processing of negatively valenced stimuli while orbitofrontal and medial prefrontal cortices as well as the nucleus accumbens were implicated in processing of positively valenced stimuli (Getov and Winston, 2015).
The spatial overlap in neural mechanisms supporting attractiveness and trustworthiness evaluation highlighted by Mende-Siedlecki and colleagues (2012) dovetails with the behaviourally-demonstrated commonalities between evaluation of different emotional expressions or social traits and the highly inter-correlated emotional and trait judgments discussed earlier (Section 1.4.1; Figure 1.6; Said et al., 2011). Similarities in the neural processing that underlies evaluation of untrustworthy and angry faces have also been outlined (Mende-Siedlecki et al. 2012). These common features among mechanisms for evaluating different social cues could reflect a core brain network for social trait evaluation, which may relate closely to the comparatively well understood neural mechanisms for evaluation of emotional facial expressions (reviewed by Vuilleumier and Pourtois 2007). Nevertheless, some of the literature reviewed in this section is also clearly suggestive of mechanisms that are specific to a particular social face trait or emotion exhibited by a face. Further work is needed to specifically explore dissociations in the neural processing of different social cues.

1.4.4 Non-conscious processing of social visual cues

Many emotional and social processes can occur automatically, without requiring awareness (Bargh et al., 2012) and the neural mechanisms underlying non-conscious processing of emotional stimuli in particular have been extensively studied (reviewed by Tamietto and de Gelder, 2010). Early and seminal studies reported that BOLD signal changes could occur in amygdala in response to fearful face images that were not consciously perceived (Morris et al., 1998; Whalen et al., 1998). Accumulating evidence from psychophysics, skin conductance measurements and fMRI has indicated that processing of visual emotional cues tends to be more rapid than for emotionally neutral stimuli. An influential suggestion, as part of evolutionary vigilance accounts, has been that emotional processing is prioritised due to its importance for survival (Sander et al., 2003; Vuilleumier, 2005). The proposed neural mechanisms for such prioritisation have been described within a dual route model (Dolan and Vuilleumier, 2003; Vuilleumier, 2005), which highlights the amygdala as a hub or central relevance detector (Sander et al., 2003; Anderson and Phelps, 2001), that can be accessed by socially relevant visual information via a rapid and coarse subcortical processing stream through the superior colliculus and pulvinar (purportedly bypassing visual cortex). According to this model the alternative slower route for social visual information makes use of the cortical visual processing systems (see Section 1.2). However, a body of relatively recent evidence supports a more complex set of mechanisms than those suggested by the dual route model, with subcortical systems including the amygdala acting as a gateway to cortical processing for emotional information, rather than as an independent subcortical stream (Pessoa and Adolphs, 2010).
A group of experimental studies have used newer methods of interocular suppression, in particular CFS (Section 1.3.3), instead of more traditional masking techniques to explore the evaluation of social stimuli that are suppressed from awareness. This work has provided evidence that faces with direct gaze overcome CFS faster than those with averted gaze (Stein et al., 2011b) and that faces with fearful emotional expressions overcome CFS faster than those with neutral expressions (Yang et al., 2007). As discussed in Section 1.3.3, breaking CFS more quickly has been used as a measure of prioritisation for awareness, and thus these findings also dovetail with evolutionary vigilance accounts. Work with both fMRI (Jiang and He, 2006) and EEG (Jiang et al., 2009) has shown that there are different signal patterns for non-conscious (i.e. suppressed by CFS) fearful-expression versus neutral-expression faces. With fMRI these differences are found primarily in amygdala and STS while using EEG they are seen at moderate latencies (>200ms) over posterior temporal sensors.

The experimental work reviewed in this section has provided substantial insight into the behavioural and neural responses to non-conscious emotional or eye gaze stimuli. However, there has not been investigation into the behavioural and neural processes associated with non-dynamic social face traits such as dominance and trustworthiness (see Section 1.4.1) when viewed outside of awareness. Notwithstanding the need to incorporate more recent findings, evolutionary vigilance remains an important framework for attempting to understand social modulation of vision and visual awareness (see Section 1.4.5). In the context of negatively valenced stimuli evolutionary vigilance can motivate opposite predictions to the cascade model used to describe defensive responses to threat (see Section 1.5). These two conceptual frameworks form the theoretical background for the design and interpretation of several experiments in this thesis (Chapters 6, 7 and 8). They are also discussed together in more detail in the final chapter (Section 9.4.1).

1.4.5 Social modulation of vision and visual awareness

In Section 1.4.4 I reviewed existing evidence that many visual social cues can be processed without awareness. From another perspective, the notion that social cues and behaviours can be important for constructing consciousness has received both theoretical and empirical support in recent years; some theories go as far as to propose that social interaction is the principle function of consciousness (Frith, 2010; Graziano and Kastner, 2011). Bistable perception has provided a powerful set of tools for investigating how conscious experience is affected by socially relevant information and it appears that such information can influence the boundaries between non-conscious and conscious visual processing. For
example, fearful faces are perceived consciously for longer than neutral-expression faces when the two are paired in binocular rivalry (Alpers & Gerdes, 2007). Moreover, BOLD signals in fusiform cortex and amygdala are stronger during awareness of emotional faces (paired with non-emotional faces in binocular rivalry), while those in frontal and parietal regions are stronger during awareness of non-emotional faces (Amting et al., 2010).

Factors relating to the observer rather than to the stimulus itself, such as depressed mood (Sterzer et al., 2011) or anxiety (Gray et al., 2009; Singer et al., 2012) can also influence the gating of emotional or threatening faces to conscious processing. Furthermore, socially relevant prior knowledge about faces (that are in their actual appearance socially neutral), can also affect the likelihood for them to be perceived consciously (Anderson et al., 2011).

Overall, while a small number of studies have explored the influence of social cues on the balance between non-conscious and conscious visual perception, many aspects of this question remain unexplored and I will attempt to address some of these in Chapter 7 of the thesis.

### 1.5 Defensive responses to threat

#### 1.5.1 Conceptual models of defence responses

Appropriate and effective behavioural responses to threat (e.g. from a predator or a rival from the same species) are vital for survival. In mammals (especially in the rat), these responses have been extensively investigated and shown to be complex and diverse (Lang and Davis, 2006). An influential group of models have structured these multifaceted behavioural and physiological changes into a so-called defence cascade, subdividing response patterns according to the proximity of danger (Blanchard and Blanchard, 1989; Lang et al., 2000). While intermediate proximity of danger is most often associated with tense and attentive immobility or freezing, increased proximity of threat tends to lead to an overt fight-or-flight response (Figure 1.8; Hagenaars et al., 2014). Defensive responses including freezing and fight-or-flight appear to be highly conserved across different mammals. In humans freezing may be triggered by some of the more diverse and subtle cues and inferences that guide the uniquely complex social interactions. Fear responses are now also known to play a role in the aetiology of human psychological illnesses including post-traumatic stress disorder and social phobia (Brandão et al., 2008; Hagenaars et al., 2014; Lang et al., 2000).
It is important to emphasise here the existence of alternative theoretical frameworks that encompass threat perception but are not clearly reconcilable with defence cascade models. In particular, evolutionary vigilance accounts, which have been applied to understanding non-conscious visual processing of emotional signals (see Section 1.4.4), posit that threatening stimuli are processed more rapidly (thereby increasing the chance of timely appropriate reaction and survival) without emphasis placed on the proximity of the threat (e.g. Vuilleumier, 2005). Arguably, evolutionary vigilance does not adequately address how perceptual processes relate to the behavioural correlates of defence, and whether perceptual processes differ between different defence modes, which, as discussed, do exist in humans (see Section 1.5.2). Dual competition models better explain deleterious effects of threatening stimuli on task performance but again do not directly describe peripheral physiological components of threat responses (e.g. associated autonomic changes; Section 9.4.1). On the other hand, neural defence distance models (McNaughton and Corr, 2004) have some similarity to defence cascade models, in that they also adopt the defensive distance dimension when characterising threat responses. However, there is additionally a defensive direction dimension with approach to threat corresponding to anxiety and avoidance of threat corresponding to fear (McNaughton and Corr, 2004; see Section 9.4.1 for further discussion). These alternative frameworks will be used to anchor the findings from some of the experimental work in this thesis, specifically when exploring influences of non-conscious socially threatening cues on defensive behaviour and awareness (Chapters 6, 7 and 8; detailed discussion in Section 9.4.1).
1.5.2 Freezing responses and their neurobiology

In rodents, the motor manifestation of freezing (immobility) is considered the core feature of post-encounter defensive behaviour in the context of moderate threat levels (Figure 1.8; Fanselow, 1994). Motor freezing can occur as part of several immobility responses that include orienting, risk assessment, tonic immobility and behavioural inhibition, although it remains under debate whether these are all distinct responses or part of a continuum of defensive behaviour (Hagenaars et al., 2014). Bradycardia is another consistent correlate of freezing (Schenberg et al., 1993; Walker and Carrive, 2003); tachycardia takes over when threat is more imminent and a more overt defensive reaction is required (Hagenaars et al.,
Much of the rodent work on freezing has been undertaken using a model of conditioned fear, which examines how associative learning mechanisms (that allow conditioned stimuli to be associated with threat) can engage threat responses. This is clearly a distinct scenario from innate (non-learned) responses to threat. While freezing responses have been shown to occur with both innate and conditioned fear, there are some differences in the underlying neural mechanisms, which will be referred to below wherever relevant (Watson et al., 2016).

The neurobiological organisation that underlies the two core components of the freezing defensive response (immobility and bradycardia) is well described in the rat. An accumulation of evidence has conferred a central role for coordinating behavioural output during freezing on the amygdala, which is in a position to receive and processes a rich complement of sensory and affective information as well as having efferent projections to various subcortical systems involved in motor and autonomic output (Lang et al., 2000). Of particular importance to freezing behaviours are projections from CeA to PAG (that mediate the motor components of freezing) and to the lateral hypothalamus and brainstem nuclei (that mediate the autonomic changes seen in freezing, including bradycardia; Fanselow, 1994). There is a degree of anatomical segregation in PAG function, with the dorsal PAG more involved in the output of innate fear responses and the ventral PAG more involved in learned fear responses (Watson et al., 2016). Upstream from the amygdala, there is also evidence for involvement of the prefrontal cortex and hippocampus in fear responses (Tovote et al., 2015).

Lesions to CeA abolish freezing indicating that this region is a crucial relay in the system (Lang et al., 2000). In addition, lesion experiments in rodents have shown that the functions of the two pathways to PAG and hypothalamus are doubly dissociable when it comes to conditioned threat of foot shock (LeDoux et al., 1988). Electrolytic lesion studies have established that ventrolateral PAG is crucial for tonic immobility in response to threat (Kim et al., 1993). However, increasingly there is evidence for differential involvement of both ventral and dorsal PAG in different subtypes of immobility response (see above; Brandão et al., 2008). Moreover, as well as descending connections to the ventral horn motor neurons of the spinal cord, different subregions of PAG send ascending projections to thalamus and amygdala. These regions in turn likely interact with cortical regions including cingulate and motor areas, which may influence other aspects of the expression of fear responses (Vianna and Brandão, 2003; discussed further in Section 1.5.3). This complex circuitry may underpin the diversity of motor freezing responses described above, and could be subject to both individual and situational variability (see Section 1.6 and Section 8.1).
Bradycardia, the other core correlate of freezing (Hagenaars et al., 2014), is mediated in the rabbit by direct anatomical projections from CeA to brainstem nuclei, which give rise to vagal cardioinhibitory control (Schwaber et al., 1982). Notwithstanding the known interactions of autonomic function with cortical processes and behaviour (Critchley and Harrison, 2013), in the context of defence responses autonomic changes may operate at a more isolated and lower hierarchical level with limited direct influence from higher processes (Behbehani, 1995). This situation contrasts with the diverse and multifaceted mechanisms for the motor output of freezing and is likely reflected in a less diversified response of HR deceleration in post-encounter contexts.

A handful of studies have recently demonstrated that defence responses consistent with freezing (specifically, bradycardia and reduced body sway) can be elicited in humans viewing strongly affective images, in particular images depicting mutilation (e.g. Azevedo et al., 2005; Stins and Beek, 2007), but also with images that are less innately aversive but more directly relevant to social interaction, namely faces (for example, faces with angry expressions; Roelofs et al., 2010). It remains unclear to what extent the neurobiological underpinnings of human freezing overlap those of rodents and other mammals, although work with fMRI and TMS is beginning to shed light on human mechanisms (see Section 1.5.3). In addition, human defensive responses specifically to social signals of threat do not have a clear equivalent in the animal literature. The behavioural characteristics and neurobiology of such responses therefore warrant further investigation and these questions are addressed in Chapters 7 and 8 of the thesis. The question of whether specific defensive behaviours (and in particular, freezing) can be triggered by subliminal visual stimuli also remains largely unanswered, and will likewise be explored in Chapters 7 and 8.

1.5.3 Neural correlates of human freezing responses

A couple of fMRI studies have begun to shed some light on the functional anatomy of defensive freezing in humans. Negative images depicting mutilation are associated with BOLD signal changes in amygdala and PAG, as well as modulation of effective connectivity between these regions (Hermans et al., 2013). Moreover, there is evidence that such images interfere with performance of visual detection tasks, and BOLD signal in mid-cingulate cortex mirrors this interference. The interpretation of the cingulate result is that this region may be the site of interaction between negatively valenced information and motor signals (Pereira et al., 2010). Mid-cingulate cortex has appropriate connectivity both to receive widespread inputs from emotion-related regions and to project independently to motor cortex and the spinal cord. Modulations of cortical activity that may closely relate to the motor components of freezing (namely reduction of BOLD signal in motor cortex) have
been reported during anticipation of aversive electrodermal stimulation (Butler et al., 2007) or in association with reaction time delay when viewing task-irrelevant threatening images (Sagaspe et al., 2011). It has been proposed that even more hierarchically advanced cortico-limbic processes (e.g. in mPFC and hippocampus) are involved in mediating the subcortically-driven responses to threat seen in freezing (Fanselow, 1994; McNaughton and Corr, 2004). In support of such proposals, fMRI findings confirm that post-encounter threat is associated with activity in subgenual ACC and hippocampus as well as amygdala (Mobbs et al., 2009). There is also converging evidence from rodent studies, where ventral medial frontal cortex lesions abolish freezing behaviours (Frysztak and Neafsey, 1991). The involvement of higher-order regions in freezing contrasts with neural responses associated with circa-strike threat, which predominantly involve hierarchically lower midbrain regions (Mobbs et al., 2009). This mapping of more proximate threat onto lower level neural mechanisms is described within neural defence distance models (McNaughton and Corr, 2004; Section 1.5.1).

Few studies have used direct peripheral measures of motor physiology (such as TMS-MEP) to evaluate freezing, and this will be discussed further in Chapter 8, where such measures are implemented. Whether the freezing-related BOLD changes described above (e.g. in primary motor cortex; Butler et al., 2007) correlate with such direct physiological measures of motor activity is also not clear.

Studies in humans have not yet clearly determined whether the mechanisms mediating motor output and autonomic changes associated with freezing are separable, as they appear to be in rodents (see Section 1.5.2). The work of Hermans and colleagues (2013) has shown that BOLD signal in PAG is correlated with bradycardia during freezing on a trial-by-trial basis, which suggests common rather than distinct mechanisms. However, it is important to note that the relationship of BOLD signal in PAG to motor freezing is far from clear. The relationship between autonomic and motor components in human freezing will also be explored in Chapter 8.

### 1.6 Individual variability in visual perception

It is of course well known that there are significant differences in various skills, dispositions and behaviours between human individuals. Such variability is jointly determined by genetic and environmental influences. In various subfields of psychology, for example studies on personality (e.g. Humphreys and Revelle, 1984) or language (e.g. Skehan, 1991) the importance of individual differences has been emphasised for many years.
By contrast in systems neuroscience, including in the fields of visual and social perception that are of particular relevance to the work in this thesis, such differences have traditionally been treated as noise and attempts have been made to minimise their impact through averaging (Kanai and Rees, 2011). A more recent and opposing trend has been to instead focus on such differences and explore what they can tell us about behaviour, cognition, and the associated neural processes. It is now known that variability in human cognition and behaviour is mirrored by striking inter-individual differences in brain structure and function (e.g. Frost and Goebel, 2012; Mueller et al., 2013; van Essen and Dierker, 2007).

There has been a recent surge in experimental studies, asking diverse biological questions, which have explored the neural basis of individual differences, for example in mood or personality traits. Such work has highlighted new ways in which behaviour is determined by the workings of the brain. Of relevance to social perception (Section 1.4), one fMRI study has shown that measures of loneliness are related to individual variability in BOLD responses to social stimuli in the striatum and visual cortex (Cacioppo et al., 2009), while another fMRI publication has demonstrated that social relationship traits are correlated with BOLD signal in prefrontal regions and amygdala when viewing social scenes (Vrtička et al., 2011). Even more specifically, of relevance to non-conscious social processing (see Sections 1.3.3, 1.3.4 and 1.4.4), BOLD activation in the amygdala and STS when viewing fearful faces masked by CFS correlates with individual differences in measures of mood and anxiety (Vizueta et al., 2012).

Another recently popular approach has been to relate variations in brain structure (rather than function) to variability in personality or cognition (reviewed in Kanai and Rees, 2011). This is facilitated by computational approaches to the analysis of structural MRI data, including voxel-based morphometry for estimating local variation in grey matter volume (Ashburner and Friston, 2000; Section 2.2.6), or surface-based approaches for derivation of local volume or cortical thickness (Fischl and Dale, 2000; Section 2.2.7). There have been characterisations of structural brain correlates of individual variability in various domains of behaviour and perception relevant to the topics of interest in this thesis (Kanai and Rees, 2011). For example, individual variability in social perception (Section 1.4) has been correlated with structural measures in posterior STS, and linked to measures of loneliness (Kanai et al., 2012). There are also relationships between focal brain structure and social perception in autistic populations (David et al., 2014). Inter-individual variability approaches have also begun to contribute to the understanding of the NCC and the mechanisms underlying visual awareness (Section 1.3). For example, individual differences in the perceived strength of visual illusions have been correlated with the surface area of early retinotopic visual regions, demonstrating how the structure of such regions can shape conscious awareness (Schwarzkopf et al., 2011). Moreover, there has been systematic
exploration of the neural correlates of the known variability in the dynamics of bistable perception (e.g. Miller, 2013; Section 1.3.3). The mean duration of ambiguous percepts is fairly stable within individuals but varies substantially across individuals. This inter-individual variability correlates with focal grey matter volume in SPL (Kanai et al., 2010), with opposite directions of correlation relating to structure in anterior and posterior portions of SPL (Kanai et al., 2011). Thus, these regions might play opposing (but perhaps complementary) roles. The functional importance of these findings has been confirmed using TMS, showing that suppressing the function of these SPL subregions has opposite effects on the dynamics of bistable perception (Kanai et al., 2011; see Section 6.1 for further details). Mechanistic explanations for this finding have been proposed and have begun to be tested empirically (Megumi et al., 2015; the implications of these findings are discussed in detail in Section 9.2.2). Such work demonstrates the utility of inter-individual variability approaches for understanding the neural mechanisms of social perception and visual awareness (see Sections 1.3.4 and 1.3.5).

A number of the studies discussed in the paragraphs above have explored neural correlates of individual variability in personality, mood or emotional characteristics. A popular way to obtain these psychological measures is through the use of self-rated questionnaires. Personality and mood questionnaires have been shown to be useful, reliable and widely applicable (e.g. McCrae and Costa Jr., 1997; Shafer, 2006). They can be comparably administered either in person or online (Pettit, 2002). However, there are also important limitations (reviewed in Fernandez-Ballesteros, 2004). Some of the experimental work within this thesis with particular focus on understanding inter-individual variability makes use of such questionnaires. In particular, in Chapter 7 I explore individual variability in the dynamics of binocular rivalry. As already discussed (Section 1.3.5) personality and mood traits can relate to the dynamics of binocular rivalry. Therefore, several questionnaires, all with proven internal validity and test-retest reliability, are used in Chapter 7 to probe personality and mood traits that may be relevant to social face evaluation. These include the Submissive Behavior Scale (Allan and Gilbert, 1997), the Propensity to Trust Survey (Evans and Revelle, 2008), the State-Trait Anxiety Inventory (Spielberger et al., 1983), and the Behavioural Inhibition/Behavioural Activation scales based on Grays’s biopsychological theory of personality (Carver and White, 1994). Additionally, the findings reported in Chapter 6 have contributed to a published study, which makes use of some of the above self-rated questionnaires (Stewart et al., 2012). Further details regarding these questionnaires can be found in Stewart et al. (2012), while the relevance of the questionnaire findings to the work within this thesis is discussed in Section 6.5.1.

Clearly, there is already substantial evidence that measures of neuronal structure and function, as well as measures of personality and mood traits, can explain some of the individual variability in behaviour and cognition, including in social perception and
conscious experience. Focus on individual variability is an enduring theme throughout much of this thesis and on several occasions it is such a focus that leads to the clearest demonstrations of relationships between behaviour or perception and anatomical measures in the brain (Chapter 6), measures of effective connectivity between different brain regions (Chapter 4) or physiological measures of peripheral motor function (Chapter 8). In addition, as part of individual difference analyses in Chapter 6, I make use of some of the structural MRI techniques mentioned above, namely VBM (see Section 2.2.6) and surface based analyses (see Section 2.2.7).

1.7 Summary and approach to present thesis

Achieving a complete scientific understanding of human consciousness remains an immensely challenging task. However, while this goal continues to appear very distant, substantial progress has been made in recent decades with help from functional imaging methodologies combined with creative manipulations in behavioural testing. A brief review of the existing literature in this field was included in Section 1.3 of this chapter. The vast majority of the work addressing the neurobiological mechanisms of consciousness has been conducted in the visual domain and the rationale for this alongside an overview of the structure of the human (and primate) cortical visual systems have been covered in Section 1.2. Since much of the work within this thesis will focus on the interactions between social visual processing (specifically of faces), responses to threat, and awareness, it has been of relevance to also undertake a review of knowledge regarding the theoretical underpinnings and neural mechanisms relevant to social face perception (Section 1.4) and behavioural responses to threat (Section 1.5). Finally, the relevance of inter-individual differences in measures of social perception and awareness, which will be explored further in the experimental chapters, has been introduced in Section 1.6.

As discussed in Section 1.3.5, a number of existing studies combining functional brain imaging (in particular fMRI) with bistable perception have shown that transitions between different perceptual experiences (in other words, changes in the contents of awareness) are associated with engagement of a group of frontoparietal cortical regions (e.g. Rees, 2007). Some of these regions may directly mediate access to awareness, although this remains under ongoing debate. The first half of this thesis addresses outstanding questions in this controversial field. I will attempt to anatomically refine our understanding of the neuronal network involved in perceptual transition (by exploring where perceptual-transition-related BOLD signal is invariant to type of bistable stimulus; Chapter 3) and then to ask more specific questions about whether effective connectivity strength between regions
involved in perceptual transition relates to behaviour (Chapter 4), or about the causal importance of specific regions in this network (namely, the frontal eye field; Chapter 5). Against the backdrop of the ongoing controversies regarding the role of this frontoparietal network in perceptual transitions, I will argue that my findings contribute significantly to the debate, and provide new detail about how these regions and their interplay influence access to awareness.

The second half of the thesis will focus on questions regarding the influence of social cues on non-conscious visual processing as well as on transition between non-conscious and conscious processing modes. The existing literature on emotional and social modulation of vision was reviewed in Section 1.4. However, there are important gaps in knowledge when it comes to understanding neural processing of social face traits, and in particular processing of social information that is not available to the observer’s awareness. Moreover, while some work on emotional modulation of awareness has been undertaken, there is minimal understanding of the influence of social cues (which may be emotionally neutral) in this context. To address these questions, I will explore behavioural responses to social cues contained in non-consciously perceived human faces and then ask how such responses relate to variations in focal brain anatomy (Chapter 6). Some unexpected findings presented in Chapter 6 (and also published in a related study; Stewart et al., 2012) resulted in the formulation of a new hypothesis, namely that an observation of delayed responses to non-conscious socially threatening cues represents a correlate of defensive freezing. This hypothesis will be tested in two further experiments where key predictions are underpinned by established knowledge regarding the neurobiology and psychophysiology of defensive responses (largely from rodent work but also in humans; Section 1.5). Specifically, in the context of viewing socially threatening faces, I will firstly attempt to verify the presence of autonomic and perceptual components of freezing in association with perceptual transitions in binocular rivalry (Chapter 7), and secondly look for the same components of freezing and their relationship to motor physiological changes in conditions of non-conscious perception during interocular suppression (Chapter 8).

At the core of the experimental work reported in Chapters 3-8 will be behavioural measures derived from visual paradigms of bistable perception in healthy human volunteers. Behaviour will be correlated with brain structure and function through concurrent use of structural and functional MRI methods (the latter performed both in the task and resting states and supplemented with connectivity analyses such as dynamic causal modelling). In addition I will use repetitive transcranial magnetic stimulation to explore causative effects of specific brain regions. Finally, psychophysiological and motor-physiological indices of defensive responses will be evaluated through measurement of
heart rate and pupillary responses, and through monitoring of TMS-induced motor-evoked potentials, respectively.
Chapter 2

General Methods

“A brain scan may reveal the neural signs of anxiety, but a Kokoschka painting, or a Schiele self-portrait, reveals what an anxiety state really feels like. Both perspectives are necessary if we are to fully grasp the nature of the mind, yet they are rarely brought together.”

Eric Kandel (1929 –)

2.1 Introduction

The bulk of the work presented in this thesis seeks to understand how brain structure and function relate to behaviour in healthy humans. This chapter provides a brief overview of the biophysical principles, hardware, image processing and analysis steps relevant to the neuroimaging methodologies (specifically structural MRI and fMRI) as well as stimulation techniques (specifically single pulse TMS and rTMS) of which I will make use in later experimental chapters. The development of these techniques over the previous decades has paved the way for dramatic progress in basic and clinical neuroscience. The ability to image the human brain (as well as other organs) non-invasively through MRI has come about through an enormous interdisciplinary scientific endeavour and has revolutionised medical science. No less impressive has been the advent of fMRI, enabling non-invasive evaluation of brain function in vivo, with profound impact on the development of cognitive and systems neuroscience (Logothetis, 2008). Processing and analysis of the rich and complex data produced by these imaging modalities has required the development of very diverse and sophisticated suites of software tools which have provided the basis for many tens of thousands of publications furthering the understanding of both structural and functional
localisation in the brain, and increasingly of both structural and functional connectivity and integration. Specific image processing and analysis methods used in this thesis, as implemented through SPM (Friston et al., 1994) and FreeSurfer (Dale et al., 1999), will be described in this chapter. In parallel with MRI, TMS has also become an indispensable investigative tool, finding use in all domains of cognitive neuroscience as well as increasingly being applied to therapeutic challenges ranging from treatment of depression to management of epilepsy and movement disorders (Bolognini and Ro, 2010; Walsh and Cowey, 2000). One of the key strengths of this methodology is its potential to complement functional imaging approaches, permitting studies of temporally precise causative effects with which to further enrich the spatially detailed correlational findings of fMRI.

This chapter provides a general methodological overview, with the final section detailing some general protocols for selection and recruitment of participants, behavioural testing, and imaging acquisition and preprocessing, which will be consistent across all experiments included in this thesis. Recurrently used bistable perceptual paradigms will also be described in this final section (see Section 1.3.3 for a general discussion of bistable perception). Where relevant, more specific methods for behavioural testing, MRI/fMRI acquisition and analysis, TMS administration, as well as recording and analysis of heart rate, eye tracking, pupillometry and MEPs will be covered in ‘Materials and methods’ sections within the corresponding experimental chapters.

### 2.2 MRI and fMRI

In its earliest incarnation, MRI was used to image two tubes of water using gradients in all three dimensions, creating images on the basis of the already known principles of nuclear magnetic resonance (Lauterbur, 1973). In the subsequent four-and-a-half decades the use of MRI has become extremely widespread in a variety of applications, but especially in the biomedical field. The modality allows the imaging of structures with high spatial resolution, and more importantly, non-invasively. Around 25 years ago, the finding that MRI could be used to measure local brain blood flow and oxygenation, and that this could provide a reliable (albeit indirect) reflection of neural function both in animals and humans (Kwong et al., 1992; Ogawa et al., 1992, 1990), led to the development of fMRI which has revolutionised systems and cognitive neuroscience (Logothetis, 2008).

Even after the technological and methodological feats of obtaining good quality MRI data have been accomplished, translating MRI or fMRI signal into statistically meaningful spatial patterns of structural variability or brain activation requires a large number of data
processing and analysis steps. There are several popular software packages that facilitate
the necessary workflows. The most widespread of these is Statistical Parametric Mapping
(Friston et al., 1994), a set of tools implemented through MATLAB, which have been
developed and are continually refined and updated within the Wellcome Trust Centre for
Neuroimaging. Throughout the experimental chapters described in this thesis I have made
use of mostly SPM8 and on some occasions SPM12 (in each case, the version used is
specified). In addition, the FreeSurfer suite of tools (Dale et al., 1999) was used for some
analyses in Chapter 6 and for displaying imaging results in some other chapters.

2.2.1 Fundamental principles of MRI and fMRI
The physical and physiological principles upon which MRI and fMRI are built are based on
several substantial and fundamental discoveries and insights in atomic physics, magnetic
resonance, image processing, biochemistry, vascular physiology and neurovascular
coupling. These developments have made MRI and fMRI possible and continue to shape
our understanding of the possibilities and limitations of these techniques (McRobbie et al.,
2006). These underlying principles and their relevance to generation and acquisition of MR
images are all described in more detail in the following five sections. There follows
discussion of how atomic nuclei behave in static and varying magnetic fields, how magnetic
resonance results in the release of radiofrequency energy that forms the basis of the MR
signal, how spatial information is encoded in this signal, how image acquisition can be
optimised depending on the target tissue, and how the MR signal relates to underlying
biochemical and neural processes.

2.2.1.1 Behaviour of atomic nuclei within magnetic fields
Protons are subatomic particles that have intrinsic angular momentum, known as spin.
Normally spins are randomly orientated. However, when certain atomic nuclei, in
particular hydrogen (which has a nucleus consisting of a single proton), are placed in a
strong static magnetic field ($B_0$) they align according to this field, much like the needles of
many tiny compasses. The nuclei can align with respect to the magnetic field in one of two
states: the slight majority align parallel to the magnetic field since this is the lower-energy
state; the remainder align anti-parallel. The nuclei resonate or ‘precess’ around the axis of
the external field at a frequency known as the Larmor frequency, which is described by the
following equation (where $\omega$ is the resonance frequency and $\gamma$ is the magnetogyratic ratio,
different for each type of atomic nucleus; see Figure 2.1):
The Larmor relationship is a fundamental principle of all nuclear magnetic resonance. Since in a static magnetic field more nuclei are aligned parallel to the field, there is a net magnetisation vector, which is parallel to $B_0$ and is especially apparent when there are a large number of hydrogen nuclei that are freely mobile in water (as in the human body, including the brain).

\[ \omega = \frac{\gamma}{2\pi} B_0 \]  

(2.1)

As shown by the Larmor equation (2.1), the size of the static magnetic field is a key determinant of the energy that can be acquired by particles placed within it, and thus also of the size of signal that can subsequently be produced. Accordingly, the field strength is a highly relevant property of each MRI scanner (the MRI and fMRI data presented in this thesis were collected using 1.5 and 3.0 Tesla scanners). Since hydrogen atoms have the largest magnetogyrionic ratio of all atoms (>42 MHz/T), they can produce the largest signal at any given field strength. The signal produced by nuclei in a static magnetic field (such a field is constantly present in an MRI scanner) can be detected by an appropriately tuned coil and receiver; however this would not be sufficient to produce an image since there would be no way of determining where different parts of the signal originated.
Spatial localisation in MRI is achieved through the intermittent application of additional smaller magnetic fields containing spatial gradients (gradient fields, $G$), which can be superimposed on $B_0$, causing variability in the strength of the magnetic field in a certain direction and thus allowing specification of spatial location within the signal (Figure 2.2; see Section 2.2.1.3 for more details).

Gradient fields can be added to the Larmor equation to specify that the position of each proton within the gradient field be encoded by resonance frequency differences:

$$\omega = (\gamma / 2\pi)B_0 + G$$  \hspace{1cm} (2.2)

**Figure 2.2:** The effects of a static magnetic field and an additional superimposed field gradient on protons and their spins

In this example the magnetic field strength is measured (and affected by the field gradient) along the $x$-axis; the distance between the horizontal lines in the upper plots signifies the local field strength. In (a) the field strength is uniform (static field) along the whole $x$-axis and all the protons therefore experience the same field and have the same frequency. In (b) the addition of a gradient field along the $x$-axis means that the field strength is variable and so the frequency of protons depends on where in the magnetic field they are found, allowing them to be localised according to their frequency. Image adapted from McRobbie et al. (2006).
2.2.1.2 Generating a signal: magnetic resonance and relaxation

As discussed in Section 2.2.1.1, when nuclei are found in a static magnetic field ($B_0$), most of their protons have spins aligned parallel with this field, and are in a lower-energy state, but some have spins that are anti-parallel and are in a higher-energy state. If a pulse of electromagnetic energy (RF pulse) is applied perpendicular to $B_0$ any of the lower-energy nuclei that have a Larmor frequency matching the frequency of the pulse can absorb energy and move into the higher-energy state (excitation). This causes the net magnetisation vector to tip away from being parallel to $B_0$ (as it is in a static field; see Section 2.2.1.1) and into the transverse plain. Longitudinal magnetisation is thus converted to transverse magnetisation; this can only occur if the RF pulse frequency equals the Larmor frequency, which is why the technique is called magnetic resonance. A large enough RF pulse at the correct frequency can flip the net magnetisation by 90° (this is known as the flip angle of the RF pulse). When the transient magnetic field is switched off, the nuclei emit radio waves as they return to their steady state condition (relaxation). The release of RF energy during relaxation can be detected by an RF receiver coil and forms the basis of the MRI signal.

Nuclear relaxation can be subdivided into two types of process. Longitudinal relaxation describes the return of the proton spins to thermal equilibrium, as they revert to the lower energy state, according to a time constant, $T_1$. During this process the energy absorbed from the RF pulse dissipates into surrounding tissues as heat and the net magnetisation vector along $B_0$ gradually grows again. Longitudinal relaxation is relatively slow ($T_1$ is typically around 1 second) but varies according to the type of tissue. Transverse relaxation describes changes in precession rates resultant from protons spinning out of phase during relaxation with associated interaction of their magnetic fields (spin-spin interaction). This leads to fairly rapid decay of the recently generated transverse magnetisation according to another time constant, $T_2$. Typically $T_2$ is around 100ms (an order of magnitude shorter than $T_1$) but as for $T_1$ this is dependent on tissue type. One important difference is that $T_1$ is affected by the strength of the static magnetic field, whereas $T_2$ is not. The fact that both $T_1$ and $T_2$ vary according to tissue type enables differential definition of tissues in the image (Section 2.2.1.4).

2.2.1.3 Spatial localisation within the MR signal

In order to construct an image from the MR signal, spatial information needs to be extracted from it. This is achieved through the use of transient spatial gradient fields (explained in Section 2.2.1.1), which are added to the static magnetic field. In MRI scanners the gradient fields are usually generated by three electromagnets in orthogonal plains: $G_x$, $G_y$, and $G_z$. 

a linear gradient along the z-axis (superior to inferior); $G_x$, a linear gradient along the x-axis (left to right); $G_y$, a linear gradient along the y-axis (anterior to posterior). Gradient fields are normally divided into discrete steps that, when the three dimensions are combined, form cubes of space (or voxels). A voxel formed by the smallest available step in the gradient field for each dimension determines the image resolution (since the signal from all protons within a voxel is combined in the reconstructed image). Together $G_x$, $G_y$, and $G_z$ can determine a unique point $(x, y, z)$ in three-dimensional voxel space. Thus, gradient coils are capable of finely tuned adjustments in field strength and direction that determine the speed and resolution of image acquisition. Gradients are applied for short periods of time only (as pulses) and produce the characteristic rhythmic noise made by MRI scanners.

There are three principal ways in which the gradient fields are used to encode spatial information. First, by combining the RF excitation pulse with a gradient field (the slice-select gradient), magnetic resonance can be restricted to a two-dimensional plane, where the Larmor frequency matches the frequency of the RF pulse. This means that only protons in this 2D plane will become excited (and subsequently release signal), forming the basis of acquisition of signal in slices. The slice select gradient is most often in the z orientation, although gradients in any of the x, y or z orientations (or any combination of these) can be used, allowing generation of coronal, sagittal, axial, or oblique slices. Two other steps allow further spatial information to be extracted from the 2D slice. Phase encoding is applied to the ‘columns’ of the slice (usually the y plane) and causes changes in the rate of spin of protons in that column. The change in spin is position-dependent, varying with distance of the protons along the column. Frequency encoding (performed in a perpendicular plane to phase encoding and thus applied to the ‘rows’ of the slice, normally in the x plane) allows measurement of the MR signal at different time points during the phase encoding gradient pulse. The combination of these two strategies enables the recording of RF waves where spatial information in the plane of a slice is encoded through varying frequencies and phases (at a resolution of individual voxels). The phase and frequency data are written in a more simple and application-friendly way into a temporary virtual space known as k-space, which is equivalent to a Fourier plane where each point is a spatial frequency component, and is covariant with physical space. Data from k-space is reconstructed into an image using a 2D inverse Fourier transform.

### 2.2.1.4 MR image acquisition: contrast and echoes

MR images need to differentiate tissue types based on a contrast mechanism that reflects anatomical differences. As mentioned in Section 2.2.1.2, relaxation time constants differ between different tissues and this is a principal basis for contrast differentiation. When the
MR signal is sampled at a time point prior to full relaxation its intensity will thus differ according to type of tissue and the timing of this signal readout can be adjusted to determine whether the contrast is primarily dependent on differences in T1, T2 or overall proton density.

Selection of certain image acquisition parameters can thus enable contrast optimisation based on relaxation time differences between tissues. These parameters include TR (the repetition time between two consecutive RF pulses) and TE (the time between an RF pulse and measurement of the signal; see below for further detail). For maximal differentiation of tissues based on T1 relaxation, which clearly highlights the differences between grey and white matter, relatively shorter TR and TE are needed. Specifically, when TR is shorter than the T1 of a tissue, the tissue will appear darker on the image (since longitudinal magnetisation does not have a chance to recover between RF pulses). For differentiation based on T2 relaxation, longer TE and TR are needed; specifically, if TE is longer than a region's T2 relaxation time some of the signal will be lost and the tissue will appear darker on the resulting image.

The observed decay in MR signal when imaging biological tissues is faster than T2 relaxation would theoretically predict. This is because the magnetic fields are not homogenous, partly due to differences in tissue composition. The time constant T2* describes MR signal decay that takes into account these field imperfections. Such imperfections importantly include magnetic susceptibility variations in blood vessels, which can fluctuate over time; these changes are particularly relevant to fMRI contrast. Thus T2* relaxation can be used to detect changes in relaxation times at the same anatomical site across different conditions (that can be relevant to brain function). T1-weighted scans form the basis of all structural images used in this thesis, whereas T2* contrast is used in the fMRI sequences described.

Echoes are additional RF pulses that allow collection of extra MR signal and enable further contrast optimisation. In spin echo an RF pulse (at 180°) follows the initial RF pulse and acts to reverse dephasing caused by B0 field inhomogeneities (while having no effect on T2, which is related to spin-spin relaxation). This results in emission of an RF echo after twice the time delay between the first and second RF pulse. The time between the initial RF pulse and the echo is known as the echo time or TE and the difference in T2-related decay at TE determines the contrast of the image. Gradient echo sequences use a combination of an RF pulse and a gradient field followed by a reverse gradient to deliberately introduce dephasing and subsequent rephrasing of the spins. An echo of the signal is emitted at TE and in this case images depend on T2* contrast.
Echo-planar imaging is a method for acquiring brain imaging data very rapidly, at the expense of spatial resolution and increased proneness to susceptibility artefacts. Briefly, additional RF pulses are delivered at 180° to the initial RF pulse and these additional pulses ‘refocus’ the signal by resetting the precession of the nuclei in the imaged tissues. This produces ‘echoes’ allowing very rapid acquisition or images based on T2* contrast. Such rapid acquisition means that all the lines of k-space can be encoded in a single pulse and a complete slice can thus be acquired in less than 100ms. In addition, the T2* contrast is quite sensitive to BOLD signal (see Section 2.2.1.5). The properties of these sequences thus make them especially suitable for measuring brain function, which varies over time. All the fMRI experiments in this thesis are carried out using EPI sequences.

2.2.1.5 Biochemical and neural basis of the BOLD signal

Blood-oxygen-level dependent signal is the measure most widely used in fMRI to probe spatially specific patterns of neural activity. However, the relationship between BOLD and physiological changes in neurons is complex and indirect (Heeger and Ress, 2002; Logothetis, 2008). It is essential to bear this in mind when attempting to interpret the findings from fMRI experiments in terms of underlying neuronal mechanisms. The key considerations that allow an understanding of the nature of the BOLD signal include firstly how this signal is generated by changes in the oxygenation state of haemoglobin, secondly how vascular physiology is linked to regional changes in neural activity, and finally what exactly BOLD can and cannot tell us about neural activity. Each of these questions in turn is briefly discussed below.

As indicated by its name, BOLD is a signal that varies according to blood oxygenation. Oxygen is transported to the tissues bound to haemoglobin (a molecule contained within red blood cells) and is required for energy metabolism, which in the brain is largely determined by neural activity. Accordingly blood flow is linked to neural activity so that there is redistribution of blood to more active brain regions. While the mechanisms for such neurovascular coupling remain incompletely understood (Heeger and Ress, 2002), the principles are fundamental to brain function and as far as imaging is concerned mean that neuronal activity in the brain can be indirectly determined through measurements of tissue perfusion, blood volume or blood oxygenation, such as BOLD (Logothetis, 2008).

When not bound to oxygen haemoglobin is paramagnetic (slightly enhancing local magnetic field, reflected as reduced transverse relaxation and consequently reduced T2 and T2*-weighted MRI signal). However, this situation is altered by oxygenation of haemoglobin, which reconfigures the molecule and it becomes diamagnetic (slightly reducing local magnetic field; Pauling and Coryell, 1936). These oxygenation-dependent
magnetic properties of haemoglobin mean that local differences in haemoglobin concentration and blood oxygenation (which are assumed to be tightly coupled to local neuronal activity) are reflected in T2* contrast changes measurable with fMRI. Thus, deoxygenated blood produces a reduced BOLD signal compared to oxygenated blood, and local increases in BOLD can be used as an indirect measure of locally increased neuronal activity. It has been known for more than 25 years that fMRI can be reliably used to measure these signal and contrast differences (Ogawa et al., 1992, 1990).

The temporal evolution of the BOLD response to transient neuronal activity is described by the haemodynamic response function (Figure 2.3), which is composed of three key elements. First there is a small decrease in signal below baseline related to neuronal activity-induced increases in metabolic demand, which result in increased oxygen consumption and thus a transient local increase of deoxygenated haemoglobin (Malonek et al., 1997). Second, there is a large increase in signal (starting around 1-2 seconds later) resulting from reactive vasodilatation and a large increase in oxygenated haemoglobin, over-compensating for local oxygen consumption. This results in an activity peak at around 6 seconds after onset of neuronal activity, which is the most identifiable peak in BOLD (Fox and Raichle, 1986). Third, there is an undershoot of signal back below baseline lasting a number of seconds, where there is re-regulation of blood flow, with reduction in the supply of oxygenated blood. The relatively slow time course of the HRF and BOLD response underpin the limited temporal resolution of fMRI. In particular, the peak in BOLD signal occurs 4-6 seconds after the onset of the neuronal activity of interest.
Exactly how neuronal activity relates to blood flow changes and therefore to the BOLD signal is a complex question, which remains under active debate and investigation. It seems that BOLD is less related to action potentials and more related to synaptic signalling (Heeger and Ress, 2002). In a landmark study where BOLD acquisition and direct recordings of neuronal spiking activity and local field potentials were performed simultaneously in primates, it was found that the latter physiological measure was more closely correlated with BOLD (Logothetis et al., 2001). The interpretation of these findings was that BOLD reflects the input and intracortical processing within a region rather than its spiking output. There are a number of other issues, including the fact that BOLD does not differentiate well between excitatory and inhibitory interneuronal signals or between bottom-up and top-down signals. In addition, haemodynamic responses are sensitive to the size of an activated neuronal population but less so to the density of activation within this region (Logothetis, 2008). Notwithstanding the enormous utility of fMRI for studying neuronal processes in vivo, it is therefore of great importance to bear in mind these limitations and to remember that BOLD is a rather indirect and incomplete reflection of neuronal function, as well as being somewhat temporally imprecise.
2.2.2 fMRI preprocessing

The raw fMRI time series data contain significant noise from various sources, some of which can be systematic and predictable (e.g. head motion). Preprocessing is concerned with accounting for these sources of noise and unwanted variance before fitting a statistical model to the data. Briefly, the general principles of preprocessing are to correct the signal for inhomogeneities in the magnetic field, to spatially transform and align each individual’s fMRI data, to warp the data so that they fit into a common anatomical framework, and finally to smooth them to reduce high-spatial-frequency noise. For the fMRI experiments contained within this thesis, preprocessing was undertaken in SPM8. The relevant steps, as performed in this version of SPM, are described in the following seven subsections, while any more specific aspects of the methodology are described in Section 2.4.8 or in the relevant experimental chapters. Figure 2.4 provides an overview of preprocessing in SPM.

![Preprocessing overview](image)

**Figure 2.4:** Summary of SPM preprocessing steps for fMRI. Image obtained from teaching slides for SPM course 2016.
2.2.2.1 Bias correction

As described in Section 2.2.1, when biological tissue (including brain tissue) is placed in a magnetic field and subjected to RF pulses, MR signals are produced. In an MRI scanner these signals are detected by a receiver head coil. While multichannel head coils allow improved signal-to-noise ratio and spatial resolution in the images (Parikh et al., 2011), MRI data acquired with a 32 channel head coil suffer from strong intensity inhomogeneities, which can lead to problems with SPM preprocessing. In particular, there is a smooth spatially varying artefact that modulates the intensity of images (bias). The purpose of bias correction is to “flatten” the intensity profiles across the acquired images. The heterogeneity of the images is estimated based on the first image of the time series using image segmentation (see Section 2.2.2.5). An inverted version of this estimated heterogeneity is then applied to the remaining images in the time series. Image intensities are changed by a simple multiplicative factor so patterns of signal within the images are unaltered.

2.2.2.2 Field maps to correct for $B_0$ field distortions

The $B_0$ magnetic field can be inhomogeneous and suffer from dropout or geometric signal distortion, particularly at boundaries between different tissue types (such as the sinuses). Geometric distortions can be measured and therefore accounted for by acquiring additional data for a field map (via some short additional imaging sequences while participants are in the scanner). Field map correction can then be implemented through a dedicated SPM toolbox (Hutton et al., 2002). Briefly, two images with different gradient echo weightings (different TE parameters; see Section 2.2.1.4) are acquired and the phase difference between them is used to estimate a voxel displacement map (this shows the magnitude of deviation in magnetic field at each voxel, which is proportional to the extent of distortion). This map is then applied to the EPI data to correct for distortions as well as movement-by-inhomogeneity interactions (see Section 2.2.2.3).

2.2.2.3 Realignment and unwarping

Image time series need to be realigned to a common reference frame. This allows correction for head movement, which even if slight (and minimised through participant cooperation and the use of foam padding in the scanner), can cause significant shifts between scans. For example, for tissues at the grey/white matter boundary the tissue type falling within a voxel could change from one scan to the next, adversely affecting data quality. To counteract this each scan in the time series can be realigned with a reference scan (often the first scan in the series) using a six-parameter rigid-body transformation.
The six parameters comprise both translation and rotation vectors for all three axes of space. Realignment does not tend to remove all movement-related variance from the data since movement can also place the same brain structure in different parts of the inhomogeneous magnetic field at different time points. Unwarping makes use of the field map data (see Section 2.2.2.2) to generate a forward model of these movement-by-inhomogeneity interactions and account for them. Additionally, movement parameters from realignment can be included as covariates in the general linear model during statistical analysis of the data (see Section 2.2.3.1). However, this will only be helpful in removing unwanted variance if movements are not correlated with the behavioural task under investigation. The realignment step also creates a mean EPI image, which is then used for co-registration (see Section 2.2.2.4).

2.2.2.4 Co-registration to a T1 structural image

This step uses a rigid-body transformation as in realignment (Section 2.2.2.3), although the algorithms used differ. The source image is a T1 structural image and the reference image is the mean EPI image created during realignment. Interpolation over the original borders of voxels is carried out where needed and the mismatch between the source and reference images is established, iteratively calculating transformation parameters to reduce mismatch until it is minimised. Thus the lower-resolution EPI image (voxels typically around 3mm) can be co-registered with a higher-resolution T1 structural (voxels typically around 1mm).

2.2.2.5 Segmentation

Segmentation is discussed in more detail for preprocessing of structural MRI data (Section 2.2.5.1). In fMRI preprocessing the mapping of grey and white matter images onto template tissue probability maps produces spatial normalisation parameters that can be applied to both the structural and functional images for the spatial normalisation step (Section 2.2.2.6).

2.2.2.6 Spatial normalisation

Spatial normalisation is important to enable anatomical interpretation of results to be standardised; signal can then be combined across participants and anatomical localisation can be compared across different studies using standard three-dimensional coordinates. Images are warped and transformed into a standard space (this can be done with the aid of the normalisation parameters obtained during segmentation as described in Section 2.2.2.5, or using a Bayesian framework to estimate the optimal warp). The MNI template (a
template scan that is the average of more than 150 normal MRI scans, created as part of the P-20 project at the Montreal Neurological Institute; Mazziotta et al., 2001) is used for normalisation and the location of each voxel is defined in three coordinates from an origin \((x = 0, y = 0, z = 0)\) at the anterior commissure. Position in the sagittal plane defines the \(x\) coordinate (medial to lateral with right being positive); position in the coronal plane defines the \(y\) coordinate (anterior to posterior with anterior being positive); and position in the axial plane defines the \(z\) coordinate (dorsal to ventral with dorsal being positive).

2.2.2.7 Spatial smoothing

The final standard preprocessing step is to convolve each EPI volume with a Gaussian-shaped kernel. This removes high-spatial-frequency noise, providing a weighted average of the local signal, which improves signal-to-noise ratio (since local averaging helps to cancel out noise) as well as reducing small-scale variability in localisation of activations across participants. In addition, smoothing is important for subsequent statistical analysis since it reduces the number of resolution elements (helping to partially alleviate the multiple comparisons problem; Section 2.2.3.2) and it helps ensure that error is normally distributed which is an important assumption for statistical parametric tests. FWHM width of the kernel used determines the extent of spatial blurring.

2.2.3 fMRI analysis and statistics

Once the acquired MR images have been adequately noise-corrected through preprocessing, statistical analysis is necessary to define voxels or clusters of voxels containing significant BOLD activation during specific experimental conditions (often when contrasting two experimental conditions). SPM makes use of a mass-univariate approach, performing a separate statistical test (based on the General Linear Model) at each voxel. The statistical results (parameters) are displayed in a voxel-wise fashion as a three-dimensional image (a statistical parametric map) which highlights significantly activated voxels. Given the typical SPM contains tens of thousands of voxels, correction for multiple statistical comparisons, which within SPM is based on Gaussian random field theory, is essential before interpreting results. An overview of these steps for statistical analysis of fMRI data, alongside the pre-processing steps already described in Section 2.2.2, is provided in Figure 2.5 below. The statistical analysis steps are then described in more detail in the following two subsections.
2.2.3.1 The General Linear Model

The GLM is a statistical framework, which involves modelling what one would expect to see in the data and then testing how well this fits with the observed findings. The framework incorporates a number of statistical models including ANOVA, linear regression and t-tests. Within SPM, the GLM is usually applied to imaging data in a univariate fashion (separately for the responses within each voxel). For fMRI analysis the model can be described by an equation that shows how \( Y \) (a vector containing a time series of BOLD signal measurements) can be determined from a linear combination of predictor variables contained in the design matrix, \( X \) (one row per observation and one column per model parameter); a vector containing the parameters to be estimated, \( \beta \); and a vector of error terms, \( \varepsilon \) (Friston et al., 1994):
The parameter estimates are the values that $X$ must be multiplied by to fit the data; in other words, they reflect the contribution of $X$ to the data. $X$ is a matrix containing vectors representing experimental manipulations, confounds, and/or covariates of no interest. Each vector (or regressor) is entered as a separate column. Regressors of interest, related to experimental events, are created by placing a stimulus (stick or boxcar) function at time points that correspond with the effect of interest. These are then convolved with the HRF (to account for the delayed and dispersed nature of the BOLD response; see Section 2.2.1.5) before being entered into the design matrix (Figure 2.6, left panel). In order to account for random low-frequency fluctuations in the fMRI data, a set of regressors capturing these low-frequency components is also entered into the GLM. This ensures that the model fits the data even when there is low-frequency drift (see Figure 2.6, right panel). The beta parameters for each voxel are estimated using multiple linear regressions to minimise the sum of squares of the differences between observed data and predictions from the model (restricted maximum likelihood algorithm; ReML).

\[ Y = \beta X + \varepsilon \]  

\hspace{10cm} (2.3)

**Figure 2.6:** Accounting for frequency drift in fMRI time series

The left hand panel shows an example fMRI time series with modelling of changes in experimental stimulus (red), stimulus changes after convolution with HRF (green) and observed time series data from a single voxel (blue). The right hand panel shows how low frequency drift must be accounted for in the GLM (green) as otherwise the model does not fit the data (red). Image obtained from teaching slides for SPM course 2016.
2.2.3.2 Statistical inference

The variance (uncertainty) in the estimation of each of the parameters from the $\beta$ vector is calculated and two types of statistical test are then possible. With an $F$-statistic test the null hypothesis is that the parameter estimates are zero. On the other hand a $T$-statistic approach tests whether some linear combination of estimates is significantly different from zero. The $T$-statistic is therefore a directional test, whereas the $F$-statistic is not. A high $T$ value suggests that the parameter estimate is strong relative to its uncertainty and thus indicates a strong contribution of that regressor to the data, relative to other regressors. $F$- or $T$-statistics are applied at each voxel producing an image of statistics covering the whole brain volume (a statistical parametric map). The SPM can be shown at a particular threshold to display which voxels are activated by a particular contrast at the chosen level of significance.

Since the mass-univariate approach involves thousands of separate statistical tests across the whole brain volume the threshold for statistical significance needs to be adjusted for these multiple comparisons in order to control for type I statistical error. Clearly with this many tests, classical family-wise error control with Bonferroni correction would come with an unacceptably high cost of loss of sensitivity and introduction of type II statistical error. Gaussian random field theory (Friston et al., 1994) can instead be employed to alleviate this problem by accounting for the spatial smoothness of the SPM. There is a strong degree of spatial correlation (in part ensured by smoothing; see Section 2.2.2.7), which allows estimation of how many clusters of voxels covary (known as resolution elements or resels). False positives are controlled at the level of these clusters, which are both much less numerous and usually much more functionally relevant than individual voxels. Inference can then be made at the cluster level (testing whether the number of activated voxels in a cluster is greater than expected by chance; Worsley, 2007). Alternatively, peak voxel-level inference tests only at the peak value in each cluster. Voxel-level inference permits better spatial localisation (Poline et al., 1997) while cluster-level inference is more sensitive to spatially extended activations that may not contain a clear peak. A combination of these two types of statistical inference (but more often peak-level inference) has been used during statistical analysis of fMRI and structural MRI data within this thesis.

Whereas the above considerations apply to multiple comparisons across the whole brain volume, another approach, which has also been used extensively in this thesis, is to restrict the volume of analysis to a predefined volume of interest usually based on previous findings or on specific a priori predictions. VOIs are usually defined either anatomically, by using a ‘mask’ circumscribing a previously reported cluster of activation, or by constructing a 3D volume around a previously reported peak activation. This small-volume correction
approach can be used as an alternative to performing correction for multiple comparisons across the whole brain volume as long as the small volume to be used was determined *a priori* based on a clear rationale rather than being selected *post hoc*. Since fewer multiple statistical comparisons need to be made as part of small-volume correction (compared to whole-brain volume correction) this method is less statistically conservative.

The procedures described so far in this section lead to the computation of *first-level* SPMs (performed at the individual participant level and permitting inference relating to *fixed effects* or intra-individual error variance). Spatial normalisation (see Section 2.2.2.6) enables *second-level* analysis (performed at the group level and permitting inference relating to *random effects* or both intra- and inter-individual variance). The possibility of performing second-level analysis is clearly important for the majority of fMRI studies where groups of participants are tested. In such analysis a *t*-test is performed on the first-level contrast estimates, essentially treating these statistics related to individual participants as new response variable vectors in the second-level (group analysis) GLM. Random effects analysis assumes that first-level parameter estimates are normally distributed in the population, allowing for relatively easy estimation of both within- and between-participant variance.

### 2.2.4 Resting-state fMRI analysis

While the majority of fMRI studies in systems neuroscience have sought to understand neural processing related to certain stimuli or tasks, a great deal of information is also contained in spontaneous non-task-related fluctuations in the BOLD signal. Studies employing such a resting-state fMRI approach were first carried out early in the fMRI era (Biswal et al., 1995) but interest in this field has grown greatly over the past couple of decades, with increased understanding that spontaneous activity does not constitute noise but rather reflects specific intrinsic functional brain organisation (Fox and Raichle, 2007).

By definition, resting-state fMRI studies are interested in spontaneous neural activity and thus seek to minimise changes in sensory input and performance of any action or task by the participant. Normally, continuous resting state conditions are employed throughout MRI scanning, such as maintaining fixation on a cross or rest with the eyes closed (Fox and Raichle, 2007). It is also preferable that there are no unplanned changes in state of alertness (for example participants falling asleep) during the scan. Experimental paradigms for resting-state fMRI are therefore straightforward. On the other hand, there is undoubtedly substantial variability in the internal mental processes as well as level of alertness when participants lie with their eyes closed for prolonged periods and these are
important caveats when interpreting results from resting-state experiments (Cole et al., 2010).

Once resting-state fMRI data have been acquired, there are important preprocessing and analysis-related issues specific to this setting. Here intrinsic (brain-based) resting-state BOLD fluctuations are the measure of interest and need to be distinguished from external influences on the BOLD signal such as inhomogeneity in the static magnetic field, head movement or breathing, which make a significant contribution to observed spontaneous signal fluctuations (Birn et al., 2006; Glover et al., 2000; Zarahn et al., 1997; see also Section 2.2.2). There are established ways to tackle this problem. Use of a high sampling rate helps reduce non-neuronal noise (e.g. Cordes et al., 2001) but the popularity of this method has been limited due to the reduced spatial coverage that is possible. Another option is to measure physiological parameters during BOLD acquisition and then remove their effects from the data by linear regression (e.g. Birn et al., 2006). This can be done in SPM by adding variables related to cardiac or respiratory fluctuations to the GLM design matrix as regressors of no interest. This was the approach used in the resting-state fMRI experiment included in this thesis (Chapter 4). A third option is to remove noise sources from the BOLD data itself by using independent components analysis to regress out signal that is common to all voxels (e.g. Macey et al., 2004).

Since one of the principal aims in resting-state fMRI experiments is to understand the intrinsic functional architecture of the brain, a second signal-analysis-related issue is how to best detect patterns of coherent or inter-related BOLD activity (van den Heuvel and Hulshoff Pol, 2010). Such functional connectivity analyses seek to define brain networks through temporal dependency of neuronal activation in anatomically separated brain regions. Broadly, this can be done in a model-dependent fashion (generally by choosing a seed region and then seeking to correlate the resting-state BOLD time series in this region against those of all regions in the brain), or by model-free methods, which are not constrained by the need for a priori definition of a seed region. More specific advantages and disadvantages of the number of available methods will not be reviewed here but have been extensively discussed elsewhere (Cole et al., 2010; van den Heuvel and Hulshoff Pol, 2010). Overall, a number of resting-state networks have been robustly outlined by these methods and are found consistently across different populations and resting-state fMRI analysis methods (van den Heuvel and Hulshoff Pol, 2010; see Figure 2.7).
In addition to the replicable findings of large-scale resting-state functional connectivity networks, it has been shown repeatedly that regions that interact in the task state are likely to be functionally connected in the resting state (van den Heuvel and Hulshoff Pol, 2010). Thus, resting-state functional connectivity approaches are revealing a great deal about the functional architecture of the brain that is also relevant to task performance. One disadvantage of functional connectivity analyses is that the direction or sequence of interactions among regions with coordinated activity cannot be established. This more detailed level of understanding of network function can be achieved through effective connectivity analyses (see Section 2.2.5 and Section 4.1). This kind of approach has only been applied to resting-state fMRI and/or to understanding the mechanisms of bistable perception and access to awareness relatively recently (discussed in Sections 4.1 and 4.4). One of the experiments in this thesis makes use effective connectivity analysis of resting-state fMRI data to probe the mechanisms of perceptual transition (Chapter 4).
2.2.5 Dynamic causal modelling

Evaluation of connectivity between different regions in the brain is a core approach in systems and imaging neuroscience (Rubinov and Sporns, 2010). Such analyses can be classified into three principal categories: those that focus on structural connections (anatomical connectivity), those that evaluate statistical dependencies between regional brain activity (functional connectivity, see also Section 2.2.4) and those that interrogate causal interactions between regions (effective connectivity; Sporns, 2007). Each of these different modes of brain connectivity has distinct characteristics (for example, structural connectivity only varies over long timescales whereas functional connectivity can vary on a second-by-second basis). At the same time, they can provide complimentary information and there is evidence for overlap between structural connections and functional/effective connectivity patterns (perhaps because the latter may to some extent be constrained by existing structural connectivity; e.g. Honey et al., 2007; see also discussion in Section 4.1). Dynamic causal modelling is a method for inferring directional effective connectivity between brain regions that involves Bayesian inversion of generative dynamical systems models (Friston et al., 2003). The DCM framework has two main components: biophysical modelling and statistical data analysis. In the case of fMRI, DCM treats the brain as a dynamic input-state-output system, with the inputs being experimental conditions, and the outputs being BOLD responses. Each brain region has a hidden neuronal state as well as hidden haemodynamic states. Neuronal state equations describe reciprocal influences between brain regions in terms of neuronal activity. Haemodynamic state equations allow transformation of neuronal activity in each region to the observed BOLD response. The earliest, and most widely utilised, implementations of DCM assume deterministic neuronal processes (influenced by known inputs) and ignore random fluctuations in neuronal states (Friston et al., 2003). Such a framework is well suited for assessing how perturbations from an experimental stimulus or task alter the effective connectivity of a studied network. However, deterministic analyses are less suited to experimental conditions where there is no stimulus input and furthermore they do not properly account for hidden endogenous fluctuations in physiological states. Conversely, a more recent stochastic DCM approach additionally accounts for endogenous fluctuations, or ‘neural noise’, in unobserved (hidden) neuronal states (Daunizeau et al., 2012). This means that stochastic DCM can be used to study effective connectivity in the resting state by accounting for spontaneous fluctuations in activity (Li et al., 2011). Stochastic DCM is a suitable model for resting-state data since there is no requirement for an input to the model (such inputs usually correspond to experimental manipulations that are not present in resting state experiments). On the other hand, accounting for random neural noise comes at significant computational cost as well as being associated with difficulties in the haemodynamic modelling necessary to apply stochastic DCM to fMRI (Daunizeau et al., 2012). More recently, a newer version of DCM
for resting-state fMRI data, based on spectral analysis, has been validated and shown to have some potential advantages over stochastic DCM, although relatively few experimental data analyses have been published to date using this technique (Razi et al., 2017, 2015). Different types of DCM are also discussed briefly in Section 4.1.

A crucial aspect of the DCM approach is to empirically determine and select the model (in other words, the architecture of effective connectivity among the regions of interest included in the analysis) that best reflects the observed data. This has traditionally been achieved through a hypothesis-driven Bayesian model selection approach where the evidence for different possible models is compared and the most likely model (that with the strongest evidence) among these competing hypotheses can then be selected (Marreiros et al., 2010; Stephan et al., 2010). Subsequently an alternative post hoc model selection method has been developed. This procedure involves inverting a fully connected model (assuming bidirectional connections between all ROIs) and then searching over all possible reduced versions of this full model before selecting the best (Friston and Penny, 2011; method described in more detail in Section 4.2.7). The post hoc method provides a much faster and less computationally demanding alternative to the hypothesis-driven variational free energy approaches used in Bayesian model selection, with the added advantage of allowing comparison between a large number of possible models, rather than having to pre-specify a small number of specific models for comparison (Friston and Penny, 2011). Simulations using both synthetic and real fMRI data show that (at least for bilinear deterministic DCMs) there is very good agreement between model evidences obtained using these two methods (Rosa et al., 2012).

2.2.6 Voxel-based morphometry

Voxel-based morphometry focuses on analysis of structural MRI data in order to identify regional differences in structure (e.g. brain tissue density or volume) in a voxel-wise manner, in an analogous way to the identification of differences in functionally relevant BOLD signal by fMRI. VBM has been used to highlight structural differences between groups (e.g. between schizophrenic patients and healthy controls; Honea et al., 2005), between different time points (e.g. before and after learning a skill, Draganski et al., 2004; or at different stages of adolescent development, Crews et al., 2007), or to define focal regions where brain structure co-varies with particular cognitive or behavioural measures (reviewed in Kanai and Rees, 2011). Increasingly, it has become apparent that individual performance of a large number of cognitive and perceptual tasks, as well as various personality traits, are correlated with local brain structure (Kanai and Rees, 2011; see also Section 1.6). VBM analysis can be performed within SPM using similar steps to those described for fMRI
analysis (see Section 2.2.3). The key differences here are that the structural image is the only image type (there are no EPI or fieldmap sequences); the imaging data itself is not task related since it contains structural measures only; and there is a single image rather than a time series. The standard analysis steps, described in the following sections, are analogous to a subset of the fMRI preprocessing steps (see Figure 2.4) and include segmentation to isolate different types of brain tissue (e.g. grey matter and white matter), warping of images into a common anatomical space, smoothing and finally measuring the tissue volume per voxel in the spatially normalised image. The amount of warping that has occurred during normalisation (related to how locally different an individual’s brain is to the anatomical template) will be reflected in the value assigned to that voxel.

2.2.6.1 Image segmentation

SPM8 makes use of unified segmentation where each voxel is assigned to one of four categories: GM, WM, CSF or other. A single probabilistic model allows for: 1) correction of image intensity inhomogeneity using a discrete cosine transformation; 2) mapping of each voxel into one of the tissue classes using a ‘mixture of Gaussians’ model which allows voxel intensities in the image to be interpreted as probabilistic evidence that the voxel belongs to a certain tissue class and then combining this with priors form tissue probability maps; and finally 3) nonlinear warping of images using transformation coefficient estimation so that they are coregistered with the tissue probability maps (Ashburner and Friston, 2005). This single streamlined model avoids experimenter decision points at each step of the process and thus minimises biasing of results.

2.2.6.2 DARTEL registration

DARTEL is an elaborate, complex and accurate registration algorithm, which is beneficial for VBM where localisation of regional volumetric differences needs highly accurate warps (Ashburner, 2007). Unlike the ~ 1000 parameters employed during warping as part of unified segmentation (Sections 2.2.2.5 and 2.2.6.1), DARTEL uses around 6 million parameters; this has been empirically shown to achieve much more accurate registration than other methods (Klein et al., 2009). DARTEL works through iteration between an average template of all maps and warping of individual images to this average, aiming to gradually minimise differences between the images. Where large deformations are needed, these are decomposed into groups of smaller deformations so that each individual deformation contains a one-to-one symmetry between the warped image and the target image. The DARTEL procedure creates a flow field, which describes how individual GM and WM tissue probability maps need to be warped to best match the average shape of the
template. A further spatial transformation is required to ensure that the template image is in MNI space, allowing standardised reporting of results and comparisons between studies, as for fMRI (see also Section 2.2.2.6). In order to retain information about how much deformation has been needed to achieve accurate registration, warped images are multiplied by the Jacobian determinants of the deformation. Therefore for structural regions that are unusual (and have thus required large deformations for registration) the local signal intensity will be modulated to a greater extent to account for this. Hence, VBM results reflect the original tissue volume per unit volume (voxel) of spatially normalised image.

2.2.6.3 Smoothing
As for fMRI data analysis (Section 2.2.2.7), images must be smoothed to remove residual inhomogeneities and registration inaccuracies as well as to help circumvent the problem of multiple statistical comparisons (Section 2.2.3.2).

2.2.6.4 Statistical analysis
Once VBM data have been preprocessed as described above (Sections 2.2.6.1 – 2.2.6.3), the aim is to construct maps of voxel intensity statistics that reflect regional variation in GM volume. This is initially carried out on a participant-by-participant basis and then individual measures of GM volume can either be averaged and compared between groups or correlated with participant-specific traits or behaviours. Statistical analysis in VBM is performed inside the GLM framework, and GRF theory is used to correct for multiple comparisons over the whole brain or within a pre-defined VOI, in the same way as described for fMRI analysis (Sections 2.2.3.1 and 2.2.3.2).

The preprocessed image of each participant is entered into a multiple regression model, which allows co-variation between regional volume change and a parameter of interest to be identified. Covariates of no interest, which may have an effect on brain volume, such as gender and age (Good et al., 2001; Smith et al., 2007), as well as individual global effects, should also be modeled in the design matrix. Compared to fMRI analysis, cluster-based inference carries an increased risk of type I statistical error in VBM since SPMs generated from tissue probability data violate the stationarity assumption. Therefore, additional correction for non-stationarity of smoothness must be undertaken (Hayasaka et al., 2004). Such procedures are carried out in Chapter 6 and described in the Materials and methods section of that chapter.
In an analogous way to the complex relationship between fMRI-BOLD and neuronal physiology (Section 2.2.1.5), interpreting the underlying anatomical basis of VBM findings can be difficult. Locally increased GM volume could reflect a combination of increases in cortical folding and increases in cortical thickness, or potentially be an artefact of tissue misclassification or suboptimal registration (Figure 2.8). Some alternative methods to VBM, described next in Section 2.2.7, permit more specific study of cortical thickness and cortical surface area.

2.2.7 Surface-based MRI analyses using FreeSurfer

While techniques like VBM are volumetric and based on dividing the brain into 3D voxels, other methodologies rather make use of models of the brain surface, which may in some instances provide more meaningful representations of the topographical organisation of the cortex (which only accounts for the ~3mm of brain tissue closest to the surface; Dale et al., 1999). FreeSurfer is the name given to a set of software tools for structural MRI and fMRI analysis and visualisation developed at Massachusetts General Hospital and Harvard University (Fischl, 2012). The initial motivation was specifically to enable construction of surface models of the cerebral cortex (Dale et al., 1999). However, multiple other
functionalities have gradually been added, including (of particular relevance to the work within this thesis), volumetric segmentation algorithms and mapping of the thickness of cortical grey matter.

Beyond the fundamental difference in approach to analysis (volumetric versus surface-based), there are some specific differences between the SPM and FreeSurfer methods, which might influence choice of methodology in particular circumstances. For example, surface-based registration, where the cortex is treated as a sheet and an attempt made to match sulcal and gyral patterns, might map cytoarchitectonic borders more accurately than volume-based registration (Fischl et al., 2008). Moreover, automated subcortical structure segmentation performed in FreeSurfer was found to provide superior results (closer to manual segmentation) than automated subcortical structural segmentation in SPM (Dewey et al., 2010). The two different segmentation methodologies are discussed in Sections 2.2.6.1 and 2.2.7.3. In another study, total brain volume measurements were found to be less accurate with FreeSurfer than with SPM but on the other hand the measurements made using FreeSurfer were more reliable across different magnet field strengths (Heinen et al., 2016). Several other relevant differences between FreeSurfer and SPM are covered in the Discussion section of Chapter 6 (Section 6.5.4).

In this thesis I have made use of FreeSurfer for cortical thickness and surface area estimation (through cortical surface reconstruction) and for automated subcortical segmentation (details of all these analyses are recorded in Chapter 6). The necessary preprocessing procedures are automated and involve various steps of spatial transformation and intensity normalisation. Cortical model generation involves reconstruction of the pial surface and grey-white matter boundary from individual MRI scans and the use of deformation procedures to compute surface based data such as maps of curvature and sulcal depth (discussed in more detail in Section 2.2.7.1). A surface-based coregistration procedure can be used to visualise data on a standard template brain, which can be inflated to better visualise sulci (Fischl et al., 1999). Curvature detail can also be added to visualisations using information obtained during segmentation and deformation procedures (Fischl and Dale, 2000). Automated segmentation is discussed in more detail below (Section 2.2.7.3).

2.2.7.1 Constructing an accurate cortical surface model

The cerebral cortex, which is responsible for much of the brain’s complex computational operations, is a thin sheet of tissue but contains a myriad of folds (sulci) and outcrops (gyri). Being able to reconstruct this surface to determine (and visualise) how different structural or functional measures inter-relate on it is therefore challenging. Current imaging
modalities (i.e. MRI) do not have adequate resolution to accurately represent the upper cortical (pial) surface. However, the folds at the lower cortical surface (the grey-white boundary) are less closely spaced and less complex. FreeSurfer models that focus on the grey-white boundary provide geometric accuracy but not estimates of the pial surface, which would be needed for morphometric measures of cortical thickness and volume.

One way of reconstructing the cortical surface is to use deformable models, which consist of curves or surfaces defined within an image domain and can move under the influence of internal forces (defined within the surface itself) or external forces (computed from the image). The deformation from a surface with a known topology (e.g. a sphere) to the brain (e.g. pial or grey-white matter boundary) surface can be modelled in this general fashion (e.g. MacDonald et al., 2000). However, it has proven extremely challenging to design algorithms that will ‘push’ the deformable surface through the many narrow, and sometimes deep, openings on the cortical surface (e.g. the Sylvian fissure). FreeSurfer takes the inverse approach (driving the cortical surface outwards towards the surface of a sphere) since it proves to be substantially less challenging computationally (Fischl et al., 2001; see Figure 2.9). Only a very small proportion of the cortical surface cannot be mapped one-to-one to a sphere; these regions are therefore classed as topological defects (e.g. holes in the cortical surface). Correcting such defects to obtain complete surface models in not a trivial problem and cannot be resolved from the surface rendering alone but requires reference back to the volumetric data (Fischl, 2012).

![Figure 2.9: Contrasting cortical surface modelling with standard deformable models and FreeSurfer](image)

Schematic examples of standard deformable models (left) and the FreeSurfer approach (right) for modelling the cortical surface. In standard deformable models a sphere (green) is driven towards the desired pial surface (red) by an energy functional; however, finding terms that will allow the surface to remain smooth and yet pass through deep sulci is difficult. Conversely, deforming the topologically incorrect surface model outwards towards a sphere is relatively straightforward computationally. Figure adapted from Fischl (2012).
2.2.7.2 Cortical thickness estimation

Cortical surface reconstruction needs to be especially accurate in order to measure morphometric differences, where changes over distances of as little as 0.5 mm can be meaningful. This is clearly a much higher resolution than that required to accurately analyse fMRI data that has been acquired using 3 mm voxels. Optimising accuracy requires some important considerations to be borne in mind. Firstly, the assumption that MRI intensity of the grey-white matter boundary is uniform seems not to hold (due to differences in histological makeup of different brain regions or image acquisition artefacts) and this can lead to inaccurate models. In FreeSurfer, MR boundary intensity is determined adaptively by modeling each point on the surface (Fischl and Dale, 2000). Secondly, surface deformation models can result in surface self-intersection (when two surfaces, for example the walls of adjacent sulci, pass through the same voxel) as well as being constrained to minimise curvature; both of these situations can lead to important loss of accuracy with respect to the cortical surface. Again, FreeSurfer can solve these problems by implementing fast triangle-triangle intersection (which prevents self-intersections) and by modelling the surface with local quadratic patches and constraining models using a second order polynomial, instead of using curvature minimisation.

FreeSurfer can therefore generate surface models that are highly accurate (able to detect variations over less than 0.25 mm) as well as being robust to image acquisition artefacts and acquisition parameter variation. These models provide a fast, accurate and automated method to obtain cortical thickness measures across the whole brain. Thickness is locally computed as the shortest distance between the pial surface and the grey-white matter boundary (Fischl and Dale, 2000); it provides a different (and arguably more specific) measure of structural variability than the ‘mixed’ volumetric indices provided by VBM (Figure 2.8; see also Section 6.5.4). Extensive validation studies have been carried out to confirm that cortical thickness measures obtained using FreeSurfer accurately reflect histological measures (Rosas et al., 2002) or manual MRI-derived measures (Kuperberg et al., 2003) from the same tissues.

2.2.7.3 Brain segmentation using FreeSurfer

The generation and analysis of cortical surface models (as discussed in Sections 2.2.7.1 and 2.2.7.2) is a major feature of FreeSurfer. However, this suite of tools also provides some non-surface-based functionalities and the whole-brain segmentation tools in particular have been used in this thesis. Instead of limiting segmentation to general classes of tissue such as GM, WM and CSF, as in SPM segmentation (see Section 2.2.6.1), FreeSurfer performs automated whole-brain segmentation into tissues based on semantically meaningful classes.
of anatomical structure (e.g. 'hippocampus' or 'amygdala'; Fischl et al., 2002). This is not a trivial problem since there is considerable overlap between different brain structures when viewed in terms of image intensity alone (Figure 2.10). In addition, there is heterogeneity in the histological structure (and resultant MR signal) within structures such as the thalamus. The principle of FreeSurfer segmentation is essentially that each MRI voxel can be assigned one of 37 labels based on a probabilistic atlas. The FreeSurfer segmentation tools employ a Bayesian approach and instead of modelling tissue classes using a small number of Gaussians, a separate model is used for each structure and each point in space. This enables accurate modelling of the heterogeneity in structures as well as more detailed and informative distributions that are unique to each structure. Priors on structure identity given spatial location are augmented with modified Markov Random Field models of typical spatial relationships between structures. Using these models, a set of manually labeled images has been used to refine and validate FreeSurfer's automated whole-brain segmentation procedure that is of comparable accuracy to manual segmentation (Fischl et al., 2002) and relatively invariant to MRI acquisition parameters (Fischl et al., 2004).

Figure 2.10: Intensity histograms from different tissue classes and brain structures

Histograms show considerable overlap demonstrating why brain structures cannot be classified from intensity information alone. The different colours show histograms for white matter (WM), grey matter (GM), lateral ventricle (IV), thalamus (Th), caudate (Ca), putamen (Pu), pallidum (Pa), hippocampus (Hp) and amygdala (Am). Figure adapted from Fischl et al. (2002).
2.3 Transcranial magnetic stimulation

TMS is a non-invasive stimulation technique that causes modulation of neural activity in the brain. The basic premise is that electrical fields can be generated in the brain by electromagnetic induction causing either instantaneous neuronal depolarisation or more lasting modulation of neuronal activity (depending on stimulation parameters). Compared to equivalent electrical stimulation protocols (which require large currents to overcome the high electrical resistance of the skull) TMS is relatively painless and therefore much better tolerated. The technique is complementary to functional imaging modalities such as fMRI, PET, EEG and MEG since its effect of modulating neural activity can probe the causative roles of certain regions or connections within a network (whereas the other mentioned techniques are correlational only). In addition, TMS (when administered as a single pulse; Section 2.3.3) has millisecond temporal precision, matched by MEG and EEG but superior to PET and fMRI. On the other hand, TMS studies impose a narrower spatial window and do not permit simultaneous mapping of the whole brain as is possible with functional imaging modalities. Of note, an alternative set of non-invasive stimulation methodologies applies weak electrical currents directly to the scalp; these methods have distinct characteristics to TMS but will not be discussed here (reviewed in Polania et al., 2018).

Early attempts to produce TMS were limited by the available technology but eventually a programme of study at the Royal Hallamshire Hospital and University of Sheffield in the 1980s was successful. Magnetic stimulation of peripheral nerves was reported in 1982 and shortly after, in 1985, the group achieved successful TMS (Barker et al., 1985). In the last couple of decades, fuelled largely by new developments of stimulation parameters with different effects on brain function, the use of TMS has hugely proliferated, both for the study of cognition and for the treatment of an increasing array of neurological and psychiatric conditions (Rossi et al., 2009).

2.3.1 Electromagnetic induction in the brain

Michael Faraday discovered the principle of electromagnetic induction in 1831, showing that an electrical current passed through one coil can induce a current in another nearby coil. This is because the current in the first coil produces a perpendicular magnetic field, which in turn causes current to flow through the second coil, again in a perpendicular direction to the magnetic field. In TMS an electrical current induces brief but intense magnetic field pulses, which in turn induce electric fields in the body (which acts as a second coil in this case); the strength of the induced fields is proportional to the rate of change of the magnetic field. In the brain such an induced electric field elicits neuronal
activity (Figure 2.11). Critically, TMS machines need to deliver a large current in a short space of time, inducing a rapidly changing magnetic field and thereby inducing an electrical field that is sufficient to stimulate neurons or interrupt normal brain activity.

TMS is best suited to studying cortical function since the cortex is closest to the scalp and the strength of induced electrical fields decreases rapidly with distance from the coil. While some deeper structures have successfully been stimulated with TMS (e.g. Zangen et al., 2005), the effects of this technique on structures deeper than the grey-white boundary remain unclear. TMS may activate corticospinal neurons indirectly through their synaptic inputs since it evokes longer-latency indirect waves (I-waves) compared to the shorter-latency direct waves (D-waves) evoked by electrical stimulation (Di Lazzaro et al., 1998).

Figure 2.11: Basic principles of TMS

The upper left panel in the figure shows how current in a TMS coil generates a magnetic field with perpendicular lines of flux, which in turn induces a perpendicular electric field in the brain. The upper right panel depicts a lateral (sagittal) view of the precentral gyrus, showing how the electric field is orientated at right angles to pyramidal axons. In the lower right panel it is shown that the electric field affects the trans-membrane potential and can thus lead to depolarisation and firing of the neuron. Figure adapted from Ilmoniemi et al. (1999).
The effects of TMS are dependent to a considerable extent on stimulation parameters (see Section 2.3.3). However, even with the same stimulation parameters there can be substantial variability in the effect of TMS and one theory is that this reflects state dependence, which could be underpinned by individual differences in neurochemistry or brain connectivity determined either genetically or from previous/ongoing brain activity (Silvanto and Pascual-Leone, 2008). Additionally, models of the effects of TMS at a neural level are complex and the virtual lesion characterisation is an over-simplification of the fact that a TMS pulse adds random activity into an area of organised neural impulses. This disrupts neural processes but can do so in a positive (inducing additional activity) or negative (inhibiting existing activity) fashion. Furthermore, noise generation hypotheses (and supporting evidence) suggest that the noise introduced by TMS is not completely random (as noise in its purest sense would be), since experimental manipulations bias relationships between stimulation and transient brain states, and this can serve to explain why in some cases TMS can facilitate neural processing (Miniussi et al., 2013).

2.3.2 TMS hardware

To generate a magnetic field pulse sufficient to depolarise neurons in underlying brain, rise time of around 100 µs, peak field of around 1 Tesla, and magnetic field energy of several hundred Joules are required. The more precise requirements for these parameters depend on factors such as local anatomy and stimulating coil geometry. Magnetic field pulses are usually generated by a capacitor discharge system, which requires a high voltage of several kilovolts to enable a rapid rise of current into the stimulating coil; this rapid rise time minimises the effect of charge leakage at nerve membranes ensuring more efficient induced stimulation. The induced fields in the brain are charge balanced over the duration of a magnetic field pulse since the induced current flowing in one direction in the tissue as the magnetic field rises is cancelled out by current flowing in the opposite direction as it falls.

The TMS pulse can be monophasic or biphasic and these pulse types differ in terms of efficacy of resulting neuronal depolarisation. Monophasic pulses have stronger short-term effects during rTMS and it may be that monophasic pulses preferentially affect one neuronal population orientated in the same direction. On the other hand, biphasic pulses have a higher peak-to-peak amplitude and longer duration and provide stronger stimulation for a given stimulation intensity when performing single-pulse TMS (Arai et al., 2005). The stimulator used for the experiments in this thesis delivered biphasic pulses.

TMS coils usually contain a spiral of wound copper or Litz wire in order to minimise resistance. The most widely used coil geometries are accordingly either circular or 'figure-
of-eight' (two attached circles) configurations. Given the principles of electromagnetic induction (see Section 2.3.1), induced electric field flows through the brain in loops parallel to the plane of the coil with the strongest current around the circumference of the coil, weaker current towards the centre, and no current at all at the centre of the coil. A figure-of-eight coil correspondingly produces maximal current density at the intersection of its two round components and this means that there is a clearer focus of maximal stimulation (Figure 2.12). The spatial resolution of TMS is limited (and in theory the magnetic field has infinite spatial extent) but the functionally relevant induced electric field can be quite focal with the optimal stimulation parameters and hardware; for example, a figure-of-eight coil has been used to stimulate regions as small as single finger representations in primary motor cortex (Ro et al., 1999). The TMS experiments described in this thesis all used a standard figure-of-eight coil.

Figure 2.12: Topographical representations of electrical field strength induced below circular and figure-of-eight TMS coils
Circular coil shown on right and figure-of-eight coil shown on left. Figure adapted from Ilmoniemi et al. (1999).

### 2.3.3 Stimulation protocols

TMS can be applied as a single stimulus (single pulse TMS), as paired stimuli (paired pulse TMS), or as trains of stimuli with a particular frequency and overall duration (repetitive TMS or rTMS). These different types of stimulation have different effects on underlying neuronal function and accordingly are used for largely different sets of applications. Single-pulse TMS involves pulses lasting around 100 µs (a duration similar to that used for conventional electrical stimulation of peripheral nerves). When applied to the scalp
overlying M1 single-pulse TMS affects synaptic inputs to corticospinal neurons (see Section 2.3.1). This results in reliable facilitation of corticospinal pathways and therefore the technique is useful for checking the integrity of these pathways, calculating central motor conduction time (Hallett, 2000), or functionally delineating motor cortex (through systematically mapping relationships between TMS pulses over different cortical locations and resulting muscle contractions or motor-evoked potentials). In addition, single-pulse TMS can be used for studying causal chronometry: if a TMS pulse is delivered over a particular brain region at precisely defined time points it is possible to determine at exactly what point in time the region contributes to task performance. One caveat with such an approach is that away from M1 it is often not clear what sort of effect the TMS pulse is having on local neuronal function (see Section 2.3.1), although it might be assumed that local neural processes at the time of the pulse are disrupted (Pascual-Leone et al., 2000). Paired-pulse TMS either involves a pair of pulses delivered in rapid succession to the same cortical target (and in this case can be used for studying intracortical facilitation or inhibition by probing interneuronal function; Chen, 2000) or involves sequential stimulation of two different cortical targets using two different coils (and in this case can be used to study causal interactions between different brain regions; Pascual-Leone et al., 2000). Despite the potential precision of the single and paired-pulse approaches, the effect (excitatory versus inhibitory) can often be difficult to predict and stimulation of interneurons as well as distant influences (probably facilitated by existing brain networks) are thought to play a role in this (Miniussi et al., 2010; Thut et al., 2005). In contrast to the brief, temporally precise and at times unpredictable effects of single-pulse and paired-pulse TMS, rTMS can be used to produce more sustained and robust neuromodulatory effects. For repeated stimuli, stimulation frequency appears to be of importance, with slow stimulus rates of <1 Hz usually being inhibitory and faster rates of >1 Hz usually facilitatory (Fitzgerald et al., 2006).

A particularly relevant stimulation protocol to the study of human cognition is the delivery of TMS in short repetitive trains (or bursts). Such stimulation configurations can facilitate or inhibit local cortical function over relatively prolonged periods (Huang et al., 2005; Nyffeler et al., 2006). For example, continuous theta-burst stimulation protocols can inhibit the function of underlying cortex for up to 20 minutes (Huang et al., 2005) and they thus provide an opportunity to induce temporary virtual lesions (more reliably and for more prolonged periods than single- and paired-pulse protocols) and thus to study the causative roles of brain regions and networks in a way that that was previously only possible by the more opportunistic study of patients with brain lesions due to previous insults. While the cTBS approach promises enticing possibilities, it is important to bear in mind that viewing the effects of TMS as a lesion is an over-simplification (discussed further in Section 2.3.1).
Another important feature of TMS protocols is the choice of whether to deliver stimulation online (during task performance, usually at a carefully chosen time point) or offline (e.g., several minutes before performance of a task). These different approaches are suited to addressing different experimental questions and also have distinct limitations (for example, online TMS risks interfering not just with activity in the targeted region but also more directly with the observed behaviours while offline TMS has limited control over the time-varying changes of the effects of stimulation). Unsurprisingly, different TMS stimulation protocols are differentially suited to online or offline stimulation; for example, the temporally precise nature of single-pulse TMS is helpful for online stimulation whereas the long-lasting effects of cTBS are better utilised for offline stimulation.

Finally, choosing the anatomical site for application of TMS is of central importance. This usually depends on the experimental question and hypotheses, which define target brain regions. The TMS coil can then be positioned over the scalp and underlying brain based on functional localisation (adjusting coil position until a particular function is elicited or disrupted; e.g., looking for a maximal muscle contraction at a given stimulation intensity), or based on anatomical localisation. Anatomical localisation is either guided by scalp landmarks or more accurately by neuronavigation. In neuronavigation a pre-acquired structural image of the participant’s brain (usually an MRI scan) is incorporated into a 3D model which also contains fiducial markers placed on the participant’s head; optical imaging systems can then determine where the TMS coil is in relation to an anatomically defined point in the participant’s brain in real time (Bolognini and Ro, 2010). In this thesis one experiment makes use of a functional target approach (based on overt motor response; Chapter 8) and another experiment makes use of an anatomical approach (finding predefined targets on a structural image loaded into a neuronavigation system; Chapter 5).

2.3.4 TMS safety

TMS has been used extensively in biomedical research and increasingly in clinical practice over more than two decades and its overall excellent record of safety is well established. Nevertheless, rare but potentially serious side effects of the technique require agreed safety guidelines to be closely adhered to (Rossi et al., 2009). Guidance on ethical and regulatory issues includes obtaining informed consent and weighing up the risks and benefits of research, as part of a formal ethical review process (in line with practice for other behavioural experimental work involving human volunteers). However, even when applying TMS to healthy human volunteers, it is advised that all studies should involve a medically responsible physician who should be familiar with the study protocol, the risks of TMS, and the treatment of any possible complications and side effects; this individual
should be locally available whenever experimental work is being carried out. Careful review has been made of the reported side effects along with guidance on ways to minimise their incidence (Rossi et al., 2009). In particular, adherence to stimulation parameters such as maximum intensities, maximum stimulation frequencies or maximum overall number of stimuli per session, depending on stimulation protocol, is advised. The most important (though very rare) complication of TMS is the provocation of epileptic seizures. While EEG aftereffects of TMS lasting around one hour are very common, a systematic review of all reports until 2008 found only 16 cases of seizures related to TMS, in the context of studies involving tens of thousands of patients in total (Rossi et al., 2009). It was found that almost exclusively these individuals were either administered TMS protocols outside of currently advised safety parameters or were taking epileptogenic medications. In addition, in several of these cases there was significant diagnostic uncertainty, for example due to the presence of clinical markers for non-epileptic or syncopal attacks. Distinction between these types of events and epileptic seizures can be challenging on clinical grounds and monitoring with EEG and/or ECG is not practically feasible for the majority of studies. Overall, the careful adherence to safe TMS parameters and the exclusion of individuals with risk factors for epilepsy (including epileptogenic medications) should be enough to minimise an already very low risk of precipitating seizures. Neurocardiogenic syncope has been reported rarely (though much more commonly than epileptic seizures) in association with TMS, and is thought to generally relate to anxiety and discomfort rather than to any more direct effects of stimulation. More commonly, headache, toothache, or paraesthesia local to the stimulation site have been reported. Notwithstanding the importance of awareness of the potential side effects as well as preparedness to deal with complications when they arise, TMS (within established safe boundaries for stimulation parameters) has proven to be a very safe technique in the vast majority of cases (Rossi et al., 2009).

2.4 Other general experimental methods

A number of specific details in the methodology for recruitment of participants and conducting behavioural and MR imaging protocols will be common to all or many experiments included in this thesis. These details are therefore included in this section to avoid undue repetition in other parts of the thesis. More specific methods unique to individual experiments are included in the relevant Materials and methods sections within the experimental chapters. Throughout the thesis all apparatus, hardware and software are mentioned by brand name. Further details regarding each of these tools are provided in Appendix 2, with a link to a URL where applicable.
2.4.1 Participant inclusion criteria and consent

In all experiments participant inclusion criteria were: age 18–45; right-handed; normal vision or vision corrected to normal with contact lenses (glasses were not permitted due to incompatibility with the testing apparatus); no history of major illness including neurological or psychiatric illnesses. In all cases participants gave written informed consent and experiments had been approved by the University College London Research Ethics Committee.

2.4.2 Procedures for behavioural experiments

Behavioural testing was always conducted in a dark and quiet room. All behavioural experiments included a visual task for which the participant’s head position was secured with a head and chin rest. This minimised head movement, and helped ensure a reliable signal for eye tracking. In binocular rivalry and CFS experiments a black cardboard divider was placed vertically between the participant’s nose and the centre of the display screen, ensuring that from the participant’s perspective each side of the screen was visible only to one eye. For both binocular rivalry and CFS experiments optimal perceptual fusion of the two stimuli on the screen was ensured prior to commencing experimental trials by adjusting relative stimulus position as necessary. In all experiments behavioural responses were recorded using a computer keyboard pad, manipulated with the participant’s dominant right hand and missing or ambiguous responses were minimised by instructing that a response button was pressed at all times, and only one button was pressed at any one time during an experimental trial (except in cases where an ambiguous percept was reported by pressing no buttons). To avoid disturbance of the natural dynamics of bistability and to optimise data quality for eye tracking (where performed), participants were also instructed to avoid forced eye blinking and to keep both eyes open equally. Any disruption of perceptual fusion was reported at the end of each experimental trial.

2.4.3 Face stimuli for BR and CFS experiments

For all BR and CFS experiments in which face stimuli were used, these were kindly supplied from the laboratory of Alex Todorov (see Contributions and Publications). The faces had been computer-generated randomly and were emotionally neutral but were manipulated, in units of s.d., along orthogonal dimensions of dominance and trustworthiness, or along a threat dimension formed by the main diagonal (Oosterhof and Todorov 2008; Figure 2.13; see Section 1.4.2).
2.4.4 Procedures for CFS and BR visual tasks

For CFS and BR experiments experimental paradigms were programmed using the Cogent Toolbox (Cogent 2000 v1.25) for MATLAB. Participants viewed two stimuli (a static face and a colourful dynamic pattern for CFS experiments; a static face and grating or two rotating gratings in BR experiments). Stimuli were placed on top of tile surrounds and shown side-by-side on a computer screen, each with a central fixation point. For rest periods (employed during fMRI experiments), the stimuli were removed from the display but tile surrounds and fixation points remained visible. Face stimuli used were taken from the model described in Section 2.4.3. Viewing was either through a pair of prism glasses (prescription lenses made with no refractive correction, four prism diopters base out on both sides, and anti-glare coating), which aid perceptual fusion of dichoptically presented images (Schurger, 2009), or through angled mirrors attached to the chin and head rest, constituting a mirror stereoscope. In CFS experiments the face stimulus was shown to the dominant eye and in BR experiments the face stimulus was shown to the right eye. Eye dominance was determined during a practice run for which the eye presented with the face stimulus was randomised on each trial and the eye resulting in more rapid detection of perceptually suppressed images thereafter became the dominant eye.
2.4.4.1 CFS experimental paradigm

In every CFS trial, the face stimulus contrast was linearly ramped from 0 - 100% over a period of 2.2 seconds. The dynamic coloured noise pattern (changing at a frequency of 10 Hz) was concurrently shown at full contrast to the non-dominant eye; this induced continuous flash suppression (CFS; Tsuchiya & Koch, 2005) while the contrast change of the face enabled this suppression to be broken after some delay and for the face to emerge into awareness (b-CFS; Jiang et al., 2007).

For each experimental CFS trial faces appeared either 1 cm to the left or 1 cm to the right of the central fixation cross and participants were instructed to make a button press (left or right arrow) on the keyboard pad as soon as they could see the face appear on one side or the other. Thus, they performed a left/right discrimination task, which was independent of the measure of interest (see below). Participants were asked to balance speed and accuracy when making responses. Correct responses provided a measure of time-to-emergence ($t_{2e}$, measured in milliseconds from onset of stimulus presentation to button press). This has previously been used as a measure of the potency of a suppressed stimulus to compete for awareness (Jiang et al., 2007; Stein et al., 2011; Stewart et al., 2012; see Section 1.3.3; Figure 2.14). If no response had been made inside 10 seconds, the CFS trial terminated.

**Figure 2.14:** CFS visual task paradigm

Faces are shown to one eye (the dominant eye, in this example the left eye) with gradually increasing contrast while a colourful dynamic CFS mask is shown to the other eye at full contrast. Initially, participants can see the CFS mask only; they then need to indicate when the face emerges into awareness ($t_{2e}$) by button press (Section 2.4.4.1).
2.4.4.2 Binocular rivalry paradigm: static face and grating

A face image (taken from the model described in Section 2.4.3) was paired with a visual grating. The latter was constructed from a Gabor patch (which is a sinusoidal grating with a Gaussian envelope) created using MATLAB. Parameters used were a Gaussian envelope of s.d. 0.3° visual angle; orientation 45°; spatial frequency 3.2 cpd; 50% contrast. The face and grating images (presented to the right and left eye, respectively) were shown at their predetermined positions on the screen simultaneously and induced binocular rivalry (see Section 1.3.3), resulting in recurrent changes of participants' perceptual experience between face, grating, or a blend of these two images. Participants held down one of three buttons (denoting which of the three percepts they were presently experiencing; Figure 2.15).

2.4.4.3 Binocular rivalry paradigm: two rotating gratings

The stimuli used for this paradigm were identical and the procedure very similar to those used in a previous study (Haynes and Rees, 2005b; Figure 2.16). A red visual grating was displayed on one side of the screen and a blue grating on the other. Throughout visual stimulation both gratings rotated at a constant speed of 1 rotation/second, remaining orthogonally orientated with respect to each other at all times. The spatial frequency of both gratings was 0.5 cpd and both were presented within a smoothed annular window. Participants were asked to maintain gaze on a central fixation dot (shown in Figure 2.16). This configuration induced binocular rivalry (see Section 1.3.3), with spontaneous changes in perceptual experience between periods of seeing exclusively the red grating, periods of
seeing exclusively the blue grating, and relatively brief periods where a mixture of the two gratings was seen. The side on which each grating colour was shown was varied in a pseudo-random manner and counter-balanced across all experimental trials. Participants held down one of three buttons (denoting which of the three percepts they were presently experiencing).

![Figure 2.16: Binocular rivalry paradigm (two rotating gratings)](image)

Two orthogonal gratings (one visible to each eye; one red and one blue in colour; both with a central fixation dot) were viewed through a mirror stereoscope. Gratings both rotated at a constant speed in a clockwise direction. This configuration induced binocular rivalry, where perceptual experience predominantly alternated between exclusive red grating and exclusive blue gating percepts, with interspersed shorter periods of mixed percepts. Figure adapted from Haynes and Rees (2005b).

### 2.4.5 Procedures for SFM visual task

SFM stimuli, programmed using Psychtoolbox3, comprised a configuration of 200 full-contrast white dots, which moved sinusoidally with an angular velocity of 151 degrees/second and were perceived by observers as a rotating sphere. Perceived rotation was about the vertical or horizontal axis, with spontaneous changes in the direction of rotation (rotating right or left and rotating up or down, respectively). A red fixation dot was presented at the centre of the stimulus (see Figure 2.17), appearing 5 seconds before each trial and beginning to flash 2 seconds before the start of the trial. After trial onset the 200 moving dots were present continuously for 48 seconds. During this time, participants provided an ongoing report of their perception of the direction of sphere rotation by holding down one of two keys (or neither key if direction of rotation was unclear).
Figure 2.17: Structure-from-motion visual paradigm

Participants viewed a single image on a computer screen consisting of a collection of white sinusoidally moving dots displayed on a black background and with a central red fixation dot (left sided image). This configuration was perceived as a sphere rotating either around its horizontal or its vertical axis, with spontaneously changing direction of rotation. In the example on the right side of the figure, the perceived sphere rotates around its vertical axis and is seen to alternately be rotating either to the right or to the left.

2.4.6 Procedures for visual tasks inside the MRI scanner

For all experiments in which dichoptic tasks were employed inside the MRI scanner specific steps were taken to facilitate such visual stimulation. With participants lying inside the scanner stimuli were projected, using a JVC DLA-SX21 projector (resolution 1400x1050 at 60 Hz for experiments using BR, and 1024x768 at 60 Hz for experiments using SFM), onto a screen (27 cm wide x 20 cm high; 62 cm from participants’ eyes) mounted at the back of the MRI magnet bore and visible through a mirror mounted on top of the RF coil. Two side-by-side images were viewed through a pair of MRI-compatible prism glasses, which facilitate image fusion (as per Section 2.4.4). A black cardboard divider was mounted in the vertical plain between the participant’s head and the projector screen to ensure each image was only visible to one eye (except in replay trials in SFM, which relied on binocular disparity cues; Section 3.2.5). Optimal image fusion was ensured prior to commencing experimental trials by adjusting relative stimulus position as necessary, and responses were made with the dominant right hand on an MRI-compatible keypad.

2.4.7 MRI and fMRI acquisition protocols

During all MRI experiments participants were instructed to lie as still as possible and small involuntary head movements were minimised by the use of foam padding.
All imaging data (except those acquired in Chapter 6; see details in that experimental chapter) were obtained on a Tim Trio 3T scanner using a 32-channel head coil. Functional images were acquired using a gradient EPI sequence with 3 mm isotropic resolution, obtaining axial slices in ascending order. Prior to EPI acquisition, fieldmaps were obtained to correct for geometric distortions in the functional images caused by heterogeneities in the $B_0$ magnetic field (double-echo FLASH sequence with a short TE of 10 ms and a long TE of 12.46 ms, $3 \times 3 \times 2$ mm, 1 mm gap; Hutton et al., 2002). Following acquisition of EPI scans, a T1-weighted structural image was acquired for each participant using a 3-D MDEFT sequence (Deichmann et al., 2004; 1 mm isotropic resolution, matrix size $256 \times 240$ (increased to $256 \times 256$ for participants with a larger head size), 176 sagittal slices, TE 2.48 ms, TR 7.92 ms, TI 910 ms). Peripheral measurements of pulse and breathing were made together with scanner slice synchronisation pulses using a Spike2 data acquisition system. The cardiac pulse signal was measured using an MRI compatible pulse oximeter attached to the left index finger. The respiratory signal (thoracic movement) was monitored using a pneumatic belt positioned around the abdomen close to the diaphragm.

2.4.8 fMRI preprocessing methods

For all fMRI experiments image preprocessing, GLM implementation, and BOLD signal extraction were performed using SPM8. Image files were converted to NIfTI format. The first five volumes from the EPI runs were then discarded (to account for equilibration effects), and the remaining functional images were independently mean bias corrected, realigned and unwarped (using voxel displacement maps generated from the fieldmaps). The functional images were co-registered with the respective anatomical MDEFT scan for each participant, normalised to MNI space, and smoothed with an 8 mm Gaussian kernel.

2.4.9 Eye tracking

In all behavioural experiments where eye tracking was used eye movements and pupillary area of the right eye were continuously tracked throughout behavioural trials using an EyeLink 1000 infra-red eye tracking system set to a sampling frequency of 1000Hz. The eye tracker was programmed using PsychToolbox3. Eye position and pupillary area were measured in pixels and pixels² respectively, on the basis of the screen resolution used for that experiment (with the lower left screen corner corresponding to position 0,0).
Chapter 3

Stimulus-invariant brain activity underlying perceptual switches and relationship to resting-state networks

“The highest activities of consciousness have their origins in physical occurrences of the brain, just as the loveliest melodies are not too sublime to be expressed by notes.”

Somerset Maugham (1874 – 1965)

3.1 Introduction

Only a small minority of the visual information processed by the human brain becomes available for conscious perception. As discussed in Section 1.3.3, binocular rivalry has been a particularly popular experimental approach for studying non-conscious and conscious visual processes and their interplay. Observers of binocular rivalry (and other bistable stimuli) experience periods when one perceptual interpretation predominates but such perceptual states are inevitably separated on either side by instances of perceptual transition, when there is a switch in the experienced conscious percept. As discussed in Section 1.3.5, underlying these perceptual transitions may be the mechanisms for access of visual information to awareness and there is converging evidence that the transitions are associated with BOLD signal changes in right superior parietal and inferior frontal cortex. The precise role and significance of these transition-related BOLD signals remains under considerable debate (see Sections 1.3.5; 3.4.5 and 9.2.3). In addition, the spatial extent of this network and the relative involvement of different regions within it remain
incompletely delineated. Further exploration of the characteristics of this perceptual-transition-related pattern of distributed neural activity could thus be one approach towards a better understanding of the mechanisms governing access to visual awareness.

There is evidence to suggest both common and distinctive features among the mechanisms underpinning different types and classes of bistable stimuli (see Section 1.3.3). For example, focusing on behavioural responses, bistable visual configurations with widely differing physical characteristics may result in highly similar reported dynamics of bistability within individuals; in particular, this has been shown for motion-induced blindness and binocular rivalry (Carter and Pettigrew, 2003). Highly correlated rates of perceptual alternation are even found across the visual and auditory modalities (Pressnitzer and Hupé, 2006). In terms of the underlying neuronal mechanisms of bistability, there are likely important differences between paradigms. For example, one dominant theory relating to binocular rivalry has suggested that bistability in this case is primarily stimulus-driven and depends on lower-level sensory processing (including competition between neurons processing monocular input from each eye in V1; Blake, 1989). On the other hand, bistability for ambiguous figures may be relatively dependent on higher attentional processes (Meng and Tong, 2004). A more recent and more integrative view regarding mechanisms of bistability is that these depend on continuous and dynamic interactions between 'low-level' (sensory) and 'high-level' (frontal and parietal) brain processes (Sterzer et al., 2009). The neural correlates of perceptual transitions provide a good example of common mechanisms across different bistable paradigms, with the pattern of fMRI-BOLD signal time locked to perceptual transitions exhibiting strikingly common features across different bistable tasks and stimulus types (Kleinschmidt et al., 2012; Rees, 2007). On the other hand, there is also evidence that perceptual-transition-associated BOLD signals have stimulus-specific properties (for example, different visual regions are activated during perceptual reversals dependent on perceived stimulus characteristics; Freeman et al., 2012). Overall, the described findings suggest that there is a combination of stimulus (or paradigm) -invariant and stimulus (or paradigm) – specific mechanisms underlying bistable perception.

With regard specifically to perceptual-transition-related BOLD signal, we may postulate that in association with a particular bistable task this signal reflects a combination of activity directly related to perceptual transition and activity related to the particular stimulus and task demands (e.g. see Freeman et al., 2012). Experiments exploring perceptual-transition-related mechanisms have attempted to minimise stimulus-specific transition-related brain activity by contrasting endogenously generated perceptual transitions occurring during bistability with a replay condition where stimulation conditions and task demands are matched as closely as possible with the bistable condition (and stimulus and task-related activity in these two conditions can thus be mutually
This approach has received some criticism and the closeness with which replay conditions can match bistable ones has been questioned (e.g. Knapen et al., 2011; see also Section 3.4.5). Another (possibly complimentary) approach to exploring neural signal specifically related to perceptual transitions would be to highlight activity that is stimulus and task-invariant; in other words activation patterns that are consistently present across widely different tasks and stimulus types. In order to address this question, in the present experiment my first key aim will be to delineate patterns of perceptual-transition-related BOLD signal that are common to two different bistable paradigms (while disregarding transition-related activation that is not common to both situations and thus may be specific to an individual stimulus or task rather than reflective of general perceptual transition-related mechanisms).

I plan to use BR (between faces and visual gratings) and SFM (using moving white dots that are perceived as a rotating sphere) since these visual stimuli differ in the following key physical properties and task demands: in BR a different image is viewed by each eye, inducing interocular competition, whereas in SFM the same image is viewed by both eyes and there is no interocular competition; the chosen BR stimulus contains components requiring visual processing of a static grating as well as object (face) perception, whereas the SFM stimulus requires evaluation of motion and three-dimensional structure (in other words the stimuli have requirements for ventral stream and dorsal stream processing respectively; see Section 1.2); finally, BR often involves periods of mixed perception (e.g. Weilnhammer et al., 2013), whereas SFM typically does not (e.g. Megumi et al., 2015). By using two paradigms with these different characteristics and demands on neural processing I hope to maximise chances that any transition-related BOLD activity common to them both would be purely related to perceptual transitions, rather than to aspects specific to the stimuli or tasks.

As discussed elsewhere (Sections 1.3.5, 3.4.3, 5.1 and 9.2.2), an understanding of the specific roles of regions in the perceptual transition-associated frontoparietal network remains incomplete. Some novel proposals, based on findings presented in this thesis, are explored in Sections 5.4, 9.2.2 and 9.2.3. One relatively well-explored area is the role of superior parietal cortex, investigated through a series of studies using rTMS for offline disruption of SPL function during different forms of visual bistability. Initial studies provided seemingly discrepant results regarding the involvement of SPL in bistable percept duration: following SPL disruption with rTMS both decreased percept duration (Carmel et al., 2010), or a contrasting increase in percept duration were reported (Kanai et al., 2010). These results have subsequently been reconciled by showing that SPL can be functionally fractionated into an anterior subregion (aSPL, which stabilises perceptual experience during bistability and where GM volume correlates positively with SFM percept duration), and a posterior subregion (pSPL, which destabilises bistable percepts and where GM volume correlates
negatively with SFM percept duration; Kanai et al., 2011). Thus different subregions of SPL have opposite causal influences on bistable perception; these can be exerted without any associated effect on spatial or sustained attention (Schauer et al., 2016). These findings have led to the proposal of theoretical models, based on predictive coding formulations, for the role of SPL in perceptual changes during bistability (Kanai et al., 2011; discussed further in Section 9.2.2). Despite the coherence of this line of work, it has been difficult to reconcile the results with others from experiments of online TMS to sites near aSPL, which have shown the opposite effect of increases in perceptual duration (Vernet et al., 2015; Zaretskaya et al., 2010). Further characterisation of how the effects of TMS on neural activity and dynamics differ for online and offline stimulation will likely be needed before such findings can be fully integrated (see also Section 2.3.3).

A separate line of evidence has added importantly to our understanding of the functional anatomy of SPL in recent years. Using diffusion-weighted MR tractography, this region can be optimally subdivided into five subregions; additionally the functional connectivity of each of these subregions with other brain areas has been defined on the basis of resting-state fMRI data (Mars et al., 2011; see Section 2.2.4). These results highlight several networks of intrinsic (resting-state) connectivity, each incorporating a distinct SPL subregion. As such, the findings may allow refinement of our understanding of perceptual-transition-related neural signal patterns in two important ways. Firstly, the data represent a framework according to which transition-related BOLD signal found in a specific SPL subregion would generate predictions about co-involvement of other parts of that region's resting state network. Secondly, relationships between intrinsic networks and task-related brain activity during bistability can be explored.

Alongside examining patterns of stimulus-invariant perceptual transition-related BOLD signals, my second key aim for this experiment will therefore be to make use of the SPL parcellation and resting-state connectivity analyses of Mars and colleagues (2011) to extend understanding of BOLD activation patterns associated with perceptual transition. Existing evidence enables me to have a strong prior hypothesis that transition-related BOLD signal will be found in right frontoparietal cortical regions, including SPL (Lumer et al., 1998; Rees, 2007; Sterzer and Kleinschmidt, 2007; Weilnhammer et al., 2013). In addition, I hypothesise that SPL will contain stimulus-invariant perceptual-transition-related BOLD signals (present across both BR and SFM paradigms). Next, I will ask specifically whether such stimulus-invariant BOLD signals in SPL overlap spatially with a particular SPL subregion, based on the parcellation of Mars and colleagues. Finally, I will explore whether other regions in the network of resting state functional connectivity coupled to the relevant SPL subregion also exhibit transition-related BOLD signal change.
If I find that BOLD activity related to perceptual transitions maps closely onto intrinsic connectivity networks present in the resting state, this would suggest that bistable perceptual dynamics can relate to intrinsic and non-task-related functional brain properties. Spontaneous fluctuations in brain activity can be closely linked to task-associated activation patterns (Section 2.2.4; Fox et al., 2006). More generally, while the functional role of resting-state networks remains under debate, these spontaneous coordinated neuronal fluctuations have been proposed to reflect either past, current or future coordinated activity among groups of brain regions that work in concert; in addition such resting state networks may relate to anatomical connectivity (Fox and Raichle, 2007). To what extent these links between intrinsic connectivity, task-related mechanisms and anatomical connections apply to bistable perception is a relevant question for current and future exploration.

In summary, the main aims of the present experiment are firstly to highlight patterns of BOLD signals related to perceptual transitions that are common to two different bistable perception paradigms (BR and SFM); and secondly to assess the relationship between such perceptual-transition-related neural activity and established SPL-associated resting state functional connectivity networks.

### 3.2 Materials and methods

#### 3.2.1 Contribution acknowledgement and reuse of data

The SFM fMRI data described below were collected and preprocessed by Megumi Fukuda as part of a separate experiment (published in Megumi et al., 2015). The remainder of the experimental work presented in this Chapter (and all associated analysis and interpretation) is my own. Full details of external contributions can be found in the Contributions and publications section at the start of the thesis. The BR fMRI data presented below were also used in the fMRI experiment described in Chapter 7.

#### 3.2.2 Participants

For the BR experiment 20 participants were recruited (15 female; mean ± SD age = 25.2 ± 3.8 years; range = 20 - 34 years; see also Section 7.3.2.1), and for the SFM experiment an independent sample of 18 participants were recruited (10 female; mean ± SD age = 26.0 ± 6.2 years; range = 18 - 39 years). All participants across both experiments were recruited according to the procedures and inclusion criteria listed in Section 2.4.1.
3.2.3 Display apparatus inside MRI scanner

The apparatus for appropriate visual stimulation and recording of participant responses during both BR and SFM tasks were set up as described in Section 2.4.6.

3.2.4 Binocular rivalry: stimuli and experimental procedures

Stimuli and visual paradigms were as described in Sections 2.4.2 – 2.4.4. Experimental procedures inside the MRI scanner were as per Section 2.4.6. An identical experimental methodology was also used in Chapter 7 (Sections 7.2.2.2, 7.2.2.3, and 7.3.2.2) where the same data were analysed for a different research question. Face images were paired with visual gratings to induce binocular rivalry (Section 2.4.4.2), with each stimulus subtending in total 2.9° visual angle and each fixation cross 0.5° visual angle. Two versions of the same computer-generated face identity (differing in dominance traits; as described in Section 7.2.2.2) were shown to all participants, with experimental trials split equally and pseudorandomly between the two face versions. Effects of facial dominance were not evaluated in the present study and I therefore collapsed across the dominance factor when analysing the data in this experiment. Trials lasted for 40 seconds (Figure 3.1A) and were separated by 20-second rest periods (see Section 2.4.4), during which participants were asked to maintain ‘loose’ fixation. To be able to measure BOLD signal specifically related to endogenous perceptual transitions, a “replay” condition was included (as per Lumer et al., 1998). Here stimuli were still presented dichoptically, with an identical screen configuration to binocular rivalry trials. However, instead of presenting a different image to each eye, the same image was presented to both eyes (Figure 3.1B). This image was changed between face, grating and a simulated mixed percept with a sequence and time course that matched a previous rivalry trial. Thus, in this condition perceptual changes were generated exogenously at the sensory input level. Replayed mixed percepts were matched in duration with the corresponding mixed percept in the paired rivalry trial and were constructed by creating an overlay of images representing the preceding and following percepts in rivalry with gradual contrast changes to achieve a blend from the former to the latter.

Each 6-minute experimental run consisted of three rivalry trials, followed by the three replay trials yoked to them (in a pseudo-random order), with rest periods after trials of both types. Participants underwent 6 runs, with 1–2 minutes of complete rest (no visual stimulation) between runs. Prior to the first run, participants underwent a practice run which contained two rivalry trials and two related replay trials.
Figure 3.1: Stimuli and experimental paradigms employed in BR and SFM experiments

For the BR experiment (Section 3.2.4) in rivalry trials (A) a face image was shown to the right eye and a grating image was shown to the left eye, whereas in replay trials (B) a face image, grating image, or blend of the two, was shown to both eyes in a sequence matching a previous rivalry trial. For both BR conditions stimuli were displayed for 40 seconds at a time; an example of the observer’s experience in each case is shown in the blue shaded boxes. Perceptual report was provided by holding one of three buttons denoting current percept (face, grating or mixed; mixed percepts are not shown here for simplicity). For the SFM experiment (Section 3.2.5) in ambiguous trials (C) an identical configuration of oscillating white dots was shown to each eye and these were perceived as a sphere rotating upwards or downwards, whereas in replay trials (D) binocular disparity cues were added to the dots enabling them to be seen to rotate non-ambiguously upwards or downwards, in a sequence matching a previous ambiguous trial. Trials lasted 31.5 seconds and, as for BR, perceptual report was provided by holding one of three buttons denoting current percept (upward rotation, downward rotation, unclear). Figures in panels C and D adapted from Megumi et al. (2015).
3.2.5 Structure-from-motion: stimuli and experimental procedures

The experimental procedure used for SFM was as described in Section 2.4.5 and as employed in a previous study (Megumi et al., 2015). The experimental setup inside the MRI scanner was as per Section 2.4.6. The perceived sphere (with a diameter subtending 3.1° visual angle) appeared to rotate about its horizontal axis in an alternating direction (either upwards or downwards). A red fixation dot (0.1° visual angle in diameter) was presented at the centre of the sphere. Two such spheres were shown side-by-side over identical square frames to facilitate perceptual fusion when shown dichoptically. In ambiguous trials participants viewed the fused spheres and reported spontaneous changes in perceived direction of rotation (Figure 3.1C). In replay trials (Figure 3.1D) the sequence of percepts reported in a previous ambiguous trial was matched and binocular disparity cues (calculated separately for each moving dot) were added to the stimuli allowing participants to view them in a non-ambiguous stereoscopic fashion.

Each trial began with a 10.5-second rest period, during which a fixation dot was visible and participants were asked to maintain ‘loose’ fixation, as in the BR experiment. There followed a stimulation period when the two moving spheres were present continuously for 31.5 seconds and participants provided continuous online report of their perception of the direction of sphere rotation by holding down one of three keys (upward rotation, downward rotation, unclear).

Each seven-minute experimental run consisted of 10 ambiguous trials and 10 replay trials yoked to them, with rest periods preceding trials of both types. The order of ambiguous trials and replay trials was pseudo-randomized. Participants underwent 4 to 7 runs (Mean ± SD: 6.4 ± 0.9), with 1-2 minutes of complete rest between runs.

3.2.6 fMRI data acquisition and pre-processing

Acquisition of fMRI data for both the BR and SFM experiments was undertaken according to the procedures and using the parameters described in Section 2.4.7, with the following exceptions and additional details. Cardiac pulse was recorded during both experiments but respiratory signal was monitored during the BR experiment only. These peripheral physiological data were not used for further analysis for the study reported in this chapter (the cardiac and respiratory data recorded for the BR experiment have been used in analyses elsewhere in this thesis; see Sections 7.3.2.5 and 7.3.2.6). Slice acquisition time was 70 ms, TE was 30 ms, and matrix size was 64 × 74 for both experiments. For the BR
48 axial slices of 3 mm thickness were obtained and volume TR was therefore 3.36 s. For the SFM experiment, 30 axial slices of 3 mm thickness were obtained resulting in a shorter volume TR of 2.1 s. To optimally reduce susceptibility-induced signal loss from regions of orbitofrontal cortex and amygdala, a slice tilt of -30° (T>C), Z-Shim of -0.4 mT/m*ms, PE direction positive/up and fieldmap parameters PE blips = -1 were used for both experiments (Weiskopf et al., 2006). A total of 115 volumes per BR run and 205 volumes per SFM run were acquired. For both BR and SFM experiments image preprocessing was carried out according to the methodology described in Section 2.4.8. Low-frequency fluctuations in the signal were removed by using a high-pass filter with cut-off at 128 s.

3.2.7 fMRI data analysis

For both BR and SFM experiments, the GLM was implemented in SPM8 to produce 3D maps of parameter estimates representing the contribution of each regressor to the data (see Section 2.2.3). In both analyses, the regressors of interest were vectors containing the timing of all transition events between possible percepts (face, grating and mixed in the BR experiment; upward rotation or downward rotation for the SFM experiment). Typically SFM does not induce mixed percepts and perceptual switches are crisp; therefore no mixed condition was included in the SFM fMRI analyses. Indeed, I found that the total duration of the 'unclear' (mixed percept) condition was very short in the SFM experiment (on average 2.23% of the total duration of stimulus presentation per MRI run; Megumi et al., 2015). In contrast, the proportion of mixed perception tends to be much higher in the paradigm used for BR (28% of the total duration of stimulus presentation in a behavioural experiment using an identical stimulus; Section 7.2.3.1) confirming the necessity of including a mixed condition in the BR fMRI analyses.

The perceptual transitions (modelled as instantaneous events with an impulse function) were defined temporally by subtracting mean latency to button press (calculated from the replay condition in each experiment) from each change in button press. For both experiments, separate regressors were used for endogenously generated transitions during ambiguous trials and exogenously generated transitions during replay trials (transitions were grouped into regressors irrespective of preceding or following percept). For both BR and SFM, additional regressors (modelled with boxcar functions) represented periods of stable perception (three regressors for face, grating or mixed percepts in BR; a single regressor for visual stimulation in SFM) as well as rest/fixation periods (a single regressor in both cases). Perceptual state regressors spanned both ambiguous and replay conditions to reduce correlation with the transition regressors and thus optimise design
orthogonality. All regressors were then convolved with a canonical HRF implemented in SPMs. For both BR and SFM head motion parameters were added to design matrices as covariates of no interest. Overall the analysis employed here for the BR experiment was similar to that described in Section 7.3.2.5 (where the same data are used for a different experiment) with some key differences. In the present experiment cardiorespiratory data were not added as covariates of no interest in order to more closely match the SFM experimental analysis (where respiratory data had not been obtained). In addition, separate regressors for dominant and neutral face trials were not modelled in the present experiment and neither were data from a functional face localiser (Section 7.3.2.3), since face processing and the effect of face dominance are not under investigation here.

In both BR and SFM analyses, maps of effect sizes for the main effect of endogenous transitions > replayed transitions were computed for each participant by using the HRF regressors. Subsequently, random-effects (second-level) analysis formed the basis for statistical inference, where one-sample t-tests on these effect size maps were performed across all participants within each experiment. This produced two group-level SPMs of regional effects for the contrast endogenous transitions > replayed transitions (one for BR and one for SFM). Next, to create SPMs representing endogenous transition > replay transition-related BOLD signal that is present across both experiments (and therefore stimulus-invariant), each group-level result was in turn inclusively masked with the other.

Statistical results were examined and subsequently displayed in tables and figures based on a threshold of $p < 0.001$ (uncorrected for multiple comparisons). In assessing statistical significance FWE correction across a volume restricted to regions of interest (defined a priori) was undertaken, thus identifying voxels where there was a significant effect. Statistical significance was determined at the peak level, with a criterion of $p < 0.05$. The regions used for these small-volume-corrected analyses, which were based on the results of previous MRI and fMRI studies of perceptual transition (Kanai et al., 2010, 2011; Lumer et al., 1998; Lumer and Rees, 1999) as well as on the resting-state SPL-associated networks outlined by Mars and colleagues (Mars et al., 2011), are listed in Table 3.1. Correction was performed within spherical volumes of 10 mm radius centred on the central/peak coordinates for each of the listed ROIs. Additionally, small-volume correction at the five SPL subregions defined by Mars and colleagues was performed (Mars et al., 2011; Figure 3.2).
Table 3.1: Results of previous studies that motivate ROIs to be used for small volume correction

See also Section 3.2.7. Coordinates refer to the peak activation or the centre of a region reported in a previous study of perceptual transition (as specified in the Source column).

**Figure 3.2:** Five subregions of SPL, proposed on the basis of diffusion-weighted tractography results. Each of these subregions is associated with a distinct pattern of resting-state functional connectivity (see Mars et al., 2011). Subregions are shown on a brain rendering performed in MRicron.
3.3 Results

I examined patterns of fMRI-BOLD signals time locked to perceptual transitions in two separate experiments each making use of a different bistable perception paradigm in an independent sample of participants. While the two paradigms differed in stimulus characteristics and neural processing requirements, scanning parameters and analysis methodology were very similar across the two experiments (see Section 3.2.6 and 3.2.7). I therefore expected that differences in results would relate primarily to differences in paradigm and task.

Behavioural results have been described elsewhere (Section 7.3.3.1 for the BR experiment; Megumi et al., 2015 for the SFM experiment). For each of the two experimental datasets the focus of fMRI analysis was to highlight BOLD signals specifically related to endogenously generated switches between the perceptual interpretations possible for that particular bistable paradigm. Thus, in both cases, BOLD signals time locked to endogenously generated perceptual switches during ambiguous trials were contrasted with activation time locked to exogenously generated perceptual switches during a non-ambiguous replay condition. The next step was then to highlight brain regions where transition-related BOLD signals were seen across both paradigms. Finally, overlap of this stimulus-invariant pattern of transition-related activation with SPL subregions and established resting-state networks (as outlined in Mars et al., 2011) was explored.

3.3.1 Binocular rivalry experiment results

BOLD signals for the contrast endogenously generated BR transitions > replayed BR transitions were found in a predominantly right-hemisphere group of regions, especially in superior parietal and inferior frontal areas, consistent with previous fMRI studies of perceptual transition (Bisenius et al., 2015; Rees, 2007; Figure 3.3A). Activation was also seen prominently in cingulate and precentral cortices. Coordinates and statistics for activated regions are listed in Table 3.2 upper section (note these differ slightly to the findings listed for the same dataset in Chapter 7, see Table 7.2, where physiological noise correction, a slightly different design matrix, and a different statistical threshold were used; however, the overall pattern of findings is highly similar).
### Table 3.2: Stimulus-specific perceptual-transition-related activation; whole brain analyses

Brain regions activated for the endogenous transitions > replayed transitions contrast are shown for the BR experiment (upper section) and SFM experiment (lower section). Region labels were produced using the AAL toolbox for SPM (Tzourio-Mazoyer et al., 2002). Coordinates and statistics apply to peak-level activations. All activations surviving a threshold of $p < 0.001$ uncorrected, and with a cluster size of at least 10 voxels, are displayed. Hem, hemisphere.

<table>
<thead>
<tr>
<th>Location</th>
<th>Hem</th>
<th>MNI Coordinates</th>
<th>Cluster Size</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regional cerebral cortex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral transitions &gt; replayed transitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral</td>
<td>Right</td>
<td>48 -10</td>
<td>48</td>
<td>434</td>
</tr>
<tr>
<td>Putamen</td>
<td>Right</td>
<td>28 24</td>
<td>0</td>
<td>1122</td>
</tr>
<tr>
<td>Middle cingulate</td>
<td>Right</td>
<td>4 22</td>
<td>30</td>
<td>584</td>
</tr>
<tr>
<td>Insula</td>
<td>Left</td>
<td>-34 8</td>
<td>6</td>
<td>294</td>
</tr>
<tr>
<td>Precentral</td>
<td>Right</td>
<td>22 -20</td>
<td>58</td>
<td>118</td>
</tr>
<tr>
<td>Calcarine</td>
<td>Right</td>
<td>28 -70</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>Precuneus</td>
<td>Left</td>
<td>-4 -58</td>
<td>18</td>
<td>140</td>
</tr>
<tr>
<td>Middle frontal</td>
<td>Left</td>
<td>-26 44</td>
<td>32</td>
<td>190</td>
</tr>
<tr>
<td>Precentral</td>
<td>Left</td>
<td>-32 -14</td>
<td>42</td>
<td>67</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>Right</td>
<td>8 -66</td>
<td>-10</td>
<td>20</td>
</tr>
<tr>
<td>Superior temporal</td>
<td>Left</td>
<td>-64 -22</td>
<td>8</td>
<td>112</td>
</tr>
<tr>
<td>Rolandic operculum</td>
<td>Right</td>
<td>56 -12</td>
<td>16</td>
<td>131</td>
</tr>
<tr>
<td>Precuneus</td>
<td>Left</td>
<td>-12 -44</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>Postcentral</td>
<td>Left</td>
<td>-22 -34</td>
<td>54</td>
<td>14</td>
</tr>
<tr>
<td>Postcentral</td>
<td>Right</td>
<td>22 -42</td>
<td>48</td>
<td>121</td>
</tr>
<tr>
<td>Calcarine</td>
<td>Left</td>
<td>-24 -70</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>Thalamus</td>
<td>Left</td>
<td>-4 -12</td>
<td>-2</td>
<td>44</td>
</tr>
<tr>
<td>Paracentral lobule</td>
<td>Right</td>
<td>20 -40</td>
<td>52</td>
<td>10</td>
</tr>
<tr>
<td>Medial superior frontal</td>
<td>Left</td>
<td>-14 50</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>Superior temporal</td>
<td>Left</td>
<td>-58 4</td>
<td>4</td>
<td>46</td>
</tr>
<tr>
<td>Cuneus</td>
<td>Right</td>
<td>16 -78</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>Superior parietal</td>
<td>Right</td>
<td>22 -52</td>
<td>50</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Hem</th>
<th>MNI Coordinates</th>
<th>Cluster Size</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regional cerebral cortex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endogenous SFM transitions &gt; replayed transitions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior frontal</td>
<td>Right</td>
<td>18 8</td>
<td>66</td>
<td>5187</td>
</tr>
<tr>
<td>Insula</td>
<td>Left</td>
<td>-34 22</td>
<td>2</td>
<td>718</td>
</tr>
<tr>
<td>Middle occipital</td>
<td>Left</td>
<td>-44 78</td>
<td>10</td>
<td>775</td>
</tr>
<tr>
<td>Postcentral</td>
<td>Right</td>
<td>56 -26</td>
<td>48</td>
<td>4888</td>
</tr>
<tr>
<td>Middle frontal</td>
<td>Left</td>
<td>-26 -4</td>
<td>50</td>
<td>820</td>
</tr>
<tr>
<td>Superior parietal</td>
<td>Left</td>
<td>-18 -62</td>
<td>52</td>
<td>1559</td>
</tr>
<tr>
<td>Middle frontal</td>
<td>Right</td>
<td>32 52</td>
<td>22</td>
<td>544</td>
</tr>
<tr>
<td>Inferior occipital</td>
<td>Right</td>
<td>38 -78</td>
<td>-14</td>
<td>62</td>
</tr>
</tbody>
</table>
**Figure 3.3:** Brain regions with increased BOLD signal for the contrast endogenous transitions > replayed transitions in the BR and SFM experiments

Results for BR experiment (A) and SFM experiment (B) shown. The corresponding results are also described in Sections 3.3.1 and 3.3.2 respectively, and displayed in Table 3.2. Masking the endogenous transitions > replayed transitions contrast image from the BR experiment inclusively with that from the SFM experiment reveals perceptual-transition-related BOLD signal that is common to both paradigms (C). The reverse order of masking (inclusive mask of BR results applied to the SFM findings) was also performed to ensure consistency of findings (D). The results for these stimulus-invariant activation analyses are described in Section 3.3.3 and displayed in Tables 3.3 and 3.4. Activations are shown in colour, corresponding to T values as indicated by the colour bar, and are overlaid on inflated brains as implemented in FreeSurfer. A threshold of $p < 0.001$ uncorrected was used for display purposes. Only the right hemisphere, where the majority of BOLD signal was found, is shown.
3.3.2 Structure-from-motion experiment results

BOLD signals for the contrast endogenous SFM transitions > replayed SFM transitions were found predominantly in the right hemisphere. There is marked similarity to the findings for binocular rivalry described in Section 3.3.1 (Table 3.2 lower section; Figure 3.3B), although transition-related signal for the SFM paradigm was generally more statistically robust and with substantially larger clusters of activation (see T value and cluster size columns in Table 3.2). Besides frontal and parietal regions, there was also extensive activation in visual areas in this case, particularly in area V5, which was likely related to the strong component of motion in this stimulus. These results have been reported elsewhere (Megumi et al., 2015).

3.3.3 Stimulus-invariant transition-related BOLD signal patterns

Masking the endogenous transitions > replayed transitions contrast image for each bistable paradigm inclusively with that for the other paradigm enabled me to highlight patterns of perceptual-transition-related BOLD signals that are common to both paradigms. When the BR findings were masked inclusively with the SFM findings, activation was seen in a predominantly right-sided and frontoparietal group of cortical regions (Figure 3.3C; Table 3.3 upper section). The activation pattern is similar but more spatially limited than that for perceptual-transition-related signal in BR alone (Figure 3.3A). Performing the reverse contrast masking procedure (SFM findings masked inclusively with BR findings) led to a very similar pattern of activation to the first masking procedure, again centred on right frontoparietal cortical regions (Figure 3.3D; Table 3.3 lower section). As before, this pattern was more spatially limited than the corresponding finding for SFM alone (Figure 3.3B). With both masking procedures, stimulus-invariant perceptual-transition-related BOLD signal was seen in right insula/putamen, left insula, right superior frontal gyrus, middle cingulate gyrus, and superior parietal/postcentral areas.
Table 3.3: Brain regions activated during endogenous (versus replayed) perceptual transitions in a stimulus-invariant fashion; whole brain analyses

Endogenous transitions in BR (masked with endogenous transitions in SFM)

<table>
<thead>
<tr>
<th>REGION</th>
<th>HEM</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>CLUSTER SIZE</th>
<th>T</th>
<th>Z</th>
<th>P VALUE (UNC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putamen</td>
<td>Right</td>
<td>28</td>
<td>24</td>
<td>0</td>
<td>425</td>
<td>6.43</td>
<td>4.58</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Insula</td>
<td>Left</td>
<td>-38</td>
<td>12</td>
<td>-2</td>
<td>49</td>
<td>5.66</td>
<td>4.24</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior frontal</td>
<td>Right</td>
<td>30</td>
<td>-6</td>
<td>64</td>
<td>37</td>
<td>4.70</td>
<td>3.75</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Middle cingulate</td>
<td>Right</td>
<td>6</td>
<td>26</td>
<td>32</td>
<td>17</td>
<td>4.63</td>
<td>3.71</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Postcentral</td>
<td>Right</td>
<td>30</td>
<td>-46</td>
<td>60</td>
<td>64</td>
<td>4.52</td>
<td>3.65</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior frontal</td>
<td>Right</td>
<td>22</td>
<td>-8</td>
<td>54</td>
<td>21</td>
<td>4.36</td>
<td>3.55</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior parietal</td>
<td>Right</td>
<td>22</td>
<td>-52</td>
<td>50</td>
<td>19</td>
<td>4.19</td>
<td>3.45</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Endogenous transitions in SFM (masked with endogenous transitions in BR)

<table>
<thead>
<tr>
<th>REGION</th>
<th>HEM</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>CLUSTER SIZE</th>
<th>T</th>
<th>Z</th>
<th>P VALUE (UNC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior frontal</td>
<td>Right</td>
<td>28</td>
<td>-2</td>
<td>64</td>
<td>37</td>
<td>7.27</td>
<td>4.84</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Insula</td>
<td>Right</td>
<td>30</td>
<td>22</td>
<td>4</td>
<td>425</td>
<td>6.41</td>
<td>4.51</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior frontal</td>
<td>Right</td>
<td>24</td>
<td>-6</td>
<td>56</td>
<td>21</td>
<td>6.19</td>
<td>4.42</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Insula</td>
<td>Left</td>
<td>-36</td>
<td>16</td>
<td>0</td>
<td>49</td>
<td>5.65</td>
<td>4.18</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior parietal</td>
<td>Right</td>
<td>20</td>
<td>-58</td>
<td>58</td>
<td>19</td>
<td>5.23</td>
<td>3.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Superior parietal</td>
<td>Right</td>
<td>26</td>
<td>-52</td>
<td>64</td>
<td>64</td>
<td>4.89</td>
<td>3.81</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Middle cingulate</td>
<td>Right</td>
<td>8</td>
<td>26</td>
<td>34</td>
<td>17</td>
<td>4.40</td>
<td>3.54</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

3.3.4 Relationship of fMRI findings to SPL subregions and associated resting-state networks

The outlined stimulus-invariant perceptual-transition-related BOLD signal patterns (Section 3.3.3) can be further characterised according to their relationship to distinct SPL subregions and associated resting-state functional networks (Mars et al., 2011; see Section 3.1 and Figure 3.2). Masking the contrast images obtained from BR and SFM paradigms inclusively with each other revealed stimulus-invariant BOLD signals in right aSPL. This result overlaps spatially with the most anterior subregion of the SPL parcellation of Mars.
and colleagues (red region in Figure 3.2). The overlap is shown graphically in Figures 3.4 and 3.5 and when this SPL subregion is used as a mask for small-volume correction, the overlap with stimulus-invariant signal (SPL cluster) is statistically significant ($x = 22, y = -42, z = 48; T = 4.65; P_{FWE-corr} = 0.02$ for BR masked with SFM and $x = 32, y = -48, z = 50; T = 5.20; P_{FWE-corr} = 0.005$ for SFM masked with BR). The results of further small-volume corrected analyses at regions that form a functional resting-state network with this anterior SPL subregion (Mars et al., 2011; these include the FEF, PMd and inferior PMd regions listed in the lower section of Table 3.1) are described in Section 3.3.5.

![Figure 3.4](image)

**Figure 3.4:** Overlap between SPL subregions and increased BOLD signal for the contrast endogenous transitions > replayed transitions in BR masked with SFM and in SFM masked with BR BOLD signal shown in hot colour scale. Contrast in BR masked with SFM (A) and in SFM masked with BR (B). The five SPL subregions are shown, as in Figure 3.2, here with the most anterior region (marked with (1) in panel A), seen to overlap most closely with transition-related BOLD signal. The three regions with reported resting-state functional connectivity with this anterior SPL subregion are also shown in the same cyan colour and have been numbered in panel A: PMd (2), FEF (3), and inferior PMd (4). It is clear that panels A and B show highly similar findings. The BOLD signal patterns (displayed at a threshold of $p < 0.001$), SPL masks (provided by Rogier Mars) and masks representing other regions (created using the Pick atlas) are shown on a brain rendering performed in MRIcron. Alternative visualisation of these data can be seen in Figure 3.5.
Figure 3.5: Overlap between anterior SPL subregion, regions with associated resting state functional connectivity and stimulus-invariant perceptual-transition-related BOLD activity. This figure shows the same data as Figure 3.4A but for more detailed visualisation they are displayed in multiple axial (A), sagittal (B) and coronal (C) slices. Numbers above each image provide MNI coordinates reflecting the position of that particular slice. BOLD signal is shown in hot colours and regions of interest are shown in cyan and numbered (1 = aSPL, 2 = PMd, 3 = FEF, 4 = inferior PMd). Colours and numbering match those used in Figure 3.4. Close association between BOLD signal and the regions of interest can be seen. In addition, BOLD signal can be seen in bilateral insula/IFG. BOLD signal patterns displayed at a threshold of \( p < 0.001 \); SPL masks provided by Rogier Mars; masks representing other regions created using the Pick atlas; results shown on brain slices visualised in MRIcron.

3.3.5 Results from small-volume-corrected analyses centred at other predetermined regions of interest

Small-volume-corrected analyses were undertaken on the stimulus-invariant contrast images (where endogenous perceptual transition-related activity for each bistable paradigm had been inclusively masked with that for the other bistable paradigm; Section 3.3.3). Correction for multiple comparisons was restricted to regions of interest relating either to neural correlates of perceptual transition or to SPL-associated resting state networks (see Section 3.2.7 and Table 3.1).

Using the perceptual transition-related regions of interest listed in Table 3.1 (upper section), for endogenous (versus replayed) BR transitions masked inclusively with
endogenous (versus replayed) SFM transitions small volume correction revealed statistically significant activations in three regions: right aSPL \((x = 30, y = -16, z = 58; T = 4.45; P_{\text{FWE-corr}} = 0.02)\); right insula \((x = 36, y = 12, z = -6; T = 6.20; P_{\text{FWE-corr}} = 0.001)\); and right IFG \((x = 44, y = 12, z = -4; T = 4.86; P_{\text{FWE-corr}} = 0.01)\); statistics also shown in Table 3.4, upper section). For the reverse contrast masking procedure (endogenous versus replayed SFM transitions masked inclusively with endogenous versus replayed BR transitions) small volume correction using the same regions of interest \((x = 32, y = -50, z = 56; T = 4.39; P_{\text{FWE-corr}} = 0.01)\); right insula \((x = 36, y = 16, z = -6; T = 5.82; P_{\text{FWE-corr}} = 0.001)\); and right IFG \((x = 42, y = 16, z = -6; T = 5.46; P_{\text{FWE-corr}} = 0.002); statistics also shown in Table 3.4, lower section). Notably, both contrast masking procedures produced very similar results. The aSPL findings have been discussed in the context of SPL parcellation frameworks and associated resting-state connectivity networks in Section 3.3.4. Importantly, a statistically significant result was not found after small volume correction at pSPL (see Table 3.1), the second SPL subregion proposed to play a role in perceptual transition in the work of Kanai and colleagues (2011).

I found statistically significant stimulus-invariant perceptual transition-related activation in aSPL (see above), which closely overlapped a particular SPL subregion from the parcellation of Mars and colleagues (2011; see Section 3.3.4). Given this result, regions in the resting state network associated with this SPL subregion were also used in further small-volume correction analyses (these include FEF, PMd and inferior PMd; Mars et al., 2011; region coordinates listed in the lower section of Table 3.1). For endogenous (versus replayed) BR transitions masked inclusively with endogenous (versus replayed) SFM transitions these analyses revealed statistically significant activations in right FEF \((x = 22, y = -10, z = 50; T = 4.02; P_{\text{FWE-corr}} = 0.04)\); and right PMd \((x = 28, y = -8, z = 64; T = 4.16; P_{\text{FWE-corr}} = 0.03); statistics also shown in Table 3.4, upper section). For endogenous (versus replayed) SFM transitions masked inclusively with endogenous (versus replayed) BR transitions the small-volume-correction analyses again revealed statistically significant findings in right FEF \((x = 22, y = -8, z = 52; T = 4.81; P_{\text{FWE-corr}} = 0.006)\); and right PMd \((x = 30, y = -8, z = 62; T = 4.39; P_{\text{FWE-corr}} = 0.01); statistics also shown in Table 3.4, lower section). Of note, inferior PMd, the third region in the aSPL-associated resting state network outlined by Mars and colleagues (2011) did not contain statistically significant activation for either contrast.
### Table 3.4: Brain regions activated during endogenous (versus replayed) perceptual transitions in a stimulus-invariant fashion; small volume corrected results

Statistically significant findings after correcting across the volume of predefined regions of interest (as per Table 3.1) are shown for endogenous (versus replayed) transition-related BOLD signal for BR masked inclusively with SFM (upper section); and for SFM masked inclusively with BR (lower section). Coordinates and statistics apply to peak-level activations. All activations surviving a threshold of \( p < 0.05 \), small volume corrected, are displayed. Hem, hemisphere.

<table>
<thead>
<tr>
<th>Location</th>
<th>HEM</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Size</th>
<th>T</th>
<th>Z</th>
<th>P Value (SVC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endogenous BR transitions masked inclusively with endogenous SFM transitions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-volume corrected results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insula</td>
<td>Right</td>
<td>36</td>
<td>12</td>
<td>-6</td>
<td>360</td>
<td>6.20</td>
<td>4.48</td>
<td>0.001</td>
</tr>
<tr>
<td>Inferior frontal</td>
<td>Right</td>
<td>44</td>
<td>12</td>
<td>-4</td>
<td>150</td>
<td>4.86</td>
<td>3.84</td>
<td>0.01</td>
</tr>
<tr>
<td>Anterior SPL</td>
<td>Right</td>
<td>30</td>
<td>-46</td>
<td>58</td>
<td>218</td>
<td>4.45</td>
<td>3.61</td>
<td>0.02</td>
</tr>
<tr>
<td>Dorsal premotor</td>
<td>Right</td>
<td>28</td>
<td>-8</td>
<td>64</td>
<td>8</td>
<td>4.16</td>
<td>3.44</td>
<td>0.03</td>
</tr>
<tr>
<td>Frontal eye field</td>
<td>Right</td>
<td>22</td>
<td>-10</td>
<td>50</td>
<td>165</td>
<td>4.02</td>
<td>3.35</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Endogenous SFM transitions masked inclusively with endogenous BR transitions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small-volume corrected results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insula</td>
<td>Right</td>
<td>36</td>
<td>16</td>
<td>-4</td>
<td>208</td>
<td>5.82</td>
<td>4.26</td>
<td>0.001</td>
</tr>
<tr>
<td>Inferior frontal</td>
<td>Right</td>
<td>42</td>
<td>16</td>
<td>-6</td>
<td>198</td>
<td>5.46</td>
<td>4.09</td>
<td>0.002</td>
</tr>
<tr>
<td>Frontal eye field</td>
<td>Right</td>
<td>22</td>
<td>-8</td>
<td>52</td>
<td>3</td>
<td>4.81</td>
<td>3.77</td>
<td>0.006</td>
</tr>
<tr>
<td>Dorsal premotor</td>
<td>Right</td>
<td>30</td>
<td>-8</td>
<td>62</td>
<td>22</td>
<td>4.41</td>
<td>3.55</td>
<td>0.01</td>
</tr>
<tr>
<td>Anterior SPL</td>
<td>Right</td>
<td>32</td>
<td>-50</td>
<td>58</td>
<td>38</td>
<td>4.39</td>
<td>3.54</td>
<td>0.01</td>
</tr>
</tbody>
</table>
3.4 Discussion

The results presented in this chapter highlight BOLD signals related to endogenously generated perceptual transitions that are common to both BR and SFM experimental paradigms (in other words, are stimulus-invariant). These signal patterns are located primarily in right aSPL, right insula and right IFG. Furthermore, signal in aSPL can be mapped to a particular SPL subregion on the basis of an existing parcellation of this brain area (Mars et al., 2011). This subregion has strong functional connectivity with FEF, PMd and inferior PMd in the resting state, and I have shown that two of these loci (FEF and PMd) also contain stimulus-invariant perceptual-transition-related BOLD signals. Overall, the current findings apply novel anatomical and functional frameworks to perceptual-transition-related BOLD activity and furthermore suggest that such activity overlaps closely with an established resting state network, implying that the neural mechanisms that support access to awareness relate to coordinated endogenous neural fluctuations.

3.4.1 Neural correlates of endogenous perceptual transitions during BR and SFM

The results I have presented for the contrasts endogenous perceptual transitions > replayed perceptual transitions for BR and SFM show substantial similarity both visually (Figure 3.3 A and B) and in terms of regions highlighted by the SPM statistical results (Table 3.2). Apart from right superior parietal and inferior frontal/insular regions, right precentral and left insular signal are also seen prominently in both cases. These findings are largely in accord with previous descriptions of fMRI signal patterns associated with perceptual transitions for various bistable paradigms (e.g. Lumer et al., 1998; Sterzer and Kleinschmidt, 2007; Weihhammer et al., 2013). However, I also found some differences between transition-associated activation patterns in BR and SFM with midline cingulate regions more prominently involved in the former and occipital regions more prominently involved in the latter. These differences may reflect the different stimulus properties and task demands of these two paradigms. Future studies could explore this possibility further by manipulating such properties and demands systematically.

Despite the use of identical statistical thresholds in the BR and SFM analyses, the activation pattern for SFM is more statistically robust and with larger clusters. This may reflect a larger discrepancy between processing demands for disambiguating direction of rotation for an ambiguous SFM stimulus and those for a non-ambiguous rotating sphere, compared to the discrepancy in processing ambiguous and non-ambiguous binocular...
images. Alternatively, the more robust activation patterns for SFM may reflect the relatively shorter TR (2.1 seconds for SFM versus 3.36 seconds for BR). It has been shown that shorter TRs improve statistical power for discriminating activated and non-activated brain tissue, although of course this usually comes at a trade-off with respect to brain coverage (Constable and Spencer, 2001).

3.4.2 Stimulus-invariant brain activity related to perceptual transitions

As discussed in Section 3.1, there is evidence for both commonality and distinctiveness in mechanisms of bistable perception across different paradigms of bistability. Perceptual transitions are associated with a consistent pattern of right frontoparietal fMRI-BOLD signal in existing literature (Kleinschmidt et al., 2012; Rees, 2007) but it remains unclear to what degree this activity reflects changes in perceptual experience per se, rather than relating to associated stimulus and task demands (see Section 1.3.5). Isolating transition-related activity that is common across two bistable paradigms with different stimulus properties and visual processing/task demands is one way to reduce the contribution of activity that may be additional to the essential perceptual transition machinery. However, it is important to note that this sort of method would not be expected to completely eliminate such additional signal components. Isolating stimulus-invariant activity across a wider range of bistable paradigms (including paradigms in different sensory modalities) may be a useful way to take this approach further in future. Other approaches to understanding more clearly the functional role of perceptual-transition-related activity include the use of causal methods such as TMS (e.g. Graaf et al., 2011; Kanai et al., 2011; this approach is taken in another experiment in this thesis presented in Chapter 5); the use of more temporally powerful methods such as EEG (e.g. Britz et al., 2009); or the use of invasive techniques in animals that permit more direct physiological intracranial recording (e.g. Thompson and Schall, 1999).

The present findings highlight a network of regions including right insula, right IFG, right aSPL, right FEF and right PMd where activity is associated with perceptual transitions across both BR and SFM paradigms. I therefore propose that these regions form a stimulus-invariant network involved in perceptual transitions. Such a finding provides clear anatomically defined targets that could be useful for testing further hypotheses about the involvement and functional roles of the implicated regions in perceptual transition.
3.4.3 Possible roles of regions within the stimulus-invariant network in access to awareness

The role in perceptual transition for the regions highlighted in the present stimulus-invariant findings remains unclear. However, there is relevant evidence in each case, which can shed some light on this or at least stimulate speculative suggestions.

The parietal lobe is a multimodal association area with anatomical connectivity to multiple sensory processing regions (Lewis and Van Essen, 2000). More specifically, SPL is heavily involved in goal-directed attention (Corbetta and Shulman, 2002), multisensory integration (Molholm et al., 2006; Wolpert et al., 1998) and episodic memory (Cabeza et al., 2008). Perceptual-transition-related activation in SPL has been reported across a number of fMRI studies of bistable perception and is robustly found in meta-analyses of such work (Bisienius et al., 2015; Rees, 2007). Furthermore, there are data from rTMS studies that support a causal role of SPL in the dynamics of bistable perception (Carmel et al., 2010; Kanai et al., 2010; Zaretskaya et al., 2010; see also Sections 1.3.5, 3.1, and 9.2.2). These strands of fMRI and TMS-derived evidence have led to a proposed subdivision of SPL into anterior (aSPL) and posterior (pSPL) regions (Kanai et al., 2011). The current stimulus-invariant activations are clearly located in aSPL, rather than more posteriorly. Moreover, having used the coordinates of the aSPL region previously highlighted by Kanai and colleagues for my small-volume corrected analyses (Section 3.2.7 and Table 3.1), my results suggest that the anterior SPL locus involved in the stimulus-invariant perceptual transition-related network is the same as the one shown to have a causal role in perceptual transition (Kanai et al., 2011). One proposal for the role of this aSPL region has been as a site where top-down predictions about the visual environment are transmitted to sensory regions, which generate perceptual representations. This predictive coding framework suggests that aSPL normally works to stabilise the current perceptual interpretation during bistability and this role is supported by existing data (Kanai et al., 2011; Vernet et al., 2015; see also Section 9.3).

The frontal eye field is involved in voluntary saccadic function (Rivaud et al., 1994) as well as priority mapping relevant to spatial attention (Ptak, 2012; see Section 5.1 for a more detailed review of relevant literature). Previous fMRI studies have also found activation in FEF at the time of perceptual transitions (Knappen et al., 2011; Sterzer et al., 2002). Disruption of human FEF function with TMS (Grosbras and Paus, 2003) or measurement of FEF activity with invasive single cell recordings in the macaque (Thompson and Schall, 1999) also suggest that this region plays a role in awareness. One may speculate that the established role of FEF in spatial attention and saliency coding (Ptak, 2012), and in
particular its proven modulatory effects on early visual areas (Ruff et al., 2006) might relate
to its role in awareness, which could be to stabilise or destabilise neural representations in
eye visual areas related to perceptually dominant (or indeed perceptually suppressed)
 aspects of bistable stimuli. These possibilities are discussed and explored more directly in
Chapter 5. It remains uncertain how any such role of FEF might interface with the
proposed functions of SPL in a single coordinated network subserving perceptual
transition; this is discussed further in Section 9.2.1.

The dorsal premotor cortex is involved in visuomotor learning, specifically for the selection
of upper limb motor responses based on learned conditional relationships with sensory cues
(Halsband and Freund, 1990; Petrides, 1997). Functional imaging both with PET (Grafton
et al., 1998) and fMRI (Toni et al., 1999) shows activation in this region in relation to
visuomotor conditional responses. The anatomical distinction of human PMd from human
FEF has been under some debate since fMRI findings suggest that regions activated
during eye movement (characterised as FEF) and regions activated during visuomotor
upper limb conditional activity (characterised as PMd) spatially overlap (Amiez and
Petrides, 2009). However, examining these foci of activation on a subject-by-subject basis
shows that PMd is reliably found to be relatively dorsocaudal to FEF, which is in line with
previous findings from primate work (Amiez et al., 2006). This is also reflected in the
currently outlined network of regions where PMd ($x = 28$, $y = -8$, $z = 64$) is dorsal to FEF
($x = 22$, $y = -10$, $z = 50$). It is not clear what role PMd may play in the context of
perceptual transition. It is possible that the region’s involvement relates to some aspect of
the motor performance involved in both the bistable tasks. However, motor performance
was equated across rivalry and replay conditions and in addition the two sets of activations
(from BR and SFM) were masked with each other so there should be minimal residual
motor-related activation present. It may be that transition-related BOLD signal in PMd
more generally reflects preparation for motor responses to changing perceptual
interpretations of the environment, although this is a speculative suggestion that will need
to be directly tested in future experiments.

Of note, no stimulus-invariant perceptual transition-related activation was found in inferior
PMd. This region also forms part of the aSPL-associated resting-state connectivity
network (Mars et al., 2011) and was therefore included in my a priori predictions (Table
3.1). Inferior PMd has previously been implicated in visuomotor exploration (Hinkley et
al., 2009) and the current findings provide no support for its involvement in perceptual
transitions.
With regards to insula and nearby IFG/opercular cortex, activation in these regions is well established in association with perceptual transitions (Lumer and Rees, 1999; Rees, 2007; Sterzer and Kleinschmidt, 2007; Weilnhammer et al., 2013). Besides the insula’s involvement in emotional processing and risk evaluation discussed elsewhere (Section 6.5.2), this region has been proposed to support amplification of visual signal or to subserve increased alertness related to stimulus or task demands, rather than to be coupled to awareness of a stimulus per se (Sterzer and Kleinschmidt, 2010). On the other hand, an influential account posits that the insula is of central importance to awareness of the immediate moment since it is in a position to integrate information about internal bodily states, emotion and perception of the external environment (Craig, 2009).

Finally, an important question is how the regions in the presently outlined stimulus-invariant perceptual transition-related network interact functionally. This question is addressed in the next chapter of the thesis (Chapter 4).

3.4.4 SPL subregions, associated resting state networks and perceptual transition

The availability of an anatomical parcellation of SPL as well as resting state networks associated with each proposed SPL subregion (Mars et al., 2011) has provided a very useful framework within which to place the present findings. The stimulus-invariant perceptual-transition-related activation in aSPL overlapped with the most anterior SPL subregion described by Mars and colleagues (2011); this was also closely aligned with the aSPL region described in other studies of perceptual transition (Kanai et al., 2011; Figure 3.6). Of note, no stimulus-invariant activation in pSPL, another region proposed to play a role in perceptual transition, was found (Kanai et al., 2011; Table 3.1). While the role of pSPL in perceptual transition was robustly established in the work of Kanai and colleagues (Kanai et al., 2010, 2011) and has been extended in more recent work exploring effective connectivity between aSPL, pSPL and V5 during bistable perception (Megumi et al., 2015), those findings have been reported specifically in the context of SFM paradigms. Indeed, perceptual transition-related pSPL activation was not found in the BR experiments presented in this thesis (Chapter 7 and this chapter). The current results therefore do not support a stimulus-invariant role for pSPL in perceptual transitions, suggesting it may be of more specific importance for bistability in the case of SFM (and perhaps other paradigms involving motion and/or interpretation of three-dimensional structure), rather than being generally important across different bistable stimuli and tasks.
The resting state functional connectivity of the most anterior subregion of SPL provides important insight into its intrinsic relations with other brain regions. Such networks can be closely related (although are of course not synonymous) with anatomical connectivity (e.g. Fox and Raichle, 2007). Moreover, there is ample evidence that resting state functional connectivity relates to task-related brain activity networks (Smith et al., 2009); one possibility is that spontaneous network dynamics relate to groups of regions that have been modulated or co-activated together in a task-dependent manner (Fox and Raichle, 2007). Thus, there are some grounds to hypothesise that the functional connectivity patterns established by Mars and colleagues (2011) may also be co-activated with aSPL during perceptual transition; this has been confirmed by the analyses presented in this chapter. The present findings support a link between intrinsic connectivity and the mechanisms driving the dynamics of bistability (in particular perceptual transitions); more generally the results also lend further support to the notion that intrinsic connectivity can reflect commonly occurring task-related activation patterns.

![Figure 3.6: Overlap between SPL subregions defined by Kanai et al. (2010, 2011) and the anterior SPL subregion from the parcellation of Mars et al. (2011)](image)

Subregions defined by Kanai and colleagues (2010, 2011) shown in red and the anterior SPL subregion from the parcellation of Mars and colleagues (2011) shown in cyan. The spatial overlap with aSPL (anterior red mask) but not pSPL (posterior red mask) is apparent. Statistically, stimulus-invariant transition-related activity was also present both in aSPL (Section 3.3.5) and in the anterior SPL subregion of Mars and colleagues (2011; see Section 3.3.4), but not in pSPL. Masks are shown on a brain rendering in MRIcron.
A related question is whether bistability is an inherent property of the brain that is nested in spontaneous neuronal dynamics, rather than (or as well as) being a task-related property. There is evidence to suggest that pre-stimulus BOLD signals predict which version of a bistable stimulus will subsequently be perceived (Hesselmann et al., 2008), and that alpha oscillation power predicts the persistence of bistable percepts (Piantoni et al., 2017), suggesting that endogenous neuronal fluctuations influence perceptual inference. Moreover, it is important to point out that bistable perception in itself is not a task but rather a spontaneous fluctuation in perception (albeit with some susceptibility to top-down control; e.g. Meng and Tong, 2004). Paradigms of bistable perception commonly include a task (such as pressing buttons to report perceptual experience) as well as visual stimulation. While characteristic perceptual transition-related patterns of BOLD signal can be seen independently of button press tasks (Lumer and Rees, 1999), in practice it is difficult to dissociate bistability completely from task demands. Peripheral physiological factors may also influence spontaneous neural dynamics; for example, the likelihood of consciously perceiving a visual stimulus at detection threshold relates to timing in the cardiac cycle (Park et al., 2014).

If the mechanisms of perceptual transition in bistable perception indeed relate to intrinsic non-task-dependent networks of functional connectivity this may explain the similar activation patterns seen across different bistable stimuli and tasks. To what degree such resting state neural connectivity patterns relate to the dynamics of bistable perception is a relatively unexplored avenue of inquiry (though see Baker et al., 2015). I will attempt to address this question next, in Chapter 4 of this thesis, by exploring individual variability in effective resting state connectivity and whether this relates to individual dynamics of bistable perception.

### 3.4.5 Is frontoparietal BOLD signal cause or consequence of perceptual transitions?

There is ongoing debate regarding the role of frontoparietal BOLD signal associated with perceptual transitions in bistable perception (discussed also in Sections 1.3.5 and 9.2.3). A body of literature suggests that these activation patterns reflect top-down mechanisms and are associated with the *cause* of perceptual reinterpretation (e.g. Sterzer and Kleinschmidt, 2007; Weilnhammer et al., 2013) while other evidence suggests that such activation is a *correlate* of perceptual transitions and is not in any way reflective of their cause (Brascamp et al., 2015; Knapen et al., 2011; evidence reviewed in more detail in Section 1.3.5).
A key argument put forward by Knapen and colleagues (2011) is that a number of previous studies (e.g. Lumer et al., 1998) have potentially overestimated frontoparietal activation associated with endogenous perceptual transitions by contrasting long periods of mixed perception in rivalry (which were modeled as transitions) with instantaneous changes during replay (also modeled as transitions; replay conditions discussed in Section 3.2.4). Knapen and colleagues (2011) have claimed that if perceptual transitions are instantaneous (and matched in duration) across ambiguous rivalry and non-ambiguous replay, associated frontoparietal activation is negligible, and such activation is therefore postulated to represent responses to mixed percepts and associated processes of sustained attention (rather than representing active reinterpretation of perceptual experience). The current findings go against such proposals since the perceptual transitions in both paradigms studied here were essentially instantaneous. In the case of the BR paradigm this was achieved by modelling a mixed percept\(^1\) (in addition to face and grating percepts); in the SFM paradigm this was achieved by the nature of the stimulus, where transitions are typically rapid (Megumi et al., 2015). The fact that these uniformly sharp transitions (when contrasted between rivalry and replay) were still associated with robust frontoparietal BOLD signal in both paradigms suggests that these signals are not merely reflective of prolonged transitions/mixed percepts. Further support for this comes from an independent study using a different type of bistable stimulus where the length and nature of rivalrous and replayed perceptual transitions were carefully matched (Weilnhammer et al., 2013). On the other hand the frontoparietal stimulus-invariant activation patterns I have demonstrated have relatively limited spatial extent and smaller cluster sizes, perhaps suggesting that at least some of the stimulus-specific transition-related activations are epiphenomenal (in the sense of not being directly related to perceptual transition \textit{per se}). The significance of the present findings in the context of this existing literature is discussed further in Section 9.2.3.

3.4.6 Summary

In this chapter two separate experiments combining fMRI with different paradigms of bistable perception independently produce similar patterns of perceptual-transition-related BOLD signal in right-sided frontal and parietal cortex. Masking these two signal patterns inclusively allowed me to highlight a stimulus-invariant network of activity associated with endogenously generated perceptual transitions during bistability. This activity was focused

\(^1\) These mixed percepts are equivalent to the prolonged perceptual transitions described in previous studies (where mixed percepts were assumed to be part of the perceptual transitions themselves; Lumer et al., 1998; Sterzer and Kleinschmidt, 2007). Presently mixed percepts were modelled as separate perceptual states, ensuring that all transitions between different perceptual interpretations were instantaneous in the current paradigm (both in rivalry and replay conditions).
mainly over right-hemisphere regions including aSPL, insula and IFG. The signal in aSPL could be mapped to the most anterior subregion from a proposed parcellation of SPL (Mars et al., 2011). I was able to show that stimulus-invariant perceptual-transition-related activity was also found in regions with strong resting-state-functional connectivity with this SPL subregion, including PMd and FEF.

The findings delineate a pattern of BOLD signals that are associated with perceptual transitions across different bistable paradigms and moreover show that this activity is closely associated with resting-state functional connectivity networks, lending support to the notion that bistability relates to intrinsic brain activity fluctuations and interactions. In addition, the findings add to the debate regarding the role of frontoparietal activity in perceptual transition, suggesting that such activity is not simply stimulus or task-related.

The regions of the presently outlined stimulus-invariant perceptual transition-related network provide targets for further exploration of the mechanisms of perceptual transition and access to awareness. In following chapters of this thesis I will use these findings to motivate study of the relationship between resting-state effective connectivity among this network and the dynamics of bistable perception at the individual level (Chapter 4) as well as to probe the causal roles of some of these regions in endogenous perceptual transition (Chapter 5).
Chapter 4

Effective resting-state connectivity and individual dynamics of bistable perception

“The brain is a world consisting of a number of unexplored continents and great stretches of unknown territory.”

Santiago Ramón y Cajal (1852 – 1934)

4.1 Introduction

The mechanisms of perceptual transition, or access to awareness, in the human brain remain incompletely understood (Section 1.3.5) and the first half of this thesis focuses primarily on their further exploration. The findings presented in Chapter 3 replicated and then refined existing findings regarding perceptual-transition-related fMRI-BOLD activation, particularly highlighting stimulus-invariant patterns of signal and their relationship to established resting-state functional networks. I was thus able to outline a group of brain regions in right inferior frontal and superior parietal cortex that are involved in perceptual transition across different visual stimuli and tasks, and which appear to share intrinsic functional connectivity (see Section 2.2.4 for a brief discussion of resting-state fMRI methods).

This demonstration of spatial concordance between fMRI-BOLD signals associated with perceptual transitions (in a stimulus-invariant fashion) and established resting-state functional networks may suggest that (intrinsic) networks of correlated resting-state activity contribute to the mechanisms of perceptual transition. Since bistable perception
dynamics show considerable inter-individually variability (see Section 1.6), a key related question that remains unanswered is whether measures of resting-state connectivity among nodes in the stimulus-invariant perceptual-transition-related network relate to bistable perception dynamics at the individual level. If so, one might be able to demonstrate that the interactions (connections) between certain brain regions are particularly predictive of the dynamics of bistability.

The loci shown to have stimulus-invariant perceptual-transition-related BOLD signals in Chapter 3 were found in right insula, right IFG, right aSPL, right FEF and right PMd. In the present experiment my aim was to measure effective connectivity between these regions in the resting state, and to explore any correlations between specific effective connection strengths within individuals and behavioural measures of bistable perception for the same individuals.

Methods for evaluating connectivity between brain regions can be divided into those that focus on structural connections (anatomical connectivity), those that evaluate statistical dependencies between regional brain activity (functional connectivity) and those that interrogate causal interactions between regions (effective connectivity; see Section 2.2.5; Sporns, 2007). Some of these different approaches to characterising connectivity have been applied to the study of bistable perception. For example, it has been shown that experience of ambiguous apparent motion relates to structural properties of transcallosal connections between motion areas in the two hemispheres (measured using DTI; Genç et al., 2011). Recent work has explored the functional connectivity of SPL and its subregions with other parts of the brain in the resting state, and the relationship of this connectivity to the dynamics of perceptual changes when viewing ambiguous figures (Baker et al., 2015). Finally, effective connectivity analyses with DCM have been used to show that perceptual switches (and the related BOLD signal) can be associated either with modulation of top-down connectivity from frontal to visual regions (Weinhammer et al., 2013) or of bottom-up coupling from visual to parietal regions (Megumi et al., 2015). While resting-state fMRI data have been used for functional connectivity analyses of bistable perception-related networks (Baker et al., 2015), measurements of resting-state BOLD signals have not been applied to effective connectivity analyses in this context and such an approach could further our understanding of how intrinsic neuronal networks relate to mechanisms of perceptual reinterpretation.

Dynamic causal modelling (Friston et al., 2003) has become one of the most established methods for evaluating effective connectivity. A brief discussion of the principles underpinning DCM and the differences between various implementations of this framework are included in Section 2.2.5. Of particular relevance to the present experiment
is the fact that the traditional deterministic version of DCM requires an input to the model and does not account for endogenous fluctuations in neuronal states, making it less suited for resting-state analyses like those planned currently. More recent stochastic DCM implementations overcome these limitations, albeit at a significant computational cost, and are thus better suited to modelling of resting-state fMRI signal. An even newer version of DCM based on spectral analysis methods has potential advantages over stochastic DCM, although the more established stochastic method will be used presently (see also Section 2.2.5).

Aiming to link behavioural performance in bistable perception with resting-state effective connectivity at an individual level, I plan to enroll the same sample of individuals into two experiments, conducted on separate days. In an initial behavioural experiment participants will perform two different bistable visual tasks (BR and SFM). The reasoning for choosing these two paradigms is similar to that discussed in Chapter 3 (Section 3.1). Briefly, the two paradigms have different stimulus characteristics and place distinct demands on visual processing that rely on different parts of the visual system (in the present case the two tasks involve perception of colour versus perception of three-dimensional structure). The aim of the behavioural experiment will be to obtain robust measures of bistable perception dynamics (in particular mean percept duration) on both paradigms for each individual. The second experiment will involve acquisition of resting-state fMRI data on a separate day to the behavioural experiment. This will enable analysis of resting-state effective connectivity between predetermined ROIs, in which I have already demonstrated perceptual-transition-related BOLD signal changes in a stimulus invariant fashion (these include insula, aSPL, FEF and PMd, based on findings from Chapter 3; see Section 4.2.6 and Table 4.1 for details regarding these ROIs and their selection). Correlations between behavioural measures of bistable perception dynamics relating to each individual, and perceptual transition network-specific resting state effective connectivity measures for the same individual will then be explored to determine whether intrinsic connectivity in this network is linked to the dynamics of bistable perception.

### 4.2 Materials and methods

#### 4.2.1 Contribution acknowledgement

Six resting-state fMRI datasets (Section 4.2.4) had previously been collected as part of a separate experiment, the results of which have been published (Urner et al., 2013). Maren Urner kindly allowed for these data to be used in the present experiment. The remainder of
the experimental work presented in this Chapter is my own. I re-recruited the six participants from the experiment of Urner and colleagues to perform the behavioural experiment only (Section 4.2.3). The remaining twenty-eight participants recruited also performed the behavioural experiment and I then collected resting-state fMRI data for each of them using the same scanner and running the same imaging sequences with the same parameters as those from the study of Urner and colleagues (2013). Instructions given to participants during the scanning sessions were also identical. Full details of external contributions can be found in the Contributions and publications section at the start of the thesis.

4.2.2 Participants

I recruited thirty-four participants (21 female; mean ± SD age = 25.8 ± 6.4 years; range = 18 to 45 years) according to the procedures and inclusion criteria listed in Section 2.4.1. A target sample size of between 30 and 40 was chosen based on samples used in previous studies exploring individual variability in functional connectivity (reviewed in Mueller et al., 2013).

4.2.3 Behavioural experiment: stimuli, display apparatus and experimental procedures

The BR and SFM experimental paradigms were programmed using Psychtoolbox. Stimuli were presented on a 22-inch LCD monitor (Samsung SyncMaster 2233RZ; spatial resolution 1680 x 1050; monitor refresh rate 60Hz) and viewed at a distance of 72 cm. I made use of a standard set of instructions and experimental apparatus, as described in Section 2.4.2. The mean duration of bistable percepts, averaged across all experimental trials, was the dependent measure for both paradigms.

4.2.3.1 Binocular rivalry experimental procedure

Procedures for the binocular rivalry experiment were as described in Section 2.4.4. A mirror stereoscope set-up was used and two rotating gratings (blue and red) were shown on the screen (each subtending 3.4° visual angle in diameter) continuously for 40 seconds at a time (Section 2.4.4.3; Figure 4.1A). The rotation helps reduce mixed percepts so that two alternating perceptual interpretations predominate, as is the case for SFM.
After two practice trials, participants completed 8 experimental trials. Prior to the practice and experimental trials each participant performed a four-trial staircase task to adjust the intensity of components from the RGB model (Hunt, 2004) that determined the colours of the red and blue gratings (a maximum intensity red grating would be expressed as \( r,g,b = 255,0,0 \) and a maximum intensity blue grating would be expressed as \( r,g,b = 0,0,255 \)). The aim was to modify the colour intensities of the two gratings until their brightness was subjectively equal for that participant. The resulting RGB values for each grating were then used for all of that individual’s binocular rivalry trials. This ensured individual dynamics of binocular rivalry were not affected by systematic differences in perceived brightness between the blue and red stimuli.

![Stimuli used in BR and SFM behavioural experiments](image)

**Figure 4.1:** Stimuli used in BR and SFM behavioural experiments

Behavioural experiments described in Section 4.2.3. In the BR experiment (A) participants used a mirror stereoscope to view two orthogonal gratings (one visible to each eye; one red and one blue in colour; both with a central fixation dot), which both rotated at a constant speed in a clockwise direction. This configuration induced binocular rivalry, where perceptual experience predominantly alternated between exclusive red grating and exclusive blue gating percepts, with interspersed shorter periods of mixed percepts. In the SFM experiment (B) participants viewed (without a mirror stereoscope) a single image consisting of a collection of white sinusoidally moving dots displayed on a black background and with a central red fixation dot. This configuration was perceived alternately as a sphere rotating to the right or as a sphere rotating to the left (C). Figure in panel A adapted from Haynes and Rees (2005b).
4.2.3.2 Structure-from-motion experimental procedure

The stimulus and procedure for the SFM session were as described in Section 2.4.5, and as employed in Chapter 3. However, in this case a single image was shown to both eyes and the configuration of dots was perceived as a sphere rotating about its vertical axis (left or right), with a diameter subtending 3.6° visual angle and a central fixation dot subtending 0.1° visual angle (Figure 4.1B-C). Following a practice trial, an experimental run of 10 trials was completed. Each trial was 48 seconds in duration.

4.2.4 Resting-state fMRI: experimental procedures

Participants lay in the scanner with only a dim light on in the scanning room. During a 10-minute resting state scanning session, they were asked to, “Lie still, relax, and keep your eyes closed, please. Try not to fall asleep”. Care was taken to provide every participant with exactly the same instructions. The camera of a non-ferrous infra red Eyelink eye tracking system installed in the MRI scanner was left on, to monitor whether participants’ eyes remained closed according to the instructions, and to closely observe behaviour immediately after the end of scanning for any visible indication that a participant may have fallen asleep (any sign of possible sleep was also evaluated based on verbal response to the question ‘How are you feeling’ asked through the MRI intercom immediately at the end of the 10-minute resting state scan).

4.2.5 Resting-state fMRI: data acquisition

Acquisition of resting-state fMRI data, and measurement of heart rate and breathing, was undertaken according to the procedures and using the parameters described in Section 2.4.7. A total of 32 axial slices were acquired per volume with slice acquisition time 68 ms, TE 30 ms, TR 2.176 s, matrix size 64 x 64, field of view 192 x 192 mm. I obtained 276 volumes in total, corresponding to 10 minutes of resting state data.

4.2.6 Resting-state fMRI: data preprocessing, GLM, and VOI signal extraction

Image preprocessing was undertaken according to the methods described in Section 2.4.8. Low-frequency fluctuations in the signal were removed by using a high-pass filter with cut-
Implementation of the GLM allowed the resting state data to be fitted with a physiological noise model constructed to account for fluctuations in the BOLD signal related to cardiac and respiratory phase, using an SPM toolbox (Hutton et al., 2011). As a result, 14 regressors representing these physiological parameters were constructed based on the cardiac pulse and breathing measurements sampled at a reference slice in each image volume. I added six further regressors representing head movement parameters, derived from the realignment preprocessing step for each image volume. The resulting 20 regressors were placed in the GLM as covariates of no interest, alongside the corresponding 276 EPI volumes. Given that there was no experimental task, no regressors of interest were added to the GLM.

Since the experimental hypotheses required me to determine resting-state effective connectivity between regions that (in the task state) contain BOLD signal related to perceptual transitions (Chapter 3), the next step was to extract local activity time series from each of these regions. The stimulus-invariant perceptual transition network outlined in Chapter 3 was highly similar irrespective of the exact procedure for combining results relating to BR and SFM (see Section 3.3.5 and Table 3.4); therefore, the region coordinates resulting from masking BR transitions inclusively with SFM transitions were used as the basis of the present analysis (see Table 3.4, upper section). The insula and IFG regions in the stimulus-invariant network were consistently very close to each other (see Table 3.4) and in order to simplify the present effective connectivity analysis (by including four interconnected regions in the connectivity model rather than five), I chose to use only the insula region (which represented a larger cluster of activation with a stronger T value than for IFG; Table 3.4, upper section).

Based on the above considerations, four regions centred at the following coordinates formed the network presently submitted to DCM analysis: right aSPL $x = 30, y = -46, z = 58$; right FEF $x = 22, y = -10, z = 50$; right PMd $x = 28, y = -8, z = 64$; right insula $x = 36, y = 12, z = -6$ (see Table 4.1 and Figure 4.5A). I used spheres with a 10mm radius around each of these coordinates for VOI signal extraction using SPM8 to obtain the principal eigenvariate time series around participant-specific local maxima of activation nearest to each of these four coordinates (Figure 4.2). In order to ensure that the time series were extracted from the correct anatomical areas at individual participant level, anatomical masks (spheres with 10mm radius, constructed using the Pick atlas) were used to constrain local peak maxima searches. All voxels contributing to the eigenvariates exceeded a significance threshold of $p < 0.05$ uncorrected. This signal extraction procedure produced
four time series vectors per participant (each corresponding to one of the four VOIs and containing 276 values, one value for each resting state EPI volume acquired).

<table>
<thead>
<tr>
<th>LOCATION</th>
<th>MNI COORDINATES</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>aSPL</td>
<td>Right 30 -46 58</td>
<td>Chapter 3 (Table 3.4, upper section)</td>
</tr>
<tr>
<td>FEF</td>
<td>Right 22 -10 50</td>
<td>Chapter 3 (Table 3.4, upper section)</td>
</tr>
<tr>
<td>PMd</td>
<td>Right 28 -8 64</td>
<td>Chapter 3 (Table 3.4, upper section)</td>
</tr>
<tr>
<td>Insula</td>
<td>Right 36 12 -6</td>
<td>Chapter 3 (Table 3.4, upper section)</td>
</tr>
</tbody>
</table>

**Table 4.1**: Anatomical location of brain regions to be used in DCM analysis

Regions obtained from the stimulus-invariant perceptual-transition-related network outlined in Chapter 3. For each participant, BOLD signal for local maxima within a 10mm-radius spherical VOI (centred at each of these coordinates), was extracted for DCM analysis (Sections 4.2.6 and 4.2.7; Figure 4.2).

**Figure 4.2**: Graphical representation of VOI signal extraction procedure using SPM8

This figure illustrates the principal eigenvariate time series (right panel) for a 10mm radius sphere around the aSPL coordinate (left panel) for one example participant. Similar time series were extracted from spheres around all regions used in the DCM analysis (Table 4.1), for each participant in turn (see Section 4.2.6).
4.2.7 Dynamic causal modelling

DCM12 (implemented in SPM12) was used to model effective connectivity between the four predetermined ROIs (Table 4.1). I chose to use stochastic DCM, which is well suited to exploring resting-state effective connectivity, since this type of approach contains no requirement for inputs (corresponding to experimental perturbations); moreover endogenous fluctuations in hidden neuronal states are accounted for in the model (Li et al., 2011). The general principles of DCM and the key differences between deterministic and stochastic versions are discussed further in Section 2.2.5; see also Sections 4.1 and 4.4.3. I undertook model selection analysis, in order to determine the pattern of effective connectivity between the regions in my model that best reflects the observed data, using a *post hoc* approach (Friston and Penny, 2011). The differences between this method and more traditional Bayesian model selection approaches are discussed in more detail in Section 2.2.5. The *post hoc* model selection method allows for comparison between a large number of possible models in a computationally efficient manner and is non-hypothesis-driven so well suited to situations where there are no strong prior hypotheses about directional interactions between brain regions.

*Post hoc* model optimisation involves inverting a fully connected model (assuming bidirectional connections between all ROIs, resulting in 12 extrinsic connections between the four ROIs used in the present case) and then searching over all possible reduced versions of this full model before selecting the best one. The procedure takes all free parameters in the fully connected model and searches over all permutations of the 8 parameters whose removal produces the smallest reduction in model evidence. This produces 256 models and from these the model with the greatest evidence is picked for the next iteration of the procedure. This continues until all parameters are retained in the best model (i.e. no more connections can be pruned) and this best model is found within the final model space of 256 models (Figure 4.5C illustrates the results of this procedure for the current dataset).

The posterior probability of each model in the final model space is compared using family partitioning (e.g. all models containing a given connection are compared with all models that do not contain that connection). Essentially the posterior probability is the probability that a model (or family of models) provides the best explanation for the measured data across participants (Penny et al., 2004). Bayesian parameter averages for each connection contained in the model with the highest evidence provide a quantitative measure of the ‘strength’ of that connection, in effective connectivity terms, and these can be evaluated using conventional frequentist statistics (Stephan et al., 2010). In the present case, I
evaluated correlations between the BPA for each directional connection and mean percept duration in the behavioural tasks (BR and SFM) at individual participant level.

In summary, for each participant, a fully connected model with 12 extrinsic connections between the four ROIs (see Figure 4.5A) was inverted using stochastic DCM with generalised filtering (Li et al., 2011). The assumption was that accounting for endogenous fluctuations in neuronal activity would allow accurate modelling of resting state activity. Since the aim was to detect individual differences in intrinsic (resting-state) effective connectivity, no bilinear or modulatory effects were modelled, and the model contained no inputs. Post hoc model optimisation was then performed and I examined BPAs (representing effective connectivity 'strength') for each of the connections retained in the final model. Specifically, I tested whether individual parameter estimates for each connection in turn were correlated with measures of bistable perception dynamics for the same individual.

4.3 Results

Seven participants were excluded from results analysis for the following reasons: four fell asleep during resting-state fMRI acquisition, two reported frequent disruption of perceptual fusion during the bistable perception behavioural tasks and one had significant periods of missing responses during the behavioural tasks (due to incomplete understanding of the instructions). The remaining sample, used for subsequent analyses, contained twenty-seven participants (18 female; mean ± SD age = 24.8 ± 5.2 years; age range = 18 to 36 years).

4.3.1 Behavioural experiment results

There was substantial inter-individual variability in mean percept duration for both BR and SFM (Figure 4.3A-B). The mean percept duration for BR across the whole sample of 27 participants was considerably shorter than that for SFM (3.2 and 9.3 seconds, respectively). There was a moderately strong positive correlation between individual mean percept durations for BR and for SFM \[ r = 0.35; p = 0.07; \] Figure 4.3C\].

4.3.1.1 Binocular rivalry results

In the binocular rivalry experiment, mean percept duration across the whole sample of participants was 3.2 seconds (SD = 0.79 seconds; range = 2.0 – 5.5 seconds). Mean
duration was 3.2 seconds for red percepts and also 3.2 seconds for blue percepts with no significant difference between the two \(t_{(26)} = 0.18, p = 0.86\]. Mean duration of percepts relating to stimuli seen with the right eye was 3.3 seconds and that for stimuli seen with the left eye was 3.1 seconds. Percepts for right eye stimuli were significantly longer \(t_{(26)} = 2.1, p = 0.05\]. Hierarchical repeated-measures ANOVA with participants nested in eye dominance (two levels) and repeated measures on eye receiving the perceived stimulus (two levels) showed that while there was a significant main effect of eye receiving the perceived stimulus \(F_{(1,25)} = 4.9, p = 0.04\], there was no main effect of eye dominance \(F_{(1,25)} = 1.9, p = 0.18\] and no perceiving eye*eye dominance interaction \(F_{(1,25)} = 0.94, p = 0.34\]. Mean duration for periods with no button press (corresponding to an unclear or mixed percept) accounted for a relatively low proportion of the overall duration of stimulus presentation (mean 8.4%, SD = 5.4%, range = 0.44 – 20%).

The intensities of the RGB colour components for the red and blue gratings were adjusted to achieve subjectively equal brightness of the two gratings for each participant (see Section 4.2.3.1). The resultant mean RGB values were \([186,0,0]\] for the red grating (intensity range for the red component was 159-207); and \([0,0,248]\] for the blue grating (intensity range for the blue component was 237-251). The red-blue component intensity difference and the mean BR percept duration for the same individual were not correlated \(r = -0.10, p = 0.61\].

### 4.3.1.2 Structure-from-motion results

For the SFM experiment, group-level mean percept duration was 9.3 seconds (SD = 4.3 seconds; range = 4.3 – 19.6 seconds). Mean duration of rightward-rotating sphere perception was 8.9 seconds and mean duration of leftward-rotating sphere perception was 9.7 seconds. The difference appeared not to be statistically significant \(t_{(26)} = -1.7, p = 0.11\]. However, this changed when eye dominance was accounted for. Hierarchical repeated-measures ANOVA with participants nested in eye dominance (two levels) and repeated measures on direction of rotation (two levels) revealed a significant main effect of direction of rotation \(F_{(1,25)} = 6.0, p = 0.02\]; no main effect of eye dominance \(F_{(1,25)} = 0.05, p = 0.82\]; and no significant direction of rotation*eye dominance interaction \(F_{(1,25)} = 3.1, p = 0.09\]. Periods when perceptual interpretation for SFM was not clear and did not clearly correspond to a leftward or rightward rotating sphere (indicated by no button press) accounted for 4.7% of overall stimulus presentation (SD = 3.8%; range = 1.6 – 19.8%).
Figure 4.3: Behavioural experiment results

See Section 4.3.1. Histograms of mean percept durations for the BR experiment (A) and SFM experiment (B). Considerable inter-individual variability can be seen in both cases. It is also apparent that percept durations were substantially longer for SFM than they were for BR; however, there was a moderate correlation between the two behavioural measures ($r = 0.35$; C).

4.3.1.3 Fit of individual percept duration distributions to a gamma function

I evaluated each individual participant's distributions of BR and SFM percept durations using inbuilt MATLAB code, based on the assumption that such data are usually distributed in close approximation to a gamma function (Blake and Logothetis, 2002; Fox and Herrmann, 1967; see Section 7.2.3.1; although see also Brascamp et al., 2005 for a critique of this premise). A shape parameter (representing the shape of the distribution and related to its skewness or kurtosis) and a scale parameter (representing how spread out the distribution is, and related to measures of dispersion) were calculated in each case. The distributions for three example participants are shown with histograms in Figure 4.4 (red dashed lines show fit to gamma distributions). I then used a chi-square goodness-of-fit test to determine whether the null hypothesis (that the data come from a gamma distribution) should be rejected at the 1% significance level. Generally, the data showed an acceptable fit to a gamma distribution in the majority of cases: the null hypothesis was rejected for only two (of a total 27) participants based on BR distributions, and for two different participants based on SFM distributions. Three of these participants did not represent statistical
outliers (according to Chauvenet’s criterion); removing the one that did from subsequent analysis did not alter the effects reported later (see Section 4.3.2.2).

The parameters describing fit of the data to gamma distributions were strongly intercorrelated within each paradigm with strong negative correlations between BR shape parameters and BR scale parameters \( r = -0.55; p = 0.003 \) as well as between SFM shape parameters and SFM scale parameters \( r = -0.55; p = 0.003 \). Across paradigms there was a strong positive correlation between BR shape parameters and SFM shape parameters \( r = 0.55; p = 0.003 \) but no correlation between BR scale parameters and SFM scale parameters.

![Figure 4.4: Percept duration histograms from BR and SFM behavioural experiments](image)

See Section 4.3.1.3. Three example participants are shown (left, middle and right; separated by blue dashed lines). For each participant BR data are shown in the upper row and SFM data in the lower row. Red dashed lines show fit of each data set to a gamma distribution. Percept duration (seconds) is shown along the x-axes and proportion of trials within each time bin is shown along the y-axes.

### 4.3.2 DCM results

#### 4.3.2.1 Post-hoc Bayesian model selection

The results of the model selection procedure (Section 4.2.7) confirmed that the model containing bidirectional connections between all four ROIs (the fully connected model; Figure 4.5B) was the winning model, with a probability of 0.42 (Figure 4.5C). Given that
the ‘best model’ approach can become brittle in post-hoc Bayesian model selection, where a large number of models are compared (Hillebrandt et al., 2013; Penny et al., 2010), I also examined the family-level inference results. These showed that the posterior probability (over parameters) for reciprocal connections between all four regions was 1 in all cases. These findings suggest that none of the alternative possible models outperformed the full model. Therefore, the fully connected model was selected for further analysis.

4.3.2.2 Relationship between individual effective connectivity and dynamics of bistable perception

I examined the BPA values for each connection in the selected fully connected model at the individual participant level and asked whether these were correlated with either of the behavioural measures (mean BR percept duration and mean SFM percept duration). I found that the BPA for directional connectivity between insula and aSPL was strongly positively correlated with mean BR duration \(r = 0.49; p_{\text{uncorrected}} = 0.009; p_{\text{Bonferroni corrected}} < 0.05\); see Figure 4.5D\(^2\). However, values relating to directional connectivity between insula and aSPL were not correlated with mean SFM duration \(r = 0.02; p_{\text{uncorrected}} = 0.91\). None of the individual BPAs for any of the other 11 unidirectional connections in the fully connected model were correlated with individual BR or SFM duration (see Table 4.2).

Next, I checked whether the gamma distribution shape and scale parameters relating to individual participant percept duration distributions in BR and SFM (see Section 4.3.1.3) were correlated with BPAs for the directional connection from insula to aSPL in the DCM model. I found that there was a moderate positive correlation with the BR distribution scale parameter \(r = 0.41; p = 0.03\). In other words, a more spread out distribution of percept durations in BR was predictive of a stronger effective connection from insula to aSPL. Neither the BR distribution shape parameter nor the two parameters for the SFM distributions were correlated with BPAs for the insula to aSPL connection. No other directional effective connections in the model were significantly correlated with shape or scale parameters for either BR or SFM distributions.

\(^2\) A partial Bonferroni correction was performed since the BPAs for the different directional effective connections were strongly inter-correlated (mean \(r = 0.53\).
The four ROIs are shown rendered on a template brain (A), and shown schematically with all 12 bidirectional connections between them on an inflated template brain (B). The results of post-hoc Bayesian model selection are shown in (C) with the left hand panel showing the range of log-posterior probabilities and the right hand panel showing the posterior probabilities for all models in the final iteration of the model selection procedure (see Section 4.2.7). Model 256, with the highest posterior probability of 0.42, was the fully connected model, which was accordingly selected for further analysis (see Section 4.3.2.1). When examining correlations between individual BPAs for each of the 12 connections in the model and individual mean percept duration (on BR and SFM tasks), I found a statistically significant correlation between directional effective connectivity from insula to aSPL and mean BR percept duration (D). The correlation plot is shown in the left hand panel and a schematic of the relevant connection in the model (highlighted in red) is shown in the right hand panel.
### Table 4.2: Statistical results for correlations between individual mean BR durations or mean SFM durations and individual Bayesian parameter average values

For each of the 12 unidirectional connections between the four regions of interest the Pearson coefficients ($r$) and $p$ values for correlations between individual mean BR durations or mean SFM durations and individual Bayesian parameter average values are shown. A statistically significant correlation between BPA for the insula $\rightarrow$ aSPL connection and BR duration is shaded.

<table>
<thead>
<tr>
<th>Connection</th>
<th>BR Duration $r$</th>
<th>BR Duration $p$</th>
<th>SFM Duration $r$</th>
<th>SFM Duration $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>aSPL $\rightarrow$ FEF</td>
<td>0.13</td>
<td>0.53</td>
<td>0.05</td>
<td>0.82</td>
</tr>
<tr>
<td>aSPL $\rightarrow$ PMd</td>
<td>0.13</td>
<td>0.52</td>
<td>0.11</td>
<td>0.59</td>
</tr>
<tr>
<td>aSPL $\rightarrow$ Insula</td>
<td>0.22</td>
<td>0.28</td>
<td>0.02</td>
<td>0.91</td>
</tr>
<tr>
<td>FEF $\rightarrow$ aSPL</td>
<td>0.13</td>
<td>0.52</td>
<td>-0.01</td>
<td>0.97</td>
</tr>
<tr>
<td>FEF $\rightarrow$ PMd</td>
<td>0.01</td>
<td>0.97</td>
<td>-0.02</td>
<td>0.94</td>
</tr>
<tr>
<td>FEF $\rightarrow$ Insula</td>
<td>0.05</td>
<td>0.80</td>
<td>-0.12</td>
<td>0.57</td>
</tr>
<tr>
<td>PMd $\rightarrow$ aSPL</td>
<td>0.18</td>
<td>0.36</td>
<td>0.07</td>
<td>0.74</td>
</tr>
<tr>
<td>PMd $\rightarrow$ FEF</td>
<td>0.06</td>
<td>0.77</td>
<td>-0.03</td>
<td>0.89</td>
</tr>
<tr>
<td>PMd $\rightarrow$ Insula</td>
<td>0.14</td>
<td>0.48</td>
<td>-0.04</td>
<td>0.85</td>
</tr>
<tr>
<td>Insula $\rightarrow$ aSPL</td>
<td>0.49</td>
<td>0.009</td>
<td>0.02</td>
<td>0.91</td>
</tr>
<tr>
<td>Insula $\rightarrow$ FEF</td>
<td>0.28</td>
<td>0.16</td>
<td>-0.05</td>
<td>0.80</td>
</tr>
<tr>
<td>Insula $\rightarrow$ PMd</td>
<td>0.32</td>
<td>0.10</td>
<td>0.02</td>
<td>0.93</td>
</tr>
</tbody>
</table>

### 4.4 Discussion

The findings presented in this chapter show substantial inter-individual variability in behavioural measures of bistable perception, in keeping with established literature (see Section 1.6). To probe the neurobiological bases of this variability in a novel way, I explored the relationship of the individual dynamics of bistable perception on two paradigms (BR and SFM) with intrinsic effective connectivity in resting-state fMRI data obtained from the same participants on a separate occasion. Using stochastic DCM, I modelled resting-state effective connectivity between four ROIs chosen on the basis of previous findings presented in this thesis (see Chapter 3). I found that the strength (BPA) of directional connectivity from insula to aSPL was correlated with two measures of behavioural performance in BR for the same individual (mean BR percept duration and the
scale parameter of the BR percept duration distribution). However, there were no correlations between effective connectivity measures and behavioural measures in SFM.

4.4.1 Behavioural dynamics during BR and SFM tasks

4.4.1.1 Mean percept durations

Mean BR percept duration across the whole sample of participants was substantially shorter than for a previous experiment using a very similar BR paradigm (Haynes et al., 2005). For SFM mean duration was similar to that previously reported on a similar paradigm (Kanai et al., 2010), although longer than that reported in another study (Megumi et al., 2015). Besides the fact that the known inter-individual variability in these measures (Sections 1.6 and 4.4.1.2) means such discrepant findings may simply reflect differences in experimental samples, there are known to be various other influences on the dynamics of bistability. Low-level stimulus features play a role (Alais and Blake, 2005) and may have been affected by differences in lighting conditions or display configurations. In addition top-down and emotional factors can play a role (e.g. Gray et al., 2009; Meng and Tong, 2004; Sterzer et al., 2008); such factors were not equated across these different studies (but should have been relatively stable within participant in the present experiment).

For both BR and SFM, certain perceptual configurations were associated with slightly (but significantly) longer mean percept durations on average across the whole experimental sample. Specifically, this related to percepts resulting from stimuli shown to the right eye for BR and to perception of a leftward rotating sphere for SFM. Eye dominance did not play a significant role in either of these findings and it remains difficult to propose an explanation for them. Nevertheless, since the focus of further reported analyses was on individual mean percept duration in both BR and SFM (collapsed across all percept types within that individual and that paradigm), it seems unlikely that group-mean differences between the durations of different percept types would importantly influence the reported findings.

4.4.1.2 Relationships between behavioural results for BR and SFM

Behavioural measures for BR and SFM were overall moderately correlated at the individual participant level. This is in keeping with previous reports of correlated dynamics across
different types of bistable stimuli (e.g. Carter and Pettigrew, 2003; Pressnitzer and Hupé, 2006; see also Section 3.1). Such findings suggest there may be common mechanisms underlying bistable perception dynamics across different stimulus and task conditions. Indeed, in Chapter 3 I proposed a stimulus-invariant network of brain activity associated with perceptual transitions; these findings may reflect such common mechanisms and they were used as a template for the regions submitted to DCM analysis in the present chapter.

However, examining the present behavioural findings more closely also reveals differences between BR and SFM. While mean percept durations for both paradigms are moderately correlated (e.g. those with relatively long percept durations on one are also likely to also have long durations on the other), the same is not uniformly true when assessing the individual percept duration distributions. The two parameters describing the gamma distribution are strongly correlated within bistable paradigm and this has been previously described (e.g. Mamassian and Goutcher, 2005). Across the two paradigms, while there is a strong correlation for the parameter representing distribution shape (e.g. skew), such a correlation was not found for the parameter representing distribution scale (e.g. dispersion). Therefore, how spread out a distribution of BR percept durations is does not predict how spread out the corresponding distribution for SFM percept durations will be. This suggests that the variability of individual percept durations within an individual is specific to each of the tested bistable paradigms, and may thus be underpinned by different underlying mechanisms.

As proposed some decades ago (Levelt, 1967), one possible mechanistic explanation for gamma-distributed data in bistable perception is that every switch between percepts occurs after a fixed number of ticks from a Poisson clock (which ticks at random intervals). The proposal was that the ticks of this clock correspond to neuronal “spikes” related to the suppressed percept and eventually cause it to be re-established as the dominant percept. In this construct, the scale parameter corresponds to the basic duration between two ticks while the shape parameter describes the number of “ticks” that lead to a change in perception (Brascamp et al., 2005). The nature of the neural mechanisms that underpin perceptual states and perceptual transitions during bistable perception remains incompletely understood, and these are likely distributed across different anatomical structures and different hierarchical levels in the brain (e.g. Sterzer et al., 2009). Nevertheless, the Poisson clock framework provides useful and enduring suggestions of how percept visibility distributions may reflect underlying neural events, and could be used to explain differences between findings related to BR and SFM dynamics (see Section 4.4.2).
4.4.2 Intrinsic (resting-state) effective connectivity and behavioural dynamics of bistable perception

I used stochastic DCM to explore resting-state effective connectivity between regions belonging to the stimulus-invariant perceptual transition-related network outlined in Chapter 3 (including insula, FEF, aSPL and PMd). The *post-hoc* model selection procedure employed to evaluate the most favourable model structure (Section 4.3.2.1) indicated that a fully connected model with bidirectional connections between all four of these regions should be used. This result suggests that the resting-state BOLD signal patterns within these regions imply effective connectivity among all of them, providing supporting evidence that they indeed act as an intrinsic functional brain network.

4.4.2.1 Resting-state effective connectivity from insula to aSPL is correlated with BR dynamics at the individual level

I performed analyses to evaluate correlations between individual effective connection ‘strengths’ (BPAs) relating to each of the 12 directional connections within the proposed network (Figure 4.5B) and the behavioural dynamics of BR or SFM. The results of these analyses revealed that for a single directional connection (specifically that from insula to aSPL) BPAs were strongly and significantly correlated with individual mean BR percept duration. Furthermore, I found that the strength of the same directional connection was also correlated with the scale parameter of the BR percept distributions, providing additional evidence to support the resulting claim that there is a clear relationship between this particular effective connection and the dynamics of binocular rivalry within individuals. A possible underlying mechanism could be that the insula interacts with parietal regions (namely aSPL) to stabilise the percept in binocular rivalry. Thus, a stronger effective connection from insula to aSPL would be associated with more prolonged (in other words, more stable) percepts in BR. Such a proposal could be well incorporated into the model put forward by Kanai and colleagues (2011), where aSPL is involved in generating a prediction (as part of a predictive coding framework) that is fed back to hierarchically lower sensory regions and acts to stabilise the current percept. The current findings could suggest that such top-down (prediction) information is fed back from insula to parietal cortex. Whether intrinsic effective connectivity with other regions, e.g. in prefrontal cortex, contributes to this stabilisation of percepts is not clear from the present findings and additional analyses may be beneficial in this regard (see Section 4.4.3). However, my work shows for the first time that the stability of perceptual experience in BR relates to *intrinsic* effective
connectivity between two specific regions (insula and aSPL) and that the individual variability of such connectivity has behavioural correlates.

The present findings are in accordance with existing literature highlighting the importance of inferior frontal/insular regions in perceptual changes in bistable conditions (de Graaf et al., 2011; Lumer et al., 1998; Sterzer and Kleinschmidt, 2007). One fairly recent study used DCM analysis for task-related fMRI data (collected while viewing perceptual changes for a rotating Lissajous figure) to show that perceptual reversals were associated with modulation of top-down connectivity from inferior frontal to visual cortex (Weilnhammer et al., 2013). Other recent work has used a combination of EEG and independent TMS to both parietal and frontal regions to show that parietal regions (specifically, IPS) are important for percept stabilisation, while frontal regions (specifically, DLPFC) modulate parietal activity, including triggering of perceptual switches (Vernet et al., 2015). Vernet and colleagues found destabilising effects on bistable perception (opposite to the effects implied by the present findings); however, the region stimulated in their study was anatomically quite distant (superolateral) from the present insular region. Moreover, the stimuli of Vernet and colleagues (2015) were presented intermittently (with likely different underlying mechanisms to continuous bistability; e.g. Sandberg et al., 2013) and single-pulse TMS was used (which does not always have predictable effects; see Section 2.3.3). Further work will be required in the future to fully elucidate the influences of different areas in frontal and insular cortices on bistability but the combination of functional imaging and interventional approaches seems a promising one.

In contrast to findings described above, which highlight the role of inferior frontal/insular regions in bistable perception dynamics, several studies have presented results that do not support an active role of inferior frontal cortex in this context (e.g. Brascamp et al., 2015; Knapen et al., 2011; these findings are discussed in more detail in Sections 1.3.5 and 3.4.5). In addition, the present findings differ from those of another recent study that demonstrated resting state functional connectivity between pSPL and prefrontal regions as well as between aSPL and sensory areas or striatum (Baker et al., 2015). Those findings were interpreted in terms of pSPL being part of a system that stabilises particular interpretations of low-level sensory input and aSPL being part of a system that allows regulation of low-level visual incongruity. Apart from differences in the highlighted anatomical regions (e.g. insular versus prefrontal cortex), the clear differences from the present findings may be accounted for by discrepant methodologies. Firstly, Baker and colleagues (2015) used a different approach for preprocessing and statistical analysis of their resting state fMRI data (based on FSL software), which relies on different methods for spatial registration and statistical inference that may lead to discrepant results (e.g. Rajagopalan et al., 2014). Secondly, they used smaller seed ROIs (with a radius of only
2mm) for aSPL and pSPL, which would have resulted in more spatially limited sampling of BOLD signal patterns than the methods in the present experiment. Thirdly, and perhaps most importantly, their study assessed functional connectivity (statistical interdependence between regional activity), whereas the present study assessed effective connectivity (directional interactions between regions where neuronal activity is evaluated using biophysical models). These analyses explore fundamentally different properties of brain networks and so unsurprisingly they may generate discrepant results (Sporns, 2007; see also Sections 2.2.5 and 4.1). I would argue that the present findings can be incorporated more readily into a growing body of existing literature where varied approaches including task fMRI (e.g. Sterzer and Kleinschmidt, 2007), task-related DCM analyses (e.g. Weilnhammer et al., 2013) and interventional methods such as TMS (e.g. Kanai et al., 2011; Vernet et al., 2015) have been used to understand the involvement of frontal, insular and parietal regions in the mechanisms of perceptual transition.

4.4.2.2 Resting-state effective connectivity is not correlated with SFM dynamics at the individual level

Although I found a biologically plausible correlation between local effective connectivity and BR dynamics (Section 4.4.2.1), the same was not true for any relationship between effective connectivity and SFM dynamics. Indeed, none of 12 directional connection strengths from the explored DCM correlated with mean SFM percept duration or with parameters describing the SFM percept visibility distributions. This null result is all the more surprising given the commonality in mechanisms underpinning BR and SFM, suggested both by the correlated behavioural dynamics across the two paradigms (Section 4.4.1.2) and by the findings of a stimulus-invariant perceptual transition-related network in Chapter 3. On the other hand, there are key differences between BR and SFM (discussed in Section 3.1; see also Section 4.4.2.1 in relation to current behavioural results) that imply separable neural mechanisms specific to each of these paradigms. Existing work exploring the mechanisms of perceptual transition in SFM highlights the importance of relatively posterior regions including pSPL and V5 (Kanai et al., 2011, 2010). Indeed V5 was not included in the present DCM analysis but previous work exploring task-related effective connectivity between aSPL, pSPL and V5 during SFM has shown that bottom up coupling from V5 to pSPL and thence onto aSPL is related to percept visibility durations (Megumi et al., 2015). Taken together with the present findings, these results suggest that, notwithstanding the importance of the stimulus-invariant network for perceptual transition in both paradigms (Table 4.1), the dynamics of bistability in SFM may be more dependent on posterior regions such as pSPL and V5 and those in BR may be more dependent on
anterior regions such as insula and IFG. Such a difference would not be altogether surprising when one considers that the information that needs to be disambiguated in these two paradigms carries rather different computational demands (see Section 3.1 for a discussion).

The correlated behavioural dynamics across the BR and SFM paradigms did not extend to the scale parameters describing individual percept visibility distributions for each of them. This suggests that the degree of spread of the percept duration distribution for one paradigm was independent of that for the other paradigm within an individual. If we refer again to the proposed Poisson clock mechanism in Levelt’s theory of binocular rivalry (1967; see Section 4.4.1.2), one could argue that the intervals between “ticks” on the clock (reflected by the distribution scale parameter) may be determined by the effective connection between insula and aSPL for BR. If a similar “clock” mechanism also underpins SFM one would speculate that the intervals between ticks are determined by distinct neural mechanisms, perhaps through interactions between V5 and pSPL. Such possibilities could be directly empirically evaluated in the future, perhaps by using combined TMS-fMRI approaches to selectively disrupt connectivity between these regions during bistable perception tasks.

4.4.3 Caveats and considerations for future analyses

Although the choice of regions to include in the DCM analyses was clearly motivated by existing literature as well as the findings presented in Chapter 3 (see Section 4.1), it must be borne in mind that other regions may have usefully and logically been incorporated into the model. I chose to include a locus in right insula in the analysis and not a nearby locus in right IFG (containing statistically weaker stimulus-invariant perceptual-transition-related BOLD signals) in order to retain relative simplicity in the DCM analysis. While insula and IFG were in close anatomical proximity in the findings in Chapter 3, it is quite possible that there are independent interactions between IFG and any of the presently included regions that are relevant to the dynamics of bistable perception. Moreover, a further lateral prefrontal region, found anteriorly to both the insula and IFG loci outlined in Chapter 3, has been associated with perceptual reinterpretation in bistability (Knapen et al., 2011; Watanabe et al., 2014). Finally, as discussed in Section 4.4.2.2, more posterior regions including V5 are likely of importance for the dynamics of bistability in SFM paradigms. Given these considerations it would be worthwhile repeating the DCM analyses with the inclusion of these additional regions to interrogate any further effective connectivity patterns that could be associated with the behavioural manifestations of bistable perception at an individual participant level. A more recently developed spectral DCM method (Razi et
al., 2015; also mentioned in Section 4.1) and post hoc model optimisation are particularly well suited to the analysis of effective connectivity within larger networks in the resting state (Razi et al., 2017). Spectral DCM appears to be more accurate than stochastic DCM in terms of root mean square error, as well as being much less computationally demanding (Razi et al., 2015). Repeating the current analyses, or indeed performing more extended network analyses, with spectral DCM therefore seems a logical next step. It must of course be borne in mind that the interpretation of findings relating to larger networks would be more challenging.

An additional caveat to bear in mind is that I chose (somewhat arbitrarily) to use loci from the stimulus-invariant perceptual transition-related network obtained by masking an endogenous (versus exogenous) BR transition fMRI-BOLD contrast image inclusively with an endogenous (versus exogenous) SFM transition fMRI-BOLD contrast image. The reverse contrast masking procedure produced very similar (though not identical) findings and it would be important to check whether constructing a DCM based on the findings of this alternative analysis would lead to similar results as those shown presently.

4.4.4 Summary

Through the experimental work presented in this chapter I sought to determine whether resting state effective connectivity among regions involved in perceptual transition was correlated with the dynamics of different bistable perception paradigms (BR and SFM) at the individual participant level. In other words, I asked whether individual differences in behaviour during bistable perception could be explained by local differences in intrinsic effective brain connectivity. The results revealed strong correlations between individual behavioural measures of bistable perception both within and across the BR and SFM paradigms. Additionally, and critically, a DCM analysis of effective connectivity between frontoparietal regions involved in perceptual transition indicated that the dynamics of BR (but not of SFM) are correlated with the strength (BPA) of the directional effective connection from insula to aSPL in the resting state. This finding suggests that directional interactions from insular to parietal regions relate to the frequency of spontaneous changes in perceptual experience in certain circumstances. The result does not elucidate whether this effective connection plays a causative role in the dynamics of bistability, although it can nevertheless motivate specific hypotheses for future studies (e.g. using combined TMS-fMRI approaches) to directly probe the causal role of such connectivity patterns.

These findings further understanding of the information flow that is relevant to the dynamics of perceptual transition, particularly in binocular rivalry, lending support to
accounts that place an active role on inferior frontal/insular cortical regions. The same directional connections were not predictive of behaviour during SFM and this suggests that the dynamics of bistability for BR and SFM may relate to at least partly distinct mechanisms. Previous studies have highlighted the importance of posterior parietal and dorsal visual regions in SFM dynamics and on balance it seems likely that such previous work and the present experiments have probed different parts of a distributed network that balances top-down and bottom-up modulatory mechanisms to facilitate bistable perception. Future studies could employ approaches with multiple paradigms (as in Chapter 3) as well as look to perform connectivity analyses on networks with larger numbers of ROIs.

The causal impacts on bistable perception of the perceptual-transition-related regions (outlined in Chapter 3), as well as the key connections between them (highlighted in this chapter), remain incompletely characterised. As discussed elsewhere (Sections 1.3.5, 3.1 and 5.1), a coherent body of work has used transcranial magnetic stimulation to establish the role of SPL in this regard (e.g. Kanai et al., 2011). However, the casual role of frontal regions remains less clear and this question will be addressed further in Chapter 5.
Chapter 5

Functional role of right frontal eye field in the dynamics of bistable perception

“Some chaos exists out there, and the brain seems to have more flexibility than classical physics in finding the order in it.”

James Gleick (1954 – )

5.1 Introduction

As reviewed in Section 1.3.5, the timepoints of transition between different percepts in bistable perception provide an opportunity to study the neural mechanisms underlying changes in the contents of awareness. There is an established correlation between such changes in perceptual experience and BOLD signal changes in a right frontoparietal cortical network. In Chapter 3 I highlighted a small group of regions (including right FEF, right aSPL, right PMd and right insula) that are active during perceptual transitions in a stimulus-invariant fashion. In Chapter 4 I then showed that at the individual level, dynamics of binocular rivalry are correlated with directional effective connectivity strength between two of these regions (specifically, connectivity from insula to aSPL). To go a step further than these correlational findings and determine whether these relationships between perceptual transition and BOLD activity are causative, and any of these cortical regions (or the connections between them) are necessary for perceptual transition, approaches that modulate cortical function are required.
In the last decade, a series of experiments have used rTMS to induce temporary disruption of focal cortical function (see Section 2.3) leading to confirmation that superior parietal cortex plays a critical role in determining the frequency of perceptual transitions during bistable perception (Carmel et al., 2010; Kanai et al., 2010, 2011; see Sections 1.3.5 and 3.1). This accumulation of evidence has motivated the proposal of a theoretical framework, which characterises the specific functions of SPL in perceptual reorganisation, as well as proposing how these functions map onto anatomical subdivisions of SPL (Kanai et al., 2011). More recent follow-up work with fMRI and effective connectivity analysis has confirmed some of the predictions instantiated by this model (Megumi et al., 2015; see also Sections 1.3.5 and 3.1; discussion in Section 9.2.2).

Regarding the role of frontal cortical regions in perceptual switches, the existing evidence is less convergent, both in terms of anatomical localisation and in terms of the experimental approaches that have been employed to explore this question. Early work on this topic has shown that permanent structural damage to frontal brain regions, in particular in the right hemisphere, impairs the ability to experience multiple interpretations of ambiguous images (Meenan and Miller, 1994; Ricci and Blundo, 1990). Additionally, changes in awareness during a change blindness task are delayed after inhibitory TMS to DLPFC (Turatto et al., 2004). One more recent study employing a bistable perception task used rTMS to show that disruption of DLPFC function impairs voluntary control over bistability but has no effect on spontaneous perceptual switches (de Graaf et al., 2011). Overall, the functional role of frontal brain regions in bistable perception and in perceptual transitions remains incompletely understood.

Given the results from Chapter 4 and the established knowledge regarding the functions of SPL (see above), exploring the causal role of right insula (and overlying operculum) in bistable perception would be a logical next step. This could in theory be achieved by temporarily disrupting function in this area with rTMS. Although TMS to IFG has been successfully performed (e.g. Sharot et al., 2012), this is known to be fraught with methodological difficulties since it requires stimulation in a region that frequently activates the temporal muscles and can cause substantial discomfort. In addition, the inferior frontal/insular findings from Chapters 3 and 4 were deep to the brain surface (largely located in insula and to a lesser extent in overlying operculum), which can cause problems with efficacy of TMS (see Section 2.3). I therefore decided to instead target FEF, another region reliably containing stimulus-invariant perceptual-transition-related activity, as outlined by the results presented in Chapter 3. Previous studies suggest that repetitive TMS over FEF is generally possible and well tolerated (e.g. Hubl et al., 2008; Muggleton et al., 2003).
The functions of FEF have been widely investigated, with a long-established role in oculomotor control and various types of saccadic function, both in primates (e.g. Bruce and Goldberg, 1985; Dias et al., 1995) and in humans (e.g. Rivaud et al., 1994). FEF is especially important in complex goal-directed (rather than reflex) saccades (Guitton et al., 1985). Indeed, it is fairly common to use experimental eye movement paradigms to define this region functionally (e.g. Amiez and Petrides, 2009; Paus, 1996).

Work in the monkey has long suggested that FEF may additionally be involved in other more perceptual aspects of visual function (e.g. Latto and Cowey, 1971; Rizzolatti et al., 1983). In humans FEF is involved in visual detection (Shulman et al., 2001) and is necessary for selection in visual search tasks (Muggleton et al., 2003). FEF neurons respond non-selectively to visual stimuli and their responses are modulated by top-down factors (for example, whether or not they - that is, the neurons - are found in the receptive field of a saccade target). This line of evidence has led to proposals that FEF (in particular in the right hemisphere) is involved in priority (or saliency) coding, which contributes to mechanisms for spatial attention (Corbetta and Shulman, 2002; Ptak, 2012; Schall and Thompson, 1999). Work employing concurrent TMS and fMRI has gone further, providing direct evidence that FEF feeds back onto early visual areas, modulating neural activity in these regions (Ruff et al., 2006). In particular, the study of Ruff and colleagues found that increasing magnetic stimulation strength to FEF resulted in progressively increased suppression of BOLD activity in early visual areas representing the central visual field and concurrent increase in BOLD activity in areas representing the peripheral visual field. An additional psychophysical experiment confirmed that these findings were perceptually relevant: contrast perception in the periphery of vision was indeed enhanced as a result of TMS to FEF (Ruff et al., 2006). These results clearly demonstrate that FEF plays a role in boosting saliency of certain visual signals and in addition suggest that the region affects perceptual experience through modulation of early visual representations.

FEF is well placed anatomically to balance bottom-up and top-down influences, with extensive connections with frontal and parietal cortex as well as high-order visual regions, including dense connections to STS and IPS in the monkey (Huerta et al., 1987). There appears to be an anatomical gradient of connectivity with different visual regions: lateral FEF is connected with ventral stream areas and areas involved in processing of foveal visual representations while medial FEF is connected with areas involved in dorsal stream processing and representations of peripheral vision (Schall et al., 1995).

Besides functions related to spatial attention, as already mentioned in Section 3.4.1, several studies have implicated FEF directly in relation to visual awareness. In the macaque, electrophysiological measurements of FEF activity can predict visual detection (Thompson
and Schall, 1999), and furthermore, microstimulation of this region can improve visual detection thresholds (Moore and Fallah, 2001). One study performed in humans using single-pulse TMS showed that stimulation of right FEF increases the probability of detecting subliminal images (presented between 40 and 100ms after TMS and followed by a backward mask; Grosbras and Paus, 2003). In this study, Grosbras and Paus speculate that the mechanism that underpins their findings is boosting of background activity in early visual areas by feedback from FEF, meaning that sensory signals that would normally remain subliminal could now reach awareness, as per predictions from accrual decision-making models (e.g. Carpenter, 1999). Although in relation to a different experimental paradigm, the work of Ruff and colleagues (2006; discussed above), has already demonstrated that FEF modulates activity in early visual areas.

The aim of the present experiment was to directly evaluate the causative role of right FEF (specifically, the region containing BOLD signal change in association with perceptual transitions across different paradigms in Chapter 3) in the dynamics of bistable perception. I therefore planned to temporarily disrupt the function of right FEF using cTBS, a variant of TMS, which is known to reliably suppress the function of underlying cortex for 10-20 minutes (Huang et al., 2005; see Section 2.3.3). Participants would perform a bistable SFM visual task (see Section 1.3.3) both before and after TMS, allowing comparison of mean perceptual durations pre- and post-stimulation (as described in Kanai et al., 2010; see also Section 2.3.3 for a brief discussion of offline TMS approaches).

Based on previously published findings, in particular those of Grosbras and Paus (2003) and of Ruff and colleagues (2006), I hypothesized that the role of right FEF in bistable perception is to reduce the strength of signal related to the currently perceived (dominant) percept in early visual areas (and to boost signals related to the suppressed percept). In this scenario, activity in FEF would normally serve to destabilise the current perceptual interpretation, increasing the chances of a perceptual switch and therefore leading to shorter periods of perception in a bistable visual context. Consequently, I predicted that inhibitory TMS to right FEF would result in stabilisation of perceptual switching (an increase in mean percept duration) during a bistable SFM task. Predicting the direction of effect of TMS to FEF was challenging in the present set of circumstances; it was therefore important to include an active control condition. I thus undertook pSPL stimulation, using the same target coordinates as those employed in a previous study that explored effects of cTBS on SFM dynamics and aiming to replicate those results (increased percept duration).

In this sort of active control the purpose is to match all aspects of the test condition (in this case, FEF stimulation), but instead to stimulate a region with a known effect on the experimental task. If the result of the active control (in this case, pSPL stimulation) replicates previous findings this confirms that TMS has been effective in general. Thus, a technical failure of stimulation cannot be blamed if there is a null result for the test condition in this case, effectively enhancing the ability to interpret such a null result.
after cTBS to pSPL; Kanai et al., 2010). In order to attempt replication of the findings of Kanai and colleagues, the experimental paradigm and methodology used presently (for both test and control conditions) were designed to be as much as possible identical to that study. I used stimulation over the vertex as an additional (passive) control condition; stimulating at this site was not expected to have any consequence on bistable perception dynamics (e.g. Kanai et al., 2010) and it thus allowed me to control for any non-specific effects of TMS.

Given the established importance of FEF for eye gaze control and saccadic function (see above; Müri and Nyffeler, 2008), inhibitory cTBS over this region could disrupt gaze function. Gaze position and saccades were therefore monitored throughout the experiment.

5.2 Methods

5.2.1 Participants

Seventeen right-handed participants were recruited according to standard procedures and inclusion criteria detailed in Section 2.4.1. As an additional criterion, individuals who had already undergone structural MRI scans at the Wellcome Centre for Human Neuroimaging were recruited (to enable targeting of specific anatomical sites for TMS; see Sections 5.2.5 and 5.2.6); each participant gave consent for their scan to be used in the study and permission was separately sought from the investigator who had originally obtained the MRI data. Participants also completed a screening questionnaire to exclude any contra-indications to TMS (Rossi et al., 2009).

The experimental sample size was determined based on similar previous experiments using TMS to study effects on bistable perception, which have shown adequate effect sizes with samples of 10-15 individuals (de Graaf et al., 2011; Kanai et al., 2010).

5.2.2 Experimental sequence

Each participant was asked to attend for three experimental sessions on separate days. No two sessions occurred on consecutive days; this was to minimise any side effects of TMS from the previous session, both to avoid contamination of experimental effects and for safety reasons (Rossi et al., 2009; Section 2.3.4). The gap between sessions was between 2 and 25 days. A different one of the three brain regions chosen for TMS (right FEF, right pSPL and vertex) was stimulated at each visit using an existing MRI brain scan and
neuronavigation software to guide location of TMS administration. The order of regions was randomised for each participant prior to the experiment. Before the first session participants were safety screened and asked to provide written consent. They then carried out two practice trials of the behavioural paradigm, followed by a block of four experimental trials (pre-TMS condition; Section 5.2.3). Participants then underwent rTMS, delivered at rest, to a site on the scalp overlying the target brain region predetermined \textit{a priori} for that session (Sections 5.2.5 and 5.2.6). After a 5-minute rest period, a further run of four behavioural trials was undertaken (post-TMS condition). Pupil position and size were tracked continuously throughout all behavioural trials (Section 5.2.4). Finally, participants completed a short questionnaire regarding their experience of TMS during that session (Section 5.2.7). The remaining two TMS target regions were stimulated at subsequent sessions (conducted at minimum time intervals as described above). Other than change of stimulated region, the experimental protocol for the second and third sessions was identical to that for the first session.

5.2.3 SFM behavioural paradigm

The behavioural paradigm and procedures were as described in Section 2.4.5 (note SFM was also used in Chapters 3 and 4). In the present instance, the stimulus was perceived as a sphere rotating about its vertical axis in an ambiguous direction (either to the right or left; Figure 5.1). Stimuli were presented on a computer monitor (resolution 1024 x 768; refresh rate 60 Hz) and viewed at a distance of 65 cm, subtending 5.3° visual angle in diameter. Each trial lasted 48 seconds with participants providing ongoing report of their perception of the direction of sphere rotation.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.1.png}
\caption{Experimental SFM stimulus}
A display of white dots moving in an oscillatory fashion (Section 2.4.5) was shown in the centre of a computer screen on a black background and with a central red fixation dot (A). This stimulus configuration could be perceived as a sphere rotating either to the right or to the left (B).}
5.2.4 Eye tracking

Eye tracking was performed throughout all behavioural trials according to the procedure described in Section 2.4.9.

5.2.5 MRI preprocessing and TMS target localisation

The target coordinates (in MNI space) for each of the three regions chosen for TMS were: $x = 22, y = -10, z = 50$ for FEF, based on stimulus invariant perceptual-transition-related BOLD signal (Chapter 3); $x = 34, y = -66, z = 34$ for pSPL, as reported in Kanai et al., 2010 (see section 5.1); $x = 0, y = -30, z = 50$ for vertex (determined manually on a template brain in MNI space). Spherical masks with 10 mm radius, centred at each of these three loci, were made using the Pick atlas. Each participant’s pre-acquired T1-weighted structural MR brain image (see Section 5.2.1) was converted to NIfTI format and segmented using SPM8. The normalization parameters obtained from segmentation were then used to transform the three Pick atlas masks from MNI space into each participant’s native brain space. Masks were then re-sliced and merged with the native structural images. Thus, each participant’s structural image now contained three ‘markers’ to guide TMS (Figure 5.2).

![Figure 5.2: Processing sequence for placing TMS targets onto individual structural MRI scans](image)

MR images were segmented (1). The normalisation parameters obtained in step (1) were then used to transform spherical masks for each of the three regions (created in MNI space) into each participant’s native space (2). Masks were then re-sliced (3) before being superimposed with the original MR image for that participant (4). Processing of the FEF mask is shown as an example in the figure.
5.2.6 TMS procedure

TMS administration was guided using an ANT Visor 2 neuronavigation system, which uses an infrared camera to track spherical markers attached to the participant’s forehead and to the TMS coil (Figure 5.3). Prior to the first experimental session for each participant, that individual’s structural MR scan (with superimposed ROI masks; see Section 5.2.5 and Figure 5.2) was loaded into the Visor 2 system. Target markers were placed in the centre of each of the three ROI masks. A head model was then created using proprietary software by first marking the position of the participant’s nasion as well as the tragus of each ear with a pointer tracking tool (Figure 5.3). This head model enabled movements of the TMS coil (I used a 70 mm figure-of-eight coil) over the participant’s scalp to be tracked precisely with respect to underlying brain anatomy (this is achieved through continuous online tracking of TMS coil position relative to head position using camera detection of a coil tracker and head tracker respectively; Figure 5.3). The TMS coil was moved over the scalp (orientated tangentially to the scalp with the handle pointing backwards and parallel to the midline) until the neuronavigation system confirmed that the coil position matched the target for that session. The coil was then held still over the participant’s scalp (while they maintained a stable head position); the neuronavigation system display was constantly monitored to confirm ongoing alignment between TMS coil and stimulation target.

TMS pulses were delivered using a Magstim Rapid2 stimulator, set at 50% of maximum stimulator output. An established cTBS protocol was employed (3 pulses at 50Hz, repeated every 200 ms for a 40 second period, resulting in a total of 600 pulses; as used by Kanai et al., 2010). Such stimulation is known to reliably suppress the function of cortex underlying the stimulated area for 10-20 minutes (Huang et al., 2005).
The pointer tool was used to localise landmarks on each participant’s head in order to guide creation of a head model; the coil tracker was attached to the TMS coil to monitor coil position during stimulation; the reference tool was attached to the participant’s forehead with an elastic headband to monitor head position during stimulation (see Section 5.2.6; figure adapted from Visor 2 user guide, Rev. 14, 8th May 2012).

5.2.7 Ratings for participant experience of TMS

To check for any differences in participant experience, specifically in relation to side effects from TMS applied to each of the three stimulation sites, each individual was asked to provide ratings of ‘intensity’ and ‘unpleasantness’ at the end of each experimental session. Ratings related only to the 40-second period of TMS and were provided by clicking with a mouse button on continuous linear scales of 0-10 displayed on the computer screen (with no time limit). Since TMS to FEF is known to occasionally cause discomfort due to stimulation of facial muscles (e.g. Neggers et al., 2007), I anticipated that intensity and unpleasantness ratings might be higher for this region than for the other two regions tested.
5.3 Results

One participant failed to attend their final experimental session and one participant complained of discomfort during TMS over pSPL and did not wish to continue with the remainder of the experiment. Therefore, 15 individuals completed the experiment (8 female; mean ± SD age = 25.4 ± 6.0 years; range = 20 - 40 years).

5.3.1 Individual variability in mean percept duration

I calculated mean percept duration for each participant across the whole experiment (collapsing across all three sessions and both pre-TMS and post-TMS conditions). There was substantial inter-individual variability in mean percept duration, as seen in previous studies (Kanai et al., 2010; Figure 5.4A). Across the whole group, the mean percept duration was 12.89 seconds (SD 6.59 seconds; range 5.36 – 28.28 seconds). Within-participant variability was modest and fairly similar across participants (Figure 5.4B).

![Figure 5.4: Inter- and intra-individual variability in mean percept durations](image)

(A) Histogram showing substantial inter-individual variability of mean percept duration across the sample of 15 participants. (B) Variability within participants (represented by the errorbars) was modest and did not vary a great deal across the sample.

5.3.2 Group-level effects of TMS on percept duration

At the group level, no significant difference in percept duration change (defined by calculating mean post-TMS percept duration as a percentage increase of mean pre-TMS percept duration; as in Kanai et al., 2010) was seen between the three stimulation conditions (Figure 5.5). Although the pattern of results suggested some increase in percept duration after TMS delivered to both right FEF and right pSPL, repeated-measures
ANOVA (with a single three-level factor of stimulated region) revealed no significant main effect \( F_{(2,28)} = 0.65; p = 0.53 \). Post-hoc comparisons revealed no significant differences between change in duration after FEF stimulation and change in duration after vertex stimulation \( t_{(14)} = 1.05; p = 0.31 \). Similarly, comparisons relating to pSPL and vertex stimulation and FEF and pSPL stimulation were not significant \( t_{(14)} = 0.82; p = 0.43 \) and \( t_{(14)} = 0.37; p = 0.71 \), respectively. One-sample t-tests showed that percept duration change was not significantly different from zero for any of the three stimulation conditions \( t_{(14)} = 1.31; p = 0.21 \); for right pSPL \( t_{(14)} = 1.47; p = 0.16 \); for vertex \( t_{(14)} = 0.05; p = 0.96 \).

![Figure 5.5: Group-level behavioural results](image)

Values on the y-axis represent increase in mean percept duration after TMS (as a percentage of the mean pre-TMS duration) for each of the three experimental conditions. There was a trend towards increase in mean percept duration after TMS over both FEF and pSPL; however in both cases this was neither significantly different to the increase in percept duration after control TMS over the vertex, nor significantly different from zero. Error bars represent SEM.

Adding in age, gender, and session order as covariates in the ANOVA revealed no significant interactions with any of these factors \( \text{stimulated region*age } F_{(2,22)} = 0.05, p = 0.95 \); stimulated region*gender \( F_{(2,22)} = 0.46, p = 0.64 \); stimulated region*session order \( F_{(2,22)} = 1.54, p = 0.24 \).

It is possible that the substantial inter-individual variability in percept duration (see Section 5.3.1; also evident in the large error bars in Figure 5.5) could have contributed to the lack of clearly observable effect of stimulated region. I therefore repeated the analysis with z-scored data, in order to minimise variability between individuals. This manipulation did not alter the pattern of results: the main effect of stimulated region remained non-significant \( F_{(2,28)} = 1.16; p = 0.33 \) and again there were no significant pairwise differences.
between the regions \[\text{change in duration after: FEF versus vertex stimulation } t_{(14)} = 1.10, p = 0.29; \text{pSPL versus vertex stimulation } t_{(14)} = 1.05, p = 0.31; \text{FEF versus pSPL stimulation } t_{(14)} = 0.43, p = 0.68\].

5.3.3 Participant ratings of the experience of TMS

As predicted, mean intensity and unpleasantness ratings were higher for FEF stimulation than for the other two stimulation conditions (mean intensity for FEF = 2.44, for pSPL = 2.14, for vertex = 2.14; mean unpleasantness for FEF = 2.09, for pSPL = 1.12, for vertex = 1.38). Repeated measures ANOVA showed no significant effect of stimulated region on intensity \[F_{(2,28)} = 0.21, p = 0.81\] or on unpleasantness \[F_{(2,28)} = 2.86, p = 0.07\]. The effect of stimulated region on unpleasantness came close to significance; pairwise comparisons showed that unpleasantness scores were higher for FEF stimulation than pSPL stimulation \[t_{(14)} = 2.10, p = 0.05\], but there was no significant difference in unpleasantness scores for FEF stimulation and vertex stimulation \[t_{(14)} = 1.77, p = 0.10\] or for pSPL stimulation and vertex stimulation \[t_{(14)} = -0.66, p = 0.52\].

5.3.4 Effects of visual fixation, pupil size and saccades

Participants were asked to fixate loosely on a red dot shown at the centre of the visual stimulus (see Section 5.2.3). Fixation differed to some extent between individuals but importantly within participants fixation did not differ substantially or systematically according to stimulated region or between pre-TMS and post-TMS conditions (Figure 5.6). Moreover, at the group level, repeated measures ANOVA with SEM for mean horizontal eye position as the dependent measure and factors of stimulated region (three levels) and TMS (pre or post-TMS; two levels) showed no main effect of stimulated region \[F_{(2,28)} = 1.45, p = 0.25\], no main effect of TMS \[F_{(1,14)} = 1.07, p = 0.32\] and no stimulated region*TMS interaction \[F_{(2,28)} = 1.96, p = 0.18\]. Likewise there were no significant findings with SEM for mean vertical eye position as the dependent measure \[F_{(2,28)} = 1.70, p = 0.21\]; main effect of TMS \[F_{(1,14)} = 0.70, p = 0.42\]; stimulated region*TMS interaction \[F_{(2,28)} = 0.96, p = 0.35\], or with SEM for mean pupil size as the dependent measure \[F_{(2,28)} = 1.70, p = 0.20\]; main effect of TMS \[F_{(1,14)} = 0.53, p = 0.48\]; stimulated region*TMS interaction \[F_{(2,28)} = 0.45, p = 0.56\]. These analyses confirm that deviation from mean gaze positions and mean pupil size were not affected by TMS administration or stimulated region.
I performed group-level repeated-measures ANOVAs with the same factors as those for the SEM measures above but this time using mean horizontal eye position, mean vertical eye position and mean pupil size as the dependent measures. For horizontal eye position there was a just-significant main effect of TMS \[F_{(1,14)} = 5.37, \ p = 0.04\]. However, the main effect of stimulated region \[F_{(2,28)} = 0.75, \ p = 0.48\] and the stimulated region*TMS interaction \[F_{(2,28)} = 0.81, \ p = 0.45\] were not significant. For vertical eye position there was no significant main effect of stimulated region \[F_{(2,28)} = 1.27, \ p = 0.30\]; no significant main effect of TMS \[F_{(1,14)} = 2.10, \ p = 0.17\]; and no significant stimulated region*TMS interaction \[F_{(2,28)} = 1.54, \ p = 0.28\]. For pupil size, as for horizontal eye position, there was a significant main effect of TMS \[F_{(1,14)} = 6.90, \ p = 0.02\]. Again, the main effect of stimulated region \[F_{(2,28)} = 0.25, \ p = 0.78\] and the stimulated region*TMS interaction \[F_{(2,28)} = 0.09, \ p = 0.86\] were not significant. Post-hoc comparisons showed that the significant main effect of TMS on horizontal eye position was underpinned by a significant difference between pre-TMS and post-TMS measures in the vertex stimulation condition \[t_{(14)} = -2.30, \ p = 0.04\] and a near-significant difference in the pSPL stimulation condition \[t_{(14)} = -1.94, \ p = 0.07\]; there was no significant difference in the FEF stimulation condition \[t_{(14)} = -0.46, \ p = 0.65\]. Post-hoc comparisons showed that the significant main effect of TMS on pupil size was reflected by significant differences between pre-TMS and post-TMS measures in the FEF stimulation condition \[t_{(14)} = -2.19, \ p = 0.05\] and vertex stimulation condition \[t_{(14)} = -2.26, \ p = 0.04\] as well as a near-significant difference in the pSPL stimulation condition \[t_{(14)} = -1.82, \ p = 0.09\]. These results are illustrated graphically in Figure 5.7.
Figure 5.6: Individual eye position heat maps

Heat maps for three representative participants are shown in (A-C). Each plot shows all eye position measurements (recorded throughout all experimental trials for that condition with a sampling frequency of 1000Hz). Eye position is represented in x and y coordinates corresponding to pixels on a screen with resolution 1024x768 (with the lower left corner of the screen and corresponding plot having coordinates 0,0). For each participant, left hand panels are for the FEF condition, middle panels are for the pSPL condition, and right hand panels are for the vertex condition; the upper row shows pre-TMS trials and lower row shows post-TMS trials. The colour bar to the right of each plot represents the number of samples at each pixel position. The plots show some differences in fixation between participants, and occasional differences within participant, but no systematic differences between stimulation sites or between pre-TMS and post-TMS measures are apparent.
Given the known involvement of FEF in saccadic eye movements, the total number of saccades recorded by the eye tracker was calculated for both the pre-TMS and post-TMS periods for each of the three stimulation conditions. Repeated-measures ANOVA with factors of stimulated region (three levels) and TMS (pre or post-TMS; two levels) showed no significant main effect of stimulated region $[F(2,28) = 1.34; p = 0.28]$, no significant main effect of TMS $[F(1,14) = 0.02; p = 0.90]$, and no stimulated region*TMS interaction $[F(2,28) = 1.59; p = 0.23]$. Therefore, there was no difference in saccade generation between the three stimulation conditions or between pre-TMS and post-TMS recording periods.

**Figure 5.7: Group level eye tracking results**

Mean horizontal eye position (A), mean vertical eye position (B) and mean pupil size (C) calculated across all 15 participants for each of the three stimulation conditions. Mean values for pre-TMS trials are shown with blue bars and mean values for post-TMS trials are shown with red bars. The y-axes represent eye gaze position on the screen in pixels counted horizontally (A) or vertically (B) from the lower left hand corner of a 1024x768 resolution monitor. The y-axis in (C) represents pupil size in pixels. Error bars represent SEM. It is apparent that there is a small systematic difference in horizontal eye position (A) and pupil size (C) when comparing pre-TMS and post-TMS conditions (Section 5.3.4).

### 5.4 Discussion

In this experiment I asked whether right FEF plays a causal role in the dynamics of perceptual experience during a bistable SFM task. Based on previous studies as well as findings I presented in Chapter 3, I predicted that inhibitory cTBS to right FEF would be associated with increased mean SFM percept duration. In addition, I sought to replicate the increased mean percept duration associated with inhibitory cTBS to right pSPL reported
previously (Kanai et al., 2010). In a control condition, cTBS was administered over the vertex, which was expected to have no effect on SFM percept durations. There was substantial inter-individual variability in mean percept durations (as shown also in Chapters 3, 4 and 6-8). Contrary to my predictions, despite a trend towards increased mean percept durations in both cases, I found no significant effects on mean percept duration after cTBS to either right FEF or right pSPL. While there was no difference in participant-rated intensity associated with TMS over the three stimulated regions, there was a borderline-significant effect of unpleasantness, with TMS over right FEF rated as more unpleasant (in particular compared to TMS over right pSPL). Visual fixation did not differ systematically between the three stimulation conditions or between pre-TMS and post-TMS periods. However, across all conditions there was a small but significant rightward shift in eye position as well as a small but significant increase in pupil size following TMS.

5.4.1 Non-replication of effect of TMS to right pSPL

A fundamental challenge with interpreting the findings from the present experiment is the fact that the TMS result reported by Kanai and colleagues (2010) was not replicated. The authors of that study demonstrated an increase of mean SFM percept durations following application of cTBS to pSPL and have replicated the effect in independent samples of participants (Ryota Kanai, personal communication). Given the present non-replication I sought to carefully re-evaluate the methodology of the present experiment, seeking any aspects that may have differed significantly from the study of Kanai and colleagues (2010), in consultation with the first author of that study. I found several potential sources of variation between the methods used in the two studies. Firstly, the methodology for converting TMS stimulation target coordinates from MNI into native space for each participant was different: Kanai and colleagues had made use of FSL rather than SPM for these calculations. I therefore followed the analysis steps performed in that study and was able to confirm that in all cases the resulting stimulation coordinates were not significantly different from those used in the present experiment. Secondly, the hardware, methodology and parameters for TMS administration were carefully reviewed and it was confirmed that there were no differences between the two studies. Given this fact, the possibility that TMS administration was generally ineffective in the present experiment seems unlikely. Thirdly the methodology for the behavioural experiment was again reviewed. It was discovered that Kanai and colleagues had in fact used three blocks (with four SFM trials in each) both before and after TMS (compared to a single block of four SFM trials both before and after TMS in the present experiment). This discrepancy had unfortunately not previously been clear despite reviewing both published methods of Kanai and colleagues (Kanai et al., 2010) and communicating personally with the lead author of that study. It is likely that this
difference affected the results in one of two possible ways: the relatively smaller number of trials used presently could have resulted in a decreased signal to noise ratio in the data, reducing the likelihood of detecting any effects; alternatively, the shorter overall duration of behavioural experimentation following TMS would have rendered the present experiment insensitive to any effects of TMS that appeared later than the end of the fourth post-TMS SFM trial. The present null effect of cTBS to pSPL on SFM percept duration could have been related to either of these factors, or perhaps a combination of them. However, the relatively low intra-individual variability in percept durations (see Figure 5.4B) would go against the suggestion that the data obtained were very noisy. A further possible explanation for the null result is that contextual differences between the two experiments (for example, the experiment room or the interaction with the experimenter) were responsible for the non-replication of the findings. It is difficult to fully evaluate this possibility but after discussions with the lead author from the study of Kanai et al. (2010), conditions had been kept as similar as possible in the present experiment, and any small differences are likely to have had a minor (and probably non-systematic) impact on the findings. Overall, given that the present results show a trend towards increase in percept durations for both pSPL and FEF conditions following TMS (and no such trend in the vertex condition), the pattern of findings seems most likely to reflect insufficient statistical power.

The mean SFM percept durations reported by Kanai and colleagues (2010) were substantially shorter than those obtained presently (range 2-14 seconds compared with range 5-28 seconds). This difference may be due to sampling factors (given that SFM duration is highly variable across individuals). Alternatively it could relate to the longer overall duration of the behavioural paradigm of Kanai et al. (indeed it has been shown that more prolonged exposure to bistable stimuli, in particular in the case of binocular rivalry, is associated with progressive shortening of percept durations; Suzuki and Grabowecky, 2007).

Finally, it is relevant to note that the neurobiological effects of TMS are complex and multifactorial (see Section 2.3.1), making findings from TMS experiments inherently difficult to replicate. In the context of bistable perception, previous studies have shown opposite effects after TMS to aSPL (a region close to pSPL but with a different proposed function; see Section 9.2; Carmel et al., 2010; Zaretskaya et al., 2010). It is not clear whether these opposite findings relate to differences in stimulated locus or in stimulation methodology. One important consideration is that the effects of electromagnetic induction can spread along neuronal networks rather than remaining local to the stimulation site. In addition TMS effects can be dependent on current neuronal states (Silvanto and Pascual-Leone, 2008), and this may be particularly relevant in bistable perception where states are
dynamic. These sorts of effects are very difficult to account for adequately without concurrent use of functional imaging modalities.

5.4.2 Absence of effect of TMS to FEF on percent durations

The main prediction for this experiment was that I would observe an increase of mean SFM percept durations after cTBS over right FEF. Instead, I found no statistically significant change in percept duration for this stimulation condition. This null result would have been more convincing had I been able to replicate the previously established pSPL effect (see Sections 5.1 and 5.4.1). Given the failure to confirm an effect of pSPL stimulation on SFM percept duration in the present experiment, it cannot be conclusively determined whether the lack of FEF effect relates to neurobiological or to technical factors. Given the considerations outlined in Section 5.4.1, it seems possible that technical factors (namely, under-sampling) are contributing to the failure to demonstrate significant effects of cTBS to both right pSPL and right FEF on SFM percept durations.

In the present experiment FEF was not defined functionally (e.g. by its involvement in saccade generation, as in some previous experiments; Amiez and Petrides, 2009; Paus, 1996) so it is possible that the target region I used is in fact another cortical area nearby. Indeed, the presently stimulated locus lies slightly medial (and a little posterior) to sites stimulated in some previous studies targeting FEF (e.g. Muggleton et al., 2003; Prime et al., 2010; Ruff et al., 2006). However, the same locus as at present has previously been implicated both in saccade generation (Grosbras et al., 2001) and in deployment of spatial attention (Giesbrecht et al., 2003). Furthermore, my target lies within the extent of previous mapping of FEF across multiple studies, where the $x$-coordinates for FEF were found to be particularly variable (Amiez and Petrides, 2009; Paus, 1996). Thus, it is likely that the presently stimulated locus lies within functionally relevant FEF, although I am unable to confirm this with the available data. Ultimately, whether or not this is the case is important for accurate nomenclature and understanding of anatomical relationships to other FEF-related functions. On the other hand whether or not this stimulated locus is called FEF does not affect its proposed role in bistable perception, which is based on the findings reported in Chapter 3. It is possible that adjacent regions in the vicinity of FEF, that do not completely overlap anatomically, subserve saccadic and perceptual reorganisation functions respectively.
5.4.3 Ratings of intensity and unpleasantness of TMS

While participant ratings of TMS intensity did not differ significantly across stimulated regions, there was a small effect of stimulated region on unpleasantness ratings (such that TMS to FEF was significantly more unpleasant than TMS to pSPL; Section 5.3.3). It is possible that the unpleasant effect of TMS in itself made a contribution to the pattern of results for percept duration change. Indeed, anxiety (which could result from an unpleasant TMS experience) is associated with increased switch rate in bistable perception (Nagamine et al., 2007), and such an effect could act in opposition to the predicted increase in percept duration, and thus potentially reduce the sensitivity for detecting an effect of TMS to FEF. However, such an interpretation would not explain the failure to observe an effect of TMS to pSPL (where TMS was not rated as more unpleasant than vertex stimulation). More generally, the higher intensity and unpleasantness ratings after FEF stimulation were expected (see Section 5.2.7) and as such this finding supports the notion that TMS was correctly administered in the experiment.

5.4.4 Possible influences of pupil size and eye position on SFM percept duration

The SFM behavioural paradigm required visual fixation (see Section 5.2.3). FEF is important in gaze control (see Section 5.1), and in particular is involved in disengagement of fixation to trigger intentional saccades (Rivaud et al., 1994); therefore there is a possibility that fewer such saccades were made following inhibitory TMS over FEF (although of note, since the experimental task requires fixation, participants should be making minimal intentional saccades if the task is being performed well). Whether a difference in number of saccades would importantly influence performance of the behavioural task (for example affecting percept duration) is not clear; however it is well established that eye movements can have an effect on percept duration in bistable paradigms, including specifically for SFM (e.g. Brouwer and van Ee, 2006). This issue may even extend beyond the FEF stimulation condition given that the parietal eye field (found in IPS) is also involved in saccadic function, triggering reflexive saccades by disengagement of fixation (Müri and Nyffeler, 2008). Considering the limited spatial resolution of TMS (see Section 2.3.2), it is of course possible that the pSPL stimulation could affect relevant regions in IPS.

There were no differences in number of saccades generated in the different TMS stimulation conditions or in the pre-TMS versus post-TMS periods. The fact that inhibitory cTBS over FEF did not result in any effect on saccade generation in the present
experiment (see Section 5.3.4) raises additional suspicion that the locus stimulated in this study may not overlap the 'classic' FEF region. In all likelihood, however, limited emphasis can be placed on this possibility given that the experiment did not involve saccade generation (but rather required visual fixation) and was therefore not designed to evaluate any disruption of saccadic function.

The eye tracking results suggested that visual fixation did not differ systematically between stimulated regions or between pre-TMS and post-TMS periods (Section 5.3.4). However, I found a main effect of TMS on pupil size (pupils were relatively dilated following TMS across all stimulated regions, and in particular after FEF and vertex stimulation with a weaker effect after pSPL stimulation). One possibility is that pupillary dilatation following TMS relates to increased arousal (e.g. associated with the sound or the sensation of the TMS pulse; Bradley et al., 2008). It has previously been suggested that reduced arousal is associated with slowing of switch rate in bistable perception (Carter et al., 2007) so the opposite effect could be postulated with increased arousal (and indeed increased switch rate is seen in the context of anxiety; Nagamine et al., 2007). If arousal and/or anxiety related to the TMS stimulation resulted in speeding of switch rate (equating to reduced mean percept duration), this may partially mask any opposite effect on bistable perception dynamics associated with TMS to FEF or pSPL. However, as for the argument related to unpleasantness ratings (Section 5.4.3), such an explanation would not account for the failure to replicate previous results obtained following TMS to pSPL (Kanai et al., 2010).

There was a main effect of TMS on horizontal eye position (consisting of a slight rightward shift in gaze following TMS, in particular after vertex and to a lesser degree pSPL stimulation but not after FEF stimulation). The reason for this finding is not clear; one could speculate that this is an effect of applying TMS to the right side of the scalp (which may draw eye gaze in the same direction), although this would then apply to FEF and pSPL stimulation but not to vertex stimulation, which is in the midline. Whether a slight rightward shift in gaze should influence SFM dynamics is not entirely clear. It has been shown that for other bistable paradigms the point of fixation may vary with currently perceived stimulus but does not seem to influence perceptual switching (van Dam and van Ee, 2006).

Overall, it seems doubtful that the observed differences in eye gaze and pupil size would be responsible for the lack of significant effect of TMS over FEF or pSPL. More detailed analysis of eye position and pupillary size to study temporally-specific changes associated with perceptual transitions (or with leftward versus rightward moving percepts in SFM)
may be an interesting development for future work and may add new perspective to the perceptual transition-locked pupil response analysis for CFS (Chapter 8).

5.4.5 Summary

Inhibitory cTBS over right FEF (the site of one of the stimulus-invariant neural correlates of perceptual transition found in Chapter 3) did not confirm a causal role of this region in the dynamics on bistable perception. An attempt to replicate an established finding of increased mean percept duration after cTBS over right pSPL was also unsuccessful. The latter null result may be due to differences in the length of behavioural task, meaning the present experiment did not have sufficient statistical power to detect effects of TMS on SFM dynamics. In support of such a possibility, trends towards the predicted post-TMS increases in percept duration were seen after both FEF and pSPL stimulation when compared to control vertex stimulation. FEF stimulation was rated as more unpleasant than that over the other two target regions. In addition minor differences were observed between pre-TMS and post-TMS eye position and pupil size measurements, although none of these were specific to any of the stimulated regions. I argue that neither of these last two sets of findings is likely to account for the lack of significant effects of TMS on percept duration. A repeat study, involving a longer testing period to measure SFM percept duration both pre- and post-TMS, may help clarify the causal role of right FEF in the dynamics of bistable perception. If such a modification of the current experiment at least enables a successful active control (replication of the established effect of cTBS to pSPL), even a null result from FEF stimulation would be interpretable. Additionally, more detailed evaluation of saccadic function could establish whether the present target FEF region overlaps with the widely reported occulomotor control region in FEF.
Chapter 6

Evaluation of social face traits outside of awareness

‘A man's face as a rule says more, and more interesting things, than his mouth, for it is a compendium of everything his mouth will ever say, in that it is the monogram of all this man's thoughts and aspirations.’

Arthur Schopenhauer (1788 – 1860)

6.1 Introduction

This chapter and the remaining two chapters in this thesis will focus on social perception, and its interaction with visual awareness. Behaviour can be influenced by social cues (for example emotional facial expressions) even when these cues are not consciously perceived. Experimental demonstrations of this notion have been reviewed in Section 1.4 where I also discuss the importance of social traits including dominance and trustworthiness, how these can be conveyed by facial appearance, and the usefulness of the computer-generated social face model of Oosterhof and Todorov (2008) as a tool to study such social trait evaluation (Section 1.4.2). While there has been some investigation of non-conscious evaluation of facial emotional expression or eye gaze direction (Section 1.4.4), little is known about non-conscious evaluation of social face traits. I therefore conducted a behavioural experiment making use of emotionally neutral face images with independently variable dominance and trustworthiness traits (according to the model of Oosterhof and Todorov, 2008; Section 1.4.2), and displaying them under conditions of CFS (see Section 1.3.3) in order to render them invisible to observers. By using the b-CFS variant of the CFS paradigm, a single measure of time-to-emergence (t2e; the time required for an image to break into
awareness) could be obtained, reflecting the potency of the suppressed image in competing for awareness (see Section 1.3.3). The specific question I aimed to answer was whether the dominance and/or trustworthiness traits of non-conscious faces influenced the t2e measure. My prediction was that dominant and untrustworthy faces would gain access to awareness more quickly (have shorter t2e) than socially neutral faces. This prediction is based on previous findings of speeding of b-CFS for socially relevant stimuli, in particular faces with fearful expressions (Tsuchiya et al., 2009; Yang et al., 2007), but also faces with direct gaze (Stein et al., 2011). The prediction is also in line with evolutionary vigilance accounts (e.g. Sander et al., 2003; Whalen, 1998), and related proposals for rapid/enhanced subcortical processing for socially relevant visual information (Morris et al., 1999; Vuilleumier, 2005; see Section 1.4.4 for more detail). I hypothesised that non-conscious dominant and untrustworthy faces would be processed in a similar fashion to fearful faces, and would therefore evoke similar speeding of b-CFS responses.

As discussed in Section 1.6, a growing literature has focused on understanding individual variability in behaviour, cognition or neurobiology by evaluating relationships between these different types of measures at individual subject level. For example, of relevance to the theme of this chapter, this approach has been used to explore between-participant differences in the impact of socially relevant visual information on awareness. As part of previous collaborative work (Stewart et al., 2012; see Contributions and Publications) my colleagues and I have shown that individual variability in evaluation of non-consciously perceived dominance and trustworthiness traits is strongly correlated with scores on self-report questionnaires that reflect inclination to submissive behaviour and propensity to trust others, respectively. The neural correlates of non-conscious evaluation of social face traits remain unknown. A growing number of studies in recent years have explored relationships between individual variability in focal brain structure and individual variability in behaviour (reviewed by Kanai and Rees, 2011; see also Section 1.6), although none have yet focused on possible correlations between brain structure and non-conscious social perception. This was the rationale for undertaking a second experiment, using MRI. Here the hypothesis was that local variation in brain structure might relate to the individual variability in non-conscious social evaluation shown in the behavioural experiment. Therefore, an individual differences approach was employed by seeking correlations between local structural MRI measures obtained using VBM (Section 2.2.6) and behavioural indices of non-conscious facial dominance and trustworthiness evaluation obtained using faces from the model of Oosterhof and Todorov in b-CFS. Using gray matter volume (from VBM) as the primary measure, and based on previous findings, I predicted that: (1) behavioural measures of non-conscious dominance evaluation would be correlated with GM volume in amygdala and insula (Chiao et al., 2008; Dannlowski et al., 2007; Whalen et al., 2001); and (2) behavioural measures of non-conscious trustworthiness
evaluation would be correlated with GM volume in amygdala, insula, fusiform gyrus and mPFC (Todorov et al., 2008; Winston et al., 2002).

6.2 Behavioural materials and methods

6.2.1 Participants

Thirty-six right-handed participants (23 female; mean ± SD age = 23.2 ± 4.6 years; range = 18 to 35 years) were recruited according to inclusion criteria and procedures as detailed in Section 2.4.1.

6.2.2 Stimuli and Display Apparatus

Apparatus, stimuli and procedures were as described in Sections 2.4.2 and 2.4.3. A single randomly generated face image was manipulated to every permutation of dominance and trustworthiness, each at -3, 0 and +3 standard deviations from the neutral, resulting in 9 versions of the same face identity (Figure 6.1). Stimuli were presented on a Sony Trinitron GDM-F520 monitor (1600 x 1200 at 85 Hz) and viewed through a mirror stereoscope at a distance of 65.5 cm.

**Figure 6.1:** Face stimuli from the model of Oosterhof and Todorov (2008)

See Section 1.4.2. Trustworthiness and dominance vary in standard deviations from a neutral face along the x-axis and y-axis respectively. The nine face images shown were all utilised in the behavioural experiment.
6.2.3 Behavioural Procedures

Procedures were as described in Section 2.4.4. A CFS paradigm as per Section 2.4.4.1 was employed (Figure 6.2). Stimuli each subtended 11.77° visual angle with the fixation cross subtending 0.6° visual angle. For this experiment the dynamic noise pattern changed at a frequency of 9.4 Hz rather than 10 Hz (due to the screen refresh rate). In each experimental trial the face appeared 0.7° visual angle to the left or right of the central fixation cross. Correct responses on the left/right discrimination task were used to calculate time-to-emergence (t2e) while incorrect-response and non-response trials (no response made by 10 seconds) were excluded from further analysis.

![Figure 6.2: Schematic representation of CFS behavioural paradigm](image)

Temporal sequence of events during each experimental trial is shown. The CFS mask presented to one eye (here, the right eye) changes with a frequency of 9.4 Hz. Subjects respond by pressing one of two buttons to indicate whether the face appears on the left or the right within the black box (in the example trial shown the correct response would be “left”).

Participants completed a total of 288 trials (8 blocks of 36 trials each) with each of the 9 face versions presented a total of 32 times (4 times in each block). Before the beginning of the experiment, a 36-trial practice block was undertaken, which enabled determination of eye dominance (see Section 2.4.4).
6.3 Structural MRI methodology

6.3.1 MRI Data acquisition

High-resolution anatomical MR images were obtained for all 36 participants on a separate occasion to behavioural testing. The images were acquired on a 1.5T Siemens Sonata MRI scanner. A T1-weighted 3-D MDEFT sequence (TR = 12.24 ms; TE = 3.56 ms; field of view = 256 x 256 mm; voxel size = 1 x 1 x 1 mm) was used (Deichmann et al., 2004).

6.3.2 VBM analysis of MRI data

VBM analysis (Section 2.2.6) was performed on the acquired imaging data using SPM8. MR images were segmented for gray matter and white matter. Subsequently, DARTEL (Ashburner, 2007) was performed for inter-subject registration of the GM images. The registered images were smoothed with a Gaussian kernel (FWHM = 8 mm) and then transformed to MNI stereotactic space using affine and non-linear spatial normalisation.

The GLM was implemented in SPM8 with both dominance and trustworthiness effects placed in the design matrix as regressors of interest. Potentially confounding factors of gender and age, which affect brain structure (Good et al., 2001; Smith et al., 2007), were regressed out by modeling them as covariates of no interest. Global nuisance effects were accounted for by including the global covariate in the GLM. Statistical maps of voxel intensity statistics that reflect regional covariation of GM volume with the measures of interest were produced (Section 2.2.6.4). Non-stationary cluster-level correction (Hayasaka et al., 2004) was undertaken to improve the reliability of cluster-level statistics. I used \( p < 0.05 \) (FWE-corrected for whole brain volume) as the criterion for significance when exploring correlations between peak voxel-level results and individual behavioural measures. Based on previous evidence, correlations between behavioural measures and GM volume in a number of brain regions had been predicted \textit{a priori}, (see Section 6.1; peak coordinates of regions and references to studies in which they were reported are listed in Table 6.1). To test these predictions, small volume correction for multiple comparisons within a sphere with 15 mm radius, centred at each of the coordinates listed in Table 6.1, was employed. For fusiform gyrus and amygdala, smaller spheres with 8 mm radius (corresponding to a volume containing 2550 voxels of 1mm\(^3\) each, roughly matching the volume of functionally relevant portions of these regions; Costafreda et al., 2008; Joseph, 2001) were used. A threshold of \( p < 0.05 \) (FWE-corrected for small volume) was used as the criterion for significance.
SV Getov, Chapter 6. Non-conscious social face traits

Table 6.1: Coordinates of loci used for small volume correction analysis

<table>
<thead>
<tr>
<th>Location</th>
<th>HEM</th>
<th>MNI Coordinates</th>
<th>Source</th>
<th>Significant Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amygdala</td>
<td>Left</td>
<td>-24 -2 -22</td>
<td>Harvard-Oxford SSA</td>
<td>No</td>
</tr>
<tr>
<td>Amygdala</td>
<td>Right</td>
<td>26 0 -22</td>
<td>Harvard-Oxford SSA</td>
<td>No</td>
</tr>
<tr>
<td>Putamen</td>
<td>Left</td>
<td>-16 13 -4</td>
<td>Todorov et al, 2008</td>
<td>No</td>
</tr>
<tr>
<td>Precuneus</td>
<td>Left</td>
<td>-1 -61 39</td>
<td>Todorov et al, 2008</td>
<td>No</td>
</tr>
<tr>
<td>mPFC</td>
<td></td>
<td>1 59 24</td>
<td>Todorov et al, 2008</td>
<td>Yes¹</td>
</tr>
<tr>
<td>Fusiform G</td>
<td>Left</td>
<td>-48 -48 -24</td>
<td>Winston et al, 2002</td>
<td>Yes²</td>
</tr>
<tr>
<td>Fusiform G</td>
<td>Right</td>
<td>44 -46 -22</td>
<td>Winston et al, 2002</td>
<td>Yes²</td>
</tr>
<tr>
<td>pSTS</td>
<td>Right</td>
<td>56 -44 4</td>
<td>Winston et al, 2002</td>
<td>No</td>
</tr>
<tr>
<td>Insula</td>
<td>Right</td>
<td>42 -4 12</td>
<td>Winston et al, 2002</td>
<td>Yes³</td>
</tr>
<tr>
<td>Lingual Gyrus</td>
<td>Right</td>
<td>27 -54 13</td>
<td>Chiao et al, 2008</td>
<td>No</td>
</tr>
<tr>
<td>Sup.Temp. G</td>
<td>Right</td>
<td>60 -51 3</td>
<td>Chiao et al, 2008</td>
<td>No</td>
</tr>
<tr>
<td>Insula</td>
<td>Right</td>
<td>39 9 15</td>
<td>Chiao et al, 2008</td>
<td>Yes³</td>
</tr>
</tbody>
</table>

6.3.3 Surface-based analysis of MRI data

Gray matter volume, as calculated by VBM, provides a mixed measure of cortical gray matter, which subsumes both cortical surface area and cortical thickness (Hutton et al., 2009; Sections 2.2.6 and 2.2.7). Therefore, further analyses were performed to separately compute cortical thickness, cortical surface area and cortical volume using surface-based methods implemented in FreeSurfer (using FreeSurfer Stable v5.3.0; see Section 2.2.7).

¹ Positive correlation between untrustworthiness-related slowing and GM volume in mPFC; see Table 6.3 and Section 6.4.2.2 for details.
² Negative correlation between untrustworthiness-related slowing and GM volume in bilateral fusiform gyrus; see Table 6.3 and section 6.4.2.2 for details.
³ Negative correlation between dominance-related slowing and GM volume in the same locus as reported in whole-brain analysis (see Table 6.2 and Section 6.4.2.2).
Reconstruction of the pial surface and GM/WM boundary were performed using each participant’s T1-weighted MR image, according to a fully automated procedure (Fischl et al., 1999). Once a complete cortical model had been generated, deformable procedures could be performed, including creation of surface based data such as maps of curvature and sulcal depth. This enabled computation of cortical thickness, surface area, and volume. Cortical thickness was defined as the shortest distance between the GM/WM boundary and the pial surface (Fischl and Dale, 2000). The pial surface and thickness data were transformed to a standard brain (fsaverage) using a surface-based coregistration implemented in FreeSurfer. The thickness and area data were smoothed with a Gaussian kernel (FWHM = 10 mm). For each of the measures of cortical thickness, surface area, and volume, a separate multiple regression model was constructed to identify cortical regions where the structural measure under examination was correlated with each of the behavioural measures of interest. As for the VBM analyses described in Section 6.3.2, both gender and age of participants, as well as behavioural measures not being evaluated, were included in the multiple regression analysis design matrix as covariates of no interest. In this way, I was able to regress out any effects attributable to these variables. Cluster-wise correction for multiple comparisons was performed using Monte Carlo simulation (Hagler et al., 2006). A cluster-wise threshold of $p < 0.05$ was used as the criterion for significance.

For calculating volume of subcortical structures (in particular, the amygdala), I also used FreeSurfer (Stable v4.0.5) to provide additional measures that were independent from the VBM analysis. I used automated subcortical structure segmentation (Fischl et al., 2002) to compute the volume of both left and right amygdala from the MR image of each participant.

### 6.4 Results

In order to assess non-conscious social face evaluation, I presented face images varying along orthogonal dimensions of dominance and trustworthiness to healthy human volunteers outside of their awareness (using CFS). In a first behavioural experiment, I recorded t2e, a reaction time measure that reflects the strength of each non-conscious face image in competing for awareness. In a second MRI experiment, analyses were performed to interrogate relationships between individual variability in non-conscious social face evaluation and local brain structure.
6.4.1 Behavioural experiment results

There was substantial inter-individual variability of mean t2e values across participants (mean t2e range = 0.63 - 3.39 seconds; Figure 6.3).

![Figure 6.3](image)

**Figure 6.3:** Frequency distribution of t2e values

Histogram shows substantial variability in mean t2e across individuals in the experimental sample. Figure adapted from Getov et al. (2015).

6.4.1.1 Main effects of dominance and trustworthiness on t2e

To explore effects of non-conscious social evaluation t2e results were entered into two-way repeated-measures ANOVA with factors of dominance and trustworthiness (3 levels for each factor). This revealed a significant main effect of dominance $[F_{(2,70)} = 8.88, p < 0.001]$, an almost-significant main effect of trustworthiness $[F_{(2,70)} = 2.45, p = 0.094]$, and no interaction between these main effects $[F_{(4,140)} = 1.34, p = 0.264]$. See Figure 6.4.
Figure 6.4: Effects of social face traits on t2e

(A) shows dominance and (B) shows trustworthiness effects. In the left hand panels, mean values of t2e across all subjects (n = 36) are plotted along the y-axis. Along the x-axis are plotted dominance (A) and trustworthiness (B), in standard deviations from the neutral. For each level of dominance in A and trustworthiness in B, t2e scores are collapsed across the three levels of the other social trait. Error bars represent standard errors of mean difference between the specific condition and a neutral face. In the right hand panels, frequency distributions of individual values for (A) dominance-related slowing [t2e(3dom) - t2e(neutral)], and for (B) untrustworthiness-related slowing (t2e(3trust) - t2e(neutral)) are shown. Substantial individual variability can be seen for both dominance and trustworthiness-related measures. Figure adapted from Getov et al. (2015).

Post hoc comparisons revealed that the main effect of dominance reflected significantly slower t2e for most-dominant faces relative to least-dominant faces [t(35) = -3.41, p = 0.002], and to neutral-dominance faces [t(35) = -4.31, p < 0.001]. The main effect of trustworthiness reflected borderline-significant slowing of t2e for least-trustworthy faces relative to neutral-trustworthiness faces [t(35) = 1.98, p = 0.056]. There were no significant differences between t2e for neutral-dominance and least-dominant faces [t(35) = 0.56, p = 0.58], between t2e for neutral-trustworthiness and most-trustworthy faces [t(35) = -0.04, p = 0.36], or between t2e for least-trustworthy and most-trustworthy faces [t(35) = 1.42, p = 0.17]. Because parametric statistical tests have been used, it was important to check that log transforming the behavioural data did not lead to any important changes in the results. Such log transformation resulted in very minimal changes, suggesting that use of
untransformed data was reasonable: main effect of dominance \(F(2,70) = 9.04, p < 0.001\); main effect of trustworthiness \(F(2,70) = 2.54, p = 0.086\); interaction \(F(4,140) = 1.29, p = 0.28\). *Post hoc* comparisons again showed significant differences between most-dominant and least-dominant faces \(t(35) = -3.64, p = 0.001\) as well as most-dominant and neutral-dominance faces \(t(35) = -4.48, p < 0.001\), and a borderline-significant difference between least-trustworthy and neutral-trustworthiness faces \(t(35) = 2.02, p = 0.051\). These behavioural data have been reported in a published paper (Stewart et al., 2012; Experiment 3) and closely replicate independent findings presented in Experiment 1 and Experiment 2 from that study.

Inclusion of gender and age into the statistical analysis of the behavioural data also did not significantly change the findings. Repeated measures ANOVA with within-subject factors of dominance and trustworthiness (3 levels for each factor), and between-subject factors of gender and age, revealed a significant main effect of dominance \(F(2,34) = 4.75, p = 0.015\), and a non-significant main effect of trustworthiness \(F(2,34) = 1.57, p = 0.22\). The main effects of gender \(F(1,17) = 0.80, p = 0.39\) and age \(F(13,17) = 0.53, p = 0.88\) were not significant. There was a borderline-significant dominance*gender*age interaction \(F(8,34) = 2.20, p = 0.053\). None of the other interaction terms reached statistical significance.

### 6.4.1.2 Individual variability of dominance and trustworthiness effects

Based on group-level effects of facial dominance and trustworthiness, and the particularly clear differences in t2e between certain levels of these traits, a representative measure for individual effects of each of these dimensions of face evaluation was calculated as follows:

1) Dominance-related slowing \([t2e(+/3dom) - t2e(neutral)]\)

2) Untrustworthiness-related slowing \([t2e(-3trust) - t2e(neutral)]\)

As with mean t2e, there was again substantial inter-individual variability in the dominance-related slowing and untrustworthiness-related slowing effects (Figure 6.4). These two measures were not correlated \((r = -0.10, p = 0.56)\).

### 6.4.1.3 Error rates in the behavioural task

Rates of incorrect responses in the left/right behavioural task were low (mean 2.2%, range 0 – 7.6%, with four of the 36 participants making no errors at all), suggesting that all
included participants understood the task and performed it with care. There were no significant correlations between error rates and t2e \( r = 0.04, p = 0.82 \), error rates and dominance-related slowing \( r = -0.23, p = 0.18 \), or error rates and untrustworthiness-related slowing \( r = 0.03, p = 0.89 \).

6.4.1.4 Testing for outliers in the behavioural results

The bimodal appearance of the distribution of individual t2e in Figure 6.3 raises the possibility that the group of four participants with fastest t2e (<1.5 seconds) exhibits a different pattern of behaviour to the rest of the experimental sample. I performed further analyses showing that only one of these four data points was classed statistically as an outlier (according to Chauvenet’s criterion) and removing this data point did not substantially alter the results. Regarding these four participants, I also confirmed that: 1) the response error rate (Section 6.4.1.3) was not different to that for the rest of the experimental sample; 2) response times were distributed similarly to the rest of the experimental sample, approximating to gamma distributions (Figure 6.3).

Figure 6.5: Frequency distributions of t2e response times

The plots show histograms for individual participants (four individuals with fastest t2e on the top row and four other randomly selected individuals on the bottom row). t2e (ms) divided into time bins is shown along the x-axis and proportion of all response times accounted for by each time bin is shown along the y-axis. A gamma function is fitted to the data in each histogram (red dotted lines). All participants’ data fit reasonably well to a gamma distribution. There is no systematic difference between the distributions of the fast-t2e individuals (top row) and other individuals (bottom row). The statistical outlier discussed in Section 6.4.1.4 is S27.
The four fast-t2e individuals are not found at the extreme ends of the distributions of dominance-related slowing and untrustworthiness-related slowing scores (which are the behavioural measures used in the MRI analyses detailed in Section 6.4.2; see Figure 6.6).

![Figure 6.6: Location of fast-t2e individuals on dominance-related slowing and untrustworthiness-related slowing distributions](image)

Frequency distributions of individual values for (A) dominance-related slowing and (B) untrustworthiness-related slowing are shown (same as Figure 6.4 but here the time bins within which the four fastest-t2e participants fall are coloured in red; it is apparent that the fast-t2e participants are not located near the extremes of either of these distributions).

Removing the four individuals with fast t2e (<1.5 seconds) from subsequent analyses did not alter the pattern of the experimental findings, although the statistical significance was partially reduced, likely due to reduction in sample size. There remained a significant main effect of dominance \( [F_{(2,62)} = 8.13, \ p = 0.001] \); the main effect of trustworthiness was reduced in strength \( [F_{(2,62)} = 1.55, \ p = 0.22] \), and as previously there was no significant interaction between these main effects \( [F_{(4,124)} = 1.10, \ p = 0.36] \). Post hoc comparisons revealed significant differences between most-dominant and least-dominant faces \( [t_{(31)} = -3.07, \ p = 0.004] \) and between most-dominant and neutral-dominance faces \( [t_{(31)} = -4.23, \ p < 0.001] \); there was a persisting but weakened trend towards a difference between least-trustworthy and neutral-trustworthiness faces \( [t_{(31)} = 1.60, \ p = 0.12] \).

Overall, the findings described in this section indicate that the four fast-t2e individuals did not make more task errors, had similarly distributed response times and did not have extreme dominance-related slowing or untrustworthiness-related slowing effects. Removing these four individuals from the behavioural analyses does not alter the pattern of experimental findings and they are therefore retained in the main experimental sample.
6.4.2 MRI experiment results

Next I tested the hypotheses that individual variability in dominance-related slowing and untrustworthiness-related slowing would be correlated with individual differences in local brain structure.

6.4.2.1 Correlations between GM volume and behavioural measures: whole-brain VBM analysis

Both dominance-related slowing and untrustworthiness-related slowing scores were entered in the same SPM design matrix for VBM analysis. There were two key findings. First, GM volume in right frontal operculum was significantly correlated with individual differences in dominance-related slowing ($x = 48$, $y = 2$, $z = 13$; $T = 6.27$; $P_{\text{FWE-corr}} = 0.016$; Figure 6.7; Table 6.2).

![Figure 6.7: Structural brain correlates of individual differences in non-conscious dominance-related slowing](image)

A locus in right frontal operculum, where GM volume is correlated significantly with behavioural effects of dominance-related slowing, is shown in colour (from brown, representing low correlation, to white representing high correlation), overlaid on a standard template brain. A threshold of $p < 0.001$ uncorrected is used for display purposes, while the statistical significance threshold was $p < 0.05$ FWE-corrected for whole brain volume. Colourbar scale represents T values. Figure from Getov et al. (2015).

According to data from the Anatomy Toolbox for SPM, the region shown in Figure 6.7 is located between insula, IFG and SII (Figure 6.8). The probabilistic cytoarchitectonic maps provided in Anatomy Toolbox indicate a 10% probability that this locus is located in IFG (Brodmann 44) and a 10% probability that it is located in S2.
Figure 6.8: Cytoarchitectonically-defined regions near the locus where GM volume is correlated with dominance-related slowing

Probabilistic cytoarchitectonic maps from the SPM Anatomy Toolbox were referenced. The locus in right frontal operculum (same as in Figure 6.7) is shown in colour (from brown, representing low correlation, to white representing high correlation), along with anatomical masks derived from the SPM Anatomy toolbox for posterior insula (cyan), IFG (red), and S2 (blue). The result and masks are overlaid on a standard template brain. Figure from Getov et al. (2015).

<table>
<thead>
<tr>
<th>Behavioural Effect</th>
<th>Corr</th>
<th>Location</th>
<th>MNI Coordinates</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Region</td>
<td>Hem</td>
<td>X</td>
</tr>
<tr>
<td>Dominance-related</td>
<td>Neg</td>
<td>Frontal operculum</td>
<td>Right</td>
<td>48</td>
</tr>
<tr>
<td>slowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untrustworthiness-</td>
<td>Neg</td>
<td>TPJ</td>
<td>Right</td>
<td>51</td>
</tr>
<tr>
<td>related slowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2: VBM analysis, whole brain statistics

Coordinates and statistical results for peak voxels in clusters where there is correlation between GM volume and dominance-related slowing or untrustworthiness-related slowing ($p < 0.05$, FWE corrected across whole brain volume). Corr, direction of correlation; Hem, hemisphere.

The second key finding was that GM volume in right pTPJ was significantly correlated with individual differences in untrustworthiness-related slowing ($x = 51$, $y = -57$, $z = 31$; $T = 6.14$; $P_{FWE-corr} = 0.022$; Figure 6.9A; Table 6.2). Both findings (right frontal operculum and right pTPJ) were statistically significant after whole-brain FWE correction. In addition, both correlations were negative, indicating that reduced GM volume in these regions predicts increased dominance-related slowing and untrustworthiness-related slowing, respectively.
Figure 6.9: Structural brain correlates of individual differences in non-conscious untrustworthiness-related slowing

Loci shown in (A) pTPJ; (B) mPFC; and (C) bilateral fusiform gyrus, are regions where GM volume is correlated significantly with behavioural effects of untrustworthiness-related slowing. Colours (from brown, representing low correlation, to white representing high correlation) are overlaid on a standard template brain. A threshold of $p < 0.001$ uncorrected has been used for display purposes. For panel A the significance threshold is $p < 0.05$, family-wise-error-corrected for whole brain volume (Section 6.4.2.1). For panels B and C the significance threshold is $p < 0.05$, corrected for small volume around coordinates predicted a priori (Section 6.4.2.2). Colourbar scales represent T values. Figure from Getov et al. (2015).

Right TPJ has been divided into three subregions based on diffusion-weighted tractography and resting state functional connectivity (Mars et al., 2011). I used mask images for each of these subregions to determine within which of them the right pTPJ cluster reported here falls. Small-volume correction using the most posterior TPJ mask (TPJp) resulted in a similar result to the above-reported whole-brain pTPJ finding ($x = 51, y = -58, z = 31; T = 5.77; P_{FWE-corr} < 0.001$), confirming that the majority of this pTPJ cluster falls within the TPJp region described by Mars and colleagues (Figure 6.10).
The present pTPJ result (Table 6.2; Figure 6.9A) and the TPJp mask of Mars and colleagues (2011) are overlaid on a standard template brain. Substantial overlap is seen between the present pTPJ finding (shown in colour; scale from brown to white, representing low to high correlation with untrustworthiness-related slowing) and the TPJp subregion (shown in blue). The TPJp subregion is known to have resting-state functional connectivity with other brain regions involved in social cognition (Mars et al., 2011). Figure from Getov et al. (2015).

6.4.2.2 Correlations between GM volume and behavioural measures: small-volume-corrected VBM analysis

Small-volume corrected analyses were performed as per Section 6.3.2. There was a significant positive correlation between GM volume in mPFC and untrustworthiness-related slowing \((x = -2, y = 54, z = 13; T = 4.15; P_{FWE-corr} = 0.027; \text{Figure 6.9B; Table 6.3; small volume correction performed around a region that shows non-linear BOLD response to changes in face trustworthiness reported by Todorov et al., 2008, see Table 6.1). GM volume in fusiform gyrus bilaterally was correlated significantly and negatively with untrustworthiness-related slowing \((x = -47, y = -45, z = -18; T = 3.52; P_{FWE-corr} = 0.023 \text{ for left fusiform; and } x = 50, y = -44, z = -20; T = 3.67; P_{FWE-corr} = 0.017 \text{ for right fusiform; Figure 6.9C; Table 6.3; small volume correction performed around loci with BOLD changes in association with changes in face trustworthiness as reported by Winston et al., 2002, see Table 6.1). GM volume in right frontal operculum was negatively correlated with dominance-related slowing \((x = 48, y = 2, z = 13; T = 6.27; P_{FWE-corr} < 0.001; \text{ small volume correction performed around loci in right insula, where BOLD signal change has previously been reported for untrustworthy faces by Winston et al., 2002 and for dominant head postures by Chiao et al., 2008). The final result in fact refers to the same locus in right frontal operculum reported for whole-brain analysis (Table 6.2) and is therefore not discussed further. However, it demonstrates that the location of the right frontal opercular
locus reported in this study (Table 6.2) is not distant from coordinates previously reported in right insula with respect to processing of dominance and/or anger (Whalen et al., 2001; Dannlowski et al., 2007; Chiao et al., 2008).

<table>
<thead>
<tr>
<th>Behavioural Effect</th>
<th>Corr</th>
<th>Location</th>
<th>MNI Coordinates</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Region</td>
<td>HEM</td>
<td>X</td>
</tr>
<tr>
<td>Untrustworthiness-related slowing</td>
<td>Pos</td>
<td>mPFC (see Table 6.1)</td>
<td></td>
<td>-2</td>
</tr>
<tr>
<td>Untrustworthiness-related slowing</td>
<td>Neg</td>
<td>Fusiform G (see Table 6.1)</td>
<td>L</td>
<td>-47</td>
</tr>
<tr>
<td>Untrustworthiness-related slowing</td>
<td>Neg</td>
<td>Fusiform G (see Table 6.1)</td>
<td>R</td>
<td>50</td>
</tr>
</tbody>
</table>

*Table 6.3: Coordinates and statistical results for small volume correction analyses*

Peak voxels shown, within 15 mm or 8 mm spheres used for small-volume correction, where GM volume was significantly correlated with dominance-related slowing or untrustworthiness-related slowing (*p* < 0.05, FWE corrected for multiple comparisons across a 15 mm or 8 mm sphere centred at coordinates specified in Table 6.1). Corr, direction of correlation; HEM, hemisphere; Fusiform G, fusiform gyrus.

### 6.4.2.3 Linearity/outliers in correlations between GM volume and behaviour

To investigate whether any of the reported correlations between dominance-related slowing and untrustworthiness-related slowing with focal GM volume could have been driven by statistical outliers in the data I examined scatterplots for each correlation (Figure 6.11). Of note, other than checking that the correlations are driven by linear relationships rather than outliers, these scatterplots cannot be interpreted further, as to do so would constitute a non-independent analysis and be an example of double-dipping (Kriegeskorte et al., 2009; Vul et al., 2009). There is an outlier in the correlation between frontal opercular GM volume and dominance-related slowing (Figure 6.11A). After removal of this outlier from analysis, this correlation is no longer significant after correction for multiple comparisons across the whole brain volume. However, small-volume correction at the *a priori* prediction of right insula (centred at a coordinate reported by Chiao et al., 2008; *x* = 39, *y* = 9, *z* = 15; see Table 6.2) still leads to a significant result at the same locus (*x* = 50, *y* = 9, *z* = 15).
SV Getov, Chapter 6. Non-conscious social face traits

= 2, z = 13; T = 4.38, \( P_{\text{FWE-corr}} = 0.014 \) as the original whole-brain-corrected result. None of the other scatterplots (Figure 6.11B-E) contain outliers. Given that the right frontal opercular result is robust to outlier removal, and the remaining findings are not affected by outliers, there is no suggestion that any of the 36 participants should be removed from the analysis.

![Scatterplots showing correlations](image)

**Figure 6.11**: Correlations between focal GM volume and dominance-related slowing or untrustworthiness-related slowing

Scatterplots show individual dominance-related slowing (A) or untrustworthiness-related slowing (B-E) along the x-axis and grey matter volume (a.u., arbitrary units) at the coordinate of peak correlation along the y-axis. The five regions in which significant correlations are found are right frontal operculum (A), right pTP (B), mPFC (C), right fusiform gyrus (D), and left fusiform gyrus (E). The purpose for reviewing these plots is to illustrate the linearity of the relationships in each case. There is a statistical outlier in A (with dominance-related slowing ~ -300 ms); this is discussed further in Section 6.4.2.3. Correlations shown in plots B-E do not contain any statistical outliers.

### 6.4.2.4 Relationship between amygdala volume and behavioural measures

VBM analysis did not reveal any significant correlations between GM volume in right or left amygdala and either dominance-related slowing or untrustworthiness-related slowing (including after small volume correction; Section 6.4.2.2). Given this surprising null result, further separate analyses were performed using automated subcortical segmentation in...
FreeSurfer to re-assess for any correlation between either behavioural measure and amygdala volume (see Section 6.3.3). There was no correlation between amygdala volume and dominance-related slowing: for left amygdala \( r = 0.11, p = 0.5 \); for right amygdala \( r = -0.04, p = 0.8 \). There was also no correlation between amygdala volume and untrustworthiness-related slowing: for left amygdala \( r = 0.02, p = 0.9 \); for right amygdala \( r = -0.001, p = 0.995 \). Furthermore, I performed multiple regression analyses with dominance-related slowing or untrustworthiness-related slowing as dependent variables and age, gender, whole brain volume and either left or right amygdala volume as independent variables. This replicated the set of variables included in the VBM analysis. Again, amygdala volume did not correlate with either behavioural measure after inclusion of these other covariates. When predicting dominance-related slowing, neither left amygdala volume \( \text{Beta} 0.13, p = 0.55 \) nor right amygdala volume \( \text{Beta} -0.12, p = 0.64 \) were significant predictors; when predicting untrustworthiness-related slowing, again neither left amygdala volume \( \text{Beta} -0.06, p = 0.79 \) nor right amygdala volume \( \text{Beta} -0.16, p = 0.51 \) were significant predictors.

### 6.4.2.5 Correlation between behaviour and surface-based measures of cortical thickness, surface area and volume

As discussed in Section 6.3.3, GM volume, as calculated using VBM, is a mixed measure, which subsumes both cortical surface area and cortical thickness. Therefore, additional cortical surface-based analyses were performed to compute cortical thickness, cortical surface area and cortical volume, and explore whether focal variability in any of these measures was correlated with the behavioural indices of dominance-related slowing and untrustworthiness-related slowing (Section 6.3.3). These analyses produced several statistically uncorrected findings that suggested some correspondence with the reported VBM results (Sections 6.4.2.1 and 6.4.2.2). Firstly, there was a negative correlation between dominance-related slowing and cortical thickness in right frontal operculum \( (x = 44, y = 3, z = 24; P_{\text{uncorr}} = 0.005; \text{Figure 6.12A}) \). Secondly, there was a positive correlation between dominance-related slowing and cortical surface area in right frontal operculum \( (x = 48, y = 14, z = 23; P_{\text{uncorr}} = 0.0002; \text{Figure 6.12B}) \). Both of these results are located in the vicinity of the VBM result in right frontal operculum \( (x = 48, y = 2, z = 13) \), but only the result for cortical thickness has a correlation in the same direction (negative).
Thirdly, with regard to untrustworthiness-related slowing, there was a positive correlation with cortical thickness in right pTPJ \((x = 51, y = -58, z = 40; P_{uncorr} = 0.004; \text{Figure 6.13A})\); a negative correlation with cortical thickness in left fusiform gyrus \((x = -35, y = -47, z = -18; P_{uncorr} = 0.001; \text{Figure 6.13B})\); and a positive correlation with cortical thickness in left superior frontal cortex \((x = -14, y = 58, z = 23; P_{uncorr} = 0.007; \text{Figure 6.13C})\). These results are located nearby to the VBM results in right pTPJ \((x = 51, y = -57, z = 31)\), left fusiform gyrus \((x = -47, y = -45, z = -18)\), and mPFC \((x = -2, y = 54, z = 13)\), respectively, although only the left fusiform and superior frontal results have the same direction of correlation as the corresponding VBM results.
Finally, again with regard to untrustworthiness related slowing, there was a positive correlation with cortical volume in right pTPJ ($x = 50, y = -58, z = 38; P_{uncorr} = 0.003$; Figure 6.14A); and a negative correlation with cortical volume in right fusiform gyrus ($x = 34, y = -46, z = -19; P_{uncorr} = 0.001$; Figure 6.14B). These results are located near to the VBM results in right pTPJ ($x = 51, y = -57, z = 31$), and right fusiform gyrus ($x = 50, y = - 44, z = -20$), respectively. Only the result in right fusiform gyrus has a correlation with the same direction as the corresponding VBM result.

Figure 6.14: Regions where cortical volume was correlated with individual untrustworthiness-related slowing
Local correlations are circled at right pTPJ, A, and right fusiform gyrus, B. Positive correlations are shown in red, and negative correlations are shown in blue. Findings are plotted on inflated template brains.

These surface-based results thus suggest some possible correspondence between the VBM findings (Sections 6.4.2.1 and 6.4.2.2) and the findings for cortical thickness in particular. Differences in direction of some of the correlations between cortical thickness and behavioural measures versus the corresponding structure-behaviour correlations found using VBM are challenging to interpret in terms of which aspects of gray matter volume may be driving the observed relationships. The correspondence between the VBM findings and surface-based analysis results for cortical surface area and cortical volume are less clear.

It is important to note that the results described so far in this section were not corrected for multiple comparisons. Following cluster-wise correction, there were no longer any brain regions where cortical thickness, cortical surface area, or cortical volume was significantly correlated with dominance-related slowing or untrustworthiness-related slowing. Overall, therefore, while the surface-based analyses show some interesting trends in correlation between brain structure and behaviour, they do not demonstrate any statistically robust relationships.
6.4.2.6 Testing whether the brain structural correlates of dominance and untrustworthiness evaluation are dissociable

The results presented thus far already give considerable support for dissociation between dominance-related slowing and untrustworthiness-related slowing and their structural brain correlates. Firstly, I have shown that the two behavioural measures are not correlated (see Section 6.4.1.2). Secondly, when performing VBM analysis, both behavioural measures were placed in the same SPM design matrix. Therefore, when VBM analysis was performed to explore GM-volume correlates of variability in one of these behavioural measures, the influence of the other behavioural measure was effectively regressed out since the multiple regression analysis performed was based on partial correlation. This makes GM volume findings related to variability in one behavioural measure unlikely to be strongly influenced by variability in the other behavioural measure.

Nevertheless, I performed two additional analyses to test more thoroughly for dissociation between the structural correlates of dominance-related slowing and untrustworthiness-related slowing. Firstly, I used small-volume correction analyses (as per Section 6.3.2; in this instance setting a lenient significance threshold of \( p < 0.05 \), uncorrected) to explore whether there was any correlation between dominance-related slowing or untrustworthiness-related slowing and GM volume at loci where a correlation had been found with the opposite behavioural measure in the main results (Sections 6.4.2.1 and 6.4.2.2). For dominance-related slowing there was a weak correlation (\( P_{FWE-corr} = 0.31 \)) at \( x = 48, y = -60, z = 34 \), which is close to the TPJ locus reported at \( x = 51, y = -57, z = 31 \) (Section 6.4.2.1), as well as a weak correlation (\( P_{FWE-corr} = 0.29 \)) at \( x = 41, y = -48, z = -17 \), near to the right fusiform result at \( x = 44, y = -46, z = -22 \) (Section 6.4.2.2). There were no suprathreshold correlations with GM volume in mPFC or left fusiform gyrus even at the lenient significance threshold. For untrustworthiness-related slowing, there was a weak correlation with GM volume in right frontal operculum (\( P_{FWE-corr} = 0.30 \)) at \( x = 56, y = -9, z = 21 \), and an even weaker correlation (\( P_{FWE-corr} = 0.68 \)) at \( x = 42, y = 2, z = 3 \). Both of these loci are some distance from the frontal opercular locus reported at \( x = 48, y = 2, z = 13 \) (Section 6.4.2.1).

For the second analysis to test for dissociation between the structural brain correlates of dominance-related slowing and untrustworthiness-related slowing, I extracted fitted responses from peak coordinates for the findings reported in right frontal operculum, right pTPJ, mPFC, right fusiform gyrus, and left fusiform gyrus (Sections 6.4.2.1 and 6.4.2.2), and calculated correlation coefficients between these and both dominance-related slowing and untrustworthiness-related slowing. For each region, the difference between the correlation with dominance-related slowing and the correlation with untrustworthiness-
related slowing was calculated using Steiger’s Z-test (Steiger, 1980). For right frontal operculum, the correlation with dominance-related slowing was significantly larger than the correlation with untrustworthiness-related slowing ($r = -0.76$ and $r = 0.076$, $Z_{\text{bar}} = 4.27$, $p < 0.001$). For the remaining four regions, the correlation with untrustworthiness-related slowing was significantly larger than the correlation with dominance-related slowing (for pTPJ $r = -0.76$ and $r = -0.076$, $Z_{\text{bar}} = -3.59$, $p < 0.001$; for mPFC $r = 0.55$ and $r = -0.055$, $Z_{\text{bar}} = 2.64$, $p = 0.008$; for right fusiform $r = -0.54$ and $r = -0.054$, $Z_{\text{bar}} = -2.80$, $p = 0.009$; and for left fusiform $r = -0.56$ and $r = -0.056$, $Z_{\text{bar}} = -2.21$, $p = 0.027$).

These results rule out the possibility that GM volume in regions correlated with one of the behavioural measures was also significantly correlated with the other, and furthermore suggest that in all cases the correlation with one behavioural measure was significantly different to the correlation with the other. These additional analyses therefore support the claim of dissociability between the structural correlates of individual differences in dominance-related slowing and untrustworthiness-related slowing.

6.4.2.7 Brain structural correlates of individual differences in t2e and experimental task error

In addition to the VBM analysis to evaluate correlations between local GM volume and behavioural measures of dominance and trustworthiness evaluation, I also performed two further separate analyses to determine (1) whether GM volume in any focal brain region was correlated with individual differences in mean t2e (collapsed across all levels of dominance and trustworthiness), and (2) whether GM volume in any focal brain region was correlated with individual error rate (Section 6.4.1.3). As in the other VBM analyses (Sections 6.3.2, 6.4.2.1 and 6.4.2.2), potentially confounding factors of gender and age were regressed out, the global covariate was included in the general linear model, and non-stationary cluster-level correction was undertaken. For both of the present analyses, a single measure was placed as the only regressor of interest in the SPM design matrix, looking for correlations with local GM that were significant after correction for multiple comparisons across the whole brain volume, or after small-volume correction using the same methodology, regions of interest, and criteria for significance as described in Section 6.3.2.

The first of these analyses found no correlations between individual differences in mean t2e and local GM volume (both with whole-brain correction and with small-volume correction). The second analysis found no correlations between individual error rate in the behavioural experiment and local GM volume after correction for multiple comparisons.
across the whole brain volume. However, with small-volume correction at the predefined regions of interest, I found a significant correlation between error rate and GM volume at a right opercular locus \((x = 44, y = 2, z = 16, P_{\text{FWE-corr}} = 0.032)\) found within a 15mm-radius sphere centred at coordinates for right insula (reported by Chiao et al., 2008; \(x = 39, y = 9, z = 15\); and by Winston et al., 2002; \(x = 42, y = -4, z = 12\)). Error rate was not correlated with small-volume-corrected GM in TPJ, mPFC, fusiform gyrus, or amygdala.

The correlation between behavioural experimental error rate and local GM volume in right insula might lead one to speculate that the finding of a correlation between dominance-related slowing and GM in nearby right frontal operculum in fact reflects individual differences in error rate, or otherwise ability to appropriately perform the task irrespective of social dominance evaluation. I would argue that this is unlikely since error rate does not correlate significantly or meaningfully with the behavioural measures of dominance or trustworthiness evaluation (Section 6.4.1.3). It is therefore more likely that GM volume in two adjacent regions in right frontal operculum correlates with non-conscious social dominance evaluation and with error rate on a left/right task (the latter finding may relate to individual differences in response inhibition, which would fit with known functional roles of nearby IFG; e.g. Aron et al., 2004).

### 6.5 Discussion

In a first behavioural experiment I explored non-conscious evaluation of facial dominance and trustworthiness traits using time-to-emergence of face images from CFS as a dependent behavioural measure, and a computer-generated face model to vary social face traits. I predicted that dominant and untrustworthy faces would gain access to awareness more quickly than socially neutral faces. The behavioural results, which have been independently replicated in separate experimental samples (see Stewart et al., 2012), indicate that dominant and untrustworthy faces emerge from interocular suppression more slowly than neutral faces. In addition, there was substantial inter-individual variability in the strength of the effects of facial dominance and facial trustworthiness on t2e.

In a second imaging experiment I asked whether focal brain structure was correlated with the individual differences in non-conscious evaluation of facial dominance and trustworthiness. Based on previous experimental findings I predicted overlap in the structural correlates of non-conscious processing for these two socially relevant facial traits. Instead, I found that these neural correlates were dissociable: non-conscious slowing of dominance evaluation was negatively correlated with GM volume in right frontal
operculum, while non-conscious slowing of untrustworthiness evaluation was negatively correlated with GM volume in right pTPJ and bilateral fusiform gyrus, and positively correlated with GM volume in mPFC. This dissociation suggests that non-conscious evaluation of dominance and untrustworthiness is linked to at least partially separable neural substrates.

6.5.1 Dominant and untrustworthy faces emerge into awareness more slowly than neutral faces

Both dominant and untrustworthy faces were associated with slower t2e than neutral faces (replicated in other experiments; Stewart et al., 2012; see Section 9.3.1 for discussion). This is inconsistent with a vigilant response to non-conscious social threat, where speeded non-conscious processing is proposed to confer a survival benefit by allowing rapid detection of predators (Vuilleumier, 2005; Sander et al., 2003; Morris et al., 1999; Whalen, 1998; Sections 1.4.4 and 1.5.1). Such evolutionary vigilance accounts are supported by findings from CFS studies, which have shown speeded non-conscious processing for fearful faces (Yang et al., 2007) or faces with direct (as opposed to averted) eye gaze (Stein et al., 2011). The fact that the current findings are at odds with the prediction (based on evolutionary vigilance) could potentially be explained on the basis of the stimulus categories used (i.e. an argument that the dominant and untrustworthy faces used here are fundamentally different to fearful faces), and the consequent engagement of distinct sets of processing mechanisms (see Section 9.3.1). Indeed, I propose here that non-consciously perceived dominant and untrustworthy faces might engage a defence cascade response (e.g. Lang et al., 2000) and that the slower emergence into awareness for such faces could be a manifestation of defensive freezing (see also Stewart et al, 2012). Evolutionary vigilance and defence cascade accounts of threat-related processing are discussed in more detail in Sections 1.4.4 and 1.5.1, and the potential relevance to the current results is discussed extensively in Section 9.4.

There is some variation in statistical significance when comparing the present results to those from Experiments 1-2 in Stewart et al. (2012); in that study the main effect of trustworthiness on t2e reached statistical significance. Nevertheless, the direction of effect of trustworthiness on t2e shown here (Section 6.4.1.2) is the same as that reported by Stewart et al. (2012). Both sets of findings indicate that the effect of face trustworthiness on t2e is highly variable across individuals, and thus, the presence of a significant group-level effect would be expected to vary according to the experimental sample. Thus it is likely that the difference in statistical significance is due to differences in the participant samples.
across these experiments. The inter-individual variability in the effects of dominance-related slowing and untrustworthiness-related slowing is a key finding that I have explored in Section 6.4.1.2 and subsequently in the MRI analysis (Section 6.4.2).

6.5.2 Structural correlates of dominance-related slowing

Reduced gray matter volume in the right frontal operculum was correlated with increased slowing of non-conscious processing of dominant faces. Previous fMRI studies have reported activation in a nearby region in the middle portion of right insula when viewing dominant head postures (Chiao et al., 2008), or angry faces (Dannlowski et al., 2007). Taken together, these results suggest that right insula is involved in dominance evaluation both when it depends on relatively invariant face traits, and when it depends on more viewpoint-specific and dynamic cues such as head posture. The human insula is involved in neural processing of emotion (Phan et al., 2002). A number of fMRI studies have also demonstrated insula activation associated with risky decisions (e.g. Clark et al., 2008; Paulus et al., 2003). Paulus et al. (2003) focused on individual differences and reported that insula activation correlated positively with self-measures of harm avoidance and neuroticism. These studies point to a nuanced role for the insula in evaluating risks and possibly balancing approach versus avoidance of risky situations and/or conspecifics. My result is found in operculum adjacent to right insula, although there is evidence to suggest that these regions may have common or at least closely related functional roles, for example in representing taste (O'Doherty et al., 2001), facilitating empathy for others' emotions (Jabbi et al., 2007) and interpreting social intentions (Gobbini et al., 2007).

The neural mechanisms for evaluation of emotional facial expressions most likely overlap those for evaluation of social face traits (see Section 1.4). One proposed mechanism for emotion recognition is the engagement of mirror systems, which enable simulation of the observed emotion in the perceiver. Both quantitative lesion mapping (Adolphs et al., 2000), and suppression of activity using repetitive TMS (Pitcher et al., 2008; Pourtois et al., 2004) have shown that right somatosensory cortex is of causal importance for facial emotion recognition. The locus in frontal operculum is very close to secondary somatosensory area SII. Whether this close proximity plays a role in bringing social and emotional evaluation of faces together remains to be tested. Alternatively, the present result may relate more closely to the role of IFG in face perception, which also has links to the proposed emotional mirror systems (Jabbi et al., 2007; Shamay-Tsoory et al., 2009).
The negative correlation between GM volume in frontal operculum and dominance-related slowing may be interpreted in two possible ways. First, if dominance-related slowing reflects a freezing phenomenon (Section 6.5.1; see also Section 1.5), then one could speculate that during such freezing complex social evaluation in insula/frontal operculum is suppressed. The motor components of freezing have been shown to be variable across individuals (see Sections 1.5.2 and 8.1) and in an individual with an increased trait-like predisposition to freezing the frontal opercular social evaluation mechanisms may be relatively under-utilised, reflected in decreased GM volume. Alternatively, the result could be understood in terms of availability of processing resources for non-conscious information. The threat conveyed by dominant faces may result in engagement of limited resources in right frontal operculum, meaning that less processing capacity is available for social appraisal. Such limitation would theoretically be more severe for individuals who have relatively less GM in this opercular region, manifesting as increased slowing of non-conscious dominance evaluation. Both possibilities are discussed further in Section 9.4.

### 6.5.3 Structural correlates of untrustworthiness-related slowing

Individual differences in slowing of non-conscious processing for untrustworthy faces were correlated with local GM volume in a distributed group of brain regions. Reduced GM volume in right pTPJ and bilateral fusiform gyrus, and increased GM volume in mPFC all predicted increased untrustworthiness-related slowing. These findings only partially confirmed the a priori predictions that untrustworthiness-related slowing would co-vary with gray matter volume in the amygdala, insula, fusiform gyrus and mPFC (Todorov et al., 2008; Winston et al., 2002).

Besides relating to previous findings of increased BOLD activation in the fusiform gyrus in the context of untrustworthy faces (Winston et al., 2002), my findings in bilateral fusiform gyrus are also consistent with studies showing that this region responds generally to facial social cues (e.g. Fox et al., 2009). However, further investigation will be required to understand whether the functional role of fusiform gyrus in trustworthiness evaluation relates to computation of differences in physical appearance and configuration of features or to evaluation of more directly socially relevant information, such as emotional state or possible goals and intentions.

GM volume in mPFC was positively correlated with non-conscious untrustworthiness-related slowing. Substantial converging evidence implicates mPFC in tasks that depend on
mentalising or “theory of mind” (the making of sophisticated inferences about the goals and intentions of others; Amodio and Frith, 2006; Gallagher and Frith, 2003; Saxe and Kanwisher, 2003; Van Overwalle, 2009). Using proposed functional subdivisions of this region as a framework (Amodio and Frith, 2006; Section 1.4.4), the structural locus in mPFC identified in the present results (Table 6.3) appears to sit within the anterior rostral subregion, which is involved in a wide variety of social cognitive tasks (Amodio and Frith, 2006; Mitchell et al., 2005).

An unexpected but statistically robust finding was the negative correlation between GM volume in pTPJ and untrustworthiness-related slowing. TPJ is involved both in theory of mind and reorienting of attention (Decety and Lamm, 2007). The present result is in the posterior portion of TPJ, which is widely implicated in processes of social cognition (Saxe and Kanwisher, 2003; Van Overwalle, 2009; Section 1.4.4) and furthermore has strong resting-state functional connectivity with regions involved in social cognitive function, including mPFC, posterior cingulate and precuneus (Mars et al., 2011). The involvement of this pTPJ subregion thus could dovetail with the separate result in mPFC. The present findings therefore extend the roles of right pTPJ and mPFC to include individual differences in non-conscious social evaluation of faces based on untrustworthiness.

The negative correlations between GM volume in fusiform gyrus and pTPJ, and untrustworthiness-related slowing may lend themselves to either of the interpretations offered in the earlier discussion of the negative correlation between dominance-related slowing and frontal opercular GM volume (Section 6.5.2; namely that they relate either to individual differences in freezing behaviour predisposition or to individual differences in availability of limited processing resources). The theoretical models that could underpin these explanations are discussed further in Section 9.4. While regions of parietal cortex are involved in deployment of attention in relation to threatening stimuli (Pourtois and Vuilleumier, 2006), and TPJ also shows increased activation to threatening faces (Kret et al., 2011), I am not aware of published evidence that altered activation in TPJ can be seen in association with slowing of behavioural performance in the context of social threat, and this would be an intriguing possibility to explore in the future.

The positive correlation between untrustworthiness-related slowing and GM volume in mPFC (as contrasted with the several negative correlations already discussed) is in keeping with the often-described inverse relationships between activity in mPFC and other brain regions involved in social cognition including amygdala (Etkin et al., 2011), TPJ (FeldmanHall et al., 2013) and insula (Thom et al., 2012). Whether this reflects differential influences on these regions in the context of freezing or inhibitory influences of mPFC on earlier levels of hierarchical processing is an important question for future investigation.
6.5.4 Findings from surface-based structural brain measures

As discussed elsewhere (Sections 2.2.6, 2.2.7, 6.3.3), GM volume, as calculated using VBM, is only one of a number of possible measures of cortical brain structure, and it incorporates both features of cortical surface folding and cortical thickness. While variation in cortical surface area may imply differences in availability of processing power, variation in cortical thickness may imply differing laminar microstructure and connectivity. Determining which of these measures contributes more to the present findings could have important implications for interpretation of the results. Additional cortical surface-based analyses, using FreeSurfer (Section 6.4.2.5) were thus performed, to explore whether individual differences in the behavioural measures were correlated with cortical thickness or cortical surface area in any focal brain regions.

These surface-based analyses provided rather mixed results, which did not survive statistical correction for multiple comparisons, and do not present a clear picture as to whether the VBM results were driven by cortical thickness or surface area variability. The correlations found between VBM-derived gray matter volume measures and the behavioural measures of dominance-related slowing and untrustworthiness-related slowing reach corrected statistical significance as well as being robust to outlier removal. The failure to extend the VBM findings by demonstrating similarly robust relationships between volume (or indeed thickness or surface area) in the cortex measured using FreeSurfer, and the behavioural measures, may be due to the difference between the SPM-based and FreeSurfer-based analysis methods. For example, with SPM inter-subject coregistration was performed by the DARTEL procedure volumetrically in voxel space, whereas FreeSurfer's coregistration is based on alignment of gyri and sulci after reconstructing cortical surface models. Volumetric measures of the same brain images made in FreeSurfer and in SPM can differ by as much as 20% (Klauschen et al., 2009). In addition, surface-based measures of cortical GM volume performed in FreeSurfer (as used here) can differ substantially from volume-based measures performed in the same package, with the latter potentially being more accurate (Klauschen et al. 2009).

Thus, it has not been possible to convincingly demonstrate any clarifying relationship between cortical thickness or cortical surface area and our behavioural measures. While this does potentially limit the depth and specificity of mechanistic explanation that the current findings can offer, it does not affect their main thrust, which is to show that local structural measures predict individual differences in specific aspects of non-conscious social perception. Future refinements in measurement of cortical structural indices will hopefully enable a clearer answer regarding which elements of brain structure relate to the
demonstrated individual differences in non-conscious evaluation.

### 6.5.5 Other important caveats

Previous relevant findings, on which both the predictions and interpretation of the present results are based, relate to *functional*, rather than structural, neuronal correlates of social perception. It is important to bear in mind that relationships between brain structure and function are not yet fully established. Thus, any direct comparisons between structural brain correlates of behaviour and functional brain correlates of similar behaviours rest on a number of assumptions that will need to be directly evaluated in the future.

The failure to find a correlation between dominance or untrustworthiness-related slowing and amygdala structure (either GM volume or volume as derived from subcortical segmentation) is surprising given the wealth of evidence linking the amygdala to processing of trustworthiness and anger, both from fMRI (Dannlowski et al., 2007; Todorov et al., 2008; Whalen et al., 2001; Winston et al., 2002), and from lesion studies (Adolphs et al., 1998; Calder, 1996). In addition, there is robust activation in amygdala for non-conscious emotionally relevant stimuli (Whalen, 1998; Williams et al., 2004). It is difficult to offer an explanation of this null finding, other than to emphasise that the lack of correlation between individual differences in my behavioural measures and variation in structural measures in amygdala in no way excludes an important role for the amygdala in non-conscious evaluation of face traits. This region may have a central role in such processes without having a significant relationship (at least as far as its GM volume is concerned) with individual differences in the associated behavioural phenomena. The use of functional imaging modalities will be an important future step in evaluating more fully the proposed role of the amygdala in non-conscious social face evaluation.

### 6.5.6 Summary

Two experiments were performed to explore how social face traits (in particular dominance and trustworthiness) influence measures of non-conscious face perception, and to assess whether individual differences in such non-conscious evaluation of social traits were correlated with measures of focal brain structure. The results demonstrate that both dominant and untrustworthy faces emerge more slowly from interocular suppression than socially neutral faces. This finding was contrary to the pre-experimental predictions. Moreover, I found that there were substantial inter-individual differences in the non-conscious evaluation of dominance and untrustworthiness, and these were associated with
individual variability in local brain structure. The behavioural results may indicate that both dominant and untrustworthy faces are treated as threatening stimuli and perhaps activate subcortical emotional and threat-response mechanisms that form part of defence cascade models and are linked to freezing. In addition, however, the findings from the MRI experiment, that GM volume in distinct cortical regions was correlated with individual effects of non-conscious dominance-related slowing (frontal operculum) and untrustworthiness-related slowing (pTPJ, mPFC and fusiform gyrus), support the notion that evaluation of these traits depends on at least partially separable neural substrates. Furthermore, the MRI results show that even when performed outside of awareness, social evaluation relates to GM volume in regions subserving high-level processes of social cognition. Further empirical work will be required to determine whether involvement of these cortical regions relates directly to mechanisms and influence of purported freezing or is a separate aspect of non-conscious social processing.
Chapter 7

Modulation of conscious perception in binocular rivalry by social face traits

“There are things known and there are things unknown, and in between are the doors of perception.”

Aldous Huxley (1894 – 1963)

7.1 Chapter introduction

Faces, and the information they carry about the owner’s character and intentions, are central to human social interaction. Facial traits comprise one component of this socially relevant information and their role is discussed in detail in Section 1.4. In Chapter 6, I focused on non-conscious processing of the social face traits dominance and trustworthiness as well as on structural brain correlates of individual differences in such processing. A key finding was that speed of access to visual awareness was affected by non-consciously perceived social face traits. The present chapter aims to deepen the investigation into the influence of social cues on visual processing and awareness by exploring in more detail how such cues influence the balance and transition between non-conscious and conscious visual perception.

In binocular rivalry (Section 1.3.3), a different image is presented to each eye, and because the strength of the two stimuli is fairly balanced (unlike in CFS), the observer’s perceptual experience continually alternates between two possibilities. As reviewed in section 1.3.5, focusing on time points when there is a switch between alternative percepts in binocular
Social dominance and binocular rivalry enables interrogation of neural processes that govern access of information to visual awareness (e.g. Sterzer et al., 2009). Following on from several fMRI studies exploring these mechanisms (Section 1.3.5), the first three experimental chapters in this thesis (Chapters 3-5) focused on outstanding questions regarding the neural mechanisms behind such perceptual transitions using some novel investigative approaches.

There is evidence to suggest that socially relevant cues, for example emotional facial expressions, affect the dynamics of binocular rivalry (see Section 1.4.5). However, some key questions remain unexplored. Firstly, previous studies utilising binocular rivalry have focused on the extended \textit{periods} of perception and asked how these are modulated by social cues. Here I instead aimed to explore how \textit{access} to awareness (i.e. the moment when the contents of awareness changes) is modulated according to social relevance of visual information. Secondly, while the influence of social cues such as emotional facial expression and eye gaze direction on visual awareness has been explored to some extent, it is not clear what effects social face traits such as dominance and trustworthiness may have in this context. Finally, little is known about how social relevance affects the dynamics of conscious access at a neural level.

In Chapter 6 (Sections 6.5.3 and 6.5.4) I proposed that the slower non-conscious processing \textit{(prolonged te)} for dominant and untrustworthy faces may reflect the perceptual components of defensive freezing. Dominant and untrustworthy faces also signal threat (Oosterhof & Todorov, 2008), and could therefore theoretically elicit such defensive responses. I therefore now hypothesised that socially threatening faces would affect perceptual transition during binocular rivalry in a manner consistent with freezing. Freezing behaviours have been extensively studied in rodents (Lang and Davis, 2006) but are common to all mammals and have been demonstrated in humans (e.g. Roelofs et al., 2010). According to one theory, these behaviours are part of a complex cascade of defence responses that include both behavioural changes (reduced motor output) and autonomic changes (e.g. bradycardia; Lang et al., 2000; see Sections 1.5.1 and 1.5.2 for further details).

It was logical to make use of the social face model of Oosterhof and Todorov (2008) once again for the present set of experiments. Besides capitalising on the already-discussed advantages of the model (see Section 1.4.2), this would enable me to make more direct comparisons with the findings presented in Chapter 6. On this occasion I chose to specifically vary facial dominance while keeping trustworthiness neutral. Despite its evolutionary importance (Adams et al., 2011), the behavioural and neural mechanisms of social dominance evaluation in humans have received little attention in comparison to those for other social traits such as trustworthiness and attractiveness (see Section 1.4.1). The impact of social dominance on bistable perception, on a behavioural and neural level, has
not previously been explored. In addition, as already discussed, facial dominance is strongly correlated with threat (and this correlation is known to also apply specifically to the model of Oosterhof and Todorov, 2008). Face dominance effects are thus a surrogate for threat-related effects, and provide a way to probe freezing mechanisms. Instead of pairing socially relevant and socially neutral faces together in binocular rivalry (as in previous studies; e.g. Alpers and Gerdes, 2007; Amting et al., 2010), I planned to pair each face type with a low-level visual stimulus (a slanted grating), in separate trials. This configuration would allow me to explore perceptual transitions in a dominant-face or neutral-face context.

I sought to empirically test three key predictions that arise from a freezing hypothesis. First, I hypothesised that social face dominance would influence the stability of conscious visual perception during binocular rivalry. Specifically, I predicted that behaviourally measured binocular rivalry switch rate would be slower in a socially dominant face context, in keeping with the findings of delayed t2e for dominant faces (see Sections 6.4.1.1 and 6.5.1; also Stewart et al., 2012). Second, I hypothesised that freezing during perceptual transitions in a dominant-face context would have specific neural correlates. Using fMRI, and focusing analysis on instances of perceptual transition, I predicted that BOLD signal in right IFG (involved both in perceptual transition, Lumer et al., 1998; and in mediating the impact of socially relevant information on behaviour, Jabbı and Keysers, 2008), would be decreased at the time of perceptual transitions, consistent with a role for this region in mediating the binocular rivalry switch rate slowing associated with dominance. Moreover, I predicted decreased activation in bilateral amygdala and right STS (both regions are involved in social evaluation of faces; e.g. Adolphs, 2010; Allison et al., 2000), in association with perceptual transitions in a dominant-face context, reflecting perceptual correlates of freezing. My third main hypothesis was that social modulation of visual awareness would be accompanied by psychophysiological changes that reflect the autonomic correlates consistently associated with freezing responses (Lang et al., 2000). Specifically, I sought to explore whether perceptual-transition-evoked HR was modulated by facial dominance. Based on HR changes reported when viewing threatening face images (Bradley et al., 2005; see also Section 8.1), I predicted that there would be relative peritransition slowing of HR when transitions occur in the context of dominant faces.
7.2 Behavioural experiment

7.2.1 Introduction

The primary aim of the behavioural experiment was to test the prediction (based on a freezing hypothesis; Section 7.1) that there would be slowing of binocular rivalry switch rate in the context of dominant faces. There were two additional objectives. Following on from the individual differences approach employed in Chapter 6 (see Section 6.1), I wanted to assess whether individual variability in effects of social dominance on binocular rivalry dynamics was correlated with self-reported measures on a set of questionnaires especially selected to assess relevant personality and mood traits. Specifically, I predicted correlations between behaviour and scores on both the SBS and BAS scales (see Sections 1.6 and 6.5.1; see also Stewart et al., 2012). In addition, given the known influence of social threat conveyed by faces on maintenance of eye contact (e.g. Terburg et al., 2011), I sought to determine whether visual fixation differed between binocular rivalry trials involving a neutral face and those involving a dominant face (predicting increased tendency to break fixation in a dominant face context), and thus to be able to either confirm or exclude a contribution of eye movement differences to any observed behavioural (or subsequently neural signal) effects.

To address the above objectives, either socially neutral or socially dominant faces (as generated by the model of Oosterhof and Todorov, 2008), were paired with visual gratings in binocular rivalry.

7.2.2 Materials and methods

7.2.2.1 Participants

All participants for both this experiment and the fMRI experiment (Section 7.3) were recruited according to the inclusion criteria and procedures detailed in Section 2.4.1. Initially I recruited 20 participants and subsequently offered to enroll eight of them into the fMRI experiment (chosen according to a specific pre-determined set of enrolment criteria; see Section 7.3.2.1). Two individuals agreed to take part in the fMRI experiment (this low rate may relate to the 5-6 month gap between the two studies). To achieve a sample size of 20 for the fMRI experiment, I then went on to behaviourally test 54 further individuals (until a total of 20 had met the fMRI enrolment criteria), leading to a total sample of seventy-four in the behavioural experiment (51 female; mean ± SD age 23.1 ± 4.4 years; range 19-38 years).
7.2.2.2 Stimuli and display apparatus

Static face and grating stimuli were presented simultaneously side-by-side in a configuration that induced binocular rivalry, as described in Section 2.4.4. The face stimuli were taken from the model described in Section 2.4.3, selecting one face that was neutral in dominance, and another, of the same identity, that was moderately dominant (3 SD from the neutral in the positive direction). Both face stimuli were neutral for trustworthiness (Figure 7.1A).

The experimental apparatus were set up as described in section 2.4.2. Stimuli were viewed on a Sony Trinitron GDM-F520 monitor (1600 x 1200 at 85 Hz) at a distance of 65.5 cm through prism glasses. Each stimulus subtended 4.0° visual angle and the fixation crosses subtended 0.6° visual angle.

Figure 7.1: Stimuli and paradigm for behavioural experiment

See Section 7.2.2.2. (A) Two face stimuli (neutral and three SD positive on dominance dimension; both neutral on trustworthiness dimension; marked with a red border) were selected from a social face trait model (Oosterhof and Todorov, 2008; Section 1.4.2). (B) Schematic of binocular rivalry trials employed in both the behavioural and fMRI experiments (Sections 7.2.2.3 and 7.3.2.2). A face image (neutral or dominant; the example in the image is a neutral face) was shown to the right eye and a visual grating was shown to the left eye. Both remained on the screen for 40 seconds. The observer’s perceptual experience alternated between perception of the face and grating (blue-shaded area). The third possible percept type (mixed) is not shown for simplicity.

7.2.2.3 Binocular rivalry experimental procedures

The experimental paradigm was as described in Section 2.4.4.2; a schematic version is shown in Figure 7.1B. Trials lasted 40 seconds and participants made ongoing report of their perceptual experience during this time. Trials in which perceptual fusion broke were
discarded, and if this occurred more than three times during any experimental session, that participant was excluded from further analysis (n = 4).

The first 20 participants each completed a total of 30 trials and the remaining 54 a total of 20 trials (in both cases split equally between dominant and neutral-face conditions in a pseudo-random order). Prior to commencing each experiment, four practice trials (split evenly between dominant-face and neutral-face trials) allowed familiarisation with the paradigm.

From the recorded behavioral responses I obtained 1) perceptual switch rate (mean frequency of switches from one percept to another); 2) mean duration of percept visibility for each of the three percept types; 3) cumulative duration of percept visibility for each of the three percept types. These measures were calculated across the whole experiment (rather than within individual blocks or trials) separately for all neutral-face trials and all dominant-face trials, enabling comparisons between these two conditions. Unlike cumulative durations, mean percept durations for the three percept types are independent of each other, enabling me to analyse these data using parametric statistics.

7.2.2.4 Eye tracking experimental procedures

For 14 of the participants, eye tracking was performed according to the procedures described in Section 2.4.9.

7.2.2.5 Self-rated personality questionnaires

At the end of the behavioural sessions, while still in the experimental room, each participant completed a set of self-report questionnaires including SBS (Allan and Gilbert, 1997), PTS (Evans and Revelle, 2008), STAI-S and STAI-T (Spielberger and Vagg, 1984), and BIS/BAS (Carver and White, 1994). These scales are discussed in more detail in Section 1.6.

7.2.3 Results

7.2.3.1 Binocular rivalry results

Five participants were excluded from analysis (four due to repeated disruption of perceptual fusion and one due to prolonged periods with no behavioural response), leaving a sample of 69 participants. Mean ± SEM binocular rivalry switch rate (across both
dominant and neutral conditions) was $0.32 \pm 0.02$ switches/second. There was a significant difference in switch rate between the dominant-face and the neutral-face conditions; a dominant-face context was associated with less frequent switches $[t_{(68)} = 2.30, \ p = 0.03; \ \text{Cohen’s } d = 0.28]; \ \text{Figure 7.2A}$. The initial sample of 20 individuals (17 after exclusion of participants; see above) showed the same effect of face dominance on switch rate $[t_{(16)} = 2.40, \ p = 0.03, \ \text{Cohen’s } d = 0.58]$.

Mean ± SEM durations of percept visibility during binocular rivalry (across both dominant and neutral conditions) were: faces = $5.45 \pm 0.28$ seconds; gratings = $1.71 \pm 0.14$ seconds; mixed = $2.63 \pm 0.28$ seconds. In terms of cumulative duration of visibility across the whole experiment, faces were visible for 63% of the time, gratings for 9%, and mixed percepts for 28% (Figure 7.2B).

**Figure 7.2:** Behavioural experiment results
See Section 7.2.3.1. (A) Significant slowing effect of face dominance on binocular rivalry switch rate. (B) Proportion of time over the whole experiment that each percept type was visible for. (C-E) Mean durations of visibility (in seconds) for each of the three possible percept types (faces, C; gratings, D; mixed, E) in neutral-face and dominant-face trials. There is a significant increase in grating duration in dominant face trials (D). Error bars represent one SEM throughout.

*Indicates $p < 0.05$ using paired-sample $t$-test.
A two-way repeated measures ANOVA with mean percept duration as the dependent variable, and factors of dominance (two levels: neutral/dominant) and percept (three levels: face/grating/mixed) revealed a significant main effect of dominance \( F_{(1,68)} = 5.22, p = 0.03 \), and a strongly significant main effect of percept \( F_{(2,136)} = 97.01, p < 0.001 \). There was no significant interaction between the main effects \( F_{(2,136)} = 1.17, p = 0.31 \). Pairwise comparisons showed no significant difference in mean face duration \( t_{(68)} = -1.52, p = 0.13 \); Figure 7.2C or mean mixed duration \( t_{(68)} = -0.07, p = 0.95 \); Figure 7.2E between the dominant-face condition and neutral-face condition, while for mean grating duration the difference reached significance \( t_{(68)} = -2.58, p = 0.01 \); Figure 7.2D. To check whether evaluation of social dominance depends upon participant gender, I added gender as a between-subjects factor in the ANOVA. This did not significantly alter the results already described; moreover, there were no significant dominance*gender or percept*gender interactions. Adding gender as a between-subjects factor in the switch rate analysis also did not alter the reported significant effect of dominance on switch rate and again there were no significant dominance*gender interactions.

Next I explored the distributions of percept visibility durations for all three percept types (face, grating and mixed) separately for neutral-face and dominant-face conditions. It is well described that percept visibility durations should be distributed according to a gamma function (e.g. Fox and Herrmann, 1967; see also Section 4.3.1.3) and I wanted to ask whether the characteristics of these distributions were different for the dominant-face and neutral-face conditions. Using an inbuilt MATLAB function I was able to calculate shape and scale parameters, which describe kurtosis and dispersion, respectively, for 62 (of the total sample of 69) individuals. There were no significant (dominant versus neutral) differences in either shape or scale parameters for any of the three percept types: shape parameter differences for faces \( t_{(61)} = -0.95, p = 0.35 \); scale parameter differences for faces \( t_{(61)} = -0.86, p = 0.40 \); shape parameter differences for gratings \( t_{(61)} = -0.64, p = 0.52 \); scale parameter differences for gratings \( t_{(61)} = -1.07, p = 0.29 \); shape parameter differences for mixed \( t_{(61)} = -0.09, p = 0.93 \); scale parameter differences for mixed \( t_{(61)} = -0.67, p = 0.51 \). Percept duration histograms for two example participants are shown in Figure 7.3.
7.2.3.2 Eye tracking results

Two of the 14 participants who underwent eye tracking had substantial sections of missing eye data and were therefore excluded from this analysis. For the remaining 12 participants there were no significant differences between the two face dominance conditions for mean horizontal eye position $[t_{111} = 0.04, p = 0.97]$; for SD from mean horizontal eye position $[t_{111} = 1.20, p = 0.26]$; for mean vertical eye position $[t_{111} = 1.02, p = 0.33]$; for SD from mean vertical eye position $[t_{111} = 1.77, p = 0.10]$; for mean pupillary surface area $[t_{111} = -0.17, p = 0.87]$; or for SD from mean pupillary surface area $[t_{111} = 0.43, p = 0.68]$. The SD results are displayed in Figure 7.4, both for individual participants (Figure 7.4 A, C and E) and at the group level (Figure 7.4 B, D and F). Individuals for whom a neutral-face and dominant-face measure was significantly different are marked with an asterisk in Figure 7.4. It is apparent that these individuals represent a minority and even in these cases there are no systematic differences across all measures. Heat maps displaying eye position data for three example participants (Figure 7.4 panels G-I) show some variability in visual fixation between individuals, but overall there is a very similar pattern of fixation within
individuals for neutral-face trials and dominant-face trials. This is the case even for the individuals where numerical differences between conditions are demonstrated. For example, one participant showed a significant difference in SD from mean vertical eye position between neutral-face and dominant face conditions (panel G; see participant 6 in panel C). A second participant showed a significant difference in SD from mean horizontal eye position between neutral-face and dominant-face conditions (panel I; see participant 12 in panel A). The third example participant showed no difference between dominant-face and neutral-face conditions on any measure (panel H).
Figure 7.4: Eye position and pupillary area data from eye tracking measurements

Panels on the top left of the figure show SD from mean horizontal eye position (A), SD from mean vertical eye position (C) and SD from mean pupillary area (E) for the 12 individual participants. Asterisks indicate where there is a significant difference between the two trial types (neutral in blue and dominant in red). Panels on the top right of the figure (B, D and F) show group means for each corresponding individual plot on the top left of the figure. Errorbars correspond to one SEM. No significant group-level difference between the neutral and dominant conditions is observed for any of the three measures. Panels at the bottom of the figure (G-I) show heat maps representing eye position (sampled throughout experimental trials at 1000 Hz; neutral-face trials in upper panels and dominant-face trials in lower panels) for three example individuals. For ease of visualisation and comparison between conditions, the plots in G-I have x and y-axes limited to 200-pixel ranges and therefore only show a small part of the 1600x1200 screen area viewed by participants. Therefore, fixation was more precise (confined to a smaller portion of the screen) than these plots would suggest. In panels G-I the colour bars represent number of eye position measurements taken at each screen location (low numbers in blue and high numbers in brown).
7.2.3.3 Self-rated personality questionnaire results

Self-rated personality questionnaire scores were collected for all 69 participants included in the binocular rivalry analysis (Section 7.2.3.1). A ‘social dominance effect’ was calculated for each participant, by subtracting the mean binocular rivalry switch rate during dominant-face trials from that during neutral-face trials. There was substantial individual variability both in binocular rivalry switch rate (Figure 7.5A), and in the social dominance effect (Figure 7.5B). However, neither of these measures was significantly correlated with any of the self-report questionnaire scores. For individual switch rate, correlation statistics were: STAI-S, $r = -0.08, p = 0.51$; STAI-T, $r = -0.04, p = 0.76$; SBS, $r = -0.07, p = 0.56$; PTS, $r = 0.03, p = 0.82$; BIS, $r = -0.13, p = 0.29$; BASD, $r = -0.06, p = 0.62$; BASR, $r = -0.08, p = 0.50$; BASF, $r = 0.16, p = 0.18$.

For social dominance effect, correlation statistics were: STAI-S, $r = 0.05, p = 0.72$; STAI-T, $r = 0.08, p = 0.54$; SBS, $r = -0.02, p = 0.90$; PTS, $r = -0.06, p = 0.61$; BIS, $r = 0.02, p = 0.89$; BASD, $r = 0.11, p = 0.38$; BASR, $r = 0.01, p = 0.96$; BASF, $r = -0.03, p = 0.83$.

I additionally checked whether the eye tracking measures described in Section 7.2.3.2 were correlated with any of the self-rated questionnaire scores. Previous work has shown that dominant traits (as indexed by BAS scale scores) predict decreased tendency to avert gaze from masked angry faces (Terburg et al., 2011), as well as that anxious traits (as indexed by STAI scale scores) predict increased tendency to avert gaze from angry faces (Rohner, 2002). However, there were no significant correlations between any of the BAS or STAI subscales (or any of the other scales measured) and any of the eye tracking measures (for this analysis difference in eye tracking indices was calculated by taking measures in the dominant-face condition as a proportion of measures in the neutral-face condition). For difference in SD from mean horizontal eye position, correlation statistics were: STAI-S, $r = 0.03, p = 0.94$; STAI-T, $r = 0.02, p = 0.95$; SBS, $r = -0.16, p = 0.63$; PTS, $r = 0.34, p = 0.28$; BIS, $r = 0.25, p = 0.44$; BASD, $r = 0.29, p = 0.36$; BASR, $r = 0.06, p = 0.85$; BASF, $r = 0.57, p = 0.06$ (not close to significance after Bonferroni correction). For difference in SD from mean vertical eye position, correlation statistics were: STAI-S, $r = 0.05, p = 0.87$; STAI-T, $r = -0.52, p = 0.08$; SBS, $r = -0.42, p = 0.17$; PTS, $r = 0.51, p = 0.09$; BIS, $r = 0.61, p = 0.03$ (n.s. after Bonferroni correction); BASD, $r = -0.14, p = 0.68$; BASR, $r = 0.10, p = 0.77$; BASF, $r = 0.15, p = 0.64$. For difference in SD from mean pupillary area, correlation statistics were: STAI-S, $r = -0.27, p = 0.40$; STAI-T, $r = 0.29, p = 0.35$; SBS, $r = 0.30, p = 0.34$; PTS, $r = -0.36, p = 0.25$; BIS, $r = 0.42, p = 0.18$; BASD, $r = 0.18, p = 0.59$; BASR, $r = 0.33, p = 0.29$; BASF, $r = -0.23, p = 0.48$.

In summary, these analyses did not reveal any statistically significant correlations between dominance effect on binocular rivalry or visual fixation/pupil size measures and self-rated
personality scores. It is worth noting, in relation to the correlations with eye measures, that the sample size of 12 individuals is rather small for this sort of analysis, carrying a risk that it is underpowered.

![Image](image.png)

**Figure 7.5:** Individual variability in behavioural performance for binocular rivalry
(A) Histogram showing substantial individual variability in binocular rivalry switch rate (switches/sec). (B) Histogram showing individual variability in the effect of face dominance on binocular rivalry switch rate (switches/sec; measured as switch rate in neutral-face trials subtracted from switch rate in dominant-face trials).

### 7.2.4 Discussion

I hypothesised that social face dominance traits (which also index threat) would be associated with a slower rate of perceptual transitions in binocular rivalry. This prediction was based on previous results reported in this thesis (Section 6.4.1.1) as well as on results reported elsewhere (Stewart et al., 2012), and the proposal that delayed emergence of dominant faces from interocular suppression is underpinned by a freezing response. I paired faces (dominant or neutral) with visual gratings in binocular rivalry, enabling me to study rivalry transitions in a dominant-face versus neutral-face context. Consistent with my prediction (and the freezing hypothesis), I found that binocular rivalry switch rate was slower in the context of socially dominant faces. Across all three reported percept types (faces, gratings and mixed) there was a main effect of social dominance, with dominant faces resulting in longer mean stimulus visibility durations. However, when examining the effect of face dominance for individual percept types *post hoc*, the effect only reached statistical significance for grating visibility durations. Mean grating visibility durations were significantly longer in dominant-face trials. In other words, faces were absent from awareness (and gratings visible) for longer in the dominant-face condition. This would be
broadly in keeping with the behavioural results presented in Section 6.4.1.1, where I showed delayed access to awareness for dominant faces, as measured by t2e from CFS. However, some caution is necessary when comparing results from different forms of bistable perception since they may well depend on distinct underlying neural mechanisms (e.g. Fogelson et al., 2014).

7.2.4.1 Visibility durations for individual percepts

The finding of longer grating visibility durations in dominant-face trials could be considered within the framework of Levelt’s second law (Levelt, 1968; see also Section 1.3.3), which states that when one monocular stimulus is reduced in strength, this leads to increased predominance of the other monocular stimulus (leaving durations of the altered stimulus unchanged). Such phenomena have also been modelled at the level of neuronal mechanisms (Wilson, 2007). One could accordingly speculate that (as a perceptual consequence of freezing) neural representations associated with dominant faces in binocular rivalry are weaker than those associated with neutral faces, resulting in longer grating durations in dominant-face trials (but being associated with no change in face durations). There is evidence for increased BOLD signal in perceptual regions in association with threatening visual cues, for example angry-expression faces (Blair et al., 1999) or faces with dominant postures (Chiao et al., 2008). I am not aware of existing evidence for reduced BOLD signal in perceptual areas in association with facial threat. However, it is worth noting that standard univariate voxel-wise analyses of fMRI-BOLD signal can fail to map distributed or subtle effects on neural activation (Davis et al., 2014). For example, studies using MVPA have shown that different emotional cues and states can be decoded from activation patterns within individual brain regions that are not distinguishable with more standard univariate analyses (Said et al., 2010; Shibata et al., 2016).

The lack of a significant difference between face visibility durations in the dominant-face and neutral-face conditions appears inconsistent with reports of increased predominance of emotionally expressive faces in binocular rivalry (e.g. Alpers and Gerdes, 2007; Amting et al., 2010). Clearly, this discrepancy might be explained by the present use of faces varying in social dominance (which are emotionally neutral), rather than stimuli varying in emotional expression. It is important also to note another key difference between the above-mentioned studies and the present experiment. Rather than pairing socially relevant and socially neutral faces together in binocular rivalry, I paired each of the two face types with a visual grating in separate trials. The focus of this study was thus on comparing binocular rivalry dynamics in a dominant-face versus neutral face context, rather than on whether one type of face predominates over the other, as examined previously (Alpers and Gerdes,
Chapter 7. Social dominance and binocular rivalry

2007; Amting et al., 2010). Given the close relationship between facial dominance and threat both generally and more specifically for the face model used (Oosterhof & Todorov, 2008), I would go further to suggest that dominant-face trials created a socially threatening context, as compared to a non-threatening (or socially neutral) context in neutral-face trials.

7.2.4.2 Visual fixation and pupil effects of face dominance

The eye tracking analysis was important for excluding any significant contribution that eye movements could have made to the reported findings. There are existing reports of altered tendency to disengage gaze from angry faces, which can depend on anxiety or aggressive behavioural traits (Rohner, 2002; Terburg et al., 2011). However, disengagement of gaze from the faces presented in this experiment was not required. The present results confirmed that participants in this experiment were able to maintain fixation equally well and there were no eye gaze differences when comparing the two face dominance conditions.

7.2.4.3 Lack of correlation between individual binocular rivalry dynamics or eye gaze fixation and personality traits

No correlations between individual variability in switch rate, or the effect of face dominance on switch rate, and a number of personality and anxiety-related self-rated questionnaires were found. This contrasts with previous findings that individual submissiveness and propensity-to-trust scores predict individual non-conscious face dominance evaluation (Stewart et al., 2012). A relevant consideration is that the study of Stewart et al. (2012) made use of CFS, while the present experiment employs binocular rivalry. These two techniques are known to suppress images from awareness by different mechanisms (e.g. Tsuchiya et al., 2006) and clarifying the relative impact of these different types of suppression on the effects under scrutiny in future work would allow clearer combined interpretation of the results from the two experiments.

There were also no correlations between the eye tracking measures and the self-rated questionnaires. This differs from previous studies that have shown correlations between gaze fixation on angry faces and STAI or BAS scores (Terburg et al., 2011; Rohner, 2002). However, it must be noted that differences in the face stimuli or behavioural paradigms, and a relatively small sample size in the present experiment, could all contribute to this discrepancy.
7.3 fMRI experiment

7.3.1 Introduction

As discussed in Section 1.4.5, the dynamics of transition between non-conscious and conscious visual processing during binocular rivalry are modulated by social cues. In the first part of this chapter (Section 7.2), I have shown, with a behavioural experiment, that face dominance is another social cue that affects the dynamics of binocular rivalry. Specifically, I found that binocular rivalry switch rate was slower in the context of socially dominant faces, which would be consistent with the freezing response hypothesised on the basis of results presented in Chapter 6, namely that socially dominant and untrustworthy faces emerge more slowly into awareness and that this threat-related slowing has correlates in local variability of GM volume (see Sections 6.5.2 and 6.5.3). A logical next question could concern the functional (rather than structural) neural correlates of social modulation of awareness.

As discussed in Section 7.1, two predictions arising from a freezing hypothesis in the context of binocular rivalry are that there would be spatially localized changes in perceptual-transition-related BOLD signals (specifically reduction in activity in right IFG, right STS and bilateral amygdala), and that there would be peri-perceptual-transition deceleration in HR when binocular rivalry occurs in a dominant-face context. Additionally to these freezing-related predictions, I expected to find BOLD signal associated with endogenous perceptual transitions in right-sided frontal and parietal regions, as described in previous studies (Lumer et al., 1998; Sterzer et al., 2002; see also Section 1.3.5 and Chapter 3). These predictions were tested through an fMRI experiment with concomitant measurement of HR. Participants performed an identical behavioural paradigm to that used in the behavioural experiment (Section 7.2). I focused specifically on neuronal activity time-locked to perceptual transitions either in the socially neutral or the socially dominant condition. In addition, I measured HR continuously throughout the experiment planning to then focus analysis on HR changes during short epochs surrounding perceptual transitions.

7.3.2 Materials and methods

The fMRI data presented here have been used for separate analyses in Chapter 3 (Section 3.2).
7.3.2.1 Participants

Twenty participants (15 female; mean ± SD age = 25.2 ± 3.8 years; range = 20-34 years) were selected from the sample of 74 participants who had performed the behavioural experiment (Section 7.2). Participants were thus effectively pre-screened for ability to achieve stable perceptual fusion using prism glasses and for presence of mean binocular rivalry switch rates below 0.4/sec (16 switches in a 40 second period). Binocular rivalry switch rate is highly variable across individuals (e.g. Carmel et al., 2010) and those with faster switch rates may have transition events that are too frequent to be clearly resolved given the temporal constraints of the BOLD-fMRI haemodynamic response function (see Section 2.2.1.5). All selected individuals had mean face visibility durations longer than four seconds, considered sufficient to enable clear comparisons between dominant-face and neutral-face specific effects.

7.3.2.2 Stimuli, display apparatus and behavioural procedures

Stimuli and procedures were identical to those used in the behavioural experiment described in Section 7.2 (Figure 7.1B) and the experimental methodology inside the MRI scanner has already been described in a previous chapter (Section 3.2.3; see also Sections 2.4.3, 2.4.4 and 2.4.6). Briefly, stimuli subtended 2.9° visual angle and fixation crosses subtended 0.5° visual angle in diameter. Experimental trials (40 seconds each) were separated by 20-second rest periods. To measure BOLD signals specifically related to endogenous perceptual transitions, rivalry trials were supplemented with a second “replay” trial type (as described in Section 3.2.4), where conditions were closely matched to rivalry trials but perceptual changes were triggered exogenously by physical changes to the stimuli as displayed on the screen (Figure 7.6). Participants carried out 6 experimental runs, each containing a mixture of rivalry and replay trials as described in Section 3.2.4. Dominant-face and neutral-face trials were pseudo-randomised across the whole experiment so that in total participants completed nine rivalry trials (plus nine yoked replay trials) with each face type. Prior to the first run, participants underwent a practice run which contained one of each of the four trial types. Following the sixth run, a functional face localiser was undertaken (see Section 7.3.2.3). Eye tracking was not possible inside the MRI scanner due to glare on the prism glasses.
Figure 7.6: Replay behavioural condition employed in the fMRI experiment

Both eyes are shown an identical sequence of face and grating images, which matches the sequence reported from a previous rivalry trial. The replay sequence lasts 40 seconds, as for rivalry trials. The observer’s perceptual experience (blue shaded area) should be identical to the corresponding rivalry trial, with a sequence of percepts that have the same order and duration. Only two of the three possible percept types are shown for simplicity.

7.3.2.3 Functional face localiser

I adapted an existing functional face localizer paradigm (Pitcher et al., 2009). Blocks of face images or object images were presented in a pseudo-random order (14 blocks in total). Stimuli comprised 14 different faces and 14 different objects. All 14 unique images of a given type were presented in each block but two of these (chosen randomly) were presented twice in succession so that blocks contained a total of 16 trials. To ensure sustained attention to the stimuli, participants performed a standard one-back task, pressing a button on the keypad to indicate any consecutive presentation of the same image. Each image was presented for 200 ms, and followed by an 800 ms blank interval. Blocks were separated by 16-second rest periods.

7.3.2.4 fMRI data acquisition and pre-processing

The methodology and parameters used are as described in Sections 2.4.7 and 2.4.8. Further details relating to EPI sequences are as described for the BR experiment in Chapter 3 (Section 3.2.6). I obtained 115 volumes for each experimental run, and 142 volumes for the functional face localiser (Section 7.3.2.3).

7.3.2.5 fMRI data analysis

Construction of a GLM design matrix in SPM 8 from this dataset has already been described in Section 3.2.7. However, instead of splitting perceptual transitions into those in
rivalry and those in replay (as in Chapter 3), here they were divided into four types (neutral-face transitions in either rivalry or replay and dominant-face transitions in either rivalry or replay), conforming to a 2x2 factorial design, with factors of transition type (rivalry or replay) and also the additional factor of social face trait (neutral or dominant). Perceptual states and rest were again modelled as in Section 3.2.7, with the key difference that dominant-face and neutral-face states were modelled separately. This resulted in nine regressors of interest overall. For the functional face localiser analysis, the design matrix consisted of two regressors, one representing blocks of face image trials and the second representing blocks of object image trials. For all experimental runs, including the functional face localiser, 20 further regressors based of cardiac, respiratory and head movement data, were added to design matrices as covariates of no interest using an in-house-developed MATLAB toolbox (Hutton et al., 2011).

For each participant, maps of effect sizes for the main effect of rivalry transitions > replay transitions, the main effect of dominant-face transitions > neutral face transitions, and the interaction between these main effects, were computed by using the above model (for the localiser data, the same applied to the contrast faces > objects). Random-effects (second-level) analysis formed the basis for statistical inference: one-sample t-tests on these maps for all participants were performed for each contrast from the first level results, providing maps of evidence at each voxel against the null hypothesis of no effect.

Statistical results were examined at the peak activation level using three sequential sets of criteria, starting with the most stringent. Firstly, FWE correction across the whole brain was applied, with a criterion for significance of $p < 0.05$. Secondly, FWE correction was applied across a restricted volume, determined according to regions of interest defined a priori, again using $p < 0.05$ as the criterion for significance. For such small volume correction spherical search volumes of 10 mm radius were used around coordinates obtained from previous relevant studies or the functional face localiser (Table 7.1; for FFA and amygdala, a radius of 8 mm was used, based on data from previous meta-analyses that suggest this more accurately reflects the volume of these regions; Costafreda et al., 2008; Joseph, 2001). Thirdly, I examined uncorrected results, where the peak-voxel $p$ value exceeded a threshold of 0.001, taking care not to draw strong inferences from these findings since uncorrected fMRI results are associated with uncertain control over type I statistical error (see Section 2.2.3.2). For tables and figures, results are displayed at a threshold of $p < 0.01$ (uncorrected for multiple comparisons) with an extent threshold of 20 voxels.
### Table 7.1: Loci used for small volume correction analysis

See Section 7.3.2.5. Several of the spheres centred at these coordinates contained significant activations for a particular contrast, as indicated in the right hand column.

<table>
<thead>
<tr>
<th>LOCATION</th>
<th>MNI COORDINATES</th>
<th>SOURCE</th>
<th>SIGNIFICANT RESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amygdala</td>
<td>Left -18 -14</td>
<td>Functional face localiser</td>
<td>Borderline¹</td>
</tr>
<tr>
<td>Amygdala</td>
<td>Right 24 -14</td>
<td>Functional face localiser</td>
<td>Yes¹</td>
</tr>
<tr>
<td>FFA</td>
<td>Right 44 -24</td>
<td>Functional face localiser</td>
<td>Yes²</td>
</tr>
<tr>
<td>pSTS</td>
<td>Left -50 16</td>
<td>Functional face localiser</td>
<td>No</td>
</tr>
<tr>
<td>pSTS</td>
<td>Right 56 8</td>
<td>Functional face localiser</td>
<td>Yes³</td>
</tr>
<tr>
<td>FFA</td>
<td>Left -41 -27</td>
<td>Joseph et al., 2001</td>
<td>Yes²</td>
</tr>
<tr>
<td>Insula</td>
<td>Right 42 -6</td>
<td>Lumer &amp; Rees, 1999</td>
<td>Yes⁴</td>
</tr>
<tr>
<td>aSPL</td>
<td>Right 36 51</td>
<td>Kanai et al., 2011</td>
<td>Yes⁵</td>
</tr>
<tr>
<td>pSPL</td>
<td>Right 34 34</td>
<td>Kanai et al., 2010</td>
<td>No</td>
</tr>
<tr>
<td>IPL</td>
<td>Right 67 -33</td>
<td>Lumer et al., 1998</td>
<td>No</td>
</tr>
<tr>
<td>IFG</td>
<td>Right 52 -6</td>
<td>Lumer et al., 1998</td>
<td>Yes⁴</td>
</tr>
<tr>
<td>IFG</td>
<td>Right 50 16</td>
<td>Carr et al., 2003</td>
<td>Yes⁴</td>
</tr>
</tbody>
</table>

7.3.2.6 Psychophysiological analysis of heart rate data

Having made measurements of cardiac pulse during fMRI data acquisition (Section 7.3.2.4), I was able to examine condition-specific differences in perceptual transition-evoked HR. To obtain a HR trace from the cardiac pulse data I calculated the interbeat interval (the delay between each pair of consecutive pulses), removed outliers using a median filter (obtained from the HRVAS toolbox; Thuraisingham, 2006), and then used spline interpolation to derive continuous smoothed supersampled HR (5Hz; Perakakis et al., 2010). Data for each of the four experimental conditions described for the fMRI analysis (Section 7.3.2.5) were

¹ Activation in bilateral amygdala for the interaction term: dominant > neutral transitions (replay) > dominant>neutral transitions (rivalry; Figure 7.11A).
² Activation in bilateral FFA for the contrast replay transitions > rivalry transitions (Figure 7.10).
³ Activation in right pSTS for the interaction term: dominant>neutral transitions (replay) > dominant>neutral transitions (rivalry; Figure 7.11B).
⁴ Activation in right insula/IFG for the contrast rivalry transitions > replay transitions (Figure 7.9A).
⁵ Activation in right aSPL for the contrast rivalry transitions > replay transitions (Figure 7.9B).
epoched into time windows of six seconds around perceptual transitions (based on timescale of known HR deceleration effects related to threat; Bradley et al., 2005). Only transitions at least four seconds from other transition events were used, following the criterion determined at the participant selection stage (Section 7.3.2.1). HR was averaged across all epochs belonging to each experimental condition, creating four data vectors each with 31 data points (corresponding to a 6-second time window sampled at a frequency of 5Hz). I used SPM8 and treated each HR vector as a single channel of EEG data, converting vectors into SPM M/EEG files and then into “image” files (NIfTI format). I could thus employ a similar design to the fMRI analysis (Section 7.3.2.5) and test for statistically significant differences in transition-evoked HR between conditions at any time point within the whole six-second epoch while correcting robustly for multiple statistical comparisons. Statistical inference was based on random-effects analysis, carrying out repeated-measures ANOVA on the four condition-specific images to explore main effects and interactions. I planned to examine results both at the cluster and peak levels, applying FWE correction across the whole six-second epoch with a criterion for significance of $p < 0.05$ (uncorrected threshold $p < 0.05$).

### 7.3.3 Results

#### 7.3.3.1 Behavioural results

One participant became anxious in the scanner and could not complete the experiment, reducing the sample size to 19. Mean ± SEM switch rate in binocular rivalry was 0.26 ± 0.03 switches/second. Unlike the larger sample from the behavioural experiment (Section 7.2; Figure 7.2A), there was no significant difference in mean binocular rivalry switch rate for the dominant-face versus neutral-face conditions across all participants inside the MRI scanner [$t_{(18)} = 0.36$, $p = 0.97$, Cohen’s $d = 0.08$; Figure 7.7A]. However, comparing dominant-face and neutral-face switch rate for the same 19 participants performing the behavioural task outside the scanner (using data from the behavioural experiment; Section 7.2), there was a non-significant slowing of switch rate in the dominant-face condition [$t_{(18)} = 1.63$, $p = 0.12$, Cohen’s $d = 0.37$; Figure 7.7B]. Individual binocular rivalry switch rates were strongly correlated across the behavioural and fMRI experiments [$r = 0.56$, $p < 0.01$; Figure 7.7C]. However, there was no correlation between the effect of face dominance on switch rate in the behavioural and fMRI experiments [$r = -0.16$, $p = 0.51$; Figure 7.7D; the effect of dominance on switch rate was calculated for each participant by subtracting the mean binocular rivalry switch rate during dominant-face trials from that during neutral-face trials]. This latter finding suggests that the effect of face dominance on switch rate was altered between the behavioural and MRI settings in a manner that differed across individuals.
Figure 7.7: fMRI experiment behavioural results

(A) Unlike in the behavioural experiment (see Figure 7.2A for comparison), there was no effect of face dominance on binocular rivalry switch rate when measured in the MRI scanner. (B) The same 19 participants from (A) showed a non-significant slowing effect of face dominance on binocular rivalry switch rate when measured in the behavioural laboratory. For both panels (A) and (B), y-axes show switches/second and error bars represent one SEM. (C) Individual binocular rivalry switch rate was strongly correlated across behavioural and MRI settings. (D) The effect of face dominance on binocular rivalry switch rate was not correlated across behavioural and MRI settings.

There was no statistically significant relationship between an individual’s switch rate and the individual effect of face dominance on switch rate. However, a weak negative correlation between these measures was present in the sample who performed the behavioural study [$n = 69$, $r = -0.22$, $p = 0.07$] as well as in the sample who undertook the fMRI experiment, both when performing the task outside the MRI scanner [$n = 19$, $r = -0.18$, $p = 0.45$] and when performing it inside the MRI scanner [$n = 19$, $r = -0.21$, $p = 0.39$; Figure 7.8]. This consistent relationship suggests that the process of selection of participants for fMRI was not responsible for the lack of a dominance effect on switch rate in the fMRI experiment.
Figure 7.8: Correlations between individual binocular rivalry switch rate and individual effect of social dominance on switch rate

Scatterplots and lines of best fit are shown for the full behavioural experiment sample (n = 69; A), the fMRI experiment sample performing in the behavioural laboratory (n = 19; B), and the fMRI experiment sample performing inside the MRI scanner (n = 19; C). A weak but consistent relationship is apparent across the three plots.

Mean ± SEM percept visibility durations during binocular rivalry (across both dominant and neutral conditions) were: faces = 6.59 ± 0.67 seconds; gratings = 1.66 ± 0.34 seconds; mixed = 3.26 ± 0.47 seconds. The results demonstrate that the aims of achieving a slower mean switch rate (on average 10 switches in a 40-second trial) as well as longer mean face durations across participants for the fMRI experiment compared to the behavioural experiment (as outlined in the selection criteria for the fMRI experiment; Section 7.3.2.1) were successfully fulfilled.

A two-way repeated measures ANOVA with mean percept duration as the dependent variable, and factors of dominance (two levels: neutral/dominant) and percept (three levels: face/grating/mixed) revealed a significant main effect of percept \( [F_{(2,36)} = 27.25, \ p < 0.001] \) but no significant main effect of dominance \( [F_{(1,18)} = 0.36, \ p = 0.56] \) and no interaction \( [F_{(2,36)} = 2.19, \ p = 0.13] \).

7.3.3.2 fMRI results

BOLD signals time-locked to rivalry transitions (versus replay transitions) were found in a number of predominantly right-sided parietal and frontal regions (Table 7.2). After FWE correction across pre-defined ROIs (Table 7.1), I found two areas of statistically significant activation. Firstly, there was activation in right insula extending into right IFG \( (x = 34, \ y = 12, \ z = -6; \ T = 4.79; \ P_{\text{FWE-corr}} = 0.01 \) for insula; \( x = 46, \ y = 14, \ z = 0; \ T = 4.17; \ P_{\text{FWE-corr}} = 0.04 \) for IFG; Figure 7.9A). The result in IFG was obtained when correcting in a region...
previously implicated in perceptual transition (Lumer et al., 1998); correction in a region previously implicated in emotion recognition (Carr et al., 2003) also resulted in significant activation ($x = 44, y = 14, z = 8; T = 4.34; P_{FWE-corr} = 0.04$; see Table 7.1). Activation is also visible in left insula, but this did not reach corrected statistical significance ($x = -34, y = 8, z = 6; T = 5.48; P_{uncorrected} < 0.001$; Figure 7.9A). Secondly, there was significant activation in right aSPL ($x = 30, y = -44, z = 58; T = 4.59; P_{FWE-corr} = 0.02$; Figure 7.9B).
<table>
<thead>
<tr>
<th>REGION</th>
<th>HEM</th>
<th>MNI COORDINATES</th>
<th>CLUSTER SIZE</th>
<th>STATISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td><strong>Precentral gyrus</strong></td>
<td>Right</td>
<td>24</td>
<td>-18</td>
<td>56</td>
</tr>
<tr>
<td><strong>Insula</strong></td>
<td>Left</td>
<td>-34</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td><strong>Anterior cingulate</strong></td>
<td>Left</td>
<td>8</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td><strong>Insula/IFG</strong></td>
<td>Right</td>
<td>26</td>
<td>26</td>
<td>-2</td>
</tr>
<tr>
<td><strong>Orbitofrontal gyrus</strong></td>
<td>Left</td>
<td>-26</td>
<td>30</td>
<td>-16</td>
</tr>
<tr>
<td><strong>Anterior SPL</strong></td>
<td>Right</td>
<td>30</td>
<td>-44</td>
<td>58</td>
</tr>
<tr>
<td><strong>Cuneus</strong></td>
<td>Left</td>
<td>-18</td>
<td>-58</td>
<td>16</td>
</tr>
<tr>
<td><strong>Middle occipital gyrus</strong></td>
<td>Left</td>
<td>-36</td>
<td>-74</td>
<td>18</td>
</tr>
<tr>
<td><strong>Cerebellar vermis</strong></td>
<td>Left</td>
<td>0</td>
<td>-54</td>
<td>0</td>
</tr>
<tr>
<td><strong>Cingulate gyrus</strong></td>
<td>Left</td>
<td>-16</td>
<td>-8</td>
<td>22</td>
</tr>
<tr>
<td><strong>Caudate</strong></td>
<td>Left</td>
<td>-10</td>
<td>-46</td>
<td>50</td>
</tr>
<tr>
<td><strong>Superior occipital gyrus</strong></td>
<td>Right</td>
<td>24</td>
<td>-72</td>
<td>30</td>
</tr>
<tr>
<td><strong>Rolandic operculum</strong></td>
<td>Left</td>
<td>-42</td>
<td>-16</td>
<td>-18</td>
</tr>
<tr>
<td><strong>Cingulate gyrus</strong></td>
<td>Left</td>
<td>-16</td>
<td>-2</td>
<td>40</td>
</tr>
<tr>
<td><strong>Caudate</strong></td>
<td>Right</td>
<td>18</td>
<td>-6</td>
<td>24</td>
</tr>
<tr>
<td><strong>Cerebellum crus II</strong></td>
<td>Right</td>
<td>6</td>
<td>-72</td>
<td>-34</td>
</tr>
<tr>
<td><strong>Precentral gyrus</strong></td>
<td>Left</td>
<td>-34</td>
<td>-10</td>
<td>44</td>
</tr>
<tr>
<td><strong>Superior temporal gyrus</strong></td>
<td>Left</td>
<td>-66</td>
<td>-22</td>
<td>12</td>
</tr>
<tr>
<td><strong>Middle frontal gyrus</strong></td>
<td>Right</td>
<td>30</td>
<td>36</td>
<td>34</td>
</tr>
<tr>
<td><strong>Precuneus</strong></td>
<td>Right</td>
<td>14</td>
<td>-56</td>
<td>22</td>
</tr>
<tr>
<td><strong>Rolandic operculum</strong></td>
<td>Left</td>
<td>-58</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>Thalamus</strong></td>
<td>Right</td>
<td>4</td>
<td>-6</td>
<td>-10</td>
</tr>
</tbody>
</table>

| **Table 7.2:** Brain regions reflecting the main effect of transition type (rivalry or replay) |
| Anatomical labels, coordinates and statistics are shown for peak voxels activated for each contrast. All coordinates are in MNI space, cluster size is in voxels, and p-values are uncorrected (unless asterisked; see below). All activations surviving a threshold of \( p < 0.01 \) uncorrected, and with a cluster size of at least 20 voxels, are displayed. HEM, hemisphere. |
| Regions where activation is significant (\( P_{FWE-corr} < 0.05 \)) after small volume correction (see Table 7.1). |
A number of regions, mainly within bilateral occipitotemporal cortex, showed stronger signal during replay transitions than rivalry transitions (Table 7.2), including significant small-volume corrected results in bilateral FFA (for left $x = -42, y = -42, z = -22; T = 3.93; P_{\text{FWE-corr}} = 0.02$; for right $x = 38, y = -44, z = -26; T = 5.58; P_{\text{FWE-corr}} < 0.01$; Figure 7.10).
Figure 7.10: Brain regions with increased BOLD signal for the contrast replay transitions > rivalry transitions

Significant small-volume-corrected increases in BOLD signal during perceptual transitions in replay were found in bilateral FFA. Activations are shown in colour, corresponding to T values as indicated by the colour bar (brown represents low T values, and white represents high T values), and are overlaid on a single subject brain normalised to the standard MNI template. A threshold of $p < 0.01$ uncorrected was used for display purposes. Loci for small volume correction listed in Table 7.1.

BOLD signal associated with rivalry transitions in a dominant-face context (versus those in a neutral-face context) was found in a single cluster of activation in mPFC, which did not reach corrected statistical significance ($x = -6, y = 64, z = 16; T = 3.82; P_{\text{uncorrected}} < 0.01$; Table 7.3). The reverse contrast (transitions in a neutral face context > transitions in a dominant face context) showed activity in frontal, temporal, occipital and subcortical regions (Table 7.3), including a borderline-significant small-volume-corrected activation in right pSTS ($x = 54, y = -40, z = 2; T = 3.84; P_{\text{FWE-corr}} = 0.06$). The activation pattern here was clarified by a significant interaction (see below).
The critical contrasts pertaining to the interaction between the factors of transition type (rivalry/replay) and social face trait (neutral/dominant) allowed me to explore brain regions where signal differed between perceptual transitions in a dominant-face context compared to transitions in a neutral-face context, depending on the ambiguity of the stimulus (ambiguous in rivalry; non-ambiguous in replay). To evaluate the interaction term, I first asked where in the brain the dominant>neutral contrast was more pronounced during the non-ambiguous replay condition (versus rivalry). I expected to find right IFG activation for this interaction but there was no significant small-volume-corrected result in this region ($x = 48, y = 36, z = -6; T = 3.71; P_{uncorrected} = 0.001; \text{Table 7.4}$). As predicted,

### Table 7.3: Brain regions reflecting the main effect of social face trait (neutral or dominant)

Anatomical labels, coordinates and statistics are shown for peak voxels activated for each contrast. Coordinates are in MNI space, cluster size is in voxels, and p-values are uncorrected (unless asterisked; see below). All activations surviving a threshold of $p < 0.01$ uncorrected, and with a cluster size of at least 20 voxels, are displayed. Hem, hemisphere; Mid., middle; Sup., superior; Inf., inferior; Parahip, parahippocampal.

* Regions with borderline-significant activation ($P_{FWE-corr} < 0.06$ small volume corrected; see Table 7.1).

<table>
<thead>
<tr>
<th>Region</th>
<th>HEM</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>T</th>
<th>Z</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dominant face transitions &gt; Neutral face transitions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mPFC</td>
<td></td>
<td>-6</td>
<td>64</td>
<td>16</td>
<td>87</td>
<td>3.82</td>
<td>3.20</td>
</tr>
<tr>
<td><strong>Neutral face transitions &gt; Dominant face transitions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posterior insula</td>
<td>Left</td>
<td>-32</td>
<td>-26</td>
<td>28</td>
<td>196</td>
<td>4.94</td>
<td>3.84</td>
</tr>
<tr>
<td>Precentral gyrus</td>
<td>Right</td>
<td>26</td>
<td>-24</td>
<td>74</td>
<td>108</td>
<td>4.60</td>
<td>3.66</td>
</tr>
<tr>
<td>Mid. occipital gyrus</td>
<td>Left</td>
<td>-24</td>
<td>-96</td>
<td>4</td>
<td>231</td>
<td>4.40</td>
<td>3.54</td>
</tr>
<tr>
<td>Sup. frontal gyrus</td>
<td>Right</td>
<td>32</td>
<td>52</td>
<td>8</td>
<td>193</td>
<td>4.26</td>
<td>3.46</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>Left</td>
<td>-26</td>
<td>-34</td>
<td>4</td>
<td>86</td>
<td>4.18</td>
<td>3.42</td>
</tr>
<tr>
<td>SPL</td>
<td>Right</td>
<td>18</td>
<td>-74</td>
<td>56</td>
<td>80</td>
<td>3.94</td>
<td>3.28</td>
</tr>
<tr>
<td>pSTS*</td>
<td>Right</td>
<td>54</td>
<td>-40</td>
<td>2</td>
<td>162</td>
<td>3.84</td>
<td>3.21</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>Right</td>
<td>24</td>
<td>-42</td>
<td>74</td>
<td>162</td>
<td>3.86</td>
<td>3.23</td>
</tr>
<tr>
<td>Anterior cingulate</td>
<td></td>
<td>4</td>
<td>6</td>
<td>26</td>
<td>44</td>
<td>3.83</td>
<td>3.21</td>
</tr>
<tr>
<td>Putamen</td>
<td>Right</td>
<td>28</td>
<td>-10</td>
<td>12</td>
<td>106</td>
<td>3.82</td>
<td>3.20</td>
</tr>
<tr>
<td>Putamen</td>
<td>Right</td>
<td>26</td>
<td>14</td>
<td>12</td>
<td>92</td>
<td>3.70</td>
<td>3.12</td>
</tr>
<tr>
<td>Thalamus</td>
<td>Left</td>
<td>-20</td>
<td>-14</td>
<td>-2</td>
<td>27</td>
<td>3.57</td>
<td>3.04</td>
</tr>
<tr>
<td>Cingulate gyrus</td>
<td></td>
<td>-16</td>
<td>-26</td>
<td>34</td>
<td>72</td>
<td>3.56</td>
<td>3.04</td>
</tr>
<tr>
<td>Cuneus</td>
<td>Right</td>
<td>4</td>
<td>-80</td>
<td>42</td>
<td>57</td>
<td>3.56</td>
<td>3.03</td>
</tr>
<tr>
<td>Anterior cingulate</td>
<td></td>
<td>16</td>
<td>36</td>
<td>26</td>
<td>21</td>
<td>3.53</td>
<td>3.02</td>
</tr>
<tr>
<td>Parahip. Gyrus</td>
<td>Right</td>
<td>30</td>
<td>-6</td>
<td>-28</td>
<td>22</td>
<td>3.42</td>
<td>2.94</td>
</tr>
<tr>
<td>Cerebellar Vermis</td>
<td></td>
<td>8</td>
<td>-34</td>
<td>-30</td>
<td>68</td>
<td>3.40</td>
<td>2.93</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>Right</td>
<td>42</td>
<td>-6</td>
<td>-24</td>
<td>35</td>
<td>3.32</td>
<td>2.87</td>
</tr>
<tr>
<td>Postcentral gyrus</td>
<td>Right</td>
<td>66</td>
<td>-6</td>
<td>22</td>
<td>29</td>
<td>3.16</td>
<td>2.76</td>
</tr>
<tr>
<td>Inf. occipital gyrus</td>
<td>Right</td>
<td>40</td>
<td>-82</td>
<td>-4</td>
<td>21</td>
<td>2.99</td>
<td>2.65</td>
</tr>
</tbody>
</table>
activations in right amygdala \((x = 24, y = 2, z = -16; T = 3.88; P_{\text{FWE-corr}} = 0.04; \text{Figure } 7.11A)\), and right pSTS \((x = 54, y = -18, z = 6; T = 4.10; P_{\text{FWE-corr}} = 0.04; \text{Figure } 7.11B)\) showed small-volume-corrected significance. In addition, activation in left amygdala bordered on significance \((x = -16, y = -6, z = -14; T = 3.27; P_{\text{FWE-corr}} = 0.08)\). For right amygdala, plots of contrast estimates by experimental condition showed a crossover interaction, with strongest responses for neutral-face transitions in rivalry and dominant-face transitions in replay. Meanwhile, right pSTS was relatively under-activated in association with dominant-face transitions in rivalry, when compared to the other three experimental conditions (see right side of Figure 7.11).

Table 7.4: Brain regions reflecting an interaction between the main effects of transition type (rivalry or replay) and social face trait (neutral or dominant).

Anatomical labels, coordinates and statistics are shown for peak voxels activated for each contrast. Coordinates are in MNI space, cluster size is in voxels, and \(p\)-values are uncorrected (unless asterisked; see below). Activations surviving a threshold of \(p < 0.01\) uncorrected, and with a cluster size of at least 20 voxels, are displayed. Hem., hemisphere; sup., superior.

* Activation is significant/near-significant \((P_{\text{FWE-corr}} < 0.05 \text{ small volume corrected; see Table 7.1})\).
Figure 7.11: Brain regions with increased BOLD signal for the interaction term: dominant transitions>neutral transitions (replay) > dominant transitions>neutral transitions (rivalry)

Increased BOLD signals are seen in regions that are relatively more active in a dominant-face context and unambiguous visual environment and relatively less active in a neutral-face context and ambiguous visual environment. (A) Significant activation in right amygdala (small volume corrected). Some weaker activation is also seen in left amygdala, but this did not reach corrected statistical significance. Plots on the right show contrast estimates (arbitrary units) for each of the four experimental conditions, with error bars representing one SEM. Upper plot shows that, relative to other conditions, activation in right amygdala is increased at the time of both neutral-face transitions in rivalry and dominant-face transitions in replay. (B) Significant activation in right pSTS (small volume corrected). Plot on the lower right shows that, relative to other conditions, activation in right pSTS is decreased at the time of dominant-face transitions in rivalry. Activations are shown in colour, corresponding to T values as indicated by the colour bars (brown represents low T values, and white represents high T values), and are overlaid on a single subject brain normalised to the standard MNI template. A threshold of $p < 0.01$ uncorrected was used for display purposes. Loci for small volume correction listed in Table 7.1.

The second comparison pertaining to the interaction term identified regions where the dominant>neutral contrast was more pronounced in in the ambiguous rivalry condition (versus replay). There were no significant activations after correction for multiple comparisons. However there was an effect at uncorrected thresholds in hypothalamus ($x = 6, y = 2, z = -12; T = 4.24; P_{uncorrected} < 0.001; \text{Table 7.4; Figure 7.12}$). This activation is of potential interest since the hypothalamus is involved in regulation of autonomic responses, including during freezing (Benarroch, 1993). I therefore explored this interaction further,
bearing in mind that solid inferences could not be made from an uncorrected result. Similar to amygdala, the response pattern resembled a crossover interaction (plot on right of Figure 7.12). However, this was inverted with respect to the amygdala result, with relatively weaker responses for neutral-face transitions in rivalry and dominant-face transitions in replay.

**Figure 7.12**: Brain regions with increased BOLD signal for the interaction term: dominant transitions>neutral transitions (rivalry) > dominant transitions>neutral transitions (replay)

Increased BOLD signals are seen in regions that are relatively more active in a dominant-face context and ambiguous visual environment, and relatively less active in a neutral-face context and unambiguous visual environment. Activation was seen in hypothalamus (uncorrected). The plot on the right shows contrast estimates (arbitrary units) for each of the four experimental conditions with error bars representing one SEM. Relative to other conditions, activation in hypothalamus was decreased at the time of both neutral-face transitions in rivalry and dominant-face transitions in replay (the inverse pattern to that seen in amygdala; Figure 7.11). Activations are shown in colour, corresponding to T values as indicated by the colour bar (brown represents low T values, and white represents high T values), and are overlaid on a single subject brain normalised to the standard MNI template. A threshold of \( p < 0.01 \) uncorrected was used for display purposes.

Finally, I checked whether HR, which is the focus of a separate analysis (Section 7.3.3.3), significantly influenced the fMRI results when included as a covariate of no interest. I repeated the analysis with an SPM design matrix that included only 13 regressors based on respiratory and head movement data as covariates of no interest (and leaving out HR-related effects). The results were very similar to the original analyses and all of the statistically significant findings reported above were robust to this change. This confirmed that the fMRI results were not secondary to HR effects (such as those reported in Section 7.3.3.3).
7.3.3.3 Transition-evoked changes in heart rate

Applying the predetermined criteria for selecting sufficiently long segments of peritransition HR traces (Section 7.3.2.6) meant that six participants had no suitable HR data for at least one experimental condition (due to frequency of their perceptual switches). I therefore analysed HR data for the remaining 13 individuals.

Peritransition slowing of HR was found for trials in a dominant-face context (versus a neutral-face context) during a time window from 0.2 seconds pre-transition to 0.8 seconds post-transition, with the difference peaking at 0.4 seconds post-transition ($T = 2.56$; cluster-level $P_{FWE-corr} = 0.053$ with six data points exceeding the cluster-forming threshold; peak-level $P_{FWE-corr} = 0.087$; Figure 7.13, upper plots). Furthermore, there was also slowing of HR in trials with replay (versus rivalry) transitions during a time window from 1.2 seconds pre-transition to 0.2 seconds pre-transition, with the difference peaking at 0.8 seconds pre-transition ($T = 3.09$; cluster-level $P_{FWE-corr} = 0.053$ with six data points exceeding the cluster-forming threshold; peak-level $P_{FWE-corr} = 0.027$; Figure 7.13, lower plots). There was no significant interaction between the main effects of social face trait and transition type.
Figure 7.13: Perceptual transition-evoked heart rate
HR is shown for a six second time window around perceptual transitions (three seconds pre-transition to three seconds post-transition). Vertical lines at time 0 represent the timing of participant button press (taken to be the timing of perceptual transition in fMRI analyses). The rectangular area inside the dotted lines represents the estimated time of actual perceptual transition (mean ± 2 SD of delay, across all participants, between replayed transitions and responses in replay trials; see Section 7.3.2.2 for a description of the replay condition). The two left hand plots show HR traces for trials with neutral and dominant faces (collapsed across rivalry and replay; neutral in blue and dominant in red; upper left plot), and for trials with rivalry and replay transitions (collapsed across neutral and dominant face trials; rivalry in green and replay in magenta; lower left plot). The two right hand plots show difference traces for the plots to their left (neutral-dominant for upper right plot; rivalry-replay for lower right plot). In the upper two plots, a difference is apparent between 0.2 seconds pre-button press and 0.8 seconds post-button press; this reflects slower heart rate for the dominant-face trials during this period. In the lower two plots, a difference is apparent between 1.2 seconds pre-button press and 0.2 seconds pre-button press; this reflects slower heart rate in replay during this period. Periods where there is a significant difference between the two conditions (or difference from zero in the subtraction plots) are shaded in orange.
Of note, some of the reported HR changes apparently precede perceptual transitions. This seems a surprising finding but probably has a parsimonious explanation: the observation likely reflects changes occurring in the delay period between perceptual change and button press (since in the fMRI experimental design described in Section 7.3.2.5 perceptual transitions were denoted by button press timings, and the mean ± SD reaction time to external replayed perceptual switches across the experimental sample was 1.1 ± 0.13 seconds). In other words, in the experiment there is a 'pre-transition' period that in fact reflects participant reaction time. It was possible to determine individual reaction times (latency from appearance of a visual stimulus to button press) from behavioural responses during the fMRI replay condition, where the timing of perceptual changes is externally controlled (see Section 7.3.2.2). The relationship between timing of perceptual changes and timing of button presses is illustrated graphically in Figure 7.13.

7.3.4 Discussion

In this fMRI experiment I examined BOLD signals time-locked to perceptual transitions in binocular rivalry and concomitantly measured HR, asking how these measures were modulated by socially dominant face traits. Based on delayed emergence from interocular suppression for threatening faces, and more specifically on the proposal that this delay is a correlate of freezing responses (see Sections 6.5.2 and 6.5.3), I made two predictions that were tested in this experiment. First, I predicted that modulation of transition-related BOLD signal by facial dominance would be reflected by signal changes in IFG, amygdala and STS. Second, I predicted that peri-perceptual-transition HR would show relative slowing in a dominant face context. An alternative ‘evolutionary vigilance’ hypothesis would argue that threatening stimuli are processed more rapidly, thereby increasing the chance of timely appropriate reaction and survival (e.g. Vuilleumier, 2005), and would lead to opposite predictions to those described above.

Perceptual transitions during binocular rivalry were associated with BOLD signal changes in a right frontoparietal group of brain regions, including insula and aSPL, replicating results from previous studies (Kleinschmidt et al., 1998; Lumer et al., 1998; Sterzer et al., 2002). Crucially, the interaction between transition type (rivalry/replay) and social face trait (neutral/dominant) was expressed in right amygdala and right pSTS. As predicted, these regions were least active at the time of endogenous perceptual transitions in a socially dominant context, supporting a freezing/avoidance mechanism, rather than an evolutionary vigilance account. My prediction of activation changes in right IFG in socially relevant contexts was not borne out; although some activation was seen in this region for one contrast pertaining to the interaction term, this did not remain significant after
correction for multiple comparisons. I found that heart rate around the time of perceptual transitions was slower in a dominant-face context than a neutral-face context, which again supports the freezing hypothesis.

7.3.4.1 Behavioural and eye tracking measures inside the scanner

Individual binocular rivalry switch rates were strongly correlated across the behavioural and fMRI experiments. However, unlike in the behavioural experiment, there was no overall effect of face dominance on switch rate in the fMRI experiment (further discussion in Section 7.4).

I was not able to monitor eye movements inside the MRI scanner since the prism glasses used to aid perceptual fusion resulted in glare that interfered with eye tracking measurements. Given the behavioural paradigm was unchanged between the behavioural and fMRI experiments I reasoned it would be unlikely that eye movement differences between conditions should have occurred in the MRI scanner; however a future study that could confirm (or refute) this assumption would of course be useful.

7.3.4.2 Neural correlates of endogenous perceptual transitions

The imaging results relating to rivalry versus replay transitions (BOLD activation in right sided frontal and parietal regions) represent a close replication of previous findings relating to BOLD activity time-locked to endogenous (as contrasted with exogenously generated) perceptual transitions (Lumer et al., 1998; Sterzer et al., 2002). Demonstrating such replication was an important first step in this experiment since it allowed me to establish the validity of the current experimental paradigm for probing BOLD signal related to perceptual transitions. The significance of transition-related frontoparietal activation remains under some debate (Sections 1.3.5, 3.4.5 and 9.2.3) and is explored further in Chapter 3 and Chapter 4.

7.3.4.3 Modulation of transition-related neural activity according to social face context

I had predicted that the interaction between main effects of transition type (rivalry/replay) and social face trait (neutral/dominant) would be reflected in BOLD signal changes in right pSTS. This region is extensively implicated in visual appraisal of socially relevant
environmental cues, in particular cues from faces such as eye gaze direction or emotional expression (e.g. Allison et al., 2000; Sections 1.2.2 and 1.4.3). In the context of purported perceptual freezing, I expected reduced pSTS activation at the time of dominant-face transitions in rivalry. Examining the interaction term confirmed that pSTS was relatively (when compared to other conditions) less activated specifically at the time of transitions in the ambiguous dominant-face condition, while being relatively strongly activated for all other conditions.

Like pSTS, the amygdala is involved in social appraisal, including interpreting socially relevant cues from faces (e.g. Adolphs, 2010; Sections 1.4.3 and 1.4.4). Again, based on the freezing hypothesis, I therefore predicted that activation here would be weakest at the time of dominant-face transitions in rivalry. As for pSTS, amygdala activation was indeed relatively weak for the ambiguous dominant-face condition. However, examining the interaction more carefully revealed a different pattern to pSTS: transition-related activity was also relatively weak for a socially neutral non-ambiguous (replay) context, making the amygdala result appear more like a cross-over interaction. This may suggest that amygdala has a broader or more complex role in mediating the modulation of perceptual transition by social cues than STS, but further work will be required to clarify the relative contributions of these regions. The present results show that both are comparatively under-active in a socially dominant and ambiguous context, consistent with perceptual freezing. Whether there is active suppression of activity in these regions, and if so, how this is mediated, would be intriguing questions for future investigation.

Based on the known involvement of right IFG both in perceptual transition (Lumer et al., 1998) and in mediating behavioural responses to socially relevant information (Jabbi & Keysers, 2008), I predicted reduced activation in this region at the time of perceptual transitions in a socially dominant context; however, this was not confirmed. The relevant imaging contrast did reveal activation in bilateral IFG, but neither focus remained statistically significant after correction for multiple comparisons (even across a small volume). The strength of the effects in these regions [*T = 4.14 for left IFG; T = 3.71 for right IFG*] was similar to the significant effects reported in right pSTS and right amygdala (Table 7.4). While in the present instance they must be treated as null findings, one could speculate that the reason for not detecting significant effects in right and left IFG may be due to slight mismatch with my *a priori* ROI and the failure to predict a bilateral result, respectively. Although both the statistically significant results reported in relation to the interaction term (pSTS and amygdala) were in the right hemisphere, it is noteworthy that activations were present bilaterally in pSTS, amygdala, and IFG (Figure 7.11; Table 7.4). This may represent a network of regions involved in modulation of access to awareness.
according to social dominance (or indeed social threat more generally). Future work could focus more closely on whether such a set of modulatory mechanisms are indeed bilateral.

### 7.3.4.4 Modulation of peritransition heart rate by binocular rivalry and social face dominance

In support of the freezing hypothesis, I found that peritransition heart rate showed slowing in a dominant-face (as compared to neutral-face) context. Cardiac deceleration occurs in association with consciously perceived unpleasant images (Bradley et al., 2005) and is an autonomic correlate of freezing in defensive cascade models (Bradley et al., 2001). Additionally, peritransition heart rate showed slowing for exogenous (as compared to endogenous) perceptual transitions, broadly in keeping with established psychophysiological literature reporting that focus on external events in the visual environment is associated with cardiac deceleration whereas internal manipulation of information is associated with cardiac acceleration (Libby et al., 1973). Further studies will be required to evaluate whether such peritransition HR changes can be used as a reliable peripheral physiological marker of perceptual transition.

### 7.4 Chapter discussion

With the experiments presented in this chapter I sought to determine whether social face dominance traits modulate perceptual transition in binocular rivalry, and specifically to explore behavioural, neuronal and psychophysiological correlates of such modulation. Based on earlier work that showed delayed non-conscious processing for threatening faces (Chapter 6; Stein and Sterzer, 2012; Stewart et al., 2012) and the proposed defensive freezing mechanism underlying this delay (Sections 6.5.2 and 6.5.3; Stewart et al., 2012), I made three predictions relating to effects on behaviour, BOLD signal and psychophysiological measures of HR (Section 7.1). Consistent with these predictions (and the freezing hypothesis), I found slowing of binocular rivalry switch rate, transition-related reduction of BOLD signal in right amygdala and pSTS, and slowing of peritransition HR, all in the context of socially dominant faces. The relevance of these findings to a proposed underlying mechanism of freezing, and their contribution to such a hypothesis in the light of the results of other experiments from this thesis is discussed further in the General discussion chapter (Section 9.4).
7.4.1 Comparing binocular rivalry switch rate effects in the behavioural and fMRI experiments

While individual switch rate in binocular rivalry was strongly correlated across the behavioural and fMRI experiments, the effect of social dominance on switch rate was not correlated across these two experiments. Unlike in the behavioural experiment, there was no overall effect of face dominance on switch rate in the fMRI experiment. The same sample of individuals that performed the fMRI experiment showed a clear trend towards an effect when performing outside the MRI scanner, suggesting that this difference was not due to experimental sample differences. There are a number of potential reasons why the behavioural effect did not replicate during fMRI. One possible explanation for this discrepancy may relate to the different contexts of the behavioural laboratory and MRI scanner. For example, state anxiety affects the dynamics of binocular rivalry (Nagamine et al., 2007), as well as modulating the effects of threatening stimuli on binocular rivalry (Gray et al., 2009). If the environment of the MRI scanner was substantially more anxiety inducing than that of the behavioural laboratory, this may explain the diminished difference between binocular rivalry dynamics in a dominant-face context and a neutral-face context. Although I recorded state anxiety in the behavioural setting, and found no correlation with switch rate or the effect of facial dominance on switch rate, investigating possible interactions between environmental contexts and binocular rivalry dynamics in the present experimental setup would require recording of state anxiety (and probably other related measures) in the MRI scanner in a future experiment. Presently, the absence of a difference in behavioural performance for the neutral-face and dominant-face conditions in the MRI scanner is puzzling, but it could be seen as advantageous, since it enabled me to interpret the fMRI findings more specifically by effectively excluding the possibility that that any of the imaging results were confounded by systematic differences in behavioural output between the two conditions.

7.4.2 Disambiguating mixed percepts and perceptual transitions

Confirming a pattern of frontoparietal activation for the contrast rivalry transitions > replay transitions was an important step in validating my experimental paradigm. In addition, the fact that I modelled mixed percepts, as well as face and grating percepts, as three separate regressors in the GLM is of some significance since it enabled me to reasonably treat all perceptual transitions as instantaneous events and avoid difficulties in interpretation related to possible conflation of perceptual transitions and periods of mixed
There were undoubtedly some differences between the appearance of mixed percepts experienced during binocular rivalry, and those displayed in the replay condition. Rivalrous mixed percepts can appear in many different ways, but often a wave of change from one percept to the other is reported, beginning at one corner or one side of the display and travelling gradually to the opposite corner/side. To avoid introduction of any spatial bias, the replayed mixed percepts in this experiment consisted of a gradual fading of the contrast of one percept, while that of the other gradually increased. It is possible that differences between the appearance of such a display and the corresponding mixed percept in rivalry may have affected the fMRI findings. Of note, a previous study (Knapen et al., 2011) used a travelling wave of change occurring from a randomly selected side of the stimulus to simulate mixed percepts during replay. It seems that this is just as likely to inadvertently introduce differences between rivalrous and replayed mixed percepts since in both cases there was no knowledge of how the paired rivalrous mixed percept was experienced (for example, if it appeared as a travelling wave of change, in which direction this occurred).

7.4.3 Freezing as a possible mechanism for social modulation of perceptual transition

A putative freezing account for the effects of social dominance on binocular rivalry, as reported in this chapter, enables substantial integration of the behavioural, imaging and psychophysiological findings. As discussed in Section 1.5, recent studies have begun to characterise freezing behaviours in humans, including in the context of social threat (e.g. Roelofs et al., 2010). Furthermore, in Chapter 6 (Sections 6.5.2 and 6.5.3) I proposed that freezing responses may also have an effect on perceptual processing (see also Stewart et al., 2012). The fMRI findings of reduced pSTS and amygdala activation reported here, both observed specifically in association with perceptual transition in the socially dominant and ambiguous experimental condition, are in keeping with a role for these regions in such perceptual freezing. However, confirming that this type of mechanism indeed underpins the observed effects will require further direct empirical evaluation. Presently, the demonstrated facial-dominance-related HR deceleration further supports a freezing mechanism, as does slowing of binocular rivalry switch rate (which could be due to slower perceptual re-interpretation, slower motor responses, or both). Moreover, that these effects should occur specifically in the context of threatening as well as unclear and ambiguous information dovetails with existing characterisations of freezing behaviours (Lang et al.,
2000). I did not find statistically significant BOLD signal changes in right IFG in a perceptually ambiguous (rivalrous) and socially threatening (dominant-face) context as predicted. This null finding may stem from the lack of demonstrable slowing of binocular rivalry switch rate in the MRI scanner. A final possible link to a freezing mechanism is the increased BOLD signal in hypothalamus in an ambiguous and socially dominant context (the reverse pattern to pSTS and amygdala). Limited emphasis can be placed on this result since it does not survive statistical correction for multiple comparisons. However, the hypothalamus is important in coordinating autonomic outputs of freezing responses (Fanselow, 1994; see Sections 1.5.2 and 1.5.3), and is activated when viewing of angry faces, for example (Hermans et al., 2008). Further work to evaluate the role of this region in this setting would therefore be of interest.

7.4.4 Summary

In this chapter I have described two experiments aimed at determining whether socially threatening dominant faces modulate various aspects of perceptual transition, and specifically to explore the behavioural, neuronal and psychophysiological correlates of such modulation. Behaviourally, perceptual transitions during binocular rivalry are less frequent in a dominant face context. The finding of endogenous perceptual-transition-related BOLD signal in right frontoparietal cortex replicates previous published work that may suggest such regions are involved in mediating access to visual awareness. Moreover, I found that endogenously generated perceptual transitions in a dominant-face context are associated with relative reduction in BOLD signal in right amygdala and right pSTS. I have discussed that this finding may reflect freezing mechanisms that induce reduction in certain aspects of perceptual processing specifically in the context of social threat and ambiguity in the visual environment. The finding of peritransition HR slowing in a dominant-face context further supports the involvement of a freezing mechanism.

The findings presented in this chapter extend the work from Chapter 6 by showing that social face traits modulate the transition from non-conscious to conscious visual processing from a behavioural, psychophysiological as well as neural standpoint. In addition, the present findings provide new evidence in support of the hypothesis that socially threatening visual information presented in visually ambiguous (or intermittently non-conscious) conditions elicits a defensive freezing response.
Chapter 8

Physiological evidence for freezing responses to non-conscious social threat

“Tears come from the heart and not from the brain.”

Leonardo da Vinci (1452-1519)

8.1 Introduction

Defensive responses to threat are largely conserved across different mammalian species. These behaviours have important implications for survival, and in humans are related to the development and maintenance of psychopathological conditions. Defence cascade models provide a useful framework according to which the specific defensive response elicited depends on the proximity of danger. For example, intermediate proximity is associated with post-encounter responses including freezing (characterised by tense and attentive immobility and HR deceleration; see Section 1.5 for a more detailed review).

Freezing occurs when threat is not very imminent or when there is no means of escape. The neurobiological underpinnings of this behaviour are well described in mammals, especially the rat (see Section 1.5.2). More recently freezing responses (defined by reduced body sway and bradycardia) have been demonstrated in humans, including for socially threatening angry-expression faces (Roelofs et al., 2010; see Section 1.5.2). Several studies have begun to explore the neuronal mechanisms of freezing in humans using fMRI (see Section 1.5.3).
Two experiments already presented in this thesis have provided findings that may reflect freezing responses to socially threatening stimuli in humans. I showed in Chapter 6 that non-consciously perceived dominant and untrustworthy facial traits (both traits are correlated with threat; Oosterhof and Todorov, 2008) are associated with slower access to visual awareness than for socially neutral faces. Similar findings have been reported for faces with angry expressions (which are also threatening), when suppressed by CFS (Stein and Sterzer, 2012). In Chapter 6 I hypothesised that this delayed access to visual awareness for threatening images reflects a perceptual freezing phenomenon. In Chapter 7 I showed that socially threatening faces placed in binocular rivalry have behavioural effects on perceptual transitions (slowing of binocular rivalry switch rate) as well as having neural effects (reduced BOLD signals in amygdala and pSTS) and psychophysiological effects (HR deceleration) in relation to such perceptual switches. This combination of reduced BOLD signals in perceptual regions and relative bradycardia, both specifically in association with perceptual transitions in ambiguous (rivalrous) and threatening contexts, again provided some support for underlying freezing of perception. It remains to be tested whether the slowing of the observed in Chapter 6 is associated with both motor and autonomic components, as would be predicted by a freezing hypothesis, and if so, whether these components might be dissociable, as they are in rodents (LeDoux et al., 1988).

The non-human models (and associated experimental evidence) for freezing responses may not be directly applicable for social stimuli, which are more broadly relevant and more readily testable in humans. It is not clear whether in humans the approach of a predatory animal or a painful stimulus elicit the same type of response as the face of a conspecific that contains threatening visual cues, for example. Moreover, while there is behavioral evidence of freezing-like responses to social threat in humans (e.g. Roelofs et al., 2010; see Section 1.5.2), these findings do not identify a physiological basis of this social fear response.

Assessment of corticospinal excitability is a temporally precise approach for probing quite directly the motor physiological components of freezing. This can be achieved by measuring, from a peripheral muscle, the amplitude (or area) of a motor potential evoked by a single pulse of transcranial magnetic stimulation delivered over contralateral primary motor cortex (TMS-MEP; Chen, 2000). This measure has been used to index freezing, defined in this context as a decrease in CSE (Avenanti et al., 2009; Borgomaneri et al., 2015); reduced CSE as measured by TMS-MEP also occurs following unexpected sensory stimuli (Cantello et al., 2000) or painful stimulation (Farina et al., 2003). Faces with fearful expressions have been associated with the opposite effect of increased CSE, which is argued to reflect preparedness for action (Schutter et al., 2008). Importantly, there are differences between fearful facial expressions and other social face stimuli with negative valence, such
as angry faces, in terms of how they signal threat (environmental versus direct threat). These differences are discussed further in Chapter 9 (Section 9.3.1) and they lead to opposite predictions for effects on CSE of directly threatening facial expressions (as used in Chapters 6 and 7 as well as the present experiment) and of indirectly threatening fearful faces.

While I have shown that responses on a perceptuomotor CFS task are slower in the context of non-conscious social threat (and proposed that this reflects perceptual consequences of freezing; Chapter 6), the question of whether more core components of freezing behaviours (namely reduction in CSE and bradycardia) can be triggered by visual stimuli not accessible to conscious appraisal remains unanswered (see Section 1.5.2). Exploring this question would allow me to determine whether defensive behaviours (and in particular freezing) might occur without any conscious evaluation of the threatening situation at hand.

For the present experiment, I therefore set the objective of measuring CSE, alongside autonomic indices including heart rate and pupillary area, during non-conscious presentation of socially threatening face images. I elected to use a minimally adapted version of the CFS paradigm employed in Chapter 6 (Section 6.2.3). This allowed visual stimuli to be hidden from awareness for several seconds and also permitted the measured CSE and autonomic changes to be related to the behavioural measure of time-to-emergence (t2e; see Section 6.2.3), which reflects perceptuomotor performance. I hypothesised that delayed access to visual awareness for socially threatening images (as indexed by prolonged t2e) is part of a broader defensive freezing response, and that the motor and autonomic correlates of freezing should therefore be detectable during exposure to non-conscious social threat.

Specifically, my main prediction was that that non-conscious threatening faces would provide the appropriate conditions for reliable induction of defensive freezing responses (see Section 7.4.3) and this would be reflected in delayed t2e (as previously shown; Stewart et al., 2012; Chapter 6) as well as reduced CSE and bradycardia for threatening relative to neutral faces. The relationships between perceptuomotor performance (t2e), autonomic changes (HR) and motor physiological changes (CSE) could also be explored. Finally, pupillary dilatation is more generally associated with novelty and arousal (Bradley et al., 2008; Libby et al., 1973), and is seen in freezing, but is not a specific correlate of freezing per se. Thus, I predicted that pupillary dilatation would be observed for both threatening and neutral faces, but that it would be more pronounced in the threatening condition (Bradley et al., 2008).
Freezing of motor output can occur as part of several defensive immobility responses, which likely differentially engage subregions of PAG and ascending projections to higher brain areas (see Section 1.5.2). This diverse repertoire of immobility responses and underlying neurobiological circuitry is reflected in substantial individual variability demonstrated in the motor components of freezing (Ly et al., 2014). Such variability is correlated with affective state (Roelofs et al., 2010; Wada et al., 2001) and relates to past personal experiences (Hagenaars et al., 2012). Given the importance of individual factors in the motor components of freezing, and the established individual variability in t2e (Chapter 6), I planned to explore individual variability in the behavioural (t2e) and physiological (TMS-MEP) measures in the present experiment. Specifically I predicted that (since these measures are hypothesised to index perceptual and motor components of freezing) effects of non-conscious threat on t2e and on CSE would be correlated within individuals. On the other hand, the autonomic correlates of freezing (in particular bradycardia) appear to depend on more direct and less diverse pathways to brainstem nuclei (see Section 1.5.2), so a less individually variable effect of threat (which would not be correlated with individual CSE and t2e measures) was predicted in this instance.

To test these hypotheses and predictions, 16 healthy volunteers performed a CFS behavioural task similar to that used in Chapter 6 (Section 6.4.3). Briefly, participants used a speeded button press to indicate the time taken for face images (threatening or neutral) to overcome interocular suppression and emerge into awareness (t2e; Section 1.3.3). In addition, CSE was assessed at precisely defined timepoints shortly after visual stimulus presentation through TMS-MEP measurement from the first dorsal interosseous muscle of the inactive left hand. Finally, I recorded ECG (enabling calculation of instantaneous HR), as well as continually tracking pupil position and size, in order to explore psychophysiological correlates of defence and arousal. For the purpose of excluding any influence of direct and non-specific effects of TMS on findings for t2e, HR and pupil changes, a second auxiliary experiment was planned, which would replicate conditions of the main experiment without including TMS or MEP measurement.

8.2 Materials and methods

8.2.1 Participants

Experimental participants were recruited according to the procedures and inclusion criteria described in Section 2.4.1. My aim was to attain a sample size of 16 for analysis of data, based on previous similar studies that have demonstrated adequate statistical power with
comparable sample sizes (Schutter et al., 2008; Urgesi et al., 2006). To ensure optimal data quality, additional a priori inclusion criteria were established based on the size and consistency of MEP signals across trials (which were critically dependent on participants’ ability to maintain a very stable head position for the duration of the experiment; see Section 8.2.3.1) as well as on participants’ ability to perform the visual task and to complete the whole experiment. From a total of 43 recruited individuals (22 female; mean ± SD age 23.2 ± 4.5 years; age range 18-38 years), I obtained complete sets of data of adequate quality for 16 participants (9 female; mean ± SD age 22.7 ± 4.4 years; age range 18-36 years). The remaining 27 participants were excluded for one of several pre-specified reasons (seven participants could not adequately perform the CFS task; fourteen participants did not fulfill the predetermined TMS-MEP-related inclusion criteria (see Section 8.2.3.1); four participants could not complete the experiment due to fatigue; MEPs could not be measured at all in four participants; and the data of two participants were corrupted).

8.2.2 Experimental procedures

8.2.2.1 Overview of experimental sequence

The study consisted of two experiments, undertaken during the same session. All 16 participants performed the main experiment, in which a dichoptic CFS task was undertaken. This required a button response to be made on each trial of the task with the participant’s dominant right hand (Section 2.4.4.1). Concurrently with these responses I also recorded MEPs (from the inactive left hand), and continuously monitored ECG and pupil size/position. Single pulses of TMS were delivered to the scalp overlying right M1 at a variable latency from the start of each visual task trial, in order to induce muscle contraction in the left hand and be able to measure the associated MEP (and thus obtain an index of CSE). A subset of participants from the main experiment (n = 12) also performed an auxiliary experiment immediately prior to the main experiment. This auxiliary experiment was undertaken in an identical environment, with the same setup for the visual task and ECG/pupil recordings. The only difference was that in the auxiliary experiment no TMS pulses were delivered and MEPs were not recorded from the left hand. The auxiliary experiment was always performed first to eliminate any possible influence of lasting effects from the TMS pulses delivered in the main experiment.
8.2.2.2 Stimuli, display apparatus and procedure for CFS

The experimental setup for the visual paradigm was similar to that in Chapter 6 (see also Stewart et al., 2012); full details are described in Sections 2.4.3 and 2.4.4; see also Figure 8.1A. Screen resolution was 1024 x 768 with 60 Hz refresh rate) and viewing distance was 65 cm. With this configuration, each visual stimulus (including its tile surround) subtended 3.5° visual angle. Face stimuli, as described in Section 2.4.3, were used, specifically selecting a socially neutral face and a socially threatening face (Figure 8.1B). Trials were grouped in blocks of 24, and split equally between threatening-face and neutral-face conditions within each block. In the main experiment, participants completed eight such blocks; those who performed the auxiliary experiment completed four further such blocks. Trials with incorrect responses were excluded from further analysis (a mean of 94.6% of trials were correct across all participants; range 77.1 – 99.5%).

**Figure 8.1:** Behavioural paradigm and stimuli

(A) Faces were shown to one eye with gradually increasing contrast while a CFS mask was shown to the other eye. Initially, participants could see the CFS mask only; they then needed to indicate when the face emerged into awareness (t2e) by button press. (B) Social face model with face traits varying along orthogonal dimensions of dominance and trustworthiness; threat is defined along the main diagonal (Oosterhof and Todorov, 2008; Section 1.4.2). The two faces marked with red borders were used in the present study.

8.2.2.3 EMG and ECG recording

Disposable wet surface electrodes were stuck on participants’ skin at specific sites: over the belly of the left FDIO muscle (active) and over the left first metacarpophalangeal joint (reference) for EMG; over the lateral third of both clavicles for ECG; over the left inferior costal margin for the ground electrode. EMG and ECG signals were recorded on a Visor2 system with a sampling frequency of 2048 Hz.
8.2.2.4 Single-pulse TMS

Single pulses of TMS were delivered using a Magstim 70mm Alpha (figure-of-eight) coil plugged into a Magstim Rapide^2 stimulator. A single pulse of TMS delivered to the scalp overlying M1 induces corticospinal pathway activation and results in a contralateral muscle contraction associated with an MEP (Chen, 2000). With participants positioned in the head/chin rest the optimal stimulation site to produce MEPs in FDIO was determined through serial stimulations (moving the coil systematically around the scalp overlying M1 and selecting the stimulation site that produced the highest MEP amplitude). The TMS coil was then fixed in this position using a camera tripod and participants were asked to keep their head completely still. I used the minimum TMS stimulus intensity required to consistently produce MEPs with mean amplitude of around 500µV (across 10 stimuli). Pulses were delivered at three different latencies from visual stimulus onset in the CFS trials (1 second; 1.2 seconds; 1.4 seconds) in a pseudo-randomised and counter-balanced fashion. On any given trial, the planned TMS pulse was only delivered if the t2e response had not already been made (in other words, all TMS pulses were delivered while the face stimulus was perceptually invisible). This resulted in omission of the TMS pulse on 2.8% (range 0 – 10%) of trials.

8.2.2.5 Eye tracking

Eye tracking was performed according to the methodology described in Section 2.4.9. Of note, depending on eye dominance (see Section 2.4.4) some participants were shown the face stimulus to the right eye throughout the experiment (right eye dominant; n = 11) and the remainder where shown the CFS stimulus to the right eye throughout the experiment (left eye dominant; n = 5).

8.2.3 Data analysis

All behavioural and physiological data were processed off-line using MATLAB. Only trials in which the CFS task had been performed correctly (see Section 8.2.2.2) were used to calculate t2e and for subsequent analysis of MEPs, ECG and pupil size.

8.2.3.1 MEP analysis

Surface EMG recordings from left FDIO were used to calculate MEP amplitude for each trial. Trials with EMG activity preceding the TMS pulse (suggesting the participant’s hand was not relaxed) were discarded from analysis (8.1% of trials across all participants;
In addition, any MEPs with amplitude below 100µV were excluded (6.7% across all participants; range 0 - 47.4%). Obtaining reliable MEP measurements (which required maintenance of a very stable coil position in relation to the participant’s scalp) while participants were performing the visual task presented a significant challenge. It was therefore determined a priori (after piloting) that for inclusion of participants at least 50% of TMS pulses needed to have resulted in MEPs with amplitude ≥ 100µV. A second inclusion criterion was for a minimum of 50% of trials to result in a TMS pulse (as described in Section 8.2.2.4 on any given trial a TMS pulse was only delivered if the t2e response had not already been made; however the TMS pulse omission rate was in fact much lower than 50% in this experimental sample; see Section 8.2.2.4).

Data were filtered with a high-pass filter set at 2 Hz, then a low-pass filter set at 1000 Hz (mean correcting at each step), and then rectified. The maximum MEP amplitude within a 150-ms time window following each TMS pulse was measured.

8.2.3.2 ECG and pupil size analysis

The ECG recordings were used to calculate HR, employing methodology similar to that used in Chapter 7 (Section 7.3.2.6). Here I implemented the Pan-Tompkins QRS detection algorithm, enabling me to calculate the position of each R-wave in the data (methodology as described in Pan and Tompkins, 1985). As in Section 7.3.2.6 outliers were removed using a median filter and continuous smoothed supersampled HR (in this case at 50Hz) was derived using spline interpolation. HR was analysed within a two-second time window preceding t2e responses, providing a measure of emergence-locked HR, which enabled effects of threat on HR to be explored in the period just before emergence of stimuli into awareness. HR values were baseline-corrected using the mean HR for a one-second epoch preceding CFS trial onset. The resultant baseline-corrected HR measure was termed bHR. Trials in which t2e was shorter than 2 seconds were excluded from analysis. This resulted in exclusion of 8.3% of trials on average across all participants (range 0 - 43.2%). I confirmed that there was no correlation between t2e and HR on an individual trial level (r = -0.08) so one would not expect removal of trials with short t2e to have had any systematic effect on HR. As well as calculating the mean HR across the whole two-second analysis window, temporally-specific effects were explored by implementing the GLM in SPM8 (similarly to the procedure described in Section 7.3.2.6). In the present analysis, regressors constructed for the threatening-face and neutral-face conditions each contained 100 data points (a two-second window sampled at 50Hz). Here statistical inference was based on random-effects analysis and one-sample t-tests were performed at each data point, testing against the null hypothesis of no effect. This allowed me to examine differences between threatening and neutral conditions at any time point within the two-second pre-
t2e period while correcting robustly for multiple statistical comparisons using family-wise error correction based on Gaussian Random Field theory, as implemented in SPM (Friston et al., 1994; see Section 2.2.3). Results were examined at both the cluster and peak levels with a criterion for significance of \( P_{\text{FWE-corr}} < 0.05 \) (uncorrected threshold \( P < 0.05 \), minimum cluster extent = 0). I checked that statistical inference was appropriate at this threshold by running a simulation, which confirmed that the expected Euler characteristic was similar to the estimated number of clusters in the data (Worsley et al., 1992; simulation run using code adapted from http://imaging.mrc-cbu.cam.ac.uk/scripts/randomtalk.m).

Pupil size analysis was performed in a similar fashion to ECG analysis. The pupillary surface area from the right eye had been sampled at 1000Hz; similarly to the procedure for HR, data were extracted from a two-second time window directly preceding t2e responses, baseline corrected, and treated to remove artefacts. The resultant baseline-corrected pupil size measure was termed bpSize. As described for HR analysis, two types of analysis were undertaken, both comparing pupillary area in threatening-face and neutral-face trials: first, mean values were calculated across the whole two-second time window; second, vectors containing all 2000 data points within the time window were submitted to analysis with SPM (using the same analysis steps and thresholds as for the HR analysis).

8.3 Results

8.3.1 Main experiment results

8.3.1.1 Effects of non-conscious face threat on behavioural and motor physiological measures

I calculated group-level mean values and performed repeated-measures ANOVAs with factors of face threat (two levels) and TMS latency (three levels). For the primary behavioural measure, t2e, there was a main effect of face threat: responses to threatening faces were significantly slower than those to neutral faces \( F_{1,15} = 19.5; p < 0.001 \); Figure 8.2A; Figure 8.4A]. This replicates the findings reported in Chapter 6 and in other work (Stewart et al., 2012). The main effect of TMS latency \( F_{2,30} = 0.92; p = 0.38 \) and the face threat*TMS latency interaction \( F_{2,30} = 0.73; p = 0.49 \) were not significant (Figure 8.4A). For the dependent measure of MEP amplitude there was no main effect of face threat \( F_{1,15} = 0.05; p = 0.83 \); Figure 8.2B; Figure 8.4B], no main effect of TMS latency \( F_{2,30} = 1.5; p = 0.28 \], and no face threat*TMS latency interaction \( F_{2,30} = 1.7; p = 0.20 \); Figure 8.4B].
Consistent with previous findings (see Chapter 6), I found substantial inter-individual variability in t2e (Figure 8.2C). Having hypothesized that individual differences in behaviour would be predictive of the neurophysiological manifestations of the freezing response measured using TMS-MEP (see Section 8.1), I defined the effect of face threat at individual participant level by subtracting mean t2e in the neutral-face condition from mean t2e for the threatening-face condition (a similar approach was employed in Chapter 6):

\[
\text{Threat effect t2e} = \text{t2e (threat)} - \text{t2e (neutral)}
\]

\[
\text{Threat effect MEP ampl.} = \text{MEP ampl. (threat)} - \text{MEP ampl. (neutral)}
\]
I found a statistically significant negative correlation between individual threat effect on MEP amplitude and individual threat effect on t2e \( r_{14} = -0.61, p < 0.05 \), Bonferroni corrected; Figure 8.2D.

### 8.3.1.2 Effects of non-conscious face threat on autonomic measures

To examine the autonomic components of the freezing response to social threat, group-level means for the baseline-corrected averages over the two-second time window preceding t2e were calculated for both pupil size and HR (bpSize and bHR; see Section 8.2.3.2). These were submitted to repeated-measures ANOVAs with factors of face threat (two levels) and TMS latency (three levels; as for analysis in Section 8.3.1.1). bHR was significantly slower in threatening-face trials \( F_{1,15} = 6.6; p = 0.02 \); Figure 8.3A; Figure 8.4C. The main effect of TMS latency \( F_{2,30} = 0.13; p = 0.79 \) and the face threat*TMS latency interaction \( F_{2,30} = 0.49; p = 0.62 \) were not significant (Figure 8.4C). For bpSize there was no main effect of face threat \( F_{1,15} = 1.0; p = 0.33 \), no main effect of TMS latency \( F_{2,30} = 1.7; p = 0.20 \), and no face threat*TMS latency interaction \( F_{2,30} = 0.97; p = 0.39 \); Figure 8.3B; Figure 8.4D.

To explore correlations between individual differences in behavioral/physiological and autonomic measures, I calculated the effect of face threat on autonomic measures in an analogous fashion to that for the measures described in Section 8.3.1.1:

\[
\text{Threat effect bHR} = \text{bHR (threat)} - \text{bHR (neutral)}
\]

\[
\text{Threat effect bpSize} = \text{bpSize (threat)} - \text{bpSize (neutral)}
\]

There were no strong or statistically significant correlations between threat effect on t2e and threat effect on bHR \( r_{14} = 0.19, p = 0.48 \); Figure 8.3C or threat effect on bpSize \( r_{14} = 0.22, p = 0.41 \); Figure 8.3D. There were also no correlations between threat effect on MEP amplitude and threat effect on bHR \( r_{14} = -0.02, p = 0.95 \); Figure 8.3E or threat effect on bpSize \( r_{14} = 0.16, p = 0.56 \); Figure 8.3F.
Figure 8.3: Group and individual effects of face threat on autonomic measures
(A) Mean baseline-corrected HR was significantly slower in threatening-face trials. (B) There was no effect of face threat on mean baseline-corrected pupil size. At individual participant level, no correlations were found between behavioural or physiological measures and autonomic indices (between threat effect on t2e and threat effect on bHR (C) or threat effect on bpSize (D)); between threat effect on MEP amplitude and threat effect on bHR (E) or threat effect on bpSize (F)). For simplicity neutral-face and threatening-face results are shown in bar plots collapsed across the three TMS latencies. Corresponding results split by TMS latency are shown in Figure 8.4. Error bars represent one SEM; circles in bar plots as well as scatter plots represent individual participants.

*Denotes a significant difference at level p < 0.05.
Figure 8.4: Group-level effects of face threat and TMS latency on behavioural, physiological and autonomic measures

Plots show mean results for each of the three TMS pulse latencies, split into neural-face (blue) and threatening-face (red) trials. For pupil size and HR results are averaged across a two second time window preceding t2e and corrected to a baseline of the average across one second preceding trial onset. (A) Results for t2e show a significant main effect of face threat (longer t2e for threatening faces) but there is no main effect of TMS latency and no face threat*TMS latency interaction. (B) Results for MEP amplitude show no main effect of face threat or TMS latency and no face threat*TMS latency interaction. (C) Results for HR show a significant main effect of face threat (relative HR slowing in threatening face trials) but there is no main effect of TMS latency and no face threat*TMS latency interaction. (D) Results for pupil size show no main effect of face threat or TMS latency and no face threat*TMS latency interaction. In all plots open circles show individual participants and error bars represent one SEM. The same results collapsed across all three levels of TMS latency are shown in Figure 8.2A-B and Figure 8.3A-B.

Next I explored more temporally-specific effects of face threat during the two-second pre-t2e analysis windows for bHR and bpSize, testing for significant differences between threatening and neutral conditions at all sampled timepoints while correcting for multiple comparisons (see Section 8.2.3.2). For bHR, there was relative bradycardia in threatening-
face trials during a 1.3-second time window directly preceding t2e (T = 2.5; cluster-level P_{\text{FWE-corr}} = 0.008 with 66 data points exceeding the cluster-forming threshold; peak-level P_{\text{FWE-corr}} = 0.06 at the time of t2e; Figure 8.5A). There were no significant effects of face threat on bpSize, although a threat-invariant bpSize increase was seen across the two-second time period (Figure 8.5B). I confirmed a significant difference in mean bpSize between the first and second halves of the time window [collapsing across neutral and threatening trials; t_{13} = -5.0, p < 0.001; Figure 8.5C]. In summary, analysis of time courses during the pre-t2e period showed a threat-related deceleration of bHR that began substantially more than a second prior to t2e and a threat-invariant increase in bpSize prior to t2e.

![Figure 8.5](image)

**Figure 8.5:** Time-varying effects of face threat on HR and pupil size in the period preceding t2e

(A) Time course plot of HR for the two seconds preceding t2e shows relative HR slowing in the threatening-face condition. The period where there is a significant difference between the two conditions is shaded in orange. (B) Time course plot of pupil size for the two seconds preceding t2e shows no difference between the threatening-face and neutral-face conditions. (C) Mean time course of pupil size (collapsed across face threat) shows an increase throughout the two-second analysis window (vertical red line shows split for comparison of early and late pre-t2e periods). Shaded areas around line plots represent one SEM. Bpm, beats per minute; pix, pixels.

### 8.3.2 Auxiliary experiment results

Of the 16 participants included in the main experiment, 12 also performed a TMS-free auxiliary experiment (see Section 8.2.2.1). This enabled me to evaluate whether the observed effects of face threat on t2e, bHR and bpSize were preserved in the absence of the
TMS-MEP component of the main experiment. Since the TMS factor was absent in these analyses, paired-sample $t$-tests were performed to check for differences between neutral and threatening-face conditions. Generally, the effects of non-conscious face threat on both t2e and bHR were replicated, with a trend for prolonged t2e for threatening faces $[t_{11} = -1.7, p = 0.11]$ and significant relative slowing of bHR for threatening faces $[t_{11} = 2.5, p = 0.03]$. Again there was no significant effect of face threat on bpSize $[t_{11} = 0.66, p = 0.52]$. SPM analysis comparing the time courses of bHR for threatening and neutral-face conditions in a two-second time window preceding t2e again closely replicated the main experiment. There was relative slowing of bHR for threatening-face trials in a 1.0 second epoch directly preceding t2e ($T = 4.3$; cluster-level $P_{FWE-corr} = 0.05$ with 50 data points exceeding the cluster-forming threshold; peak-level $P_{FWE-corr} = 0.004$ occurring 400 ms prior to t2e). As in the main experiment, SPM analysis revealed no significant differences in bpSize between neutral-face and threatening-face time courses at any point in the two-second pre-t2e time window. Again there were no correlations between threat effect on t2e and threat effect on bHR $[r_{10} = -0.14, p = 0.66]$ or threat effect on bpSize $[r_{10} = 0.05, p = 0.89]$. Overall, the auxiliary experiment allowed me to confirm that all of the t2e, HR and pupil size effects reported for the main experiment persisted in the absence of TMS-MEP measurement.

Next I checked whether re-analysing the main experiment data but including only the reduced sample of participants who performed the auxiliary experiment ($n = 12$) would produce similar effects as those already reported for the main experiment with the full sample of $n = 16$. I found that the results for the reduced sample closely resembled those for the full sample in all cases: for t2e there was a significant main effect of face threat $[\text{main effect of face threat } F_{1,11} = 12.7; p = 0.004]$; main effect of TMS latency $F_{2,22} = 0.94; p = 0.41$; face threat*TMS latency interaction $F_{2,22} = 0.30; p = 0.75$]; there were no significant effects for MEP amplitude $[\text{main effect of face threat } F_{1,11} = 0.10; p = 0.76]$; main effect of TMS latency $F_{2,22} = 0.57; p = 0.58$; face threat*TMS latency interaction $F_{2,22} = 1.7; p = 0.21$]; for bHR there was a borderline-significant main effect of face threat $[\text{main effect of face threat } F_{1,11} = 4.4; p = 0.06]$; main effect of TMS latency $F_{2,22} = 0.85; p = 0.40$; face threat*TMS latency interaction $F_{2,22} = 0.44; p = 0.65$]; finally, there were no significant effects for bpSize $[\text{main effect of face threat } F_{1,11} = 0.56; p = 0.47]$; main effect of TMS latency $F_{2,22} = 0.68; p = 0.52$; face threat*TMS latency interaction $F_{2,22} = 3.1; p = 0.06$]. For this reduced main experiment sample, the correlation between threat effect on t2e and threat effect of MEP amplitude remained significant and of similar strength and direction $[r_{10} = -0.69, p = 0.01$, although no longer surviving Bonferroni correction$]$ while the correlations between threat effect on t2e and threat effect on bHR $[r_{10} = 0.27, p = 0.40]$ or threat effect on bpSize $[r_{10} = 0.39, p = 0.21]$ remained non-significant. Again, there were no correlations between threat effect on MEP amplitude and threat effect on bHR $[r_{10} = -0.03, p = 0.92]$ or threat effect on bpSize $[r_{10} = -0.40, p = 0.20]$. 
8.3.3 Effects of visual fixation and eye dominance

It is possible that any systematic differences in visual fixation between the threatening-face and neutral-face conditions could have influenced the results (especially for t2e). To test for this I calculated the mean and SEM for horizontal and vertical eye positions, which had been monitored throughout the main experiment using the eye tracker (Section 8.2.2.5). Averaging across all measurements and across the whole sample of participants, there were no threat-dependent differences in mean eye position (mean horizontal eye position for both neutral-face and threatening-face trials was 662 pixels \([t_{15} = -0.20, p = 0.85]\); mean vertical eye position for both neutral-face and threatening face trials was 392 pixels \([t_{15} = 0.01, p = 1.00]\)). I also found no differences in SEM for eye position (SEM for horizontal eye position was 14.8 for neutral-face trials and 14.5 for threatening-face trials \([t_{15} = -0.79, p = 0.45]\); SEM for vertical eye position was 15.0 for neutral-face trials and 14.9 for threatening-face trials \([t_{15} = -0.38, p = 0.71]\)). These results demonstrate that there were no systematic differences in visual fixation between the neutral and threatening conditions so eye movements could not account for the findings. The results of these analyses can also be illustrated using heat map plots (shown in Figure 8.6 for three representative participants).

**Figure 8.6:** Eye position heat map plots
Data are shown for three representative individuals (S09, S10 and S17). Neutral-face trials are shown in the upper row and threatening-face trials are in the lower row. Columns represent individual participants. Each panel shows eye position measurements plotted at multiple time points during all main experiment trials for that condition, with a sampling frequency of 1000Hz. Eye position is measured in pixels along the x and y axes of the display screen (screen resolution 1024x768 with the lower left hand corner of the screen having coordinates 0,0). The colour bar to the right of each plot represents the number of samples recorded at each pixel position.
Given that the CFS mask is consistently high in contrast and dynamic, while the face image is initially shown at zero contrast in the visual paradigm, I checked whether eye dominance (which varied across participants and determined whether the face or CFS stimulus was shown to the right eye; see Section 8.2.2.5) had any influence upon pupil size effects (pupil size was always measured from the right eye). I performed hierarchical repeated-measures ANOVA with bpSize as the dependent variable, participants nested in eye dominance (two levels), and repeated measures on face threat (two levels). There were no significant main effects or interaction [main effect of face threat $F_{(1,14)} = 0.35, p = 0.56$; main effect of eye dominance $F_{(1,14)} = 0.72, p = 0.41$; face threat*eye dominance interaction $F_{(1,14)} = 0.76; p = 0.40$]. I thus concluded that the pupil size effects were not significantly influenced by eye dominance.

8.4 Discussion

The experiments described in this chapter were aimed at further characterising the physiological basis of human defensive freezing responses. I sought to replicate findings already reported in this thesis (Chapter 6), of slower perceptuo-motor performance for socially threatening faces suppressed from awareness. I hypothesised that this effect is a reflection of defensive freezing and to test this hypothesis I probed motor physiological and autonomic changes characteristic of freezing (specifically, corticospinal excitability at precisely defined time points after visual stimulus presentation was assessed by measuring the amplitude of transcranial magnetic stimulation-induced motor-evoked potentials from an inactive peripheral muscle; both heart rate and pupil position/size were also monitored). The relationship of these indices of freezing with the perceptuo-motor t2e effect was also determined. I replicated the slowing of emergence from CFS (Chapter 6; Stewart et al., 2012) and found both a change in corticospinal excitability and a reduction in heart rate in association with non-conscious socially threatening faces, consistent with freezing. Crucially, the reduction of corticospinal excitability varied across individuals, was correlated with perceptuo-motor performance, and was unrelated to the heart rate slowing. Additionally, I found that non-conscious presentation of faces was associated with pupillary dilatation, which was invariant to facial threat. In an auxiliary experiment, performed by a subset of the participants, I replicated the conditions of the main experiment but did not perform TMS-MEP measurement. This enabled me to show that the t2e, HR and pupil effects described could be observed independently of TMS-MEP administration and were therefore not due to non-specific effects of these experimental procedures. Finally, I used the eye tracking measurements obtained to show that visual fixation did not differ significantly between the threatening-face and neutral-face conditions.
It is established that threatening images (e.g. of spiders) or images previously paired with foot shock can trigger non-specific defence responses such as changes in skin conductance even when the images are hidden from awareness (Esteves et al., 1994; Ohman and Soares, 1994). Fully visible socially threatening images (e.g. angry human faces; Roelofs et al. 2010) can elicit freezing-like behaviours in humans. However, the present findings show for the first time that non-conscious social threat is associated with motor physiological and autonomic changes specifically indicative of defensive freezing. Moreover, the results demonstrate a direct link between the motor component of freezing and the established and replicable effect of slowing of time-to-emergence for non-conscious threatening stimuli (Stein and Sterzer, 2012; Stewart et al., 2012; Chapter 6). The biological relevance of this social threat effect on t2e is supported by links to individual differences in local brain anatomy (Chapter 6) and to self-reported personality traits (Stewart et al., 2012). The present results enable a synthesis of this previous work by providing a basic physiological level of description for human defensive behaviors indexed by slowing of time-to-emergence for threatening faces. This allows me to propose a functional biological basis for the slowing of t2e. In addition, the current approach to characterizing human responses to threat could be more broadly applied to understanding other social interactions or the development and maintenance of psychopathology (Sections 9.5.3 and 9.5.4).

In general, it has been suggested that moderately threatening visual stimuli shown in the experimental laboratory environment create especially favourable conditions for the intermediate threat levels and inaccessible means of escape that promote freezing responses (Lang et al., 2000, 2013). Along these lines, I would further speculate that non-conscious stimulus presentation favours freezing since the non-conscious image is more difficult to interpret and freezing is known to be more likely in the context of degraded or ambiguous information (Lang et al., 2000; see also Section 7.4.3). Such a claim would need to be evaluated in future studies that directly compare freezing responses to consciously and non-consciously perceived images.

### 8.4.1 Slowing of t2e for non-conscious threatening faces

Emergence into awareness was slower for non-conscious threatening faces, both with a significant group-level effect but also with substantial inter-individual variability. This is the same pattern of findings as reported in Chapter 6, where an independent sample of participants was studied (see also Stewart et al., 2012). Whether the slowing of t2e for non-conscious threat is entirely accounted for by a slower motor response in the context of reduced CSE is unclear. However, while the t2e effect is observed at the group level, the CSE effect is not. I would therefore hypothesise that the slowing in t2e is underpinned by
at least two components: slowing of perceptual processing and additional slowing of motor processing (related to changes in CSE). The possibility of perceptual freezing is discussed further in Section 9.4.2.

8.4.2 Effect of non-conscious threat on CSE

While there was no group-level effect of non-conscious threat on CSE, I found that threat-dependent CSE changes were closely related to threat-dependent perceptuomotor performance changes (on the CFS task) at the individual participant level. Specifically, slower responders to non-conscious threatening faces also exhibited greater suppression of CSE. I interpreted these findings as showing a close link between individually variable motor physiological (CSE) and perceptuomotor (t2e) components of freezing. As discussed earlier (Section 8.1), inter-individual variability in motor freezing has been demonstrated in a number of previous studies using supraliminal stimuli so the present finding of individually variable changes in CSE for threatening (relative to neutral) faces is in line with existing literature. Together, these findings suggest that the motor components of freezing are deployed to a different extent in different individuals both when the threatening visual stimulus is perceived consciously and when it is perceived non-consciously.

As already mentioned in Sections 1.5.2 and 8.1, immobility responses are diverse and it is not clear whether all of these should be categorised as part of freezing or whether they are better classed as distinct threat-related immobility behaviours (Hagenaars et al., 2014). For example, orienting is a transient and novelty-dependent immobility response, which occurs soon after threat detection and habituates rapidly. Clearly, the ability to demonstrate effects on repeated stimulus presentations (as in the present experiment) would not be in keeping with such a response subtype. Another example is tonic immobility, which is thought to occur as a rather late component of defence when fight, flight or freezing have become of unlikely benefit. In this setting, reduced responsiveness and eye closure have been reported; this is again clearly out of keeping with the present experimental context. There is as yet insufficient understanding of the neurobiology of these different immobility responses to determine confidently whether they represent distinct behaviours or different aspects of one complex and diverse set of defence responses. I would favour the latter interpretation, which would invoke a single integrated system capable of producing the prominent and multifaceted individual variability seen with motor freezing responses and justify the complex neurobiological circuitry described in mammal (and increasingly human) studies (see Sections 1.5.2, 1.5.3, 8.1 and 8.4.6).
TMS was delivered (and resultant MEPs measured) at three different latencies from visual stimulus onset. The purpose of this was to explore whether any change in CSE might be transient, which would be more in keeping with some subtypes of immobility response (Hagenaars et al., 2014). I did not find any effect of the timing of TMS pulses on CSE (or indeed the other behavioural and autonomic measures), which suggests that the recorded variables were stable during the time-window of TMS pulse delivery. However, this null finding does not exclude temporally-specific effects on CSE outside of the measured time window and such a possibility would need to be explored in future experiments. Of note, the latencies to TMS pulse used presently (1.0 – 1.4 seconds) are longer than those in other experiments where latencies of ~500ms have been used (e.g. Borgomaneri et al., 2015). The rationale for this was that the present experiment made use of non-conscious stimuli, initially presented at zero contrast, which at 1.1 seconds had only reached 50% contrast.

### 8.4.3 TMS-MEP as a measure of CSE and motor freezing

As discussed in Section 8.1, TMS-MEP measurement is an established method for probing CSE (Chen, 2000), with reduced MEP amplitudes (or area) interpreted as a correlate of freezing and increased MEP amplitudes (or area) indexing preparedness for action (Avenanti et al., 2009; Borgomaneri et al., 2015). Some studies have shown increased CSE in response to aversive stimuli but this has been in the context of motor response preparation (Coombes et al., 2009), or been specific to the hand involved in responding (van Loon et al., 2010). In addition, these studies used affective stimuli for which negatively valenced exemplars had implied dynamic properties (e.g. an image of a person wielding a knife) while neutral stimuli were static. This may be an important confound, promoting action preparation in the context of negative images (van Loon et al., 2010). Indeed, images showing implied motion are associated with increased CSE (Urgesi et al., 2006). Such circumstances clearly differ to the present experiments where stimuli were balanced for dynamic content (only differing in threatening facial traits) and measurements were taken from an inactive limb (thus being independent of focal preparedness to respond).

Freezing is described by a tense and immobile posture, which can be indexed by reduced body sway, and a number of human studies have employed this measure to probe the motor component of freezing (e.g. Azevedo et al., 2005; Roelofs et al., 2010). However, reduced initiation of movement is another core feature. Indeed, the ventrolateral PAG, the major output centre for motor freezing (see Section 1.5.2), has been characterised as an immobility centre or a brake which temporarily halts the cascading of defence responses
towards more vigorous fight-or-flight behaviours (Hagenaars et al., 2014; Walker and Carrive, 2003). These considerations support the use of CSE reduction as a motor physiological freezing correlate. The TMS-MEP method used here has the advantages (over body sway) of being a more direct measure of activation of the corticospinal system as well as affording high temporal precision (for example, allowing analyses with respect to events such as stimulus presentation or emergence into awareness, and thus helping build a more integrated understanding). Ideally, the measures used in the present experiment would be supplemented with posturographic measurement, although a standing position with body sway during the experiment may influence performance on visual tasks (in particular, dichoptic tasks require stable viewing conditions) and thus necessitate substantial experimental re-design.

Importantly, TMS-MEP provides a composite measure of activation of the cortical, subcortical and spinal motor pathways. While the motor components of freezing are thought to be primarily executed via descending projections from PAG to the motor neurons in the ventral horns of the spinal cord (via cuneiform and caudal raphe nuclei; Vianna and Brandão, 2003), there is also evidence for associated cortical effects (see Section 8.4.6), including in primary motor cortex (Butler et al., 2007; Pereira et al., 2010). Such findings are supported by the existence of ascending anatomical projections from PAG to cortical regions (Vianna and Brandão, 2003; see also Section 1.5.2). Thus a measure such as TMS-MEP that simultaneously probes spinal, subcortical and cortical processes may provide a useful broad overview of the motor effects of freezing. Of course, measures that interrogate individual parts of the motor system may shed further light on the underlying mechanisms. For example, cortical excitability can be measured more specifically and directly through paired-pulse TMS protocols (Chen, 2000). Such measurements have been used to show that there is early modulation of motor cortical excitability when viewing negative body postures, suggestive of suppression of motor readiness (Borgomaneri et al., 2015). These approaches could usefully be applied to further explore responses to non-conscious social threat, although to do so with the present experimental paradigm would represent a considerable additional technical challenge.

8.4.4 Non-conscious threat association with bradycardia

A key finding, seen consistently in both the main and auxiliary experiments, was the group-level effect of facial threat on pre-emergence HR. Relative HR slowing in the threatening-face condition was detected early, at around 1.3 seconds prior to t2e, which supports the assertion that this is an effect seen during non-conscious visual perception. Given the established and consistent association of bradycardia with defensive freezing
SV Getov, Chapter 8. Physiological evidence for freezing

(Hagenaars et al., 2014; Lang et al., 2000; Roelofs et al., 2010; Section 1.5.2), this result provides important support to the interpretation of the t2e findings and the relationship between TMS-MEP and t2e as components of a freezing phenomenon (see Section 8.4.1). The pattern of threat effect on HR (a group-level effect) contrasts with the individually variable effects of threat on TMS-MEP and t2e; moreover, the effects of threat on HR and the effects of threat on TMS-MEP were not correlated. These observations are of relevance for the possible underlying neurobiological mechanisms, discussed further in Section 8.4.6.

8.4.5 Threat-invariant pupillary dilatation

The finding that pupil dilatation preceded emergence into awareness in a similar fashion for both threatening and non-threatening faces was not entirely consistent with my predictions. Novel attention-grabbing stimuli are associated with pupillary enlargement (Libby et al., 1973) and the fact that I observed such pupil changes prior to reported emergence of stimuli into awareness provides additional evidence (converging with the HR findings) that the stimuli provoked an autonomic response while still non-conscious. It has also been proposed that pupillary dilatation preceding perceptual switches in binocular rivalry (Einhäuser et al., 2008), or for ambiguous moving stimuli (Hupé et al., 2009) is linked to the impending change in perception. While such dilatation also occurs with replayed perceptual changes that are exogenously generated, and has been shown to be largely accounted for by motor responses during bistable tasks (Hupé et al., 2009), there are also components of this response that seem to depend on the perceptual properties of the stimuli involved (Kloosterman et al., 2015). It is thus quite possible that the pupillary dilatation seen presently is reflective of an impending change in perception. Increased pupillary dilatation for unpleasant images (compared to neutral images) has been reported (Bradley et al., 2008) but was not demonstrated in the present results. This may reflect different pupil responses to non-consciously and consciously perceived stimuli, or (perhaps more likely) could be due to reduced ability to detect pupillary changes during CFS, which is a bright and dynamic stimulus that likely itself elicits a robust pupillary response.

8.4.6 Proposed neurobiological basis of the effects on CSE and autonomic indices

The present findings demonstrate a link between slowing of responses to non-conscious threatening stimuli and defensive freezing. Furthermore, they highlight specific patterns of effect for the different components of the observed freezing responses. It is noteworthy that the effects of threat on t2e and HR were observed at the group level, while CSE was
correlated with t2e when assessed across individuals. This difference is likely reflective of the underlying neurobiological basis of defensive freezing responses to social threat. Specifically, as discussed in Sections 1.5.2 and 8.1, the autonomic outputs of freezing depend on direct anatomical projections from CeA to brainstem nuclei and likely operate on a low hierarchical level. In line with this relative simplicity of anatomical and functional underpinnings, these responses exhibit limited inter-individual variability. In contrast, the motor components of freezing are diverse and can be fractionated into several subtypes, linked to differential involvement of PAG subregions and associated with both descending interactions with spinal cord and ascending ones with cortical regions (Vianna and Brandão, 2003; Brandão et al., 2008; see also Section 1.5.2). This behavioural and functional complexity appears to be reflected in inter-individual variability in the motor outputs of freezing both in the present findings and in previous publications (Section 8.1). My demonstration of a group-level effect of non-conscious threat on HR and an individually variable effect on CSE is thus consistent with different underlying mechanisms for these components of freezing, and supported by the existing literature. Moreover, the effects of threat on HR and CSE were not correlated across individuals in the presently reported data, suggesting that autonomic and motor components of freezing may be dissociable in humans, as they are in rodents (Section 1.5.2; LeDoux et al., 1988). A seemingly contrasting piece of evidence is that BOLD-fMRI activation in PAG is correlated with bradycardia during human freezing (Hermans et al., 2013), although how such activation relates to motor components of freezing is unclear. Finally, the presently reported t2e result shows both a group-level effect and a correlation with CSE at individual level. This may reflect the dual perceptual (emergence into awareness) and motor (button response) components of the t2e task. Future work will need to address whether there is freezing of perception in addition to motor freezing in the context of non-conscious threat.

8.4.7 Summary

In this chapter, I studied the presence and interaction of perceptuomotor, motor physiological and autonomic correlates of defensive freezing in healthy human observers when viewing socially threatening faces suppressed from awareness. My experimental approach combined visual psychophysics, TMS-MEP recordings and psychophysiological measurements. In the main experiment, HR and pupil size were continuously monitored while participants performed a CFS visual task; CSE was measured using TMS-MEP from the non-dominant left hand. As predicted, and replicating previous findings, non-conscious threatening faces emerged into awareness more slowly than neutral faces and the degree of this threat-related slowing was variable across individuals. Non-conscious threatening faces were also associated with bradycardia, and this relative slowing of HR for threat
began more than a second before participants reported perceiving faces consciously. Moreover, there was a close relationship between individual performance on the perceptuo-motor CFS task and individual measures of corticospinal excitability: slower responders to non-conscious threatening faces exhibited greater suppression of CSE in threatening-face trials. Autonomic and TMS-MEP measures were not correlated. Finally, non-conscious faces were associated with threat-independent pupillary dilatation, likely related to novelty/arousal or reflecting an impending change in perception. An auxiliary experiment served to confirm that the heart rate, pupillary and behavioural effects were conserved in the absence of the TMS-MEP component of the paradigm.

Overall, the findings characterise responses to non-conscious social threat, outlining an autonomic (HR) component that is detectable early, is relatively invariant across individuals, and is unrelated to perceptuo-motor responses reflected by t2e and TMS-MEP. In addition, there is evidence for a motor component that is conversely highly variable across individuals and is closely related to perceptuo-motor performance as indexed by the t2e measure. This set of findings provides, for the first time, converging evidence for a freezing response to non-conscious social threat in humans. The “non-conscious” and “social” aspects of the results go beyond what is known about freezing responses from animal models. In addition, the findings of different patterns of threat effect on corticospinal excitability, heart rate and perceptuo-motor performance suggest that at least partly dissociable mechanisms underpin these components of freezing in humans; this is supported by existing literature from animal models. Whether distinct components of perceptual and motor freezing can be shown to underlie different aspects of threat-related slowing (as shown on the perceptuo-motor t2e task) is a question that should be addressed in a future experiment.
Chapter 9

General discussion

‘It’s not what you look at that matters, it’s what you see’

Henry David Thoreau (1817 – 1862)

9.1 Overview

In carrying out the experimental work contained in this thesis I have attempted to shed new light on the behavioural, psychophysiological and neuronal signatures of non-conscious and conscious visual perception. A major focus has been to advance understanding of the transition between non-conscious and conscious processing for highly ambiguous visual information (where there are spontaneous switches between these modes of processing). Although such profound ambiguity is not always found in natural visual environments, vision is nevertheless a highly inferential process: to shape perception, retinal input is subject to a multitude of complex computations which occur in parallel and hierarchically organised processing streams, but ultimately in a highly integrated fashion. The relevance of studying mechanisms of perceptual transitions is to provide a template for understanding how the brain selects and gates a subset of the (highly processed) data handled by the visual system for conscious awareness. The second major theme of the thesis has been the social modulation of visual processing (in particular non-conscious processes) and visual awareness. From an evolutionary perspective the survival value of effective interpretation of social signals has been understood for some time (e.g. Darwin, 1872). In modern human culture such signals remain a central part of successful interaction with others. Some theories posit that the need for complex social cognition is an important reason why humans require consciousness in the first place (Frith, 2010). Thus, visual
social cues would be expected to have an important impact on awareness and moreover can act as an example for how visual processing and awareness can be modulated by different types of environmental (bottom up) signals. A special case of this is visual threat, which can elicit defensive responses that have neural, behavioural and autonomic correlates. Such responses are vital for survival in animals and many of them are conserved in humans where they play an important role in social interaction and the development of psychopathology (Hagenaars et al., 2014; Lang et al., 2000). In summary, the work in this thesis has sought to contribute to our understanding of neural mechanisms of visual awareness, specifically by studying the perceptual transition network as well as by exploring non-conscious processing of social cues and shedding new light on the neural, behavioural and psychophysiological mechanisms that are engaged when non-conscious social cues are threatening.

This final chapter will build on discussion from the individual experimental chapters. I will attempt a synthesis of the experimental findings that relate to each of the main themes explored in the thesis, placing them in the broader context of key existing theories and outstanding questions in the field, and outlining logical future hypotheses and research directions that arise from them.

9.2 Neural mechanisms of perceptual transition

9.2.1 Discussion and context of findings from Chapters 3–5

The experimental results described in Chapter 3 extend the findings of several previous studies that have used fMRI to delineate the group of brain regions where activity is time-locked to perceptual transitions induced by bistable visual stimuli (e.g. Kleinschmidt et al., 1998; Lumer et al., 1998; see also Section 1.3.5). My findings suggest that there is a core group of regions involved in perceptual transitions for very different visual stimuli and paradigms (specifically, binocular rivalry between faces and gratings as well as SFM constructed from moving dot stimuli; see Section 3.1). Although there are some stimulus- and paradigm-specific differences in the associated perceptual-transition-related patterns of BOLD signal (for example, there is prominent activation in V5 for the SFM stimulus), an activation pattern that is common across both paradigms occurs in several regions in frontal and parietal cortex. This stimulus-invariant perceptual-transition-related activity is
seen in aSPL, FEF, PMd, and insula and it provides an anatomically specific set of targets for further study of mechanisms of perceptual transition.

In Chapter 4 I asked a new set of questions about the functional interactions among regions in the perceptual transition network outlined in Chapter 3, exploring the effective connectivity between them using DCM. Unlike many other experimental tasks, bistable perception is a spontaneous phenomenon, and the participant’s objective is to provide ongoing report of their changing perceptual experience. Therefore, I chose to focus on spontaneous fluctuations in brain activity (in the resting state), specifically in relation to the possible existence of intrinsic directional effective connectivity between the four regions highlighted in Chapter 3. I adopted an individual differences approach, asking whether individual dynamics of bistable perception were correlated with effective connectivity strength among particular regions in the perceptual transition network. I found that such a relationship existed specifically with directional connectivity from insula to aSPL. In other words, the strength of the intrinsic effective connection between these two regions was predictive of how flexibly (or how often) an individual switched between alternative perceptual interpretations. The implication of a specific directional interaction within intrinsic effective connectivity patterns for determining the dynamics of bistable perception is novel; moreover, the finding fits with proposed roles of inferior frontal/insular regions in initiating perceptual reversals (Sterzer et al., 2009) and sheds new light on current models of perceptual transition mechanisms (see Section 9.2.2 and Figure 9.1).

A logical next step following the findings from Chapter 4 could be to explore the functional role of right insula (and overlying operculum) in bistable perception, perhaps by disrupting function with rTMS over this area (of note the functional role of right SPL has already been evaluated in this way; see Section 3.1 and discussion below). However, this is likely to be technically challenging (Section 5.1). I therefore targeted FEF instead, this being more accessible to TMS and another region reliably containing stimulus-invariant perceptual-transition-related activity, as outlined by the results presented in Chapter 3. The rationale was that focusing on this region, besides helping to understand the perceptual-transition network, would allow me to refine an rTMS paradigm and technique before moving onto more challenging stimulation in insula/IFG in future experiments. Unfortunately, I was neither able to demonstrate a robust effect on bistable perception dynamics following rTMS over right FEF, nor to replicate findings for rTMS over right pSPL (Kanai et al., 2011). Possible reasons for these null results are discussed in Sections 5.4.1 and 5.4.2.
Proposing extended functional networks contributing to perceptual transition

There are established experimental findings and theories relevant to some parts of the perceptual-transition-related network outlined in Chapters 3-5. In particular, the role of superior parietal cortex in bistable perception has been closely investigated using a combination of individual variability approaches (linking variability in behaviour with focal brain structure) and rTMS (Carmel et al., 2010; Kanai et al., 2011, 2010; Zaretskaya et al., 2010). Taken together, the findings from these studies, which are reviewed elsewhere in this thesis (Sections 1.3.5, and 3.1) support functional fractionation of right SPL into an anterior subregion (aSPL) where GM volume correlates with more prolonged or stable percepts in bistability and a posterior subregion (pSPL) where GM volume correlates with shorter or more unstable percepts (though pSPL findings may apply specifically to the SFM paradigm). Temporarily disrupting the function of these regions with rTMS has the expected opposite effects of reducing and increasing mean percept duration, respectively.

An attempt has been made to interpret these opposite roles in bistable perception for the aSPL and pSPL subregions from the perspective of Bayesian network theory and predictive coding (Kanai et al., 2011). In this sort of framework it is postulated that feedback connections from higher-order areas send predictions about the likely current state of the external environment, while feed-forward connections from lower-order areas carry error signals representing the difference between these predictions and the actual sensory input (e.g. Rao and Ballard, 1999). This prediction error can then be used to continually refine future predictions. Kanai and colleagues (2011) argue that such a framework can usefully be applied to their findings in superior parietal cortex. Thus aSPL could be involved in prediction generation (with influence from insular/inferior frontal regions as suggested by the findings in Chapters 3 and 4) and thus support the current perceptual interpretation in a bistable context (disrupting the function of aSPL would correspondingly lead to weaker predictions and increased chance of perceptual switching). Conversely, pSPL and associated networks could be involved in transmitting prediction error signals (guided by input from sensory areas) and thus challenge the prediction signals generated via aSPL and increase the probability of perceptual switching in the context of bistability (disrupting the function of this region would therefore impair prediction error signals allowing the current perceptual interpretation to remain and reducing the chance of a perceptual switch; see Figure 9.1A).

Effective connectivity (DCM) analysis of fMRI data has more recently been performed to shed new light on the information flow between aSPL and pSPL during bistable perception (Megumi et al., 2015). The authors showed that with a moving-dot SFM paradigm, there
were reciprocal interactions between V5, pSPL and aSPL, and behavioural perceptual dynamics were predicted by bottom-up coupling between these regions. However, another fMRI study combining bistable perception and DCM approaches has instead found that top-down models, where inferior frontal regions modulate the activity of visual areas, were more predictive of the observed perceptual-transition related BOLD signal changes (Weinhammer et al., 2013). The presently reported findings from Chapter 4 highlight resting-state effective connectivity from insula (a region in close proximity to the inferior frontal regions previously implicated in top-down interactions) to aSPL. It is also important to bear in mind that the findings in Chapter 4 apply specifically to effective connectivity in the resting state that predicts the individual dynamics of binocular rivalry, and are not dependent on stimulus-related (bottom-up) input. These findings thus shed light on the intrinsic rather than task-related architecture of perceptual transition mechanisms. Nevertheless, the resting and task-related mechanisms may well overlap (see Section 4.4.2).

My demonstration of effective resting-state connectivity from insula to aSPL could be incorporated into the framework proposed by Kanai et al (2011), highlighting regions involved in generating predictions further upstream from aSPL (see Section 4.4.2.1 and Figure 9.1B, red colours). The result is broadly in keeping with the task-related findings of Weinhammer and colleagues (2013) but is seemingly discrepant with the task-related findings of Megumi and colleagues (2015). On the other hand, taking all these results together might suggest that both higher-order (frontal/insular) regions and earlier (visual) regions combine to influence the dynamics of bistable perception. This is broadly in line with popular accounts of the mechanisms underlying bistable perception (Sterzer et al., 2009), the proposed predictive coding framework, and various demonstrations of both top-down and bottom-up influences on bistability (see Sections 1.3.3 and 1.3.4). The balance between the contributions of higher and lower-level mechanisms may differ for different forms of bistability, as discussed in Section 4.4.2.2. While the causal role of insula in the dynamics of bistable perception needs to be established, and I was not able to prove a causal role for FEF (Chapter 5), there is substantial existing evidence for involvement of these regions in mediating access to visual awareness (Sections 3.4.3 and 5.1). Finally, it is important to bear in mind that the details of mechanisms discussed are based on a combination of findings relating to structural brain correlates of individual differences in behaviour, disruption of function with rTMS, and correlational fMRI approaches (at times combined with effective connectivity analyses). Recent work suggests that functional measures of large-scale brain dynamics (based on attractor-network models) can link individual differences in behaviour with individual differences in brain structure, supporting a set of unifying mechanisms for functional, structural and behavioural correlates of bistable perception (Watanabe et al., 2014). Attractor networks may represent
an alternative model for understanding bistable perception to the predictive coding framework discussed above. There are yet more approaches to the neural mechanisms of bistable perception, such as understanding them through more attention-inspired bottom-up/top-down frameworks, as per interpretations of some of the studies discussed above; see also next paragraph. Another perspective on perceptual transition relates to possible engagement of mechanisms for spatial attention and cognitive control, the neural correlates of which overlap closely with the frontoparietal perceptual-transition-related networks already described (Corbetta and Shulman, 2002; Zanto and Gazzaley, 2013; see Sections 1.3.5 and 3.4.5).

While attention and consciousness are undoubtedly very closely related, there is experimental evidence that attentional selection can occur without conscious experience and conversely that it is possible to be conscious of information that is outside the focus of attention (Tononi and Koch, 2008). Selective attention can bias bistable perception for ambiguous figures (increasing the chances of perceiving the attended or expected percept; Toppino, 2003) but not for binocular rivalry (Meng and Tong, 2004). On the other hand, complete inattention seems to abolish binocular rivalry altogether while having less effect on other forms of bistability (Dieter et al., 2016). Expectation of certain predominance in directionality of structure-from-motion stimuli can also influence perception (Sterzer et al., 2008). Thus both exogenous bottom-up capture of attention and endogenous top-down factors may play roles in bistable perception. There are striking similarities in frontoparietal BOLD signal patterns associated with attentional processes and shifts in awareness; it has been argued that these patterns cannot simply reflect shifts in spatial attention during binocular rivalry since perceptual transitions during these paradigms do not have a spatial component and are not subject to top-down influences (Lumer et al., 1998). On the other hand, work with bistable figures describes fMRI evidence suggesting that voluntary shifts in perception are mediated by spatial attention (Slotnick and Yantis, 2005). Similarly, for structure-from-motion stimuli, disruption of dlPFC function with rTMS affects voluntary control over bistability but not spontaneous bistability (de Graaf et al., 2011). Overall, it seems that attention is very important in perceptual transitions during bistable paradigms but is nevertheless not the only mechanism driving these phenomena or the one solely responsible for transition-related frontoparietal BOLD signal.

My results have highlighted one potentially important interaction in the perceptual-transition-related system (the effective connection from insula to aSPL) but the interactions between other regions in the network remain unclear. For example, it is uncertain how FEF may link, in both anatomical and functional terms, with the aSPL/pSPL system and the newly proposed functional connectivity from insula. It is known that FEF is widely connected anatomically (including to parietal regions) and some of its functional properties
also suggest a role in balancing top-down and bottom-up influences, including proposed roles in spatial attention (through saliency coding) and in modulating early visual representations (see Section 5.1). Given these anatomical and functional underpinnings, and in particular the clearly demonstrated modulatory effects on early visual areas (Ruff et al., 2006), FEF may thus act to regulate the changes in visual representation strength that determine whether a particular percept is dominant or suppressed during bistability (see also Section 5.1), under influence from bottom-up and top-down factors and guided by interactions with the aSPL/pSPL system. This proposal is in line with others made previously, that both FEF and SPL are prime candidates for controlling the feedback of ‘higher-order’ areas onto ‘lower’ visual areas, providing a sensory gating mechanism that could select information for conscious appraisal (Grosbras and Paus, 2003). Therefore, I would propose that alongside the aSPL/pSPL system, the mechanisms of perceptual transition depend also on an FEF-centred system that interacts with but is also to a degree independent from the aSPL/pSPL system (see Figure 9.1B for a putative sketch of a more extended network including both these systems).
Several studies using TMS and bistable perception (Carmel et al., 2010; Kanai et al., 2011, 2010) have revealed functional roles for subregions of SPL; this has led to development of a theoretical framework inspired by predictive coding theory. (B) Findings within this thesis have revealed a functional interaction between insula and aSPL (red; Chapter 4) as well as proposed involvement of FEF (purple; Chapter 3) and PMd (grey; Chapter 3) that require further clarification. The dotted lines represent putative connections (in many cases anatomical connectivity exists but whether this is relevant in bistable perception remains unclear). Figure adapted from Kanai et al., 2011.

It is possible that the roles of FEF and SPL in bistable perception depend to different degrees on the top-down and bottom-up influences already discussed in this section. In the field of visual attention, the differing contributions of parietal regions (LIP) and frontal
regions (FEF) have been explored with invasive recordings in the macaque (Buschman and Miller, 2007), showing that frontal contributions to a visual pop-out/visual search task reflect top-down (endogenous) aspects, whereas parietal activity is more reflective of bottom-up (exogenous aspects). A subsequent study in humans using combined TMS and fMRI (Ruff et al., 2008) similarly showed that while TMS to parietal regions (IPS) affected visual cortex in a manner dependent on the current visual stimulus, TMS to frontal regions (FEF) affected visual signal in a stimulus-independent fashion. Again, this pattern of findings was interpreted as showing primarily bottom-up contributions from parietal regions and top-down contributions from frontal regions. A series of future studies, perhaps along similar lines, would be required to clarify the roles and interactions relevant to FEF, SPL, insula/IFG and other regions in the networks proposed in Figure 9.1, and no doubt to create more accurate and refined versions of these putative networks. Avenues for future work in this area are discussed in Section 9.5.1.

9.2.3 Significance of transition-related frontoparietal BOLD signals

A central debate surrounding the now reliably reported patterns of inferior frontal and superior parietal BOLD signal time locked to perceptual transitions (see Section 1.3.5 for a summary of the relevant literature) concerns the exact role of the underlying neural mechanisms in changes between alternative perceptual experiences. More specifically, a key question is whether the observed transition-related activity reflects the initiation of changes in the contents of awareness, represents an epiphenomenon (or consequence) of shifts in perceptual experience, or is an artefact of poorly balanced perceptual transitions across perceptually ambiguous and non-ambiguous (control) conditions (see Section 3.4.5 for a more detailed discussion of this debate and the key previous experimental findings). Results reported in Chapters 3 and 7 contribute to this debate by showing that despite equating the length of perceptual transitions across binocular rivalry or SFM and a control replay condition, there is nevertheless enhanced frontoparietal BOLD signal in association with endogenous perceptual transitions. These findings contradict reports suggesting that transition-related frontoparietal BOLD signal reflects between-condition differences in perceptual transition duration (e.g. Knapen et al., 2011) and are supported by independent results obtained using other bistable visual paradigms with instantaneous perceptual transitions (Weilhhammer et al., 2013; see also Section 3.4.5). The results reported in Chapters 3 and 7 of this thesis are not sufficient to establish a direct causal role of perceptual-transition-associated BOLD signal in changing the contents of awareness, but
they nevertheless support a role for such activity that goes beyond mere responsiveness to gradual changes in perceptual experience, or to prolonged mixed percepts.

A more recent study has taken an alternative and innovative approach to probing the role of transition-related neural activity, by designing a binocular rivalry stimulus in which switches between monocular inputs are so subtle as to be unnoticed by the observer (Brascamp et al., 2015). When such unreportable switches are examined using fMRI, there is apparently no detectable associated frontoparietal activity. The authors' interpretation of this finding is that switches between two alternative monocular inputs can occur independently from frontoparietal mechanisms. However, it is highly debatable whether non-experienced changes in stimulus processing are the same (or as relevant to understanding access to awareness) as changes in conscious perception. Indeed, the evidence of Brascamp and colleagues (2015) does not directly argue against the importance of frontoparietal regions in (access to) visual awareness and neither does it rule out the possibility that consciously perceived switches could be driven by frontoparietal regions (perhaps by feeding back to lower visual regions). All the evidence cited in this section (including the results presented in Chapters 3 and 7) suggests that when the visual inputs from binocular rivalry (or other forms of bistability) are experienced as conscious percepts, the switching between these percepts is reliably associated with activity in frontoparietal regions. That this activity should disappear when the changes in perception from bistable stimuli are rendered unconscious in fact provides further support for the role of frontoparietal regions in conscious experience (including changes between alternative conscious experiences). A related but more nuanced explanation for findings like those of Brascamp and colleagues is that switches between alternative perceptual interpretations during bistable perception can be brought about by varying contributions of higher and lower level processes. The balance of these contributions could differ according to whether the changes in perceptual interpretation are experienced consciously or not and this changing balance may thus be reflected in the degree of frontoparietal activation during perceptual transition.

## 9.3 Social modulation of visual awareness

Experiments reported in the second half of this thesis made use of faces with varying social traits from the model of Oosterhof and Todorov (2008; see also Section 1.4.2) where the principal dimensions of social evaluation are dominance (explored in Chapters 6 and 7) and trustworthiness (explored in Chapter 6), although the main diagonal of the model also indexes threat (explored in Chapter 8). Since both dominant and untrustworthy faces are
also rated as threatening (Oosterhof and Todorov, 2008; see also Section 1.4.2), findings from experiments evaluating perception of these traits are also discussed in terms of what they might tell us about threat perception and the effects of threat on neural processing and behaviour. There are thus two alternative ways in which the results from these experiments can be interpreted: as reflective of specific processes for evaluation of the particular social trait in question, or with reference to more general processes for evaluating threatening or non-threatening stimuli as a whole. Since gender differences in social evaluation are an important factor to consider (see Section 1.4.3), gender was included as a covariate in all statistical analyses in Chapters 6-8 but on no occasion was this factor found to be making a significant contribution to the observed effects.

9.3.1 Effects of non-conscious social face traits on behavioural responses

In Chapter 6 I presented behavioural findings demonstrating that dominant and untrustworthy (or in both cases socially threatening) faces show delayed emergence from interocular suppression, which could be interpreted as slowing of non-conscious processing. This effect was opposite to that predicted on the basis of previous findings for emotional faces paired with CFS (e.g. Yang et al., 2007) and according to evolutionary vigilance accounts, which predict prioritisation of non-conscious emotionally relevant cues (e.g. Sander et al., 2003; see also Sections 1.4.4, 6.1 and 6.5.1). However, the findings from Chapter 6 have been independently replicated (Chapter 8; Stewart et al., 2012). The delayed emergence from interocular suppression is specific to non-conscious faces (abolished when faces are presented in full awareness), indicating that it does not reflect processing that occurs during conscious perception (i.e. after emergence of images from CFS). Furthermore, the degree of response slowing specifically for non-conscious dominant or untrustworthy faces is related meaningfully to individual differences in self-rated personality, being correlated with self-rated submissiveness scores (SBS questionnaire) and self-rated propensity-to-trust scores (PTS questionnaire), respectively (Stewart et al., 2012; these questionnaires are briefly reviewed in Section 1.6). These links to character and dispositions indicate that the observed effects of non-conscious social face traits are unlikely to reflect solely low-level image differences (even though in theory some low-level stimulus features can ‘leak’ through CFS; Yang and Blake, 2012).

As briefly discussed in Chapter 6 (Section 6.5.1), an important distinction between previous findings showing faster emergence into awareness (e.g. Yang et al., 2007) and the results from Chapter 6 showing slower emergence into awareness relates to differences in the
emotional and social content of the stimuli used. As already covered in Sections 1.4.1 and 1.4.2, previous studies showing faster t2e for non-conscious emotional faces have exclusively used fearful-expression stimuli and it is not clear to what degree such findings should generalise to other facial emotions or social cues. Dominance and trustworthiness judgments are correlated with those for anger, but not with those for fear (Said et al., 2011). Non-conscious angry faces are associated with slowing of emergence from suppression, as for dominant and untrustworthy ones (Stein and Sterzer, 2012). While dominant and untrustworthy faces are a direct signal of threat, fearful faces are not in themselves threatening but rather signal the presence of threat in the environment (Adams & Kleck, 2003; Stewart et al., 2012). Such contrasts are supported by evidence for opposite response tendencies associated with these facial expressions, with approach being faster for fearful faces and withdrawal being faster for angry faces (Marsh et al., 2005). Moreover, evidence from MEG shows even more directly that there are differences in the temporal evolution of processing with fearful faces being associated with much earlier event-related gamma band synchronisation in thalamus and amygdala than for angry faces (Luo et al., 2007). It therefore seems likely that these different types of social cue have opposite effects on non-conscious perception due to differential engagement of threat-processing mechanisms. Such a difference can be readily explained within the framework of defensive responses, which include freezing and fight/flight behaviours. The proposal here could be that threatening (angry, dominant, untrustworthy) faces are more likely to elicit a freezing response whereas fearful faces are more likely to elicit a fight/flight response. Accordingly, a further prediction would be that alongside the demonstrated opposite effects on speed of non-conscious processing there should also be opposite effects of fearful and threatening faces on corticospinal excitability, with the former associated with increased CSE and the latter with decreased CSE. This is indeed the case, as shown previously for fearful faces (Schutter et al., 2008) and as I have shown for threatening faces in Chapter 8 (see Sections 1.5 and 8.1 as well as Stewart et al., 2012 for further relevant discussion). The proposal that defensive responses may explain the delayed non-conscious processing observed in Chapter 6 led to the formulation of new hypotheses regarding engagement of freezing mechanisms by non-conscious social threat, which were directly explored in Chapters 7 and 8. The findings from these experiments and their implications are discussed in Section 9.4.
9.3.2 Structural and functional neural correlates of social visual cues: non-conscious processing and access to awareness

In the behavioural experiment presented in Chapter 6 I found that the delayed emergence into awareness for non-conscious dominant and untrustworthy faces was highly variable across individuals. I then performed a structural imaging experiment and showed that this variability was correlated with local gray matter volume in frontal operculum (for dominance) and in right TPJ, bilateral fusiform gyrus and mPFC (for untrustworthiness). The importance of these regions to social perception has been discussed elsewhere (see Sections 6.5.2 and 6.5.3), while the interpretation of the results in relation to dual competition and defensive cascade frameworks is discussed further in Section 9.4. It appears from the findings in Chapter 6 that modulation of non-conscious perception by threatening visual cues relates to the structure of high-order cortical regions involved in complex social appraisal. This might suggest that social evaluation (for example specific evaluation of dominance and/or trustworthiness) can occur without awareness, although this needs to be tested more directly with functional imaging and/or neuromodulatory (e.g. TMS) approaches. The relevance of individual variability in these neural underpinnings for non-conscious threat processing also requires further elucidation: in particular, one might ask whether this variability relates to personality traits or life experiences, and whether the situation can be modified through training. At present the findings allow speculative suggestions that appraisal of non-conscious socially threatening cues may engage complex cortical systems for social evaluation centred on mPFC, TPJ and insula/IFG (reviewed in Sections 6.5.2 and 6.5.3). This would imply a broader account than one focusing on subcortical mechanisms for promotion of survival in the context of threat (as emphasised by evolutionary vigilance and dual route accounts; Section 1.4.4), although it is already known that mPFC is involved in the production of defensive freezing behaviours (see Section 1.5.3 and further discussion in Section 9.4).

In the fMRI experiment described in Chapter 7 I moved focus from non-conscious processing of social cues to the effect of social cues on perceptual transitions. As already discussed in Section 9.2, exploring the mechanisms of perceptual transition (for socially neutral stimuli) was the major theme of the first half of this thesis. Whether such transitions, and the well established associated BOLD signal changes (see Section 9.2) are modulated by social cues was a key question for the experimental work in Chapter 7. I found that the dynamics of binocular rivalry were altered in the context of social face dominance (slower mean BR switch rate in a dominant-face context compared to a neutral-face context). However, this effect was not replicated during fMRI scanning and the
possible reasons for this were discussed in Section 7.4.1. Further results obtained from this fMRI experiment revealed relative reduction in heart rate and relative reduction in BOLD signal in right amygdala and right pSTS in association with perceptual transitions in a dominant-face context. These findings suggest that the neural underpinnings of perceptual transition are modulated by the presence of social facial cues, and this occurs with engagement of regions implicated in social face evaluation. I propose that the relative signal reduction in amygdala and pSTS in association with perceptual transitions in a dominant-face context relates to the engagement of defensive freezing mechanisms and this is discussed further in Section 9.4.

9.4 Defensive freezing in humans

One of the key findings reported in Chapter 6, namely the delayed behavioural response to non-conscious dominant and untrustworthy faces, was opposite to the pattern of findings I had predicted. The implications of this unexpected result, in the context of previously reported findings related to non-conscious processing of social cues, have been discussed in the previous section (Section 9.3). I proposed that these findings (presented in Chapter 6) might reflect the consequences of a defensive fear (or freezing) response on perceptual processing. Although freezing has traditionally been described in smaller mammals (in particular the rat), it appears to be a conserved pattern of responses also observed in humans and of relevance to human social interaction and psychopathology (see Section 1.5). A major aim for the experiments presented in Chapters 7 and 8 was therefore to test predictions and characterise the mechanisms relevant to the hypothesis that non-conscious social dominance and untrustworthiness cues from faces elicit human defensive freezing.

9.4.1 Placing findings in the context of evolutionary vigilance, dual competition and defence cascade frameworks

In Chapter 6, after demonstrating slowing of emergence from interocular suppression for dominant and untrustworthy faces, I explored the focal structural brain correlates of inter-individual variability in this slowing using VBM. I found that dominance-related slowing was negatively correlated with GM volume in right frontal operculum while untrustworthiness-related slowing was negatively correlated with GM volume in right TPJ and bilateral fusiform gyrus as well as positively correlated with GM volume in
mPFC. These findings can be interpreted with reference to different existing theoretical frameworks, as mentioned in Sections 6.5.1 – 6.5.3, and as discussed in more detail presently.

As reviewed elsewhere (Sections 1.4.4, 6.1 and 9.3.1) emotional stimuli are frequently associated with enhanced (or prioritised) perceptual processing; this observation has been used to inform evolutionary vigilance models (e.g. Sander et al., 2003). Such accounts do not readily offer an explanation for the threat-related slowing of non-conscious processing reported in Chapter 6. Moreover, there are previous demonstrations in the literature that emotional stimuli can be associated with deleterious effects on task performance, especially when the emotional content is strong and thus distracts from the task (e.g. Verbruggen and Houwer, 2007). This effect has been shown with perceptual tasks in the visual domain (e.g. Erthal et al., 2005). The dual competition model proposes an explanation for this phenomenon, suggesting that stimuli with intense negative emotional qualities (e.g. significantly threatening stimuli) automatically engage attention and thus recruit enhanced perceptual and executive processing in a way that uses up ‘common pool’ resources and can therefore impair irrelevant task performance (Huang et al., 2011; Pessoa, 2009). This account is largely informed by mechanisms of attention but also involves the integration of affective and perceptual information to guide action. A pivotal structure in these mechanisms is proposed to be the anterior cingulate cortex. In the monkey this region receives convergent input from limbic regions (Morecraft and Van Hoesen, 1998) as well as projecting to motor cortex (Morecraft and Van Hoesen, 1992) and spinal cord (Morecraft et al., 1997); it has also been cast as a premotor area in humans (Koski and Paus, 2000). BOLD activity in ACC tracks the effect of aversive stimuli on motor behaviour (Pereira et al., 2010; Pessoa, 2009). A number of other regions including anterior insula (Pessoa, 2009) and FFA (Anderson et al., 2003) also seem to be involved in dual competition mechanisms; stronger activation in these regions can be associated with delayed responses on an unrelated task in the presence of negative emotional stimuli. This implies that attentional effects of threat, which use up limited resources and impair performance, are fairly widespread. Of note, such findings potentially dovetail with the VBM findings reported in Chapter 6, where reduced GM volume in frontal operculum (adjacent to insula) and FFA was correlated with slowing of t2e for dominant and untrustworthy faces, respectively. The threat conveyed by these non-conscious faces may have engaged limited processing resources located in these regions, and the effect on behavioural performance may logically be more profound in cases where these regions already have a relatively lower GM volume (Sections 6.5.2 and 6.5.3).

Another explanation proposed for the VBM findings in Chapter 6 could be that they reflect underutilisation of certain perceptual processes (which depend on insula/frontal operculum}
and FFA) in individuals more prone to defensive freezing (Sections 6.5.2 and 6.5.3). As outlined in Section 1.5, freezing represents a particular constellation of responses that form part of defense cascade models, which were originally developed mainly on the basis of rodent experiments (e.g. Lang & Davis, 2006), and characterise responses to threat based on its proximity. There have been relatively recent demonstrations of freezing responses in humans (characterised by bradycardia and immobility; e.g. Azevedo et al., 2005) as well as exploration of associated neural mechanisms (e.g. Hermans et al., 2013). In addition, freezing in response to social visual stimuli (angry faces) has been demonstrated, which is a finding unique to the human literature (Roelofs et al., 2010). Based on these existing lines of evidence, I hypothesised that the slower non-conscious processing for dominant and untrustworthy (in other words, threatening) faces (Chapter 6) is specifically reflective of defensive freezing. The next steps taken were to systematically test predictions relevant to a putative defensive freezing response in the context of perceptually ambiguous or non-conscious socially threatening visual information (Chapters 7 and 8). In Chapter 7 I focused on perceptual transitions in the context of socially dominant faces and found that (compared to a socially neutral context) such perceptual transitions occurred less frequently, and were associated with relative reduction in BOLD signal in amygdala and pSTS (with possible increase in signal in hypothalamus that did not reach corrected statistical significance) as well as relative bradycardia. In Chapter 8 I instead focused on non-conscious social face processing and aimed to measure the psychophysiological (heart rate) and motor physiological (TMS-MEP) correlates of freezing in this instance, as predicted by defense cascade models. My findings indicated that both psychophysiological and motor physiological modulation, in the direction predicted for freezing, occurred for non-conscious threatening faces but the two effects were independent. The psychophysiological effect was observed at the group level, while the motor physiological effect was variable across individuals and correlated with perceptuomotor performance (on a CFS task identical to the one used in Chapter 6, where threatening faces take longer to overcome suppression from awareness than neutral faces). Bringing these findings together, there are clear demonstrations of motor system (Chapter 8) and autonomic changes (Chapters 7 and 8) consistent with freezing in the absence of awareness of the threatening stimulus. Moreover, the findings from Chapter 8 in particular further the understanding of human freezing by clarifying the relationship between different components of this response. Finally, the fMRI findings in Chapter 7 occur specifically in an ambiguous and threatening context, which is also consistent with freezing. The reduced BOLD signals in amygdala and STS shown in Chapter 7 further hint at the possibility of freezing of perception (initially proposed in Chapter 6, and discussed further in Section 9.4.2; see also Stewart et al., 2012). Meanwhile, the putative finding in hypothalamus is of potential relevance since the hypothalamus is known to play a central role in coordinating the autonomic outputs of freezing responses (Fanselow, 1994; see Section 1.5.2).
While the predictions and interpretations in Chapters 7 and 8 were based on a defence cascade framework, it remains possible that other theories can also contribute to a unifying explanation of the findings from Chapters 6-8. Although dual competition accounts fit very well with the brain structural correlates of slower non-conscious social processing described in Chapter 6, they have no direct way of addressing the peripheral physiological correlates of threat demonstrated in Chapters 7 and 8. Conversely, the neurobiology of defence cascade models emphasises subcortical processes, although there is also evidence for engagement of cortical mechanisms and structures (namely mPFC) particularly in the context of freezing (Frysztak and Neafsey, 1991; Mobbs et al., 2009; see Section 1.5.3). MPFC is thought to act as a monitor of internal state, context and emotional stimulus significance before coordinating the level of response (via amygdala; Mobbs et al., 2009). In particular, ventral mPFC appears to be critical for freezing responses; lesions here reduce freezing to conditioned threat (Frysztak and Neafsey, 1991). The ACC may also act as a link between complex evaluation of environmental cues and defensive behaviours, and this has been explored within dual competition frameworks. Specifically, one study found that emotional interference of performance on a visual detection task is mirrored by BOLD signal in the cingulate (Pereira et al., 2010). While this fits well with dual competition models, the duration of the behavioural interference exceeded what could be expected from an attentional effect, prompting the authors to speculate on a possible defensive freezing mechanism. It may therefore be that the ACC bridges attentional effects on executive processes and behavioural output through autonomic and motor system changes in the setting of social threat. In defence distance models (McNaughton and Corr, 2004; see also Section 1.5.1), the proximity of threat is accounted for as in defence cascade models but it maps more clearly onto the hierarchical level of corresponding neural mechanisms with more proximate threats invoking low-level subcortical mechanisms and more distant threats engaging high-order systems in prefrontal cortex. The complex functions of higher forebrain and limbic regions such as mPFC and ACC could thus represent links between dual competition, defence cascade and defence distance frameworks, prompting new more comprehensive theoretical accounts and facilitating the integration of larger pools of evidence to help understand human responses to threat.

9.4.2 Proposed role of perceptual freezing

A number of the findings reported in this thesis support the possibility of perceptual (and perhaps social perceptual) contributions to freezing behaviours. In two independent experiments (Chapters 6 and 8) I showed that performance on the perceptuoemotor t2e task is slower in the presence of non-conscious socially threatening faces. In addition, in Chapter 6 I showed that the degree of slowing of t2e is correlated with the grey matter volume of
brain regions involved in social perception and cognition. There is also evidence that slowing is correlated with personality traits such as submissiveness and propensity to trust others (Stewart et al., 2012). Moreover, in Chapter 7, the findings of reduced perceptual-transition-related BOLD signal in amygdala and pSTS (and possible increase in such BOLD signal in hypothalamus) in association with ambiguity and social threat, as well as the threat-dependent deceleration of heart rate at the time of perceptual transitions provide additional support for both autonomic and perceptual contributions to freezing.

It has been proposed that responses to intermediately proximate threat (which include freezing) involve complex evaluation of contingency and contextual information, dependent on higher brain regions, including forebrain and limbic areas (e.g. mPFC and hippocampus), while responses to more imminent threat (such as fight-and-flight behaviours) are more inflexible and hard-wired through subcortical mechanisms (McNaughton and Corr, 2004; see also Sections 1.5.3 and 9.4.1). Thus, perceptual freezing may be part of a complex and diverse set of neural processes associated with post-encounter defensive responses.

An important consideration of relevance to the proposal of perceptual freezing is that while dual competition models (see Section 9.4.1) predict enhanced sensory processing associated with threatening stimuli, the prediction based on a defence cascade framework is less straightforward. Typically freezing is associated with reduced motor output and increased attention/arousal (Lang et al., 2000). However, it has also been proposed that there could be suppression of specific stimulus-related processes during freezing (Fanselow, 1994). Based on the findings presented in this thesis and elsewhere (Stewart et al., 2012), novel hypotheses that aspects of perceptual processing could be suppressed during defensive freezing have been generated. The function of such ‘freezing’ of certain perceptual processes could be to promote reduction in associated behavioural output in the interest of avoiding detection by a predator, or perhaps to enable resources to be concentrated on other brain processes that are most relevant for survival. For example, once an individual encountered in the environment has been deemed threatening, and appropriate defence responses initiated, further more sophisticated social processing may become irrelevant. In this instance, a mechanism to reduce activity related to social processing in regions such as amygdala and STS may be quite helpful. These are speculative suggestions, which require direct empirical evaluation (see Section 9.4.4).
9.4.3 Freezing responses elicited by non-conscious images

To my knowledge, the findings presented in Chapters 7 and 8 provide the first direct evidence of freezing in humans when faced with threat that is perceptually ambiguous or not consciously perceived. That this defense response can occur independently of awareness extends the known limits of non-conscious processing. As discussed in Section 1.4.4 there is accumulating evidence that mechanisms for processing non-conscious emotional and social signals are more distributed than previously thought, involving both subcortical and cortical mechanisms, with the amygdala acting as a modulator to allow appropriate allocation of processing resources (Pessoa and Adolphs, 2010). Such evidence may dovetail with the known substrates of post-encounter responses (and perhaps also non-conscious post-encounter responses, which remain unexplored). Such mechanisms again incorporate both subcortical and cortical processes (see Sections 1.5.3 and 9.4.1).

Previous work shows that the effect of slower responses for dominance and untrustworthiness in a t2e CFS paradigm is abolished when the same paradigm is adapted so that the face stimuli remain in fully conscious viewing conditions (Stewart et al., 2012; see also Section 9.3.1). This may be a floor effect due to faster response times when the faces are not hidden from awareness by CFS. However, it has also been argued that more favourable conditions for threat-related slowing of t2e (freezing) are engendered by non-conscious viewing conditions during this CFS task (Stewart et al., 2012). In possible support of such a proposal, defence cascade models posit that post-encounter responses occur when threat is either of intermediate proximity or ambiguous (e.g. Lang et al., 2000). Whether interocular suppression and bistable paradigms create ambiguity in this sense, to make threat more intermediate, is a relevant empirical question for future investigation (see Section 9.5.3).

9.4.4 Possible mechanisms for non-conscious freezing

In the context of discussion within this section (Section 9.4) and based on a combination of existing knowledge, the findings presented in this thesis, and some hypotheses for future exploration, I have devised a putative framework for understanding human freezing responses to non-conscious social threat. This framework includes proposals of relevant structures, their activation (e.g. in terms of BOLD signal) in socially neutral and socially threatening settings, and possible interactions between them (Figure 9.2). While some parts of this framework remain highly speculative, the aim is to provide motivation for testing future hypotheses and to link new findings to more established existing knowledge.
Figure 9.2: Speculative frameworks for the neural mechanisms (brain regions and their possible interactions) involved in defensive freezing responses to non-conscious socially threatening stimuli. See Section 9.4 for relevant discussion. When a visual stimulus is neutral (left hand panel), autonomic and motor output (dependent on hypothalamus and M1/PAG, respectively) occurs with influence from visual regions (Vis; this could include advanced visual regions such as STS), subcortical structures including the amygdala (Amy) and higher prefrontal regions (PFC). When the visual stimulus is threatening (right hand panel) there is proposed to be reduced signal in visual regions (Vis) and motor cortex (M1). There are also proposed changes in amygdala with some nuclei showing reduced activity but others, particularly CeA, being more active, in the course of coordinating motor and autonomic outputs of freezing (reduction in heart rate and corticospinal excitability, via hypothalamus and PAG respectively). Regions shaded in red are proposed/know to be more active and those shaded in blue less active in the threatening condition (right hand panel). Solid lines represent known connections/interactions between regions with functional relevance to defence responses. The dotted lines represent putative connections/interactions (in some cases anatomical connectivity exists but whether this is relevant to freezing is unknown). Blue dotted lines (right hand panel) indicate proposed modulation of certain interactions in the network in the context of threat (e.g. altered connectivity between PFC and Amy or between PFC and M1). Vis, visual areas (in occipital and temporal cortex); Amy, amygdala; PFC, prefrontal cortex (including mPFC and ACC); Hyp, hypothalamus; PAG, periaqueductal grey; M1, primary motor cortex.
9.5 Future research directions

The findings that have arisen from the experimental work included in this thesis, and their subsequent interpretation, place new questions into focus and thus open up a number of avenues for further scientific enquiry and experimentation. In particular, I will argue there is potential to gain much from multimodal strategies that incorporate neuroimaging, connectivity analyses and neuromodulatory techniques such as TMS. In addition, improved understanding will undoubtedly be gained from more spatially refined approaches focused on studying mechanisms at the cellular or molecular scale (often requiring invasive work in non-human animals), including more recently emerging modalities such as optogenetics. Some future directions relevant to each of the three main themes covered in the previous sections of this chapter, as well as possible extension of the work within this thesis into clinical populations, will presently be discussed.

9.5.1 Future work on neural mechanisms of perceptual transition

Extending existing understanding of how transitions between alternative perceptual interpretations of the visual environment are mediated in the brain has been a major focus in this thesis (in relation to this topic, see Section 1.3.5 for a summary of previous literature and Section 9.2 for discussion of the work in this thesis). There is now a substantial accumulation of knowledge on this subject, but many crucial aspects remain to be explored. Having outlined a putative extended network of brain regions involved in perceptual transition during bistable perception provides clear anatomical targets for future hypotheses and experimental work. In the first instance, a wider range of bistable paradigms could be used to obtain a more refined version of a stimulus-invariant perceptual-transition-related network. A logical next step would be to continue to test the causal roles of key regions in this network (including insula and FEF) in the dynamics of perceptual transition, and to add this knowledge to the existing understanding of the functional role of superior parietal cortex (Section 3.1). TMS would be a suitable approach for ongoing work on this question. For FEF, where a null result was obtained (Chapter 5), a number of steps could be undertaken in a re-attempted experiment, including functional localisation of FEF (by confirming effects on saccadic function) and adjustment of the temporal characteristics of the experimental paradigm (see Sections 5.4.1 and 5.4.2). For targeting insula, which is a deeper brain region not accessible to traditional TMS techniques, initial studies have suggested that adapted coil designs and stimulation
protocols can be safe and effective (Ciampi de Andrade et al., 2012). Online TMS delivered in a state-dependent fashion (for example using fMRI to ensure stimulation is only delivered during particular perceptual states; Silvanto and Pascual-Leone, 2008) may be a particularly useful approach for a dynamic process such as bistable perception. Alternatively, longer-term modulation with tDCS could be attempted. This method uses small electric currents applied for prolonged periods over the scalp to initiate plastic neuromodulatory effects in underlying neural tissue by inducing a steady state extracellular electric field (Paulus, 2011). Emerging evidence suggests that this method can also be useful for modulating function of deeper structures (To et al., 2018).

Another important question related to understanding the neural correlates of perceptual transition is how the causal roles of frontal and parietal regions in bistable perception dynamics, as established by TMS and other neuromodulatory approaches, actually relate to the BOLD activity that accompanies perceptual switches. Further experiments with combined imaging and virtual lesion approaches, or analysis of intracranial EEG recordings (see Section 9.5.4) are two possible ways to tackle these issues.

Beyond the functional roles of individual regions in the perceptual transition network, the interactions and flow of information between them is likely at least as important for understanding the mechanisms of access to awareness. Some of these interactions have been explored in this thesis (Chapter 4) and other studies (discussed in Sections 1.3.5 and 9.2.2). However, a great deal remains unclear, and the sketch of possible interactions within this network attempted in Figure 9.1 remains highly speculative. A series of future experiments could help clarify these questions further through the use of functional and/or effective connectivity analyses. One approach, based on the work presented in this thesis, could be to evaluate to what degree top-down and bottom-up manipulations of bistability affect the currently proposed FEF-centred system (Figure 9.1B; purple) and SPL-centred system (Figure 9.1B; red/blue). Alternatively, combined functional imaging and neuromodulation approaches (TMS/fMRI or TMS/EEG) may be particularly useful for determining how the relevant regions and networks interact in relation to the dynamics of bistable perception. One fairly recent study used concurrent EEG and TMS to show that disruption of function in parietal regions (IPS) with TMS influences perceptual-switch-related activity and destabilises percepts, but if this is shortly preceded by TMS-induced disruption of function in frontal regions (DLPFC) the effect is abolished (Vernet et al., 2015). While an insufficient number of experimental conditions were included, the results of this study do suggest that frontal and parietal regions play both causal and distinct roles, with respect to maintaining and switching of percepts during bistable perception.
The above approaches could substantially advance the ongoing debate about the significance of the patterns of frontoparietal BOLD signal seen in association with perceptual transitions (Section 9.2.3). A further step, in light of the intriguing findings reported by Brascamp and colleagues (2015), would be to systematically manipulate observers’ experience of bistability (using the method described by those authors) while studying the causative influences of frontal and parietal regions on bistable perception. Such an approach could further disentangle the role of these brain regions in generating the perceptual switches themselves and/or mediating the observer’s experience of them.

9.5.2 Future work on the social modulation of visual awareness

There are a number of outstanding questions regarding social modulation of non-conscious visual processing. It will be important to determine to what extent the behavioural findings reported in Chapters 6 and 8, obtained using face stimuli with traits varying along dimensions of dominance/trustworthiness or threat, can be generalised to other types of social cues. It is already established that faces with fearful expressions and angry expressions affect emergence from interocular suppression in opposite directions, corresponding to the known differences in approach/avoidance tendencies and generation of defensive responses (see Section 9.3.1). However, it has not been explored whether body postures or biological motion affect non-conscious processing in similar ways (in particular with reference to their threat content). In terms of the neural mechanisms associated with these effects, it would be important to verify whether the structural correlates of non-conscious dominance and trustworthiness evaluation have functional importance. For example, one may hypothesise that there are BOLD signal changes in TPJ during non-conscious evaluation of facial trustworthiness, and furthermore that the degree of signal change (in TPJ or elsewhere) varies parametrically with the degree of trustworthiness-related slowing in performance. It is well established that BOLD signal changes associated with non-conscious visual stimuli can be difficult to detect, especially in hierarchically advanced regions, so multivariate approaches to analysis may be more successful in these types of questions than univariate ones (see Sections 1.3.4 and 1.4.4). In addition, the causal importance in non-conscious social processing of regions shown to have structural co-variation with such processing could be evaluated by temporarily disrupting their function using rTMS. TPJ would be a particularly attractive target region from a methodological perspective, although rTMS to mPFC has also been administered successfully in the past (e.g. Mattavelli et al., 2011). Whether the neuronal mechanisms underlying delayed non-conscious processing of threatening stimuli correlate with both subcortical and cortical
patterns of neural activity consistent with systems supporting defensive freezing (along with associated autonomic and behavioural correlates of freezing) will be another important area to explore, and will be discussed further in Section 9.5.3.

In terms of social modulation of perceptual transition, there is evidence that emotional faces predominate over neutral faces in binocular rivalry (see Section 1.4.5). However, whether social stimuli, when paired with another non-social class of stimulus, have an effect on binocular rivalry dynamics, has not been previously investigated to my knowledge. In Chapter 7 I showed that pairing socially dominant faces with visual gratings in binocular rivalry was associated with slower perceptual switch rates (when compared to making the same pairing with socially neutral faces). An additional question is whether a similar effect could be elicited with angry-expression faces (and an opposite effect with fearful-expression faces); this would be in line both with the theoretical frameworks discussed in Section 9.3.1 and with the results for non-conscious processing mentioned above. The neural mechanisms underlying social modulation of perceptual transition could also be explored further. One hypothesis may be that perceptual transitions in the context of happy or fearful faces are associated with increased (rather than decreased, as with threatening faces) signal in certain perceptual regions. Moreover, connectivity analyses (for example, with DCM) could be used to probe whether there are directional interactions between perceptual regions and the established frontoparietal perceptual-transition-related network (Section 9.2.1). Such interactions might represent the modulatory signals that allow social cues to influence access to awareness, and the causal role of these interactions could even be explored with the addition of carefully timed TMS administration during bistable perception (e.g. Pascual-Leone et al., 2000; Silvanto and Pascual-Leone, 2008). Subcortical or cortical neural signatures of freezing could again be explored in this context, and this is discussed further in Section 9.5.3.

Finally, the limited existing knowledge regarding neural processing of social dominance derived from face cues could be addressed with further experiments (likely focusing on fMRI and M/EEG methods) aiming to delineate the mechanisms through which dominant faces are differentiated from neutral faces. It would equally be interesting to investigate the non-conscious processing of facial dominance and trustworthiness at a neuronal level, possibly with the use of fMRI and multivariate pattern analysis methods (see Section 1.3.4). The findings of such studies could be placed in the context of existing knowledge about neural processing of facial emotional expressions and facial traits such as trustworthiness and attractiveness (see Section 1.4.3 for discussion of the existing literature).
9.5.3 Future work on mechanisms of human defensive freezing

Exploration of the neural mechanisms of human freezing is at an early stage with many unanswered questions (see Section 1.5.3). If, as discussed in Section 9.4, there are indeed both perceptual and motor components to freezing responses, disentangling these will likely require the use of neuroimaging modalities with precise temporal resolution. With the goal of resolving putative perceptual and motor components of human freezing, I have designed an experiment employing MEG and a modified version of the CFS paradigm used in Chapters 6 and 8. The intention is to compare separately stimulus- and response-related signals in non-conscious neutral and threatening-face conditions through the analysis of both event-related fields and frequency domain data (in the latter case particularly to study movement preparation through beta-band desynchronisation). I have collected MEG data on this paradigm, which will be analysed as a next step following completion of this thesis.

Specific and direct focus on putative perceptual freezing would also be a relevant future endeavour. Substantial further empirical work would likely be required to determine how sensory processing is modulated under different threat conditions, and to clarify whether this can be reasonably characterised as perceptual freezing. In particular, it will be important to map out associated behavioural and autonomic states, measure impact on task performance, and relate these measures to neuronal processes in such experiments in order to place findings in the context of defence cascade and dual inhibition models, and to be able to relate them more readily to existing literature. Given peripheral autonomic changes have well recognised impact on cortical processes (Critchley and Harrison, 2013), one intriguing question could be whether freezing-related perceptual changes occur in parallel to the peripheral autonomic changes, or in fact may be subsequent and/or dependent on them.

Another approach to understanding the neural mechanisms of human freezing could be to employ techniques with relatively high spatial resolution (e.g. fMRI) and evaluate effective connectivity between key regions within the neural systems subserving freezing in threatening and non-threatening experimental conditions. For example, dedicated exploration of the role of ACC (see Section 9.4.1) could be undertaken, in particular testing for changes in effective connectivity with other key regions such as primary motor cortex or PAG. Imaging findings could readily be correlated with autonomic indices such as heart rate. For example, with reference to findings from Chapter 7, one could explore whether peri-perceptual-transition bradycardia correlates with possible hypothalamic signal changes. Such a study would ideally make use of fMRI hardware and sequences focused methodologically on detecting BOLD signal in subcortical structures including
hypothalamus (also allowing study of other subcortical and brainstem signal changes, including in PAG). A further even more methodologically challenging approach could be to measure TMS-MEP concurrently with fMRI and heart rate, allowing correlation of imaging findings with changes in peripheral motor and autonomic physiology (as well as testing whether motor cortex activation differs when TMS is delivered in threatening-face versus neutral-face trials).

The findings from Chapter 8 suggested that inter-individual variability was a prominent feature in the perceptual and motor physiological correlates of freezing and investigating the neural underpinnings of these effects, either in the structural or the functional domains, would also be of importance.

It is of course the case that emerging experimental approaches are opening up whole new avenues for inquiry and at times leading to substantial adaptations of the frameworks supporting our understanding of the neural mechanisms in question. For example, a novel colliculo-thalamic-amygdalar neural pathway for defensive responses to innate fear has relatively recently been described in mice using optogenetic techniques (Wei et al., 2015).

I have proposed that freezing is readily elicited by socially threatening images in conditions of binocular rivalry or continuous flash suppression because such conditions make the threatening stimuli more ambiguous and therefore potentially less proximate (from a defense cascade perspective) than the same stimuli seen in full awareness (Section 9.4.3). This possibility should be evaluated empirically in the future. An initial step could be to compare behavioural and autonomic correlates associated with the same threatening stimuli when viewed in full awareness or under binocular suppression. However, for a full understanding of the relevant effects, measures of perceived distance and clarity of the threatening stimulus would also need to be incorporated into more detailed experiments. More generally, the peripheral responses associated with human freezing could be characterised further by adding measures of skin conductance and body sway to those of heart rate and corticospinal excitability. Finally, it will also be important to confirm that the freezing effect generalises to other types of threatening stimulus when hidden from awareness or made ambiguous.
9.5.4 Future clinical relevance of work presented in this thesis

A fundamental aspect of the future development of this work is its potential to be translated into clinical populations. Epilepsy makes up a substantial proportion of neurological illness and represents a significant disease burden (de Boer et al., 2008; Sander, 2003). Many epileptic seizures are associated with transient loss of awareness. Functional imaging experiments (using fMRI, SPECT or EEG) have shown that seizures accompanied by such alterations in consciousness are associated with changes in brain activity both in midline subcortical structures (brainstem and thalamus) and in widespread areas of frontal and parietal cortex (Blumenfeld, 2012; Tononi and Koch, 2008). Applying the frameworks of perceptual transition and access to awareness (Section 9.2) to epileptic seizures could shed new light on the mechanisms of loss of consciousness in epilepsy. Arguably changes or loss of awareness represent the most disruptive symptom for epilepsy patients, both in terms of risk of adverse events and in terms of impact on quality of life (Blumenfeld, 2012). Patients with poorly controlled and drug-resistant epilepsy may undergo detailed evaluation of seizure semiology and seizure onset localisation using prolonged EEG monitoring. Sometimes invasive insertion of intracranial electrodes is necessary as part of these investigations, to explore deeper structures or provide optimal spatial resolution for planning localisation-dependent treatment approaches. In these cases there could be added opportunities to gain enhanced understanding of human mechanisms of perceptual reinterpretation and changes in awareness with exquisite spatial and temporal resolution. For example, if electrodes have been implanted (due to reasons purely related to clinical care) in relevant regions of frontal, parietal or temporal cortex, the nature and spatiotemporal patterns of neural activity could be explored both during loss of consciousness induced by seizures (e.g. Englot et al., 2010), and during experimentally induced changes in awareness (e.g. Kreiman et al., 2002). Such experiments of course represent significant challenges in terms of patient recruitment, experimental data collection and analysis methods. While studying distributed areas of cortex is often not possible in such circumstances (since electrode implantation is restricted by clinical safety considerations, normally resulting in rather limited brain coverage), these types of study could potentially make invaluable contributions to understanding the function of more spatially confined systems, such as the aSPL and pSPL system for perceptual transition (Kanai et al., 2011). Such work also has the potential for focal and precise neuromodulation (e.g. Valentín et al., 2017). The growing understanding of the functional anatomy of the perceptual-transition-related neural systems (towards which the work in this thesis has contributed) improves the chances of success with this sort of approach.
Improved understanding of defensive responses to threat could also have direct and important clinical applications, in this case in developing models and ultimately treatments applicable to post-traumatic stress disorder (and probably other anxiety and panic disorders). Post-traumatic stress, like epilepsy, has a high prevalence in the general population and accounts for a substantial disease burden (Helzer et al., 1987; Kessler, 2000). Accumulating evidence suggests that freezing-like behaviours (i.e. immobility) during trauma are important in the pathophysiology of post-traumatic stress disorder (Bovin et al., 2008; Hagenaars et al., 2010; Heidt et al., 2005). Some early mechanistic proposals about the maintenance of stressful memories under influence from glucocorticoids (Brinks et al., 2008) and several neurotransmitters (Aboufatima et al., 1999) have also been made. Moreover, the control function of the ACC in coordinating the effect of aversive stimuli on behaviour (see Section 9.4.1) seems to be disrupted in anxiety disorders (Bishop et al., 2004). There is ongoing debate about the degree to which freezing behaviours are helpful or maladaptive in different situations (Hagenaars et al., 2014). The enhanced understanding of freezing responses resulting from the work in this thesis may usefully contribute to future developments in this field. Clearly a better characterisation of the neurobiological underpinnings of freezing may help a more in depth account of the pathophysiology and maintenance of post-traumatic and panic/anxiety disorders. In addition, the findings presented here could lead to the consideration of an expanded range of stimulus types (e.g. more subtle social stimuli) and contexts in evaluations relevant to these illnesses. In general, an improved understanding of the pathophysiology and neurobiology underlying these conditions could enable development of more effective treatment strategies in the future.

9.6 Concluding remarks

The experiments contained in this thesis have contributed to knowledge relevant for three key themes in research on visual awareness and social perception. The significance of these contributions and the related outstanding questions, which open up future avenues for experimental work, have been discussed in this final chapter. My work has added new insights into the anatomical distribution and functional interactions among brain regions involved in spontaneous transition between competing perceptual interpretations of the visual environment. I have then focused on the impact of social cues on visual processing and behaviour, particularly if these cues cannot be perceived consciously or are only intermittently available to the observer’s awareness. Using measures of focal neural structure and function, my findings have provided new evidence for the importance of high-order brain regions involved in complex social appraisal when evaluating social cues.
outside of awareness. Focusing on behavioural measures, my experiments have generated replicable findings indicating that non-conscious socially threatening stimuli are delayed from reaching conscious awareness (rather than prioritised), a phenomenon that I have argued is a reflection of human defensive freezing. I have gone on to explore the motor physiological, autonomic and neural correlates of this putative defensive response, gathering substantial supporting evidence for human freezing and enabling understanding of its components and their interactions.

Through the presented findings and their interpretation my work has contributed to understanding how the gating of visual information towards conscious appraisal occurs in the brain, and how this gating is influenced by socially relevant visual cues. Moreover, I have been able to advance knowledge on human defensive responses to social threat. Besides their relevance to the neuroscientific understanding of mechanisms of social perception and visual awareness, my findings can also motivate future work of importance to clinical populations. These could particularly include patients with disorders of consciousness, such as epilepsy, or those suffering from anxiety and panic disorders.
# Appendix 1: List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>anterior cingulate cortex</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>BAS</td>
<td>Behavioural Activation System</td>
</tr>
<tr>
<td>BASD</td>
<td>Behavioural Activation System: Drive subscale</td>
</tr>
<tr>
<td>BASF</td>
<td>Behavioural Activation System: Fun Seeking subscale</td>
</tr>
<tr>
<td>BASR</td>
<td>Behavioural Activation System: Reward subscale</td>
</tr>
<tr>
<td>BIS</td>
<td>Behavioural Inhibition System</td>
</tr>
<tr>
<td>BOLD</td>
<td>blood-oxygen-level-dependent</td>
</tr>
<tr>
<td>BPA</td>
<td>Bayesian parameter average</td>
</tr>
<tr>
<td>BR</td>
<td>binocular rivalry</td>
</tr>
<tr>
<td>b-CFS</td>
<td>breaking continuous flash suppression</td>
</tr>
<tr>
<td>CaA</td>
<td>central nucleus of amygdala</td>
</tr>
<tr>
<td>CBF</td>
<td>cerebral blood flow</td>
</tr>
<tr>
<td>CFS</td>
<td>continuous flash suppression</td>
</tr>
<tr>
<td>cpd</td>
<td>cycles per degree (of visual angle)</td>
</tr>
<tr>
<td>CSE</td>
<td>corticospinal excitability</td>
</tr>
<tr>
<td>CSF</td>
<td>cerebrospinal fluid</td>
</tr>
<tr>
<td>DARTEL</td>
<td>differomorphic anatomical registration through exponentiated lie algebra</td>
</tr>
<tr>
<td>DCM</td>
<td>dynamic causal modelling</td>
</tr>
<tr>
<td>deoxyHb</td>
<td>deoxygenated haemoglobin</td>
</tr>
<tr>
<td>DLPFC</td>
<td>dorsolateral prefrontal cortex</td>
</tr>
<tr>
<td>DTI</td>
<td>diffusion tensor imaging</td>
</tr>
<tr>
<td>EEG</td>
<td>electroencephalography</td>
</tr>
<tr>
<td>EPI</td>
<td>echo-planar imaging</td>
</tr>
<tr>
<td>FDIO</td>
<td>first dorsal interosseous</td>
</tr>
<tr>
<td>FEF</td>
<td>frontal eye field</td>
</tr>
<tr>
<td>FFA</td>
<td>fusiform face area</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>FWE</td>
<td>family-wise error</td>
</tr>
<tr>
<td>FWHM</td>
<td>full-width/half-maximum</td>
</tr>
<tr>
<td>GLM</td>
<td>general linear model</td>
</tr>
<tr>
<td>GM</td>
<td>grey matter</td>
</tr>
<tr>
<td>GRF</td>
<td>Gaussian random field</td>
</tr>
<tr>
<td>HR</td>
<td>heart rate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HRF</td>
<td>haemodynamic response function</td>
</tr>
<tr>
<td>IFG</td>
<td>inferior frontal gyrus</td>
</tr>
<tr>
<td>IPL</td>
<td>inferior parietal lobe</td>
</tr>
<tr>
<td>IPS</td>
<td>intraparietal sulcus</td>
</tr>
<tr>
<td>IT</td>
<td>inferior temporal cortex</td>
</tr>
<tr>
<td>LGN</td>
<td>lateral geniculate nucleus</td>
</tr>
<tr>
<td>MEG</td>
<td>magnetoencephalography</td>
</tr>
<tr>
<td>MEP</td>
<td>motor-evoked potential</td>
</tr>
<tr>
<td>M1</td>
<td>primary motor cortex</td>
</tr>
<tr>
<td>MDEFT</td>
<td>modified driven equilibrium Fourier transform</td>
</tr>
<tr>
<td>MNI</td>
<td>Montreal Neurological Institute</td>
</tr>
<tr>
<td>MR</td>
<td>magnetic resonance</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>MT</td>
<td>middle temporal area</td>
</tr>
<tr>
<td>MVPA</td>
<td>multivariate pattern analysis</td>
</tr>
<tr>
<td>NCC</td>
<td>neural correlates of consciousness</td>
</tr>
<tr>
<td>NIfTI</td>
<td>Neuroimaging Informatics Technology Initiative</td>
</tr>
<tr>
<td>OFA</td>
<td>occipital face area</td>
</tr>
<tr>
<td>OFC</td>
<td>orbitofrontal cortex</td>
</tr>
<tr>
<td>oxyHb</td>
<td>oxygenated haemoglobin</td>
</tr>
<tr>
<td>mPFC</td>
<td>medial prefrontal cortex</td>
</tr>
<tr>
<td>n.s.</td>
<td>not statistically significant</td>
</tr>
<tr>
<td>$p$</td>
<td>$p$-value</td>
</tr>
<tr>
<td>$p_{FWE-corr}$</td>
<td>$p$-value with family-wise error correction</td>
</tr>
<tr>
<td>PAG</td>
<td>peri-aqueductal gray</td>
</tr>
<tr>
<td>PC</td>
<td>principal component</td>
</tr>
<tr>
<td>PCA</td>
<td>principal components analysis</td>
</tr>
<tr>
<td>PET</td>
<td>positron emission tomography</td>
</tr>
<tr>
<td>PMd</td>
<td>dorsal premotor cortex</td>
</tr>
<tr>
<td>PTS</td>
<td>Propensity to Trust Survey</td>
</tr>
<tr>
<td>RF</td>
<td>radio frequency</td>
</tr>
<tr>
<td>rsfMRI</td>
<td>resting state functional magnetic resonance imaging</td>
</tr>
<tr>
<td>ROI</td>
<td>region of interest</td>
</tr>
<tr>
<td>SII</td>
<td>secondary somatosensory cortex</td>
</tr>
<tr>
<td>SBS</td>
<td>Submissive Behaviour Scale</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SEM</td>
<td>standard error of mean</td>
</tr>
<tr>
<td>SFM</td>
<td>structure-from-motion</td>
</tr>
<tr>
<td>SMA</td>
<td>supplementary motor area</td>
</tr>
</tbody>
</table>
SPECT  single-photon emission computerised tomography
aSPL  anterior superior parietal lobule
pSPL  posterior superior parietal lobule
SPL  superior parietal lobule
SPM  Statistical Parametric Mapping/statistical parametric map
STAI  State-Trait Anxiety Inventory
STAI-S  State-Trait Anxiety Inventory: State subscale
STAI-T  State-Trait Anxiety Inventory: Trait subscale
pSTS  posterior superior temporal sulcus
STS  superior temporal sulcus
t2e  time to emergence
TBS  theta burst stimulation
cTBS  continuous theta burst stimulation
TE  echo time
TI  inversion time
TMS  transcranial magnetic stimulation
rTMS  repetitive transcranial magnetic stimulation
TMS-MEP  transcranial magnetic stimulation-induced motor-evoked potential
TPJ  temporoparietal junction
pTPJ  posterior temporoparietal junction
TR  repetition time
V1  visual area V1/primary visual cortex/striate cortex
V5  visual area 5
VBM  voxel-based morphometry
VOI  volume of interest
WM  white matter
Appendix 2: Details of apparatus and software

Anatomy toolbox  http://www.fz-juelich.de/inm/inm-1/EN/Forschung/_docs/SPMAnatomyToolbox/SPMAnatomyToolbox_node.html

Cogent  http://www.vislab.ucl.ac.uk/cogent.php

Eyelink eye tracker  http://www.sr-research.com/

FaceGen  https://facegen.com/modeller.htm

FreeSurfer  https://surfer.nmr.mgh.harvard.edu/

FSL  https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL

HRVAS toolbox  https://github.com/jramshur/HRVAS

MRI keypad  http://www.curdes.com/

MRI pulse oximeter  http://www.nonin.com/PulseOximetry

MRI scanner  https://www.healthcare.siemens.co.uk/magnetic-resonance-imaging/for-installed-base-business-only-do-not-publish/magnetom-trio-tim/technical-details

MRICTron  http://people.cas.sc.edu/rorden/mricron/index.html

Pick atlas  http://fmri.wfubmc.edu/software/pickatlas

Prism glasses  www.bolle-safety.com/

Psychtoolbox  http://psychtoolbox.org/

Social face stimuli  http://tlab.princeton.edu/databases/

Spike 2 system  http://ced.co.uk/products/spkoin

SPM  http://www.fil.ion.ucl.ac.uk/spm/

TMS neuronavigation  https://www.ant-neuro.com/

TMS stand  https://www.manfrotto.co.uk
TMS stimulator and coil  http://www.magstim.com/
Visor 2 recording system  http://www.ant-neuro.com/products/visor2-product-range
Appendix 3: References


References


Schwaber, J.S., Kapp, B.S., Higgins, G.A., Rapp, P.R., 1982. Amygdaloid and basal forebrain
direct connections with the nucleus of the solitary tract and the dorsal motor nucleus. J.
Neurosci. 2, 1424–1438.

Schwarzkopf, D.S., Song, C., Rees, G., 2011. The surface area of human V1 predicts the


Shafer, A.B., 2006. Meta-analysis of the factor structures of four depression questionnaires:

double dissociation between emotional and cognitive empathy in inferior frontal gyrus
versus ventromedial prefrontal lesions. Brain 132, 617–627.

17062.

Shibata, K., Watanabe, T., Kawato, M., Sasaki, Y., 2016. Differential activation patterns in
the same brain region led to opposite emotional sates. PLOS Biol. 14, e1002546.


Acquis. 15, 275–298.


Slotnick, S.D., Yantis, S., 2005. Common neural substrates for the control and effects of

and gender effects on human brain anatomy: a voxel-based morphometric study in


Turatto, M., Sandrini, M., Miniussi, C., 2004. The role of the right dorsolateral prefrontal

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N.,
using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.

Urgesi, C., Moro, V., Candidi, M., Aglioti, S.M., 2006. Mapping implied body actions in the


Valentín, A., Selway, R.P., Amarouche, M., Mundil, N., Ughratdar, I., Ayoubian, L.,

Valentine, T., 1991. A unified account of the effects of distinctiveness, inversion, and race in


van Dam, L.C.J., van Ee, R., 2006. The role of saccades in exerting voluntary control in


van Loon, A.M., Wildenberg, W.P.M. van den, Stegeren, A.H. van, Ridderinkhof, K.R.,

30, 829–858.


Verbruggen, F., Houwer, J.D., 2007. Do emotional stimuli interfere with response


