A neurocomputational account of Self-Other distinction: From cell to society

Samuel Philip Aaron Ereira

UCL

Submitted for a PhD in Cognitive Neuroscience
Declaration

I, Samuel Philip Aaron Ereira, 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.

Sam Ereira

25th June 2019
Abstract

Human social systems are unique in the animal kingdom. Social norms, constructed at a higher level of organisation, influence individuals across vast spatiotemporal scales. Characterising the neurocomputational processes that enable the emergence of these social systems could inform holistic models of human cognition and mental illness.

Social neuroscience has shown that the processing of ‘social’ information demands many of the same computations as those involved in reasoning about inanimate objects in ‘non-social’ contexts. However, for people to reason about each other’s mental states, the brain must be able to distinguish between one mind and another. This ability, to attribute a mental state to a specific agent, has long been studied by philosophers under the guise of ‘meta-representation’. Empathy research has taken strides in describing the neural correlates of representing another person’s affective or bodily state, as distinct from one’s own. However, Self-Other distinction in beliefs, and hence meta-representation, has not figured in formal models of cognitive neuroscience.

Here, I introduce a novel behavioural paradigm, which acts as a computational assay for Self-Other distinction in a cognitive domain. The experiments in this thesis combine computational modelling with magnetoencephalography and functional magnetic resonance imaging to explore how basic units of computation, predictions and prediction errors, are selectively attributed to Self and Other, when subjects have to simulate another agent’s learning process. I find that these low-level learning signals encode information about agent identity. Furthermore, the fidelity of this encoding is susceptible to experience-dependent plasticity, and predicts the presence of subclinical psychopathological traits.

The results suggest that the neural signals generating an internal model of the world contain information, not only about ‘what’ is out there, but also about ‘who’ the model belongs to. That this agent-specificity is learnable highlights potential computational failure modes in mental illnesses with an altered sense of Self.
Impact Statement

The work presented in this thesis can be put to beneficial use both within and outside of academia. The ‘probabilistic false-belief task’ developed as part of this research is a novel behavioural paradigm that can be used to assess Self-Other distinction in humans with an unprecedented level of experimental control. The combination of this behavioural task along with state-of-the-art analysis techniques in multi-modal neuroimaging enabled me to answer questions about how Self and Other are represented in the brain.

These results have implications in understanding how the brain supports Theory of Mind, mental time-travel and a representation of ‘Self’ more generally. The findings presented in Chapter 3 have already been disseminated in an open-access peer-reviewed scientific journal (Ereira et al., 2018) whilst the other findings are currently being prepared for submission to journals. These findings make up a significant contribution to social and cognitive neuroscience.

This thesis also makes a significant theoretical contribution to the cognitive sciences more generally. By integrating these novel findings into a review of the existing social psychology and social neuroscience literatures in the context of learning and decision making, the thesis attempts to unify some seemingly disparate fields. The results provide evidence that domain-general learning processes may underpin complex features of social cognition, an idea which has received much theoretical attention but has been difficult to assess experimentally. The conclusions from this thesis may help to inspire novel theories and experimental approaches that are less restrictive in the hypotheses they set out to test.

Outside of academia, the work in this thesis has implications for clinical psychiatry. Dysfunctional social cognition, particularly Self-Other distinction, is a transdiagnostic dimension of many mental health disorders, including but not limited to autism spectrum disorder (ASD), schizophrenia and borderline personality disorder (BPD). The insights
Impact Statement

gained, from this thesis, into how the brain supports Self-Other distinction, illuminate potential pathophysiological pathways behind these disorders. These pathways could be diagnostically tested using the kinds of techniques developed in this thesis and could be useful targets for treatment.

The probabilistic false-belief task, developed as part of thesis, is currently being used in a group of BPD patients participating in a drug trial at The University of Chicago. The data from this trial will provide the first indication of whether this tool carries any diagnostic value, and will show whether an effective drug treatment is mediated by changes in behaviour in this task.

From a broader perspective, the work presented in this thesis takes steps to frame complex psychological constructs like ‘Self’ and ‘Other’ with formal quantitative measures. This has implications for the field of computational psychiatry more generally, which seeks to develop diagnostic and therapeutic tools for mental illness by considering how information processing in the brain becomes disrupted.
Acknowledgements

Would it be seen as so perverse,
To give my thanks in form of verse?
A little late.
That was already a rhyme.
To poetry, I will resign!

Ray, I'm here because of you,
The opportunity you gave me was huge, it's true.
I'm sometimes lost, let me be straight.
But the imposter syndrome dissipates.
The lab you've built is diverse and teeming,
With wonderful ideas. I'm sure I'm dreaming.

Zeb you're the best teacher I've ever known.
That might sound trite or overblown.
But I really cannot stress enough,
The clarity you give to this neuroscience stuff.
I'll always be inspired by your intrigue and passion,
Your generosity with your time, that you never even ration.
What you've taught will stick like glue,
I've learned general things all the way through.
You've trained me to think with an independent routine,
Like a prefrontal cortex, doused in dopamine.

All the fellows and staff at the FiL and MPC,
You're an amazing bunch, anyone can see.
You're not just colleagues but true dear friends,
Oh, and Jochen - A special mention - You're a God-send!
You've been a great friend and mentor for years,
Acknowledgements

You’ve helped my confidence when I needed it.
Really. Just want to say cheers.

Mum and Dad, thanks for all your love,
Your motivating words, your wisdom from above.

Eilis, through this time
You’ve been my rock, that’s all I’ve known.
I love you dearly.
I couldn’t have got through this on my own.

There’re plenty others who I’d like to credit,
But there’s no space here to describe their merits¹.
But here’re my thanks for making me
The me who did a PhD.

¹ Ok, I found a bit more space…Rachel and Ben, thanks for being delightful desk buddies, who kept me sane in times of insanity. Tobi, Rani and Giles thanks for being awesome collaborators, and giving me guidance on various things that I didn't understand. Thanks again Tobi for fostering such a fun atmosphere in the lab, particularly with your innovation of weird movie club! Megan, Clive, Elisa, Elaine and Yoshi, you were all fantastic radiographers to work with. Your contribution to my PhD was invaluable, both for making the science actually happen and for being fantastic company during my many hours in the basement.
# Contents

Declaration .................................................................................................................. 2
Abstract ......................................................................................................................... 3
Impact Statement ......................................................................................................... 4
Acknowledgements ....................................................................................................... 6
Contents ......................................................................................................................... 8
Figures .......................................................................................................................... 12
Tables ............................................................................................................................ 14
Abbreviations ............................................................................................................... 15

Chapter 1: Introductory material .................................................................................. 18

1.1 Philosophical foundations ...................................................................................... 18
  1.1.1 A hallmark of human cognition ...................................................................... 18
  1.1.2 The mind in a hierarchical biological system .................................................... 20
  1.1.3 Looking for the social brain ............................................................................. 23

1.2 Processing ‘my’ world .......................................................................................... 27
  1.2.1 Feature extraction ........................................................................................... 27
  1.2.2 Reward detection ............................................................................................ 28
  1.2.3 Learning about new rewards .......................................................................... 29
  1.2.4 Learning beyond rewards .............................................................................. 36

1.3 Processing ‘your’ world ....................................................................................... 42
  1.3.1 What is a social system? ................................................................................ 42
  1.3.2 Behaviour reading .......................................................................................... 45
  1.3.3 Mind reading .................................................................................................. 51
  1.3.4 Am I reading your behaviour or your mind? ................................................... 57

1.4 My world or yours? ............................................................................................... 62
  1.4.1 Feeling like me ............................................................................................... 62
  1.4.2 Thinking like me ............................................................................................. 64
  1.4.3 Failed Self-Other distinction in mental illness ................................................. 67
3.4 Results .................................................................................................................. 120
  3.4.1 Behaviour ........................................................................................................ 120
  3.4.2 Modelling simulated belief updates with simulated PEs ......................... 121
  3.4.3 Neural representations of PEs and simulated PEs ........................................ 124
  3.4.4 Decoding agent identity from learning signals .............................................. 126
  3.4.5 Neural decoding of agent identity predicts Self-Other distinction .......... 129
  3.4.6 Neural decoding of agent identity predicts subclinical traits ................. 130

3.5 Discussion .............................................................................................................. 133

3.6 Supplementary material ...................................................................................... 139

Chapter 4: Meta-representational associative learning ............................................ 144

4.1 Abstract ................................................................................................................ 144

4.2 Introduction .......................................................................................................... 145

4.3 Methods ................................................................................................................. 150
  4.3.1 Participants .................................................................................................... 150
  4.3.2 Modified false belief task ............................................................................. 151
  4.3.3 Visual perspective-taking task ...................................................................... 153
  4.3.4 A subject’s eye view of the experiment ....................................................... 159
  4.3.5 Model fitting .................................................................................................. 161
  4.3.6 Scanning protocol .......................................................................................... 161
  4.3.7 qMRI data preprocessing .............................................................................. 163
  4.3.8 fMRI data preprocessing .............................................................................. 163
  4.3.9 Mass-univariate analysis ............................................................................... 164
  4.3.10 Searchlight analysis ...................................................................................... 164
  4.3.11 Pseudotrial analysis ..................................................................................... 166
  4.3.12 Cross-decoding analysis .............................................................................. 167
  4.3.13 Analysis of questionnaire data .................................................................... 167

4.4 Results ..................................................................................................................... 168
  4.4.1 Self-Other leak emerges in training and persists ......................................... 168
  4.4.2 Training transfers to a different cognitive domain ...................................... 173
  4.4.3 Functional localisation of prediction errors ................................................. 176
  4.4.4 Plasticity of neural Self-Other distinction .................................................... 178
  4.4.5 Prefrontal microstructure and meta-representational knowledge ........... 181

4.5 Discussion .............................................................................................................. 184

4.6 Supplementary material ...................................................................................... 190
Chapter 5: Self-Other distinction in intertemporal choice

5.1 Abstract ........................................................................................................... 196
5.2 Introduction ..................................................................................................... 197
5.3 Methods ........................................................................................................... 199
  5.3.1 Intertemporal choice task ..................................................................... 199
  5.3.2 Delay discounting models ..................................................................... 200
  5.3.3 Model fitting and model comparison ...................................................... 201
  5.3.4 Logistic regression analysis .................................................................. 203
5.4 Results ............................................................................................................. 205
  5.4.1 Common computation of social and temporal distance ....................... 205
  5.4.2 vmPFC microstructure is associated with discounting behaviour ........ 209
5.5 Discussion ....................................................................................................... 210
5.6 Supplementary material ............................................................................... 213

Chapter 6: General discussion

6.1 Summary of findings ...................................................................................... 214
6.2 Interpretations and implications .................................................................... 215
6.3 Limitations ...................................................................................................... 220
6.4 Future directions ............................................................................................ 222
6.5 Concluding remarks ...................................................................................... 224

References ............................................................................................................ 226
Figures

Figure 1-1: Red blood cells of a 39-year-old healthy male.........................................................21
Figure 1-2: Relationship between neocortex size and social group size ....................................24
Figure 1-3: Literature search (1960-2014) conducted on Scopus................................................25
Figure 1-4: Illustration of dopaminergic projections in the human brain ......................................28
Figure 1-5: Error-driven learning in classical conditioning ............................................................35
Figure 1-6: State representation in rodents and humans .................................................................41
Figure 1-7: Detecting and learning about social agents .................................................................49
Figure 1-8: Simulating Other's computations in the vmPFC .........................................................53
Figure 1-9: Simulating Other's computations in midbrain and striatum ........................................57
Figure 1-10: A proposed multilayer framework for social cognition ..........................................61

Figure 2-1: The haemodynamic response function .........................................................................81
Figure 2-2: Examples of artefact correction in MEG .................................................................84
Figure 2-3: Neurocomputational workspace ..................................................................................86
Figure 2-4: Pattern-based analysis of neuroimaging data .............................................................90

Figure 3-1: Overview of behavioural paradigm ............................................................................97
Figure 3-2: Illustration of how Pseudotrial MEG data were constructed .....................................117
Figure 3-3: Subject performance in behavioural task .................................................................121
Figure 3-4: Behavioural model comparison ..................................................................................123
Figure 3-5: Regressing PE magnitude against MEG ERF .......................................................125
Figure 3-6: Decoding agent identity from neural patterns of learning signals ............................127
Figure 3-7: Relationship between neural SOD and behavioural SOD .........................................130
Figure 3-8: Relationship between neural agent decoding and subclinical traits .........................133
Figure 3-9: Qualitative differences between group A and group B models ................................139
Figure 3-10: Parameter recovery .................................................................................................140
Figure 3-11: Posterior Dirichlet distributions from Bayesian model selection ...........................141
Figure 3-12: Patterns of PE encoding in two exemplar subjects ...............................................142
Figure 3-13: Decoding the visual stimuli ......................................................................................142
Figure 3-14: Spatial topography of neural Self-Other distinction .............................................143
Figures

Figure 4-1: Modified false belief task.............................................................. 153
Figure 4-2: Visual perspective-taking task.................................................... 156
Figure 4-3: Factorial design of visual perspective-taking task ....................... 158
Figure 4-4: Three-day training procedure...................................................... 160
Figure 4-5: Performance is worse in the Hi-Share condition ......................... 169
Figure 4-6: Model fits and parameter estimates ........................................... 171
Figure 4-7: Correlations between Self- and Other-attributed beliefs................ 172
Figure 4-8: Analysis of drift rates in visual perspective-taking task ............... 175
Figure 4-9: Group-level clusters of PE-related activity.................................. 177
Figure 4-10: Evidence of plasticity of Self- and Other-attributed signals ......... 180
Figure 4-11: vmPFC microstructure is associated with plasticity and subclinical traits ...... 183
Figure 4-12: Main effect on Self-Other correlation with different bin sizes ....... 192
Figure 4-13: Regressors for fMRI analyses are not correlated .................... 193
Figure 4-14: Cross-validation scheme for cross-decoding ............................ 195

Figure 5-1: Discounting curve and model comparison.................................. 206
Figure 5-2: Performance of different logistic regression models ...................... 208
Figure 5-3: Ventromedial prefrontal white matter MT is associated with discount factor.... 210
Figure 5-4: Validating the discount factor..................................................... 213
Figure 5-5: Correlations between Self-Other distinction and discount factors .... 213
Tables

Table 3-1: Summary of all models fit to behavioural data ........................................... 109
Table 4-1: List of models fit to false belief task .............................................................. 192
Table 4-2: Summary of significant clusters ..................................................................... 194
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>Anterior cingulate cortex</td>
</tr>
<tr>
<td>ACCg</td>
<td>Anterior cingulate gyrus</td>
</tr>
<tr>
<td>ASD</td>
<td>Autism spectrum disorder</td>
</tr>
<tr>
<td>BDI</td>
<td>Beck depression inventory</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian information criterion</td>
</tr>
<tr>
<td>BIDR</td>
<td>Balanced inventory of desirable responding</td>
</tr>
<tr>
<td>BOLD</td>
<td>Blood oxygen-level dependent</td>
</tr>
<tr>
<td>BPD</td>
<td>Borderline personality disorder</td>
</tr>
<tr>
<td>BPQ</td>
<td>Borderline personality questionnaire</td>
</tr>
<tr>
<td>CA</td>
<td>Classification accuracy</td>
</tr>
<tr>
<td>CAPE</td>
<td>Community assessment of psychic experience</td>
</tr>
<tr>
<td>CLF</td>
<td>Choice likelihood function</td>
</tr>
<tr>
<td>CS</td>
<td>Conditioned stimulus</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>DDM</td>
<td>Drift diffusion model</td>
</tr>
<tr>
<td>dmPFC</td>
<td>Dorsomedial prefrontal cortex</td>
</tr>
<tr>
<td>E-M</td>
<td>Expectation-Maximisation</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>EPI</td>
<td>Echo planar imaging</td>
</tr>
<tr>
<td>EQ</td>
<td>Empathy quotient</td>
</tr>
<tr>
<td>ERF</td>
<td>Evoked response field</td>
</tr>
<tr>
<td>FLASH</td>
<td>Fast low angle shot</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional MRI</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full width at 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>hMRI</td>
<td>In vivo histology using MRI</td>
</tr>
<tr>
<td>HRF</td>
<td>Haemodynamic response function</td>
</tr>
<tr>
<td>iBIC</td>
<td>Integrated Bayesian information criterion</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent component analysis</td>
</tr>
<tr>
<td>ICU</td>
<td>Inventory of callous-unemotional traits</td>
</tr>
<tr>
<td>IRI</td>
<td>Interpersonal reactivity index</td>
</tr>
<tr>
<td>ITI</td>
<td>Inter-trial interval</td>
</tr>
<tr>
<td>LASSO</td>
<td>Least absolute shrinkage and selection operator</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum a posteriori</td>
</tr>
<tr>
<td>MBT</td>
<td>Mentalization-based treatment</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetoencephalography</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MLE</td>
<td>Maximum likelihood estimation</td>
</tr>
<tr>
<td>MNI</td>
<td>Montreal Neurological Institute</td>
</tr>
<tr>
<td>mPFC</td>
<td>Medial prefrontal cortex</td>
</tr>
<tr>
<td>MPM</td>
<td>Multi-parameter mapping</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MT</td>
<td>Magnetisation transfer</td>
</tr>
<tr>
<td>MVPA</td>
<td>Multi-voxel pattern analysis</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NSV</td>
<td>Non-social version</td>
</tr>
<tr>
<td>OFC</td>
<td>Orbitofrontal cortex</td>
</tr>
<tr>
<td>PAI-BOR</td>
<td>Borderline scale of the Personality assessment inventory</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal components analysis</td>
</tr>
<tr>
<td>PD</td>
<td>Proton density</td>
</tr>
<tr>
<td>PE</td>
<td>Prediction error</td>
</tr>
<tr>
<td>PFC</td>
<td>Prefrontal cortex</td>
</tr>
<tr>
<td>PSP</td>
<td>Post-synaptic potential</td>
</tr>
<tr>
<td>qMRI</td>
<td>Quantitative MRI</td>
</tr>
<tr>
<td>rACC</td>
<td>Rostral anterior cingulate cortex</td>
</tr>
<tr>
<td>RF</td>
<td>Radiofrequency</td>
</tr>
<tr>
<td>RL</td>
<td>Reinforcement learning</td>
</tr>
<tr>
<td>rPFC</td>
<td>Right lateral prefrontal cortex</td>
</tr>
<tr>
<td>RPE</td>
<td>Reward prediction error</td>
</tr>
<tr>
<td>rSMG</td>
<td>Right supramarginal gyrus</td>
</tr>
<tr>
<td>RT</td>
<td>Response time</td>
</tr>
<tr>
<td>RW</td>
<td>Rescorla-Wagner</td>
</tr>
<tr>
<td>SBH</td>
<td>Social Brain Hypothesis</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>SOD</td>
<td>Self-Other distinction</td>
</tr>
<tr>
<td>SPE</td>
<td>State prediction error</td>
</tr>
<tr>
<td>SQUID</td>
<td>Superconducting quantum interference device</td>
</tr>
<tr>
<td>ST</td>
<td>Simulation Theory</td>
</tr>
<tr>
<td>SV</td>
<td>Social version</td>
</tr>
<tr>
<td>TD</td>
<td>Temporal difference</td>
</tr>
<tr>
<td>tDCS</td>
<td>Transcranial direct current stimulation</td>
</tr>
<tr>
<td>TE</td>
<td>Echo time</td>
</tr>
<tr>
<td>ToM</td>
<td>Theory of mind</td>
</tr>
<tr>
<td>TPJ</td>
<td>Temporoparietal junction</td>
</tr>
<tr>
<td>TR</td>
<td>Repetition time</td>
</tr>
<tr>
<td>TT</td>
<td>Theory Theory</td>
</tr>
<tr>
<td>US</td>
<td>Unconditioned stimulus</td>
</tr>
<tr>
<td>vmPFC</td>
<td>Ventromedial prefrontal cortex</td>
</tr>
</tbody>
</table>
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP</td>
<td>Ventral pallidum</td>
</tr>
<tr>
<td>VTA</td>
<td>Ventral tegmental area</td>
</tr>
<tr>
<td>WM</td>
<td>White matter</td>
</tr>
</tbody>
</table>
Chapter 1: Introductory material

1.1 Philosophical foundations

On rare occasions I have lucid dreams. I was about five or six when I had the earliest lucid dream that I remember. In the dream, I was playing in my school courtyard, with some friends. One of my friends interrupted the game to proudly assert “This is my dream, you know!”. Frustrated by the patently inaccurate claim, I tried to correct him: “Actually, this is my dream. I know it is”. He wasn’t having any of it, and the immature exchange continued for a few more back-and-forth confident claims about whose dream we were living, until I woke up. “Well that’s that settled”, I thought.

1.1.1 A hallmark of human cognition

Humans have a remarkable ability to read minds. Our brains can represent not only physical objects but also abstract mental states, enabling us to attend to the thoughts and feelings of ourselves and of others. Cognitive scientists and philosophers have described this as the ability to represent representations, or rather ‘meta-representation’ (Wimmer & Perner, 1983, Leslie, 1987, Proust, 2007, Carruthers, 2009). Take, for instance, Homer Simpson. The patterns of neural activity in his brain constitute a ‘first-order’ representation of icing sugar when he looks at a doughnut (Kamitani & Tong, 2005) or indeed remembers one (Kurth-Nelson et al., 2015). Theoretically, this representation may itself be represented in a different activity pattern, a ‘second-order’ representation (Rosenthal, 2000, Cleeremans et al., 2007, Lau, 2008). This second pattern will encode not only information about the icing sugar, but also information about the first-order representation. Together, these informational components produce a propositional attitude such as ‘I see that the icing sugar is pink’ or ‘I
remember that the icing sugar is pink’. These two states, whilst phenomenologically distinct, are both meta-representations. Critically, they both contain truth conditions. In other words, the statements about the icing sugar can be objectively true or untrue (Shea, 2012). They also both invoke the existence of an agent, a Self or an Other\(^2\), to whom the first-order representation is attributed. This thesis will interrogate the latter feature. How do we recognise an experience, a belief, or a dream, as our own or someone else’s?

Dedicated research fields exist for exploring how humans read their own minds and the minds of Others. Self-attributed mental states are invoked in the studies of metacognition (Carruthers, 2009) and mental time-travel (Suddendorf & Corballis, 1997, Redshaw, 2014, Mahr & Csibra, 2017), whilst Other-attributed mental states are invoked in the study of Theory of Mind (ToM) (Leslie, 1987, Baker et al., 2017, Burge, 2018). This thesis will also explore the evidence for and against treating Self- and Other- attributed mental states as being qualitatively distinct constructs.

There is some evidence that humans share mind reading abilities with other species (Whiten, 2013). Researchers have observed dolphins (Smith et al., 1995, Browne, 2004), the corvid family of birds (Dally et al., 2006, Clayton et al., 2007), apes (Premack & Woodruff, 1978a, Krupenye et al., 2016, Buttelmann et al., 2017, Devaine et al., 2017) and other non-human primates (Smith et al., 1997, Devaine et al., 2017) behaving in ways that are easily explained by appealing to meta-representations. However, these experiments are not conclusive and have also been interpreted with explanations that do not rely on meta-representations (Proust, 2007, Heyes, 2014b, 2015, 2016). Nevertheless, most agree that mind reading emerged relatively recently in phylogenetic history. This recency is recapitulated in human ontogeny. One popular view is that humans develop meta-

\(^2\) In this thesis, the first letters of ‘Self’ and ‘Other’ will always be capitalised when referring to different intentional agents. This is to avoid confusion when these words are used as adjectives in non-agentive contexts (e.g. ‘These other analyses were consistent with the main analysis’).
representational skills, particularly ToM, around age four, after more basic cognitive skills have been acquired (Wimmer & Perner, 1983, Wellman et al., 2001, Rakoczy, 2017). However, some researchers, who employ nonverbal, ‘implicit’ measures of mind reading, claim to have observed ToM in much younger infants (Kovács et al., 2010, Scott et al., 2010, Király et al., 2018).

If mind reading is a recently evolved adaptation then there are two important implications. Firstly, it would make mind reading a candidate feature for a hallmark of human cognition, something that sets humans apart from other animals. Secondly it might be subserved by architectures that are particularly vulnerable to maladaptation and environmental insults. From this perspective, improving our understanding of mind reading seems invaluable for explaining human cognition in both its physiological and pathological forms.

To be consistent with existing literature, mind reading will hereafter be referred to as ‘mentalising’ (Fonagy, 1991). The central role of mentalising in human cognition becomes more intuitive when one adopts a systems-biological stance.

1.1.2 The mind in a hierarchical biological system

A biological system can be viewed as a hierarchy of levels (Ehresmann & Vanbremeersch, 1987, Ingber, 2003, Alcocer-Cuaron et al., 2014). Every biological system can be described in terms of either its structure or its function. The structure refers to simpler subsystems at the level below, whilst the function refers to a more complex suprasystem, that emerges at the level above.

Take, for instance, a human erythrocyte (red blood cell). Homer Simpson, against all the odds, takes on a PhD in cellular biology and asks himself: ‘Why does the erythrocyte have a biconcave shape (Figure 1-1)?’

---

3 Note that ‘mentalising’ is often spelled ‘mentalizing’ in the scientific literature. These are two different spellings of the same word.
Chapter 1: Introductory material

Figure 1-1: Red blood cells of a 39-year-old healthy male

Fixation in 1% glutaraldehyde (0.1M phosphate buffer). Red blood cells are characterised by their biconcave disc-like shape. Adapted from (Tokunaga et al., 1969).

Homer is asking two different questions here. Daniel Dennett describes these as the ‘how come’ question and the ‘what for’ question (Dennett, 2017). ‘Why does the erythrocyte have a biconcave shape?’ could either mean, ‘How does an erythrocyte acquire and maintain its biconcave shape?’, or it could mean ‘For what purpose does an erythrocyte have a biconcave shape?’. If Homer means the former, ‘how come’, question, he will attempt to explain the erythrocyte’s shape using a reductionist account. He could appeal to the cell’s cytoskeleton and membrane proteins. This is the cell’s structure.

If Homer means the latter, ‘what for’, question, he will describe the selective advantage that the erythrocyte’s shape endows upon humans. This is the cell’s function. This latter explanation necessitates a description of the organism’s oxygen demands and of the entire cardiovascular system. This is a more complex system, at a higher level, that emerges from the activity of biconcave erythrocytes as well as other components.

Michael Polanyi has elegantly described how a system that is nested within a hierarchy is under the ‘dual control’ of both simpler subsystems and a more complex suprasystem. Further, he argues that the existence of the
more complex suprasystem cannot be inferred from simply interrogating the subsystems:

‘You cannot derive vocabulary from phonetics; you cannot derive grammar from vocabulary; a correct use of grammar does not account for good style; and a good style does not supply the content of a piece of prose [...] Each [level] reduces the scope of the one immediately below it by imposing on it a boundary that harnesses it to the service of the next higher level [...] The presence of a mechanism is not revealed by its physical-chemical topography. We can say the same thing of all higher levels: their description in terms of any lower level does not tell us of their presence.’ (Polanyi, 1968)

In other words, an explanatory model of a system cannot reveal the full picture without considering the function of that system at the level above. Making assumptions about function allows hypotheses about structure to be constrained. If Homer assumes that erythrocytes are biconcave in order to provide an increased surface area for gaseous exchange, then he could interrogate their structure by asking well-posed, tractable questions. Instead of asking ‘How are they biconcave?’, he could ask ‘In what way are they biconcave such that gaseous exchange is enhanced?’. This question refers to both the cell’s structure and its function. The system is now constrained within a hypothetical biological hierarchy, which will make it easier for Homer to interpret his results.

Relative to other biomedical fields, cognitive neuroscience has developed slowly and is particularly prone to replication issues (Lieberman & Cunningham, 2009, Sochat et al., 2016, Cohen, 2017). Beyond methodological limitations, this may partly result from cognitive scientists expressing a reductionist bias (Krakauer et al., 2017), paying scant regard to Polanyi’s ‘higher-level control’. The most complex features of human cognition are often implicitly assumed to be at the metaphorical peak of evolutionary development, with no higher levels to constrain them. This often leaves ‘awareness’, ‘consciousness’ and ‘mentaling’ as mere linguistic constructs, obscure and inaccessible to scientific enquiry. However, it is
helpful to treat these constructs like any biological system that is nested within a hierarchy, such as a red blood cell.

More complex systems have indeed emerged from the collective minds of individuals, including political institutions, nations, religions and economies. These are social systems comprised of multiple individuals who are able to represent each other’s mental states. There has been a recent effort to model the dynamics of social systems like other biological systems (Suzuki & Akiyama, 2008, Bahrami et al., 2010, Auvray & Rohde, 2012, Schilbach et al., 2013, De Martino et al., 2013, Smith et al., 2014, Friston & Frith, 2015) and social neuroscience is a field devoted to the idea that social systems and neural systems evolve in tandem (Tennie et al., 2009, Thompson et al., 2016).

It is natural to assume that the emergence of these social systems is enabled by certain features of human cognition. In other words, the function of these cognitive features is in the generation of social systems and, taking Polanyi’s argument forward, these cognitive features are under the higher-level control of social systems. If cognitive neuroscientists seek to explain how the brain processes information about the world then a well-posed question for cognitive neuroscience would then be:

‘How does the brain process information about the world, such that social systems emerge?’

The ability to represent Others’ mental states seems a likely feature of cognition that would be necessary for social systems to emerge.

1.1.3 Looking for the social brain

The idea that mentalising is a defining characteristic of human cognition is not new. The ‘Machiavellian Intelligence Hypothesis’ (Whiten & Byrne, 1989) argued that humans and other great apes evolved large brains in order to predict and outwit each other in a cognitive arms-race of social intelligence. This line of reasoning was galvanised by an intriguing observation that the relative sizes of different primate species’ neocortices is
strongly associated with the size of those species’ social groups, as shown in Figure 1-2 (Dunbar, 1992). This was the justification behind the ‘Social Brain Hypothesis’ (Dunbar, 1998). Accumulating evidence suggests that natural selection has driven the prefrontal cortex of humans and other great apes to grow in size beyond what allometric growth would predict (Smaers et al., 2017). In other words, its size cannot be explained merely by the size of the organism but has seen ‘exceptional enlargement’. The Social Brain Hypothesis states that this adaptation was driven by the need to process ‘social’ information, about predators, prey and conspecifics, both collaborative and competitive.

![Figure 1-2: Relationship between neocortex size and social group size](image)

*Figure 1-2: Relationship between neocortex size and social group size*

Mean group size for individual genera plotted against neocortex ratio (relative to rest of brain; i.e. total brain volume less neocortex). (●) Polygamous anthropoids; (+) monogamous anthropoids; (○) diurnal prosimians; (□) nocturnal prosimians; (Δ) hominoids. Reproduced from (Dunbar, 1992).

Social systems, the Social Brain Hypothesis states, are enriched with such high-dimensional information that genetic traits for increased computational power (i.e. a relatively large neocortex) have been selected for. This flavour of reasoning has motivated the social neuroscience community to look for neural architectures that evolved specifically to process ‘social information’.

24
Figure 1-3 shows the number of publications in recent decades that contained phrases related to social neuroscience in their titles or abstracts. These publication rates grew substantially after Dunbar published the Social Brain Hypothesis in 1998. These can be compared with the publication rates of article abstracts that mentioned ‘visual cortex’ or ‘motor cortex’, which show no sudden change after 1998. This literature search does not indicate whether the Social Brain Hypothesis actually contributed to the growth of social neuroscience. However, it is evident that the field is still in its infancy and moreover, that the Social Brain Hypothesis was proposed at a time when very little social neuroscience research had been conducted.

![Figure 1-3: Literature search (1960-2014) conducted on Scopus](image)

The y-axis shows the number of publications each year that included, in their title or abstract, a specific keyword or phrase. The publication rates for articles including the three social phrases all increased following the publication of the Social Brain Hypothesis (SBH) in 1998. Publication rates for articles including the phrases ‘visual cortex’ and ‘motor cortex’ are shown for comparison.

This thesis is motivated by the foundational question stated on page 23 and restated here: ‘How does the brain process information about the
world, such that social systems emerge?’.

The Social Brain Hypothesis shares with this question a view that sociality is a defining feature of human cognition. However, the Social Brain Hypothesis and much of the work that it inspired, make additional, stronger assumptions. The idea that the evolution of the neocortex was specifically driven by ‘social’ selective pressures and is thus adapted for processing ‘social’ information, invokes a special peculiarity about ‘social’ information. The work presented in this thesis, on the other hand, is agnostic about the selective pressures that sculpted human cognition, whilst acknowledging that the way human brains process information enables the emergence of social systems.

The current perspective is compatible with the Social Brain Hypothesis, but provokes a slightly different line of enquiry, questioning which computations might enable the emergence of social systems, rather than asking how social information is processed. The following chapters, will explore what it means for an environment to be ‘social’, what it means for a mode of cognition to be ‘social’, and the relationship between representations of physical states and meta-representations of mental states. A deconstruction of ‘mentalisation’ will reveal that the ability to distinguish Self from Other is necessary for ‘social’ cognition, and that Self-Other distinction is amenable to neurocomputational modelling. Framing Self-Other distinction as a fundamental computation will provide an entry point for exploring the above foundational question, and in so doing will inevitably invoke a discussion of what is meant by the ‘Self’. Before wandering into the hazy boundaries between ‘social’ and ‘non-social’ or Self and Other, it will be useful to review what is known about information processing in non-social environments.
1.2 Processing ‘my’ world

1.2.1 Feature extraction

To be energy efficient, the brain ought only represent features of the environment that are minimally sufficient for adaptive behaviour. Representing only the salient features of an environment has the additional benefit of helping the animal generalise to novel situations. Work in computational neuroscience has shown that the brain, rather than constructing a full-blown simulation of reality, represents latent features of the environment, such as ‘value’, in the service of adaptive decision making. One of the most influential models of learning and decision making has been reinforcement-learning (RL) (Sutton & Barto, 1998), which describes animals as agents seeking to maximise their future ‘reward’.

There are, however, other ways to explain how animals identify salient features of their environment. The ‘active inference’ framework, for example, describes organisms as agents that are not reward-maximising but rather, free-energy-minimising (Friston et al., 2009, Friston et al., 2016). In this framework, the brain seeks to maximise the evidence for its internal model of the world.

‘Value’ in the RL framework, and ‘model evidence’ in the active inference framework, are both latent variables that the brain might compute, and depending on how a researcher wants to define their model, these numerical quantities might be closely related (Friston et al., 2015, Gershman, 2019). The subsequent sections will review the literature on value-based decision making to explore how information is processed in the brain of an actively behaving animal. The choice of focus on RL here is due to a century of research in learning and decision making that evolved from work using rewards and punishments. ‘Value’ has thus been a central component in developing accounts of learning and decision making. However, as will become clearer, predictability or uncertainty is also fundamental to information processing in animals.
1.2.2 Reward detection

Animals are born with an innate ‘library’ of appetitive and aversive stimuli, which they are exquisitely sensitive to, such as food, sex and pain. These primary reinforcers are intrinsically rewarding and punishing. They are not learned during the lifetime of an organism, but over evolutionary timescales by natural selection.

Dopaminergic projections (Figure 1-4) from the midbrain to the medial prefrontal cortex (mesocortical pathway) and to the ventral striatum (mesolimbic pathway) are necessary for animals to detect these latent features of the environment.

![Illustration of dopaminergic projections in the human brain](image)

**Figure 1-4: Illustration of dopaminergic projections in the human brain**

This illustration shows a midline sagittal section of the human brain. Dopamine neuron cell bodies reside in the red ventral tegmental area (VTA). These neurons project to the ventral striatum, which includes the ventral pallidum (VP) and nucleus accumbens (NAc). This neuronal projection is called the mesolimbic pathway. Midbrain dopaminergic neurons also project to cortical regions, including the ventromedial prefrontal cortex (vmPFC), dorsomedial prefrontal cortex (dmPFC) and anterior cingulate cortex (ACC). These projections are collectively called the mesocortical pathway. Reproduced from (Husain & Roiser, 2018).
The significance of dopamine in reward detection was originally made clear through the use of intracranial self-stimulation studies. Rodents will repeatedly press a lever to stimulate dopaminergic regions of their own brain (Olds & Milner, 1954). The positive reinforcement of this behaviour is weakened when drugs are administered that reduce synaptic levels of dopamine and noradrenaline (Wise & Stein, 1969) or when dopamine and noradrenaline cells are chemically ablated (Breese et al., 1971).

The mPFC, and particularly a subregion called the orbitofrontal cortex (OFC), is known to track primary reinforcers in macaque monkeys (Rolls et al., 1990) and also in humans (Francis et al., 1999, O'Doherty et al., 2001b). Furthermore, damage to the prefrontal cortex impairs the ability to keep track of rewarding stimuli in both humans (Bechara et al., 1994, Rolls et al., 1994, Hornak et al., 2004) and monkeys (Izquierdo et al., 2004).

This evaluation network brings animals, not only a pre-computed library of primary reinforcing stimuli, but also a repertoire of innate behaviours called ‘Pavlovian responses’ (Dayan & Seymour, 2008, Rangel et al., 2008, Dolan & Dayan, 2013). These responses are reflexive, stereotyped and rigid, and also depend on dopaminergic transmission (Guitart-Masip et al., 2014, Swart et al., 2017). Humans and other animals are often unable to resist eliciting an approach behaviour in response to an appetitive stimulus, like food (Hershberger, 1986, Guitart-Masip et al., 2012). This is the kind of appetitive response that led Kumbuka, a male silverback gorilla at London Zoo, to consume five litres of undiluted blackcurrant juice when he broke into the zookeepers’ area (Guardian, 2016).

1.2.3 Learning about new rewards

Studies of value-based decision making in humans often use money as a reward to motivate behaviour. However, money is a not a primary reinforcer. A one-year-old will not show a Pavlovian approach response to a twenty-pound note on the floor like an adult would. The value that an adult attributes to a twenty-pound note on the floor has to be learned through
experience. With sufficient learning, money can become a ‘secondary reinforcer’. Like food and other primary reinforcers, secondary reinforcers induce striatal dopamine release (Koepp et al., 1998, Delgado et al., 2000), are tracked in the OFC (Thut et al., 1997, O'Doherty et al., 2001a), provoke Pavlovian responses (Guitart-Masip et al., 2012) and can motivate other forms of value-guided behaviour. Secondary reinforcers are thus convenient for investigating value-based decision making in humans; it’s often more practical to motivate subjects with money than with food or sex.

The process by which a neutral stimulus becomes a secondary reinforcer, like money, is called associative learning. Classical conditioning is a form of associative learning that Ivan Pavlov demonstrated in dogs in the 1890s (Pavlov & Anrep, 1927). Before the experiment, the dogs would salivate in response to food, an innate response to a primary reinforcer. Pavlov repeatedly exposed the dogs to the sound of a metronome before feeding them. Eventually, the dogs began to salivate in response to the metronome sound alone. By the end of the experiment, the dogs had learned to produce a reflexive, conditioned response (salivating) to a previously neutral stimulus. The ability to associate neutral stimuli with reward (or punishment) enables animals to learn how to deploy innate, stereotyped responses to previously unfamiliar stimuli e.g. ‘That tree usually has tasty insects inside: Climb it and eat!’ or ‘Last time I heard that noise my leg got bitten off: Run!’

Conditioning will only take place if the animal is unable to predict the forthcoming reward or punishment.4 In a blocking paradigm (Kamin, 1969), the researcher pairs a neutral stimulus (CS1) like a light, with a rewarding unconditioned stimulus (US) like food, so that the animal produces a conditioned response to CS1, as in classical conditioning. Then, there’s a

4 In some ways this is a trivial statement. Classical conditioning is a form of learning, and if the delivery of a reward is entirely predictable then there is nothing left to learn so of course no conditioning will take place. However, the idea that prediction errors, per se, serve as fundamental units of computation was a nontrivial insight that informed models of biological learning as well as developments in artificial intelligence.
second phase of conditioning trials, where the CS1 is paired with a new neutral stimulus (CS2), like a tone, before delivering the rewarding US. The animal will not acquire a conditioned response to CS2. This is thought to be due to the fact that the CS2 is not predicting anything over and above what can be predicted with CS1 alone. Because the CS1 renders the US entirely predictable, no more learning takes place, even though the CS2 is novel.

The role of predictability in associative learning was made explicit with the Rescorla-Wagner (RW) model (Rescorla & Wagner, 1972). The RW model operationalised the ‘surprisingness’ of a stimulus, which seemed so important to explain the phenomenon of blocking. The model assumes that the strength of a conditioned response is determined by the ‘associative value’, $V$, of a to-be-conditioned stimulus (CS) and tries to predict how $V$ will change during conditioning. The model proposes that after one instance of CS-US pairing, $V$ will change as follows:

$$V_{t+1}^1 = V_t^1 + \alpha \beta (\lambda - \sum_{i=1}^{n} V_t^i)$$

**Equation 1-1**

Here $V_t^1$ denotes the associative value of CS1 on trial $t$ and $V_{t+1}^1$ denotes the associative value of CS1 after one trial of CS1-US pairing. $\alpha$ and $\beta$ are free parameters denoting the salience of CS1 and US respectively, which together regulate the learning rate. $\sum_{i=1}^{n} V_t^i$ denotes the sum of associative values on trial $t$ for all $n$ stimuli in the environment, and $\lambda$ represents the asymptote of learning (the maximum associative value that any CS can acquire).

With each instance of CS1-US pairing, $V_t^1$ steps closer to $\lambda$, and becomes more likely to elicit a conditioned response. The $(\lambda - V_t)$ term can be thought of as a prediction error term, the magnitude of which describes how surprising the US is. The associative value is updated in proportion to this surprise signal (Figure 1-5A). As this term is minimised with learning, the US becomes more predictable. If $\lambda = 1$, and after a few conditioning trials $V_t^1$ is close to 1 (i.e. CS1 almost always elicits a conditioned response), then a
compound stimulus of CS1-CS2 followed by a US will not change $V_i^2$. This is because the prediction error term uses the sum of all stimuli, including the previously learned CS1, thus the prediction error is close to 0, and no learning takes place. This is how the RW model can explain blocking.

The neural correlates of conditioning can also be explained in terms of predictions and prediction errors. As discussed in section 1.2.2, midbrain dopaminergic neurons track reward. These neurons are also implicated in conditioning. This was first made clear when it was shown that dopamine antagonists prevent the acquisition of a conditioned response (Beninger & Hahn, 1983). Subsequent in vivo electrophysiological studies in monkeys provided evidence that midbrain dopamine neurons only fire in response to unpredicted rewards (Mirenowicz & Schultz, 1994) and have a reduced firing rate when a predicted reward is not received (Schultz et al., 1993). These neurons were also found to have a distinctive temporal signature (Figure 1-5B) whereby, before conditioning, they fired at the onset of the US, but as conditioning took place, neuronal firing shifted back in time towards the onset of the CS (Ljungberg et al., 1992). Collectively, these findings supported the idea of an error-driven form of associative learning.

However, the RW model cannot predict exactly when, within a trial, reward will be delivered, nor can it explain the precise onset of dopaminergic firing, which systematically moves backwards in time throughout conditioning. A more generalised model called temporal difference (TD) learning can explain this phenomenon. TD learning is a class of reinforcement learning (RL) methods (Sutton & Barto, 1981, 1987, Sutton, 1988) that updates estimates about the world from experience at each time step, with a finer degree of temporal resolution than that offered by the RW model. RL models couch predictions about future reward in a ‘state value function’, $V(s)$. A ‘state’ can be considered a signal that provides the agent with information about what the environment is like at any moment in time. More formally, states are defined as nodes in a stochastic process that the agent is navigating. This stochastic process is characterised by the Markov property (Equation 1-2). The property asserts that current state captures all
the relevant information, from the history, about which states the agent might transition to next. In other words, the future is independent of the past, given the present.

$$P[s_{t+1}|s_t] = P[s_{t+1}|s_1, s_2, \ldots, s_t]$$

**Equation 1-2**

In RL, the stochastic process is a ‘Markov reward process’, where each state has an estimated state value function $V(s)$. A state value function is similar to an associative value in RW learning, but unlike an associative value, which only depends on immediate reward, a state value function also depends on the value of what might happen after being in that state. This prediction is encapsulated in the Bellman equation (Equation 1-3).

$$V(s) = \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots]$$
$$= \mathbb{E}[R_{t+1} + \gamma (R_{t+2} + \gamma R_{t+3} + \cdots)]$$
$$= \mathbb{E}[R_{t+1} + \gamma V(s_{t+1})]$$

**Equation 1-3**

$R_{t+1}$ is the reward received at the time point immediately after leaving the state being evaluated. The value of what might happen in the future is computed as the sum of discounted future rewards, where $\gamma$ is a discount factor between 0 and 1. Discounting rewards further away in the future is both mathematically convenient and consistent with the behaviours of animals and humans, which suggest a preference for immediate rewards over delayed rewards (Ainslie, 1974, Thaler, 1981). TD learning is just one way of estimating a value function, in which the estimate is gradually brought closer to the ‘TD target’, $R_{t+1} + \gamma V(s_{t+1})$, by minimising the TD error, or reward prediction error (RPE), the term in square brackets in Equation 1-4. Here $\alpha$ is a free parameter that governs the rate of learning.
TD learning can account for within-trial fluctuations of state value during conditioning (Sutton & Barto, 1981). If the onset of a conditioned stimulus (CS) is unpredictable, and it is consistently followed by a reward (US), then eventually there will only be a positive RPE at the time of the CS. This is because the state the animal is in immediately before the onset of the CS will have a value of zero, whilst all time points between the CS and the US will have a state value of one. If, however, the US is ever omitted, then a large negative RPE will occur at the time that the reward was expected.

Predictive learning models, like TD learning, map well onto biological phenomena. Most famously, TD learning can explain the activity profiles of primate dopamine neurons during conditioning (Montague & Sejnowski, 1994, Schultz et al., 1997). These neurons are activated in striking concordance with the RPE, increasing their firing rate when the model predicts a positive RPE and reducing their firing rate when the model predicts a negative RPE. The TD model of conditioning proposes that a sensory cue (CS) is represented with a vector of signals, which represents the stimulus at each point in time. The representation of the state value at each point in time can be adjusted by its own unique weight (Figure 1-5C). This temporal representation is called a ‘complete serial-compound stimulus’ (Schultz et al., 1997).
Figure 1-5: Error-driven learning in classical conditioning

A) The associative value (V) of a CS increases after each CS-US pairing, in a RW model. As the US becomes more predictable, the rate of change of V gets smaller, producing a logarithmic learning curve. Adapted from (Rescorla & Wagner, 1972). B) The firing rates of midbrain dopamine neurons from behaving monkeys during conditioning. Firing rate increases in response to an unexpected reward (R) but after the monkey learns that the CS predicts the reward, firing rate only increases in response to the CS. If the predicted reward is omitted, there is a dip in firing rate at the time when the reward was expected. Adapted from (Schultz et al., 1997) with permission from AAAS. C) The temporal representation of a CS for TD learning in the brain. A stimulus is represented at multiple temporal delays \( x_n \) from its initial time of onset and each delay is associated with a separate adjustable weight \( w_n \). These weights are adjusted according to the correlation of activity \( x_n \) and prediction error, and through training come to act as predictions. The TD error is the sum of the surprise signal (\( \dot{V}(t) \)) and the reward (\( r(t) \)) and can explain changes in the firing rates of dopaminergic neurons in the ventral tegmental area (VTA). Adapted from (Schultz et al., 1997) with permission from AAAS.
This correspondence between theory and biology is now accompanied by causal evidence that RPEs are encoded by dopamine per se. In a typical blocking paradigm, there is no learning of the CS2 when a compound CS1-CS2 stimulus is presented. However, when dopaminergic neurons are optogenetically stimulated at the same time as the presentation of the compound stimulus, blocking is prevented and the animal will learn that the CS2 predicts reward (Steinberg et al., 2013). This is consistent with the idea that, in blocking, there is no learning of the CS2 because there are no RPEs. Furthermore, by using chronically implanted carbon microelectrodes, rapid changes in extracellular dopamine concentrations can be measured (Clark et al., 2010) and it has been shown that changes in dopamine concentration are sufficient to encode both positive and negative RPEs (Hart et al., 2014).

Functional neuroimaging studies in humans also support the TD learning model of dopamine. The ventral striatum, which is densely innervated by midbrain dopaminergic neurons, activates more strongly in response to unexpected reward than expected reward (Berns et al., 2001). This signal can be predicted with the prediction errors from an explicit TD model, and shows the same temporal pattern as that seen in electrophysiological studies in monkeys, shifting from the US to the CS throughout conditioning, and diminishing when an expected reward is omitted (O’Doherty et al., 2003). These prediction error signals in human imaging studies are also thought to be dopaminergic, as drugs that enhance or reduce dopaminergic function, increase and reduce the PE signal respectively (Pessiglione et al., 2006).

1.2.4 Learning beyond rewards

Reinforcement with reward or punishment is not necessary for learning. This was most famously demonstrated by Edward Tolman, who exposed rats to a spatial maze for ten days. One group was rewarded every time they successfully completed the maze, whilst a second group received
no rewards. On the eleventh day, reward was introduced to the second group of rats. He found that these rats were now even quicker and more accurate than the rats who were rewarded from day one (Figure 1-6A), suggesting that the second group had learned something about the structure of the maze, despite their behaviour not being reinforced (Tolman & Honzik, 1930). Tolman thus introduced the idea that rats were using ‘latent learning’ to form a cognitive map of the environment.

Learning an explicit model of the world is thought to be required for goal-directed behaviour, whilst learning purely from reinforcement without a model of the world is thought to underlie habitual behaviour (Dolan & Dayan, 2013, Lee et al., 2014). Goal-directed behaviour is characterised by decisions that lead to outcomes currently valuable to the animal, whilst habitual behaviour is characterised by actions that have been previously reinforced by rewards, but might not necessarily lead to rewarding outcomes anymore. It is therefore useful to have an explicit model of any environment that is prone to change, to enable behavioural flexibility.

Whereas TD learning makes use of RPEs to learn about the values of states and actions, other forms of error-driven learning can enable animals to learn associations between neutral states in order to simulate a model of the environment. These kinds of learning signals can be operationalised as ‘state prediction errors’, which allow an animal to update estimates of a transition matrix that specifies the probabilities of transitioning from one state of the environment to another. In humans, the intraparietal sulcus and lateral prefrontal cortex show activity correlated with state prediction errors (Figure 1-6B) during structure learning in a task analogous to Tolman’s maze (Glascher et al., 2010).

Model-free and model-based systems were once thought to be parallel controllers, implemented in separate neural circuitries, but midbrain dopamine neurons are becoming increasingly implicated in both (Gardner et al., 2018). In complex learning tasks humans employ a mixture of model-based and model-free control. Accordingly, striatal activity in humans can be partially explained with RPEs that would be generated by a model-based
learner (Daw et al., 2011) and also with state prediction errors (Guo et al., 2016). Furthermore, L-DOPA can enhance the degree of model-based control (Wunderlich et al., 2012), suggesting that dopamine is not only involved in signalling RPEs but may be used in other computations.

Several recent experiments have called into question the primacy of value in dopaminergic signalling, by showing that dopamine is necessary for ‘identity prediction errors’, where the value of a reward is predictable, but the identity of the reward (the sensory features of the rewarding stimulus) is surprising. One electrophysiological study in rats found that midbrain dopaminergic neurons signal prediction errors when a sensory feature of a reward, like flavour, changes unexpectedly, but value remains constant (Takahashi et al., 2017). An equivalent result has since been found in humans, using midbrain blood-oxygen-level dependent (BOLD) response as a proxy for dopaminergic signalling (Howard & Kahnt, 2018). An optogenetic study using unblocking has also demonstrated that dopamine is in fact necessary for this identity prediction error (Figure 1-6C). Unblocking is a paradigm similar to blocking but where the compound (CS1-CS2) stimulus is followed by a larger reward than in the first conditioning phase, so that learning is unblocked and the animal develops a conditioned response to the CS2. Unblocking is inhibited by optogenetic inactivation of dopamine neurons. Remarkably, the same result is also seen in a paradigm where the compound stimulus is followed, not by a change in reward magnitude, but by a change in a sensory feature of the rewarding stimulus, like flavour (Chang et al., 2017). This paradigm demonstrates that learning can be unblocked by pairing the compound stimulus with a reward that has unexpected sensory features, and that this unblocking is dependent on dopamine. These results bring attention to the fact that many previous reward-learning experiments had confounded changes in value with changes in sensory features, prompting the question of whether dopamine signals a generalised error for event prediction, of which value is merely one dimension.

These recent results can be explained by incorporating state inference into the RL model. By training mice on conditioning tasks where there is state
uncertainty resolved by the receipt of reward itself, dopaminergic neurons can be shown to encode RPEs that incorporate a pre-computed state inference. In one study, a CS was followed by reward after a variable delay 90% of the time, but was not followed by any reward 10% of the time. On rewarded trials, as time elapses, the mouse infers that it is less likely to be in a ‘pre-reward’ state and correspondingly, dopaminergic RPEs are larger after longer time delays (Starkweather et al., 2017). A subsequent lesion study demonstrated that the mPFC is necessary for the time-varying effect in this ambiguous paradigm, but not in an altered paradigm where reward is always delivered (Starkweather et al., 2018). Further evidence for belief state representation in dopaminergic RPEs comes from a similar study where ambiguous states were defined by reward magnitude rather than reward onset (Figure 1-6D). After training mice to learn that a neutral cue was randomly followed by a block of five small rewards or five large rewards, new trials with intermediate-sized rewards were introduced. A conventional RL account predicts that dopaminergic RPEs will increase monotonically with reward size, whilst an RL model that uses state-inference predicts that the mouse will use the size of the first reward to resolve uncertainty about whether it is in a high- or low-reward block. For instance, if the animal receives 4 μl water, it will infer that it is in a low-reward state and therefore produce large RPEs as it continues to receive more reward than expected, whilst if the mouse receives 6 μl water, it will infer that it is in a high-reward state and therefore produce small RPEs as it continues to receive less reward than expected. Both the behaviour and the dopaminergic activity of the mice were better explained by this latter account (Babayan et al., 2018).

In light of these findings, an identity prediction error could be interpreted as an RPE conditioned on a state inference, where the state is defined by the sensory features of the reward itself. For instance, in the optogenetic study described on page 38, learning is unblocked when a novel flavour reward is introduced. From the state inference perspective, if the animal infers that it is in a new ‘chocolate’ state, it will generate an RPE because this is an unexpected and rewarded state.
Representations of sensory states, in a cognitive model of the environment, are fundamental in guiding animal behaviour. Neural responses to unexpected sensory events, unrelated to reward, have long been seen in humans, using electroencephalography (Courchesne et al., 1975, Fabiani & Friedman, 1995) and functional magnetic resonance imaging (fMRI) (Opitz et al., 1999, Strobel et al., 2008). More recently, researchers have found evidence for hierarchical, precision-weighted prediction error signals in humans and monkeys, during audio-visual learning, in brain areas specific to the sensory modality in which learning is taking place (Iglesias et al., 2013, Meyniel & Dehaene, 2017, Chao et al., 2018). These findings support ‘predictive coding’ accounts of cortical function (Rao & Ballard, 1999, Bastos et al., 2012) whereby microcircuits in cortical columns provide a bidirectional channel for feed-forward sensory information and top-down expectations.
A) ‘Latent learning’ in rats trained on a spatial maze. Rewarded rats learn the maze more quickly than unrewarded rats. However, rats that are only rewarded from day eleven perform best, supposedly because they exploit a cognitive map acquired from latent learning in the first ten days. Adapted from (Tolman & Honzik, 1930). B) Evidence of a state prediction error (SPE) and reward prediction error (RPE) in separate brain regions (right) in humans navigating a stochastic non-spatial environment (left). SPEs are thought to enable the kind of latent learning that allows animals to learn a model. Adapted from (Glascher et al., 2010). C) An illustration of unblocking with identity prediction errors, as demonstrated in (Chang et al., 2017). In phase 1 an animal associates a neutral light stimulus with a banana-flavoured reward and acquires a conditioned response to the light. In phase 2, a compound light-tone stimulus is followed by a chocolate-flavoured reward that has the same subjective value as the banana-flavoured reward. Due to the unexpected change in reward identity, learning is unblocked and the rat acquires a conditioned response to the tone. Unblocking is prevented if dopamine neurons are inactivated, suggesting that dopaminergic prediction...
errors signal unexpected changes in stimulus properties other than value. Adapted from (Keiflin & Janak, 2017). D) Evidence for state-inference in mice. In this classical conditioning task, mice associate a neutral odour-tone compound with a reward, which was unpredictably large or small. After twenty days of training, new blocks with intermediate sized rewards were introduced. The activity of dopamine neurons can be explained by an RL model that incorporates inference about hidden states. Adapted from (Babayan et al., 2018) with permission under a Creative Commons Licensecreativecommons.org/licenses/by/4.0/.

Error-driven learning is the common denominator underlying these models and data. Prediction errors allow animals to learn about states and values, so that they can execute behaviours along a whole spectrum of sophistication, ranging from rigid Pavlovian responses to reinforced habits to flexible goal-directed choices. Predictions and prediction errors appear to be fundamental units of computation for processing the world. The next chapters will explore how this computational architecture can be used to process information in such a way that social systems could emerge.

1.3 Processing ‘your’ world

1.3.1 What is a social system?

Sociologists have advocated many different theoretical frameworks for defining and studying social systems. Structural functionalism, for instance, is inspired by the notion that social systems are subject to Darwinian evolution (Spencer, 1876). Here, social systems are treated like biological systems, composed of constituent structures that enable the system to function as a stable whole (Parsons, 1961, Durkheim & Parsons, 1974). Conflict theories, on the other hand, describe social systems in terms of conflicting goals between different groups, defined by features such as social class (Marx & Engels, 1848), race (Du Bois, 1910) or gender (Addams, 1902). Symbolic interactionism is another school of thought that emphasises the role of micro-scale interactions between individuals, in order to construct a shared understanding of reality (Mead, 1934).
Chapter 1: Introductory material

A feature common to these frameworks is that multiple individuals behave in a co-ordinated fashion. Regardless of whether they are collaborating or competing, the individuals track and respond to each other’s behaviours and this co-ordination results in an emergent social system.

The ability for an individual to predict another’s behaviour is thus necessary for the emergence of human social systems, but perhaps not sufficient. Insects are adept at monitoring each other (Worden & Papaj, 2005, Leadbeater & Chittka, 2008, Battesti et al., 2012, Leadbeater & Dawson, 2017) and are so efficient at co-ordinating their behaviour that humans have even looked to insects to inspire solutions to their own design challenges (Holbrook et al., 2010). It would be surprising if insects had all the social cognitive capacities that humans have.

Humans can also represent each other’s minds; they have a ‘Theory of Mind’ (Premack & Woodruff, 1978b, Wimmer & Perner, 1983, Baron-Cohen et al., 1985). More recently, this has been framed as an ability to predict each other’s mental states (Koster-Hale & Saxe, 2013, Tamir & Thornton, 2018, Thornton et al., 2019c). Not only can humans co-ordinate their behaviour, but they can co-ordinate their beliefs and desires, or in the parlance of RL, their value functions. At the social systems level, this is most evident in the emergence of norms and morals. These forces, which guide human behaviour, provide the type of ‘higher-level control’ that Polanyi described (see section 1.1.2). The emergence of norms and morals requires humans to incorporate other people’s expectations into their world model.

Growing evidence points to the role of the prefrontal cortex in incorporating social norms into choice behaviour. The right lateral prefrontal cortex (rPFC), in particular, shows stronger activity when humans engage in norm-compliant behaviour triggered by threats of social punishment (Spitzer et al., 2007). More recent studies have provided causal evidence that this brain region is involved in arbitrating between decisions that lead to personal gain and decisions congruent with social norms, rather than merely encoding the risk of social punishment. Transcranial direct current stimulation (tDCS) is a non-invasive neurostimulation procedure, thought to modulate cortical
excitability, in a manner dependent on the polarity of the current. Cathodal stimulation is thought to reduce excitability whilst anodal stimulation is thought to increase excitability (Nitsche & Paulus, 2000). Cathodal stimulation of the rPFC reduces people’s propensity to engage in sanction-induced norm-compliance, whilst anodal stimulation increases the amount of norm-compliant behaviour (Ruff et al., 2013). This suggests that the rPFC is required for decisions motivated by social norms, in contrast to the desires of the Self. Consistent with this, anodal tDCS over the rPFC has been seen to enhance honesty in situations where dishonesty would be consistent with material self-interest (Marechal et al., 2017).

In most contexts, it is difficult to disentangle norm-compliance from ‘prosocial’ behaviour intended to benefit another. One recent study attempted to tackle this issue by imposing arbitrary rules on participants, which, if followed, could be beneficial or deleterious to Self or Other (Gross et al., 2018). This design enabled the researchers to orthogonalise rule-following from prosociality. The results showed that anodal tDCS over the rPFC led to more rule violations, when rule-following would lead to behaviour incongruent with what the participant would freely choose. Cathodal tDCS led to more rule-following, even in situations when rule-following led to monetary losses for Self or Other. Taken together, these results suggest that the PFC does not simply suppress selfish behaviour and promote prosocial behaviour, but rather integrates the costs and benefits of internal goals and external rules and restrictions, which are highly dependent on experimental and ecological contexts.

A broader view of PFC function looks at the resolution of conflict between competing decision attributes, of which Self-interest and Other-interest are merely one exemplar pair of attribute types (Nunez et al., 2005, Liston et al., 2006, Hunt et al., 2014, Maier et al., 2018). Nevertheless, a neural circuitry that incorporates a cost-analysis of competing attributes into choice behaviour, may be one component of the cognitive architecture required to enable the emergence of social norms, and social systems more generally. The temporo-parietal junction (TPJ) is another region that has
been causally linked to resolving the conflict between morality and self-interest (Obeso et al., 2018). The TPJ and various regions of the PFC are engaged in many other social contexts, so are often referred to as components of a ‘social brain’ (Amodio & Frith, 2006, Blakemore, 2012).

The subsequent sections will review the cognitive components that are necessary for humans to track each other’s behaviour and mental states, and in so doing, will discuss the role of the so-called ‘social brain’. These sections will appraise claims that ‘social information’ is of a distinct category. The chapter will lead to a discussion about which low-level computations are necessary to compare Self-attributed goals with Other-attributed goals, a high-level processing stage that contributes to the emergence of social systems such as norms and moral principles.

1.3.2 Behaviour reading

Consistent with the idea that social systems emerge from a co-ordination of behaviour and mental states, there has recently been a growing call for social cognition to be studied using interactionist paradigms (Schilbach et al., 2013, Schilbach, 2016). A conventional experimental approach is to test a subject whilst they observe and predict another agent’s behaviour, in a ‘third-person’ manner, or put themselves in the other agent’s shoes and directly simulate their experience, in a ‘first-person’ manner. An alternative ‘second-person’ approach, focuses on the interaction between a pair (dyad) or group of individuals and examines their collective behaviour, rather than the individuals in isolation. This is arguably the most ecologically valid approach to investigating social cognition.

Interactive social paradigms have shed light on many aspects of social cognition. For instance, such paradigms have revealed that collective decision making is superior to individual decision making only when the observers have similar perceptual sensitivities (Bahrami et al., 2010). They have also found neural predictors for the emergence of leader-follower roles in dyadic interactions (Konvalinka et al., 2014) and shown that dyads are
more likely to benefit from social facilitation\textsuperscript{5} if they exhibit a higher inter-subject correlation of task-related neural activity (Szymanski et al., 2017). However, any claim that social cognition can only be understood with an interactionist approach rests on the assumption that cognitive processes during interactions are of a distinct ‘social’ mode, which cannot be measured in ‘non-social’ contexts.

An alternative stance is that there is a set of computations, which, when solved together, enable the emergence of social systems, but are independently useful in non-interactive contexts. Many researchers implicitly adopt this latter position, investigating specific cognitive processes that are intuitively necessary for a full-blown social interaction. This approach sacrifices ecological validity in return for a tighter experimental control over a specific cognitive process.

In order to engage in social interactions, individuals must be able to track each other’s behaviour, and thus detect the presence of another person. Considerable research has been dedicated to investigating how humans process sensory stimuli related to other agents, including face-like stimuli, gaze direction and biological motion. Humans have long been thought to detect and process faces differently from objects (Yin, 1969, Bruce et al., 1991), an idea galvanised by the discovery of a region in the human fusiform gyrus, which shows preferential activation in response to faces (Kanwisher et al., 1997). Newborn infants, as young as 9 minutes, show a preferential orienting bias towards face stimuli (Goren et al., 1975). Furthermore, a recent attempt to present perceptual stimuli to foetuses in utero, by projecting light through the uterine wall, found that foetuses made more head turns towards a top-heavy, face-like light pattern than an inverted pattern (Figure 1-7A), suggesting that postnatal experience is not required for this face-orienting bias (Reid et al., 2017). Despite these striking findings, there is considerable debate as to whether these behaviours reflect an innate

\textsuperscript{5} Social facilitation (a form of ‘audience effect’), is a phenomenon whereby people perform differently in the presence of others. This was first reported by Norman Triplett, who observed that cyclists were faster when accompanied by a pacemaker (Triplett, 1898).
bias towards face stimuli (like the innate library of primary reinforcers described in section 1.2.2) or whether they stem from general properties of visual processing in infants, that eventually lead, through learning, to attentional biases in face processing (Reynolds & Roth, 2018). Similar findings have been reported in the domain of biological motion, in that infants will preferentially look at point-light displays that follow the kinematics of biological versus non-biological motion (Simion et al., 2008). Finally, by six to nine months of age, infants can generally direct or follow the visual gaze of social partners (Mundy & Jarrold, 2010).

Evidently humans have specialised neurocognitive apparatuses for detecting social stimuli. Whether this specialisation emerges over phylogenetic or ontogenetic timescales is unclear. Regardless, it appears that sensory stimuli associated with other agents have dedicated input channels, enabling a rapid detection of social stimuli. By virtue of the fact that social stimuli have unique, idiosyncratic features, it is unsurprising that humans have learned to leverage these features to classify stimuli as social (e.g. ‘that’s a face’, ‘that’s a person walking’) or non-social (e.g. ‘that’s a mask’, ‘that’s some random dots moving’).

Once a social agent has been detected, behaviour reading requires the ability to learn, and form expectations, about what the agent will do next. Many instances of this kind of ‘social learning’ have be explained with the typical computational architecture of prediction error minimisation, that underlies more general learning theories (de Bruin & Michael, 2018). For instance, regions of the so-called ‘social brain’ such as the superior temporal sulcus signal prediction errors when humans view other people behave in unexpected ways (Saxe et al., 2004, Costantini et al., 2005, Saygin et al., 2012, Heil et al., 2019). Beyond action observation, and considering more abstract features of other social agents, regions of the ‘social brain’ also exhibit surprise signals in response to viewing someone respond in an unexpected way to an emotional stimulus (Pegado et al., 2018).

The neural correlates of learning about another agent’s behaviour appear similar to those of learning about environmental states. Indeed,
conceptualising other social agents as basic environmental stimuli has provided an elegant explanation for observational learning. Many animal species, including humans, engage in this apparently social behaviour, whereby the individual learns not about another agent’s behaviour, but about the environment, from another agent’s behaviour. Bumblebees, for instance, can identify which unfamiliar flowers will yield food by observing the behaviour of conspecifics (Worden & Papaj, 2005). However, this copying behaviour is absent in bees that haven’t previously associated conspecifics with appetitive stimuli, and is actually reversed (i.e. the bees avoid popular flowers), if they have previously associated conspecifics with aversive stimuli (Dawson et al., 2013). This can be explained with second-order conditioning (Figure 1-7B). First, the bees associate a neutral stimulus (conspecific) with an appetitive stimulus (food). Then subsequently, in the second-order conditioning phase, they will associate the conditioned stimulus (conspecific) with new flowers and so will choose to approach these flowers. It might then be the case that any animal that can form associations between environmental states can learn ‘socially’ (Heyes, 2012).

The mechanisms by which humans learn about abstract traits of Others also conform to fundamental principles that govern value learning and latent learning of inanimate states in the environment. In an economic exchange game where subjects learn about how trustworthy another participant is, the striatum shows a stronger signal in response to benevolent behaviour than malevolent behaviour (Figure 1-7C). Furthermore, this signal occurs immediately after the Other makes their decision of how much to donate to the subject, but as the game progresses, the signal transfers back in time and becomes anticipatory of the imminent decision of the Other (King-Casas et al., 2005). This finding echoes the temporal transfer of RPE signals from the onset of reward to the onset of a conditioned stimulus (see section 1.2.3), where the conditioned stimulus here is the appearance of the social partner, whose behavioural traits are being learned about. Furthermore, associative learning signals, encoded in ‘social brain’ regions, are correlated
with learning about another person’s volatile trustworthiness (Behrens et al., 2008), as their motives change during the course of an experiment.

In addition to predicting unnatural, task-based measures, associative learning signals can also predict high-level, abstract facets of social cognition. ‘Generosity prediction errors’ pertaining to the behaviour of another person, drive generosity representations in the vmPFC, which predict how likely a subject is to choose to interact with the Other in a novel, out-of-task context (Hackel et al., 2015). Additionally, another form of ‘social prediction error’ relating to how positively or negatively another person evaluates the subject, has been shown to drive moment-to-moment changes in reported self-esteem (Figure 1-7D), which correlate with signal changes in the vmPFC (Will et al., 2017).

Figure 1-7: Detecting and learning about social agents

A) Mature human foetuses make preferential head turns towards face-like (top-heavy) light stimuli, projected through the uterine wall. Adapted from (Reid et al., 2017) with permission under a Creative Commons License creativecommons.org/licenses/by/4.0/. B) Observational learning in bumblebees can be explained by second-order conditioning. In an experiment by (Dawson et al., 2013), bees are first conditioned to associate conspecifics
with either appetitive (left top panel) or aversive stimuli (left bottom panel). In the next phase (middle panel), bees observe demonstrators on green flowers. In the choice phase, only bees trained to associate conspecifics with appetitive stimuli in phase one will copy the demonstrators. Reproduced from (Leadbeater & Dawson, 2017). C) BOLD response in the human caudate during an economic exchange game. If the Other is benevolent to the subject, resulting in greater future trust, a larger signal change is seen than for decisions that result in reduced future trust. In early rounds (top panel), this signal difference occurs after the decision of the Other is revealed (second vertical grey bar). In later rounds, this effect transfers to a timepoint before the decision is revealed, echoing patterns seen in temporal-difference learning. Adapted from (King-Casas et al., 2005) with permission from AAAS. D) Social prediction errors in the ventral striatum and anterior cingulate cortex (top left) allow a subject to learn about what another person thinks of them, and explain moment-to-moment changes in reported self-esteem (bottom). Changes in self-esteem are correlated with signal changes in the vmPFC (top right). Adapted from (Will et al., 2017) with permission under a Creative Commons License creativecommons.org/licenses/by/4.0/.

By explaining behaviour reading in terms of non-social\(^6\) computations, many of these results dilute the notion of a distinct ‘social information’. It appears that expectations about Others’ fleeting movements and enduring personality traits can both be learned by treating the Other just like an inanimate object in the environment, albeit with strong affordances that attract attentional biases. Whilst the neural signals that correlate with these learning variables are often localised to regions of the ‘social brain’, it is unclear whether this reflects a distinction in how information about Others is processed, or simply that social information tends to arrive through a specific set of sensory input channels, specialised for detecting social stimuli, like faces, gestures and voices. If a dedicated computational architecture is indeed required for learning about Others, perhaps this will be found not in the behaviour reading apparatus, but in the mind reading apparatus.

---

\(^6\) The term ‘non-social’ here refers to the fact that, of the models used to explain the behavioural and neural data in these experiments, none invoke any algorithmic ingredients over and above those required for learning in contexts where there are no other agents present.
1.3.3 Mind reading

Beyond generating predictions about another agent’s external behaviour, the ‘social brain’ also generates predictions about another agent’s internal mental states (Thornton et al., 2019c). Indeed, it is largely their role in mentalising through which these brain regions have gained their status as ‘social’. The importance of the PFC for social cognition in humans was discovered through the behaviour of patients with frontal lesions (Eslinger & Damasio, 1985, Price et al., 1990). These people had deficits in higher order cognition, including the ability to reason about morals and engage in ordinary social interactions. Subsequent functional imaging studies, in healthy adults, showed that the mPFC had greater activity when subjects either recognised ‘mind-related’ terms (Baron-Cohen et al., 1994), or took the point of view of another person (Goel et al., 1995). Many studies have since reported activity in the mPFC, temporal poles and TPJ when subjects read stories that reference mental states, such as beliefs or desires (Fletcher et al., 1995, Saxe & Kanwisher, 2003).

Adults are generally very good at representing another person's beliefs, even when those beliefs differ from their own. This is a skill, so frequently employed by humans, that it’s routinely exploited for artistic effect. Plenty of Shakespeare plays rely on ‘dramatic irony’, a theatrical device where the audience understands more about the situation than the characters. For instance, in Romeo and Juliet, Romeo takes his life based on a false belief that Juliet is dead, whilst the audience is painfully aware that she is still alive. The frustration and distress that the audience experiences are due to the fact that they are able to represent Romeo’s false beliefs in addition to their own accurate beliefs. This ability, to represent another person’s false belief, is typically tested by researchers as a gold standard measure of mentalising ability (Wimmer & Perner, 1983, Baron-Cohen et al., 1985, Saxe et al., 2006).

Cognitive scientists and philosophers of mind have historically been divided, in their opinions of how humans mentalise, into two camps, ‘Theory Theory’ (TT) and ‘Simulation Theory’ (ST) (Gordon, 1986, Gopnik &
Wellman, 1992, Henderson, 1996, Arkway, 2000, Apperly, 2008). More recently, this dichotomous framework has spilled over into the design and interpretation of social neuroscience experiments (Vogeley et al., 2001, Ramnani & Miall, 2004, Grezes et al., 2004, Mitchell et al., 2006, Suzuki et al., 2012). Broadly, TT proposes that humans reason about other minds through an innate, or learned, set of rules. For instance, if an observer sees someone crying, they might deploy a rule such as ‘people who are crying are sad'. ST, on the other hand, proposes that humans reason about other minds by putting themselves into the shoes of the other person. ST requires the subject to represent the mental states that they themselves would have in that situation, and then attribute those mental states to the Other.

TT and ST make very general claims, grounded in qualitative descriptions of cognition, which are open to various interpretations. A strong form of ST, for instance, requires that the same neural mechanisms that underpin the formation of a mental state, such as a belief, are co-opted for inferring the mental state of an Other, and that the outputs of this mechanism must serve as the basis of mental state attribution (Saxe, 2009).

The discovery of mirror neurons in the macaque premotor cortex (di Pellegrino et al., 1992) provided ammunition for the strong form of ST. These neurons fire when a monkey executes an action, but also when the monkey observes another agent execute that same action. Thus, the neurons appear to encode some abstract property of a specific action, that is independent of the agent performing it. There is some evidence that this abstract property may indeed be a mental state, such as an intention. One study reported evidence for mirror neurons that could discriminate between two types of observed grasping movements. The first type of movement was made with the intention to place a piece of food into a container. The second type was made with the intention to eat. The neurons had different activity patterns in response to observing these two types of action sequences, even before they were visibly different (Fogassi et al., 2005). A further study found that mirror neurons only respond to object-directed actions, not mimed actions,
and will respond to object-directed actions even if the object is occluded (Umilta et al., 2001).

Figure 1-8: Simulating Other’s computations in the vmPFC

A) Evidence from fMRI studies in humans. Top left and right: BOLD signal in a vmPFC cluster covaries with the magnitude of an observational RPE, in an observational learning task. Adapted from (Burke et al., 2010). Bottom left and right: BOLD signal in a vmPFC cluster covaries with the magnitude of a simulated RPE signal in a mentalising task. Adapted from (Suzuki et al., 2012). B) Evidence from single cell recordings in human rostral ACC (rACC) during an observational learning task. Regressing firing rate on expected value (exp val) and outcome (amount) at multiple timepoints, timelocked to the outcome for Other, yields a timecourse of t-statistics. The marker labelled ‘1’ shows that, before the outcome, firing rate is positively correlated with expected value. The markers labelled ‘2’ and ‘3’ show that, after the outcome, firing rate is positively correlated with outcome and negatively correlated with expected value, respectively. This activity profile is consistent with a simulation of the other agent’s TD error. Adapted from (Hill et al., 2016) with permission under a Creative Commons License creativecommons.org/licenses/by/4.0/. C) Evidence from an optogenetics study in mice. An observer mouse (labelled red) experiences an uncued shock and then undergoes observational conditioning, by observing a demonstrator mouse (labelled blue) receive shocks, cued by a neutral stimulus. The extent of conditioning
can be tested by measuring how much the observer mouse freezes in response to the CS, in absence of shock. Freezing on test day is significantly reduced in observer mice which had optogenetic inhibition (NpHR) of neurons projecting from the ACC to the basolateral amygdala, during conditioning, compared with control mice (eYFP). Adapted from (Allsop et al., 2018).

These results strengthen the hypothesis that mirror neurons encode information about the intentions of Other. However, as Rebecca Saxe notes (Saxe, 2009), the different intentions of the observed actor are confounded by subtle differences in action sequences; indeed it is the small kinematic differences in action sequences that the subject is leveraging to decode the intention of the Other. Thus, the jury is still out on whether these co-opted neural resources are encoding mental states per se, or merely physical features of the action sequence.

A weaker form of ST is agnostic about the neural implementation of mental state inference, whilst specifying that the same algorithms be used for computing Self-attributed mental states and Other-attributed mental states (Suzuki et al., 2012). Work on observational learning has provided compelling neurobiological evidence for this form of simulation. Neurons, recorded in the anterior cingulate gyrus (ACCg) of macaque monkeys, fire in response to observing another monkey receive a food reward (Chang et al., 2013). Medial prefrontal neurons in the macaque brain also encode the value of a conditioned stimulus that predicts the delivery of reward to another monkey (Noritake et al., 2018). Similarly, fMRI studies in humans show increased ACCg activity, in subjects observing an actor receive monetary reward (Lockwood et al., 2015, Shimada et al., 2016). Furthermore, this signal is modulated by the relative costs and benefits incurred by the actor in obtaining the reward (Appas & Ramnani, 2014).

Causal evidence that the mPFC is actually necessary for observational learning is seen in a human lesion study and a mouse optogenetics study (Figure 1-8C). The lesion study found that patients with vmPFC damage are less influenced by the observed outcomes for Others, in their future choice behaviour, than healthy controls (Kumaran et al., 2015).
The mouse study found that a neural projection from the ACC to the basolateral amygdala is necessary for acquiring a conditioned fear response from observing another mouse receive foot shocks. Optogenetic inhibition of this pathway prevented the acquisition of the observationally learned, conditioned fear response but did not prevent the acquisition of a classical conditioned fear response formed from directly experiencing the shock (Allsop et al., 2018). Taken together, these observational learning studies suggest that the mPFC, and particularly the ACC, is involved in encoding the values of states experienced by Other.

Computational modelling has enabled researchers to show that, in addition to simulating the value of a state experienced by Other, the mPFC can simulate the Other’s RPEs per se (Figure 1-8A-B). These simulated RPEs have been observed in humans engaged in observational learning tasks, using both fMRI (Burke et al., 2010) and single-cell recordings (Hill et al., 2016), explicit mentalising tasks (Suzuki et al., 2012) and social teaching tasks (Apps et al., 2015). Additionally, one recent fMRI study investigated whether humans can infer another agent’s action values by observing their choices but not outcomes, assuming some knowledge about the other agent’s preferences (Collette et al., 2017). This form of learning, dubbed ‘inverse RL’, requires subjects to recover the reward outcome distribution for which the other agent’s choices are optimal (Ng & Russell, 2000, Abbeel & Ng, 2004). Collette et al. showed that the dmPFC covaried with the ‘inverse RL’ outcome predictions. Thus, the dmPFC was representing action values attributed to Other. It seems then, that humans can simulate the computations of Others by representing decision variables, such as value expectations and RPEs, in the mPFC.

Simulated learning signals have also been detected in the basal ganglia. Activity in the ventral striatum is inversely related to the simulated RPE of another agent (Burke et al., 2010) (Figure 1-9C). This finding has been mirrored in recordings from midbrain dopamine neurons in macaques (Figure 1-9A-B), which fire more strongly in response to cues that are of lower value to the Other (Noritake et al., 2018). However, in one
observational learning task in humans, dorsal striatal activity was positively correlated with simulated RPEs (Cooper et al., 2012). Noritake and colleagues have proposed that striatal RPEs are multiplexed signals, incorporating information about objective value for both Self and Other, to produce a ‘subjective’ RPE (Noritake et al., 2018). The way in which value for Other is incorporated into this signal is likely to depend to subtle design features of the cognitive task (Cooper et al., 2012). Beyond model-free reward learning, one human fMRI study has also provided evidence for simulated sensory surprise signals in the intraparietal sulcus, consistent with a simulation of another agent’s model-based computations (Dunne et al., 2016).

Despite the large body of neural evidence in favour of an ST account of mentalising, some human behaviours speak in favour of TT. The fact that children and adults both make systematic errors in mentalising tasks suggests that many humans use similar ‘theories’, which lead to predictable errors (Saxe, 2005a). For instance, four-year-old children systematically conflate ignorance with incorrectness (Ruffman, 1996a). In Ruffman’s experiment, four-year-olds observed a researcher take a green bead from a dish of red and green beads. Another agent was present, who knew that the dish contained red and green beads but did not see the colour of the chosen bead. When asked to comment on the other agent’s belief about the colour of the chosen bead, children usually said that the Other believed the bead was red. Systematic mentalising errors, such as this, have been interpreted as evidence for an incomplete, developing folk theory of Others’ minds. In this case, a crude theory that ‘ignorance begets incorrectness’ is supposedly employed. The ambiguous definitions underlying TT and ST, and the evidence in favour of both, have prompted calls for a hybrid of the two accounts (Mitchell, 2005, Saxe, 2005b, Apperly, 2008, de Bruin & Michael, 2018, Tamir & Thornton, 2018).
Figure 1-9: Simulating Other's computations in midbrain and striatum

A) Two macaque monkeys learn about six neutral stimuli in a classical conditioning task. In the ‘Self-variable’ block, each CS predicts reward with equal probability \( P(r) \) for Other but variable probability for the subject. In the ‘Other-variable’ block, each CS predicts reward with equal probability for the subject but variable probability for Other. In the Self-variable block, subjects show more appetitive licking behaviour in response to cues that signal a high probability of reward. In the Other-variable block, subjects show more appetitive licking behaviour in response to cues that signal a low probability of reward. Adapted from (Noritake et al., 2018).

B) Recordings from midbrain dopamine neurons, during the same experiment as in A), show that dopamine spike rate follows the same pattern as licking behaviour. Spike rate increases with reward probability in the Self-variable block and decreases with reward probability in the Other-variable block. Dopamine neurons thus encode an inverse simulated value signal for the Other. Adapted from (Noritake et al., 2018).

C) A similar fMRI result from a human observational learning task. Ventral striatal BOLD signal is inversely related to the observational RPE timelocked to the outcome onset for the Other. Adapted from (Burke et al., 2010).

1.3.4 Am I reading your behaviour or your mind?

The debate on whether mentalising relies on ‘theories’ or ‘simulation’ is related to the question of whether infants and animals represent mental states at all. Despite a large body of research seeking to test whether ‘simpler’ organisms can attribute mental states to Others (Kovács et al.,
2010, Bugnyar, 2011, Buttelmann et al., 2017), even the most carefully designed experiments tend to fall short of distinguishing between mental state attribution and mere ‘smart behaviour reading’ (Butterfill & Apperly, 2013, Heyes, 2015, Borg, 2018, Burge, 2018). In reference to mentalising, Robin Dunbar stated in his paper outlining the Social Brain Hypothesis that:

'A more conventional behaviourist account based on simple associative learning can invariably be given for almost all examples in the literature.' (Dunbar, 1998)

In other words, most social behaviours can be explained in two ways. The first way is mentalistic and invokes representations of mental states. The second way is not mentalistic. For example, one recent study pioneered the use of infrared eye-tracking in apes to measure their anticipatory looking behaviour whilst they watched a video of a pair of intentional agents acting towards an object (Krupenye et al., 2016). In the video, character X saw character Y place an object into box A. Then character X left the scene and character Y moved the object into box B, and then took the object out of box B and left the scene with the object. Finally, character X re-entered the scene and the video stopped. The authors found that, upon X’s re-entry, apes made more saccades towards the box where character X had last seen the object. A mentalistic explanation would describe the apes as representing character X’s belief, and therefore anticipating that they would look inside box A. A non-mentalistic explanation might describe the apes as simply learning some associations between spatial configurations of features in the video. For instance, the reappearance of character X on the screen may have ‘acted as a retrieval cue, activating a memory of the object’s location when that character was last present’ (Heyes, 2016).

Although TT has been described as a way of representing and predicting Others’ mental states, it shares more similarities with the kind of non-mentalistic behaviour reading described above. Indeed, the only experimental approach that has been widely accepted as a suitable test for mental state attribution, is itself a test for simulation. This strategy makes use of a ‘self-informed’ belief induction variable (Heyes, 2014a). The technique
involves pretraining subjects so that they learn, through direct experience, that some environmental conditions beget true beliefs and other environmental conditions beget false beliefs.

In one study (Senju et al., 2011), for instance, 18-month-olds were split into two groups. Each group was familiarised with one of two identical-looking blindfolds. One blindfold was opaque and the other was translucent. The infants in each group learned that if they wore the blindfold, they could either see through it or not see through it. In the test phase, the infants watched a video where character X observed character Y place an object into box A next to an empty box B. Character X then put on a blindfold that looked identical to the one that the infants had been pretrained with. Character Y then removed the object from the box, and left the scene with it. Finally, character X removed the blindfold. Infants pretrained with the opaque blindfold made more anticipatory saccades to box A than infants pretrained with the translucent blindfold. Supposedly, this is because the infants could recall their own personal experiences (Self-attributed mental states) to infer character X’s belief. The group trained with the opaque blindfold infers that character X has a false belief that the object is still in box A, whilst the translucent group infers that character X has a true belief that the object had left the scene. The behaviour is thus explained by appealing to the infants’ memories of their own mental states, which can be used in the test phase, to attribute new mental states to the Other. A self-informed belief induction was also used in the observational conditioning study described in Figure 1-8 (Allsop et al., 2018). Here, mice only acquired the observationally conditioned fear response if they had already experienced the shock directly.

That the most convincing test of mental state attribution relies on recall of one’s own prior mental states, suggests that ‘simulation’ is entwined with the very concept of mental state attribution. Humans may learn folk ‘theories’ about Others, but without ‘simulation’, those theories treat other agents as mere objects rather than subjects.

Diana Tamir and Mark Thornton recently proposed a model (Figure 1-10) to ground behaviour reading and mind reading in a single predictive
processing framework (Tamir & Thornton, 2018). This framework includes a superficial layer that represents the observed behaviour of the observee, and a middle layer that represents the hidden mental states of the observee, which can be leveraged to make predictions about behaviour. The model also includes a third layer, which represents the traits of the observee, which can be used to make predictions about both mental states and behaviour.

The authors also suggest some specific dimensions along which mental states and personality traits can vary. Using existing theories about personality structure to explain distributed patterns of brain activity while subjects made social judgements about Others, the authors claim to find that people spontaneously encode three dimensions of personality traits: power, valence and sociality (Thornton & Mitchell, 2018). A similar approach taken during a mental state inference task also suggested that people encode three dimensions of mental state variability: rationality, social impact and valence (Tamir et al., 2016). Tamir and Thornton use their framework to reconcile the TT and ST approaches. They suggest that a TT style of cognition might involve making predictions about another agent by exploiting probable transitions within and between layers in the hierarchical state space, whilst an ST style of cognition might involve sampling less frequently traversed paths to explore alternative behaviours and mental states.
Tamir and Thornton propose that observers can make predictions about other agents by leveraging a hierarchical predictive model. They observe action, and can make predictions about subsequent action by representing the hidden mental states of the observee. Furthermore, in representing the hidden traits of the observee, predictions can be made about both mental state and action. Reproduced from (Tamir & Thornton, 2018).

One limitation of Tamir and Thornton’s framework is that it does not acknowledge any difference between representations and meta-representations. In other words, there is no qualitative distinction between representations of subjective mental states and representations of objective features of the environment. A key feature of a meta-representation is that there is an agent to whom the lower-level representation is attributed. Thus, there should be some way in which these representations of hidden mental states are selectively attributed to different agents, like Self or Other. Simply put, a simulation of another agent’s mental state is not useful, unless that state can be identified as belonging to Other and not to Self.
The rest of this thesis will be concerned with Self-Other distinction as a fundamental computation, necessary for the emergence of constructs that have been previously described as ‘meta-representation’, ‘Theory of Mind’ and ‘mentalising’. The question has shifted away from ‘How does an observer infer the mental states of an observee?’ to ‘how is it that a mental state can be selectively attributed to a particular agent?’.

1.4 My world or yours?

Humans can represent each other’s mental states, but it is not clear how the brain attributes those mental states to Other, rather than Self (Steinbeis, 2016). It is not clear how the brain encodes information about agent identity. As discussed in section 1.3.1, the emergence of social norms and morals likely depends on the ability to co-represent the goals of Self and Others. Whilst the work presented in this thesis will not be concerned with goals per se, it will investigate how other kinds of decision variables, namely predictions and prediction errors, are attributed to Self and Other. Most previous research on Self-Other distinction has been concerned with the physical, embodied states of Self and Other. The next section will provide a brief introduction to this field.

1.4.1 Feeling like me

Observing another person in a painful state can elicit feelings similar to being in the painful state oneself. Empathy for pain is one of the best studied areas of social neuroscience. fMRI studies in humans have shown that some areas of the distributed ‘pain network’ of the brain are active not only when experiencing pain, but also when observing another person in pain, in a manner reminiscent of the mirror neuron system (Singer et al., 2004, Lamm et al., 2011). Furthermore, subjective reports of the intensity of another person’s pain are reduced if subjects are given analgesic drugs (Mischkowski et al., 2016), or even placebo analgesia (Rütgen et al., 2015a).
This placebo ‘empathy analgesia’ is associated with changes in neural pain signatures that are similar to those associated with first-hand placebo analgesia (Rütgen et al., 2015a, Rütgen et al., 2015b), and is also modulated by opioidergic antagonism (Rütgen et al., 2015b).

Despite some shared neural circuitry for first-hand experience of pain and empathy for pain, there is a degree of egocentricity that biases empathy, sometimes described as an ‘empathy gap’ (Loewenstein, 2005). In other words, the simulation of another person’s bodily state is partly dependent on the subject’s own bodily state. Bridging this ‘empathy gap’ and overcoming biased empathic judgements is associated with activity in the right supramarginal gyrus (rSMG), a region next the TPJ (Silani et al., 2013). Silani and colleagues also found that disruption of the rSMG with transcranial stimulation increased the degree of egocentricity bias in a task where subjects had to consider another person’s affective state whilst having their own affective state independently manipulated.

Self is often considered to be an embodied construct, that is learned from experience with one’s own interoceptive inputs (Fotopoulou & Tsakiris, 2017, Friston, 2018), such that people come to predict their own bodily states using a ‘Self’ model (Apps & Tsakiris, 2014). Palmer and Tsakiris, for example, recently speculated that it is the contrast between high-precision interoceptive inputs as compared with lower-precision exteroceptive inputs, that allows humans to represent Others as agents that are distinct from Self (Palmer & Tsakiris, 2018).

However, over and above a bodily Self-Other distinction, humans can selectively attribute mental states to different agents. By representing a friend’s differing opinion, for instance, a more abstract kind of Self-Other distinction is engaged. Whilst humans do show some egocentricity bias when simulating another person’s bodily state, Self-Other distinction in this domain is relatively trivial, due to the divergent input channels that carry information

---

7 Interoception refers to perception of one’s own internal bodily states, for instance via afferent neurons that innervate visceral organs. It can be contrasted with exteroception, the perception of stimuli that are external to the subject’s body.
about Self and Other (Babo-Rebelo et al., 2019). In pain empathy studies, ‘my’ pain state is transmitted via nociceptive afferents from the site of injury, whilst ‘your’ pain state is transmitted via the visual system. Indeed, despite some overlap in the neural activations for first-hand pain and vicarious pain, these two experiences also have unique elements in their neural signatures, making them easily separable (Krishnan et al., 2016, Lopez-Sola et al., 2017). Self-Other distinction here is thus confounded by the interoceptive-exteroceptive distinction. In classic mentalising tasks, like the false belief task, the subject can represent two different belief states and selectively attribute them to Self and Other, without the help of a bodily Self-Other distinction. Thus, while research in the affective empathy domain can clarify the neural mechanisms underlying how it is that ‘I feel like me’, they don’t go far enough to clarify how it is that ‘I think like me’.

### 1.4.2 Thinking like me

The view that Self-attributed mental states are only meaningful with respect to Other-attributed mental states is not new. Peter Carruthers makes a particularly strong claim that humans don’t have exclusive introspective access to their own thoughts. His ‘Interpretative Sensory Access’ (IPA) theory suggests that humans evolved the ability to read Others’ minds for social purposes, and then learned to reflect on their own thoughts by turning mind reading in on the Self. In this sense ‘mind reading is prior to metacognition’ (Carruthers, 2009, Carruthers, 2011). The theory tries to explain why some neuropsychiatric patients confabulate the reasons behind their own actions. However, the theory also makes incorrect predictions about the behaviour of some rarer cases, such as a patient with combined ‘alien hand syndrome’ and ‘utilisation behaviour’ (Peters, 2013). Following Carruthers’ theory should lead researchers to use mind reading in social contexts as a starting point for understanding introspection. Regardless of whether ‘mind reading is prior to metacognition’, Self-Other distinction is a computation that is necessary for both.
Chapter 1: Introductory material

Functional and structural neuroimaging work in humans point to an overlap in the neural resources used for metacognition and mentalising (Saxe et al., 2006, Northoff et al., 2006, Valk et al., 2016, Vaccaro & Fleming, 2018). The most consistent finding here is that the vmPFC is involved, both in self-referential thought, and in thinking about other people’s thoughts. Beyond simply representing value, the vmPFC is known to play a role in comparing reward values to guide choices that are consistent with subjective preferences (Boorman et al., 2009, Levy & Glimcher, 2011, Lim et al., 2011, Hunt et al., 2012). The vmPFC also encodes information about other people’s preferences, even when they differ from one’s own (Mitchell et al., 2006, Jenkins et al., 2008, Ames et al., 2008, Tamir & Mitchell, 2010, Nicolle et al., 2012, Garvert et al., 2015).

The vmPFC has been shown to encode the distance between one’s own preferences and another person’s, suggesting that people use their own preferences as a starting point for estimating other people’s (Klucharev et al., 2009, Tamir & Mitchell, 2010). Some studies have reported evidence for an anatomical gradient, whereby the vmPFC encodes the preferences of the Self and similar others, whilst the dmPFC encodes the preferences of dissimilar others (Mitchell et al., 2006, Jenkins et al., 2008, Heleven & Van Overwalle, 2018). However, more recent work has shown that this anatomical gradient might not reflect a gradient in agent identity per se. Instead, this could be a gradient in decision relevance, ranging from decision variables relevant for current choices, encoded in the vmPFC, to decision variables irrelevant for current choices, encoded in the dmPFC (Nicolle et al., 2012). By asking subjects to learn about another person’s preferences and then make alternating choices on behalf of Self and Other, Nicolle and colleagues orthogonalised the dimension of agent identity from the dimension of decision relevance. The result suggests that preference encoding in the mPFC might be agent-independent, further complicating how Self-Other distinction is achieved.

The discovery that decision variables in the mPFC might be agent-independent provides fertile ground for a mechanistic account of how people
adopt each other’s values. Specifically, learning about another person’s values might induce plasticity in one’s own value representations. This can explain a behavioural phenomenon whereby people’s preferences shift towards the learned preferences of Others (Klucharev et al., 2009, Campbell-Meiklejohn et al., 2010, Garvert et al., 2015, Moutoussis et al., 2016). Indeed, after learning about another person’s preferences, vmPFC representations of one’s own preferences become more similar to representations of the Other’s preferences (Garvert et al., 2015). This kind of social conformity can modulate the extent to which the ventral striatum differentially responds to rewards (Campbell-Meiklejohn et al., 2010). Similarly, when subjects learn about other people’s risk-seeking propensities, they shift their behaviour to be more like the Other (Suzuki et al., 2016, Reiter et al., 2019). This conformity effect is associated with a change in risk perception signals in the caudate nucleus of the dorsal striatum (Suzuki et al., 2016). Thus, repeatedly activating representations of Others’ preferences can induce changes in how fundamental decision variables are computed for Self.

These studies provide a neurobiological framework for understanding how humans incorporate other people’s mental states into their own. They invoke a Self-Other distinction that is more abstract than the bodily Self-Other distinction discussed above. However, studies of preference simulation are not well suited to investigating Self-Other distinction, because ‘my’ preferences are value functions that ‘I’ have honed over a long period of time whilst, in typical experiments, ‘your’ preferences are entirely novel. Even, when the preferences of a novel agent have been learned, such that the agent is now familiar, there is a subjective value differential between decisions that lead to rewards for Self and decisions that lead to equivalent rewards for Other, confounding the Self-Other distinction. In other words, a decision that leads to reward for Other will be underpinned not only by signals that simulate the Other’s computations, but also by the subject’s own value function. Indeed, observing another agent receive a reward can carry negative reward signals for the subject, whilst diminishing appetitive behaviour (as shown in Figure 1-9). For these reasons, beliefs about novel
environments, engendered by nonreinforced learning, might be mental states that are better suited for probing Self-Other distinction. Indeed, these are the kinds of mental states that subjects have to infer in classic mentalising tasks like false belief tasks.

Like subjective preferences, refutable beliefs are also malleable under social influence (Asch, 1956). Subjects will incorporate another person’s beliefs into their own, weighted by the Other’s subjective confidence (Campbell-Meiklejohn et al., 2017) and objective validity (Li et al., 2019). In this case, there is evidence that a specific region of the vmPFC (area 10) is involved in integrating another person’s beliefs, conditioned on their confidence (Campbell-Meiklejohn et al., 2017). This implies a degree of agent specificity, or at least a degree of specificity in how different information sources are encoded. In order to investigate Self-Other distinction in beliefs however, the sensory input channels providing information about ‘my’ belief and ‘your’ belief should be matched.

Self-Other distinction appears to be impaired in a number of mental health illnesses. These examples of dysfunctional Self-Other distinction further justify the assumption that it is a fundamental computation in human cognition, and that it might be a relevant target for diagnosis and treatment of some conditions.

1.4.3 Failed Self-Other distinction in mental illness

From the feelings of worthlessness experienced by someone with major depression (Mojtabai & Olfson, 2004), to the incoherent sense of Self experienced by a person with borderline personality disorder (BPD) (Gold & Kyratsous, 2017), many symptoms have descriptions that rely on the notion of an altered sense of Self. Social cognitive deficits are also frequently observed in many psychopathologies. They can manifest as an inability to maintain stable relationships (Skodol et al., 2002), or more abstractly, an inability to understand Others (Palmer et al., 2015). Social neuroscientists have managed to state their field’s relevance to virtually all common mental
illnesses (Cacioppo et al., 2007, Cacioppo et al., 2014, Bicks et al., 2015). Beyond merely being a way to secure research grants in a highly competitive era, the statements are intuitive under the assumption that sociality is a defining feature of human cognition. As social cognition relies on many constituent processes, such as attention, learning and perhaps Self-Other distinction, one should expect there to be many pathophysiological routes to a dysfunctional social cognition.

The growth of computational neuroscience has led to recent calls for a computationally inspired nosology of mental illnesses (Montague, 2007, Huys et al., 2011b, Montague et al., 2012, Friston et al., 2014). For instance, symptoms of depression can be described in terms of aberrant reinforcement learning (Maia & Frank, 2011, Huys et al., 2015, Adams et al., 2016), whilst some symptoms of autism spectrum disorder (ASD) or schizophrenia can be described in terms of aberrant Bayesian inference (Friston et al., 2014, Adams et al., 2016, Lawson et al., 2017, Powers et al., 2017). Computational models provide a range of putative ‘failure modes’ – different ways that cognition can become dysfunctional – and hence a range of potential diagnostic and therapeutic targets.

Given that so many mental health disorders have symptoms that refer to alterations in how Self and Other are perceived, it might be useful to incorporate Self-Other distinction into computational frameworks. As Self is generally a fuzzy concept, it is tempting to couch a computational model, of cognition or mental illness, in terms that eschew reference to Self. This is a reasonable approach for modelling some features of cognition, but it can’t be guaranteed that something that looks like Self will simply jump out of the model. An alternative would be to cast the model at a level of explanation where Self and Other are explicitly defined. Self-Other distinction may be a good feature of a computational model that can provide formal descriptions of aberrant Self perception and social cognition. The aim here is not to supplant existing models, but to supplement them with the construct of agent identity, in order to avoid unnecessary reductionism.
Chapter 1: Introductory material

Using schizophrenia, ASD and BPD as examples, the subsequent sections will briefly review reports of disrupted Self-Other distinction in mental illness.

1.4.3.1 Schizophrenia

People with schizophrenia often show a marked difficulty in daily social functioning (Green et al., 2015). This may emerge from impairments in a wide range of cognitive processes. These range from basic perceptual discrimination of different facial (Kohler et al., 2010) and vocal (Leitman et al., 2007, Kantrowitz et al., 2013) stimuli, to more complex processes such as empathic accuracy (Lee et al., 2011, Harvey et al., 2013) and mentalising (Frith & Corcoran, 1996, Doody et al., 1998, Wang et al., 2015).

It is difficult to identify a deficit in mentalising, over and above deficits in other cognitive processes like working memory and attention. However, many delusions and hallucinations experienced by schizophrenia patients, or indeed other patients experiencing psychosis, have a ‘fundamentally social nature’ (Bell et al., 2017). These experiences often involve falsely inferring the presence of an agent, such as in auditory verbal hallucinations or persecutory delusions. These ‘positive’ symptoms have been described in terms of an inability to accurately represent mental states, or rather an inability to meta-represent appropriately (Frith & Frith, 1991, Gambini et al., 2004).

Recent task-based imaging studies have shown that schizophrenia patients (Bliksted et al., 2018) and people with subclinical schizotypal symptoms (Acosta et al., 2019) both show abnormal activation in mentalising brain regions. Compared with healthy controls, people with schizophrenia are more likely to attribute mental states to non-social stimuli, and show more activation in the anterior mPFC, whilst also being less likely to attribute mental states to social stimuli (Bliksted et al., 2018). These hyper- and hypo-mentalising effects point to a dysregulation of the processes that govern the selective attribution of mental states to different agents. Erroneously attributing one’s own mental states to Other could theoretically create
experiences such as auditory verbal hallucinations, persecutory delusions and thought insertion\textsuperscript{8}. Conversely, the incorrect attribution of an Other’s mental state to the Self could induce the experience of thought broadcasting\textsuperscript{9}.

Recent neurocognitive accounts of schizophrenia refer to dysregulated Self-Other distinction (van der Weiden et al., 2015). Most results point to a downregulation of Self-Other distinction. Compared with healthy controls, people with schizophrenia show more overlap in the neural activations associated with hearing one’s own voice and another person’s voice during an incidental task (Jardri et al., 2011). Some patients also show enhanced motor imitation (Simonsen et al., 2019a) and conformity to Others’ preferences (Simonsen et al., 2019b). On the other hand, there is some evidence in favour of an upregulated Self-Other distinction in schizophrenia. The Simon effect is the phenomenon whereby, after learning two stimulus-response mappings, response times are faster if a stimulus is presented closer to the location of the response, even though the location of the stimulus is irrelevant to the task (Simon & Rudell, 1967). The effect disappears if subjects are told to ignore one of the stimulus response pairs, in what is effectively a go/no-go task (Hommel, 1996). However, the effect is reinstated if another social partner takes control of the irrelevant stimulus response pair (Sebanz et al., 2003). This ‘social Simon effect’, is thought to reflect an implicit representation of the Other’s action demands. The social Simon effect is absent in some people with schizophrenia (Liepelt et al., 2012), possibly suggesting a reduced integration of Other’s mental states into one’s own. Modelling the distinction between Self-attributed and Other-attributed mental states may help to reconcile these findings and provide a formal framework for explaining disordered Self and Other processing in schizophrenia.

\textsuperscript{8} Thought insertion is the sensation that someone else’s thoughts are being put into one’s mind.

\textsuperscript{9} Thought broadcasting is the sensation that other people are aware of one’s own thoughts.
1.4.3.2 Autism spectrum disorder

The significance of impaired social cognition in ASD has been long recognised. Abnormal social and communication skills are core features of its diagnosis. Children with ASD show an arrested development in mentalising ability, as indicated by an impaired performance on false belief tasks (Baron-Cohen et al., 1985). Whilst healthy controls show TPJ activity that discriminates between true and false belief conditions, people with ASD don’t show this pattern of neural activity (Nijhof et al., 2018). It’s recently been shown, with the help of a computationally inspired learning task, that people with ASD can in fact keep track of another person’s beliefs, but are unable to correctly infer the Other’s intentions from their actions (Rosenthal et al., 2019). Whilst the picture is far from complete, it seems clear that people with ASD find it harder to represent other people’s mental states than healthy controls.

Less well known are the features of ASD that relate to impaired Self-referential processing. In addition to the difficulties in inferring Other’s mental states, people with ASD also perform worse than controls on tasks that engage mental time travel (Adler et al., 2010, Terrett et al., 2013, Ciaramelli et al., 2018), metamemory\(^\text{10}\) (Grainger et al., 2014) and judgements of their own agency (Zalla et al., 2015). Furthermore, whilst people with ASD have lower empathy scores than controls (Baron-Cohen & Wheelwright, 2004), they also often have alexithymia (Hill et al., 2004). This is a reduced ability to interpret one’s own emotions, and it is associated with the strength of empathic pain response in the anterior insula (Bird et al., 2010). Finally, in a mentalising task where subjects had to make judgements about Self and Other, neurotypicals showed more vmPFC activation during Self-reflection than Other-reflection whilst people with ASD showed the same vmPFC response to both conditions. Those who showed more Self-Other distinction

---

\(^{10}\) Metamemory is a subtype of metacognition that specifically refers to awareness of one’s own memory capabilities.
in this neural response had the least severe social impairment in childhood (Lombardo et al., 2010).

It thus seems that ASD involves a general disruption in meta-representation, beyond the social impairment (Frith & Frith, 1991). Incorporating Self-Other distinction into models of ASD might be useful for reconciling this range of impairments in mental state attribution (Lamm et al., 2016).

1.4.3.3 Borderline personality disorder

People with BPD (also known as emotionally unstable personality disorder or EUPD) are often described as having an impairment in maintaining and discriminating between representations of Self of Other (Beeney et al., 2015). This may be one of the reasons why people with BPD suffer from unstable relationships with close Others (Hallquist et al., 2017). Whilst there have been efforts to model aberrant decision making in people with BPD, for instance with a game-theoretic approach (King-Casas et al., 2008), there has been little work to formally describe the disturbances in Self-Other processing.

Mentalising tasks that require the integration of multiple perspectives are challenging for people with BPD and the extent of this impairment is correlated with the severity of childhood experiences of punishment (Petersen et al., 2016). Indeed, ‘Mentalization-based treatment’ (MBT) is an effective psychotherapy designed to improve mentalising in people with BPD (Fonagy & Bateman, 2006, Vogt & Norman, 2018). Grounding mental state attribution in a formal model might help to explain what patients are actually learning when they undergo MBT (Moutoussis et al., 2017), and pave the way for more targeted treatments.

1.5 Situating the current work

These ideas have inspired the original work that will be presented in the subsequent chapters. The assumption driving this research is that the
encoding of agent identity is a computational problem, which if solved, can enable meta-representation and hence the emergence of human social systems. Of note, this work is not guided by an assumption that there is a distinction between ‘social’ and ‘non-social’ cognition. Rather, the assumption is that the encoding of agent identity is a computation that gives rise to a whole suite of human cognitive features, some of which enable the emergence of social systems.

Research at multiple levels of organisation is starting to point to some unifying principles of brain function and cognition. One of these principles is that of predictive coding, whereby learning and inference rely on predictions that can be updated in the face of new evidence. The first experiment asks how the signals that drive these computations are selectively attributed to different agents. At what processing stage is a Self-Other distinction encoded? The second experiment investigates plasticity in Self-Other distinction, by asking whether and how experiences can shape the fidelity of agent identity encoding. The third experiment explores whether agent identity encoding might be relevant for cognition in non-social contexts.
Chapter 2: Neurocomputational methods

The work presented in this thesis makes use of two different neuroimaging techniques: magnetic resonance imaging (MRI) and magnetoencephalography (MEG). This chapter will briefly review how signals are obtained from MRI and MEG, and how behavioural data can be combined with these signals to inform neurocomputational models.

2.1 Magnetic resonance imaging

2.1.1 Biophysics of MRI

MRI relies on the principle of nuclear magnetic resonance (NMR) (Rabi et al., 1938). The magnetic spins of atomic nuclei will align when subjected to a constant magnetic field ($B_0$). This can be thought of as a net magnetisation vector that is parallel to the applied field. If this alignment is disturbed by a weak, oscillating magnetic field ($B_1$), also known as a radiofrequency (RF) pulse, the atomic nuclei emit an electromagnetic signal. This signal induces a current in an electromagnetic receiver coil in the MRI scanner, which is sensitive to the properties of different atomic nuclei.

In MRI the RF pulse is delivered perpendicular to the axis of the $B_0$ field, and at the resonance frequency of hydrogen atom nuclei. MRI thus typically maps the location of hydrogen nuclei, which are abundant in water and fat. The RF pulse tilts the vector of magnetisation from the longitudinal axis ($z$ axis) of the $B_0$ field to the transverse plane ($x$, $y$ plane). When the RF pulse is removed, the net magnetisation vector ‘relaxes’ towards the longitudinal axis of the $B_0$ field. This is determined by $T_1$, the time taken for magnetisation to return to the
longitudinal axis, and $T_2$, the time taken for transverse magnetisation to decay.

In addition to the magnet that induces the strong $B_0$ field, MRI scanners have weak gradient fields, induced by three gradient coils, one for each of the x, y and z axes. The gradient fields cause predictable and spatially specific distortions to the $B_0$ field, so that the magnetisation of protons is slightly different in each spatial location. This is a way of encoding spatial information so that a three-dimensional image can be constructed.

MR images are based on contrasting properties (e.g. $T_1$ and $T_2$) of different tissue types. Images can be weighted in their sensitivity to different contrasts, by varying different parameters of an RF pulse sequence. For instance, a long echo time (TE) between RF transmission and electromagnetic signal detection will produce a signal with a larger contribution from contrasts in $T_2$ ($T_2$-weighted image). A short TE, on the other hand, is used in acquiring $T_1$-weighted images and proton-density (PD) weighted images. $T_1$-weighted images use a short repetition time (TR), between one RF pulse and the next, whilst PD-weighted images use a long TR.

### 2.1.2 Quantitative MRI

The weighted images produced with MRI are qualitative in nature. They are useful for visualising different tissue types within the same image but the absolute values within each voxel (x, y, z location) are arbitrary. These values depend not only on tissue properties but also on the magnetic fields used in the scanner. Thus, they cannot be compared between different images. However, with biophysical models, quantitative maps can be estimated from qualitative weighted images.

In quantitative MRI (qMRI), the scanner is no longer being used like a camera, but rather like a measuring device, such as a thermometer. Using a multi-parameter mapping (MPM) sequence, several quantitative NMR properties can be measured. These properties include effective proton
density (PD*), transverse relaxation rate (R_1), effective longitudinal relaxation rate (R_2*), and magnetisation transfer (MT). These properties are measured in absolute values, which are comparable across scanning sessions, subjects and even scanner sites (Weiskopf et al., 2013).

These measures are often used as proxies for microstructural tissue properties, such as macromolecular content (MT, R_1), water content (PD) and iron content (R_2*). These relationships between tissue microstructure and NMR properties have been validated with ex vivo animal tissues (Stanisz et al., 1999), post-mortem human brain tissues (Stuber et al., 2014), and known anatomical variations in microstructure (Marques et al., 2017). These quantitative measurements thus provide early steps towards in vivo histology using MRI (hMRI) (Edwards et al., 2018, Tabelow et al., 2019).

In a typical MPM sequence, used for instance in (Callaghan et al., 2014), three multi-echo 3D fast low angle shot (FLASH) images are acquired, with predominant PD, T_1 or MT weighting. These weightings are produced by varying the TR and the shape of the RF pulse, which in turn modulates the flip angle, the extent to which the net magnetisation vector tips away from the B_0 field. The ‘multi-echo’ aspect of this sequence refers to the fact that, for each weighting, multiple images are acquired, each one with a different TE. The images acquired at each TE are then averaged together to increase the signal to noise ratio (SNR). The ‘FLASH’ aspect of this sequence refers to the fact that transverse (B_1) components of the magnetisation are removed, or ‘spoiled’, prior to each RF pulse, leaving only the longitudinal magnetisation. Whilst spoiling occurs naturally if long TRs are used, this makes image acquisition very slow. FLASH sequences enable much quicker imaging protocols at a high spatial resolution (Haase et al., 1986).

---

11 R_1 is simply 1/T_1 and R_2 is simply 1/T_2. MT refers to the transfer of magnetisation from one population of atomic nuclei to another. In MRI, MT is concerned with the transfer of magnetisation between three different pools of hydrogen nuclei: Free water, bound water and macromolecules. It is largely determined by the macromolecular content of different tissues.
The weighted images can be converted into quantitative maps using signal models, which have been extensively documented elsewhere (Helms et al., 2008b, Helms et al., 2008a, Weiskopf et al., 2014, Tabelow et al., 2019). The maps can then be used to assess relationships between tissue microstructure and behaviour or functional activity.

2.1.3 Functional MRI

2.1.3.1 Basis of the BOLD signal

Functional MRI (fMRI) makes use of a T$_2^*$ contrast, which is affected not only by the transverse relaxation time but also by local inhomogeneities in the magnetic field. Haemoglobin, the iron-containing protein found in red blood cells, is diamagnetic in its oxygenated form, but paramagnetic in its deoxygenated form (Pauling & Coryell, 1936). Deoxyhaemoglobin therefore dephases the spins of nearby protons to introduce inhomogeneities and causes a reduction in T$_2^*$ decay time. In a T$_2^*$-weighted image, contrasts in haemoglobin oxygenation can be detected (Ogawa et al., 1990). This is conventionally described as a blood-oxygen level dependent (BOLD) contrast.

The BOLD signal can be used as a proxy for neural activity. The rationale here is that electrically active neurons have large metabolic demands and generate an energy debt. This energy debt is signalled to arteriolar smooth muscle cells, via many putative molecular messengers, such as neurotransmitters and metabolic waste products, to induce local vasodilation (Attwell et al., 2010). Thus, the theory goes, when a pool of neurons is more active, the local blood supply becomes flooded with oxyhaemoglobin, reducing the T$_2^*$ decay time and increasing the BOLD signal intensity. Indeed, by simultaneously recording intracortical activity and BOLD signal, it has been shown that BOLD responses correlate with local field potentials (Logothetis et al., 2001).
Caution is required when using BOLD as a marker for neuronal activity because a large amount of neural signal is filtered out by the slow haemodynamic response. Whilst neural events can occur on the order of milliseconds, the BOLD response typically peaks five to ten seconds after an increase in neural activity. Therefore, high-frequency neural signals are naturally filtered out of the BOLD signal. Furthermore, it is unclear, from an increase in BOLD signal, which kinds of neurons have become more active. As microcircuits usually include a mixture of excitatory, inhibitory and neuromodulatory cells, it is difficult to make inferences about the neuronal computations that are actually taking place. All that can be said is that a region is more or less metabolically active (Hall et al., 2016).

2.1.3.2 fMRI data pre-processing

The fMRI experiment reported in this thesis used a scanner with a magnetic field strength of 3 Tesla with a resolution of 3 mm$^3$ voxels. This makes it possible to make inferences about which brain regions are active under different conditions at the mm scale. A conventional analysis involves constructing a general linear model (GLM) to explain the observed BOLD timecourse in each voxel (Friston et al., 1994). In order to do this, several pre-processing steps are required, to correct for in-scanner motion and also to account for intersubject differences in brain position, size and morphology.

In fMRI it is desirable to sample the BOLD signal as frequently as possible. Thus, a fast sequence such as echo-planar imaging (EPI) is typically used. In an EPI sequence a whole stack of 2D slices, or a ‘volume’, can be acquired in a matter of seconds, reducing the risk of significant motion during a single volume acquisition, which is difficult to correct for. Usually the first few volumes of a scanning ‘run’ are discarded prior to the analysis. These ‘dummy’ volumes are excluded to allow for T1 equilibration effects.

Despite the relatively fast acquisition of an EPI sequence, there is still a delay of a few seconds between acquiring of the bottom slice and the top slice of a volume. By interpolating each voxel’s data between sampled time
points, slice-timing correction enables an estimation of what the data would have looked like in each slice, were they sampled simultaneously. In SPM12 (the fMRI analysis toolbox used in this thesis), a Fourier transform is performed and then a phase shift is applied to the sine waves that make up the signal.

In order to correct for any head motion that occurs during a scan, each volume undergoes a realignment procedure. Without realignment, the signal recorded in a given voxel might reflect one brain region at the beginning of a scan and a different region by the end of a scan if the subject has moved by a few millimetres. These movements are modelled with six parameters, which denote translation along and rotation around the x, y and z-axes. A reference image, such as the first volume acquired, is selected, and then each successive volume is realigned by minimising its difference from the reference image.

Even after realigning the volumes, there is usually still variance in the BOLD timeseries that is attributable to movement. This is partly due to a ‘susceptibility-by-movement’ interaction (Hutton et al., 2002). Spatial variations in magnetic susceptibility, which are particularly sharp at air-tissue boundaries, produce inhomogeneities in the magnetic field. These inhomogeneities lead to a warped image of the object being scanned. Furthermore, the pattern of these warps will vary as a function of spatial location within the scanner, such that the shape of the image varies with motion. By acquiring a field map before an EPI sequence, the inhomogeneities can be mapped and the images can be ‘unwarped’ to correct for these motion-dependent distortions.

In order to make group-level statistical inferences, voxels must be anatomically aligned across subjects. First the functional EPI images are co-registered to the subject’s own structural anatomical image, and then the structural and functional images of all subjects undergo a spatial normalisation procedure. Here the images are warped into the standardised Montreal Neurological Institute (MNI) space through a series of affine transformations.
Finally, the EPI images are spatially and temporally filtered. In the spatial domain, a 3-dimensional Gaussian kernel is applied to all voxels to smooth the images. The wider the kernel, the more information from neighbouring voxels is incorporated into each voxel, making the image smoother. Spatial smoothing reduces noise and increases the sensitivity to detect regional effects at the group level, albeit with a reduced spatial resolution. In the temporal domain, a high-pass filter is applied to the timeseries to exclude slow signal drifts that are unrelated to signals of interest.

When constructing the statistical model to explain the BOLD timeseries, knowledge about the haemodynamic response can be leveraged. Typically, a canonical haemodynamic response function (HRF) is assumed. This is a mixture of two gamma functions, one to capture a peak in the signal at around six seconds and one to capture an undershoot and a return to baseline after around thirty seconds. Regressors are convolved with the HRF so that they reflect realistic BOLD timeseries (Figure 2-1).

In reality, haemodynamic responses are variable across different individuals and different brain regions. To accommodate this variability, more complex and flexible models can be used. In this thesis, the ‘informed basis set’ (Friston et al., 1998) was used. This basis set includes the canonical HRF, along with its partial derivatives with respect to delay and dispersion (Figure 2-1). The derivative with respect to delay is essentially the difference between the canonical HRF and a slightly temporally delayed HRF. By including this additional regressor, the model can flexibly accommodate deviations from the canonical HRF in the onset of the BOLD response. The derivative with respect to dispersion is essentially the difference between the canonical HRF and a slightly wider HRF. By including this additional regressor, the model can flexibly accommodate deviations from the canonical HRF in the width of BOLD responses. Additional ‘nuisance’ regressors can be added to the model as covariates. These might include motion parameters, and parameters related to physiological noise such as breathing and heartbeat.
Chapter 2: Neurocomputational methods

2.2 Magnetoencephalography

2.2.1 Basis of the MEG signal

Post-synaptic potentials (PSPs) and action potentials both generate electrical currents and therefore magnetic fields. MEG uses highly sensitive magnetometers to detect small changes in these electrically induced magnetic fields at the scalp. The fields that are detectable with MEG are assumed to arise largely from post-synaptic potentials at the dendrites of pyramidal neurons in the sulci of the cortex (Lopes da Silva, 2010).

Figure 2-1: The haemodynamic response function

Left panel: An illustration of the canonical HRF (purple) and the two gamma functions (red and blue) that it is composed of. Middle panel: An illustration of the informed basis set. The canonical HRF (purple solid line) is shown along with its partial derivative with respect to delay (pink solid line) and its partial derivative with respect to dispersion (purple broken line). Right panel: Convolving a series of ‘events’ with the informed basis set. The grey bars on the top represent stimulus onsets. These events are modelled with stick functions, with ones at the time points when stimuli are presented, and zeros at all other time points. The events are convolved with all three basis functions in the middle panel to produce, in the right panel, a predicted BOLD timeseries (purple solid line) along with its temporal (pink solid line) and dispersion (purple broken line) derivatives. All three timeseries are included as regressors in a linear model designed to predict the real BOLD timeseries. A different regression coefficient will be obtained for each of the three regressors. The relative amplitudes of the functions have been arbitrarily set to aid visualisation.
PSPs are more likely than action potentials to contribute to the MEG signal because action potentials have a very short duration, on the order of one or two milliseconds. This means that an extremely high synchrony across different cells would be required to generate a current large enough to induce a detectable magnetic field. PSPs, on the other hand, have a much broader timecourse, and are more likely to summate to produce a detectable magnetic field.

The apical dendrites of pyramidal neurons are likely candidates for the source of the MEG signal. This is because they are aligned in parallel, with electrical ‘sinks’ (inward currents) located near the tops of the dendritic trees, and electrical ‘sources’ (outward currents) located near the cell bodies (Kutas & Dale, 1997). This asymmetric distribution of sinks and sources is an open-field configuration, meaning that an observable current dipole is generated. This is in contrast to a region where neurons are organised radially or randomly, such that sinks and sources are symmetrically distributed, and the resultant dipoles cancel each other out.

Finally, MEG is more sensitive to PSPs in cortical sulci than in cortical gyri. Radially-oriented dipoles, which are perpendicular to the scalp surface, generate tangentially-oriented magnetic fields, which are parallel to the scalp surface and are largely contained within the head. Thus, pyramidal neurons at the crowns of gyri and depths of sulci are unlikely to contribute to the MEG signal. Tangentially-oriented dipoles, generated by superficial sulcal neurons, will generate magnetic fields outside the head that can be detected with MEG.

The MEG system used in this thesis uses a helmet consisting of 275 sensors. Each sensor consists of a flux transformer (axial gradiometer), which is sensitive to gradients in magnetic fields. Each sensor also contains a superconducting quantum interference device (SQUID), which generates a current at a frequency dependent on the magnetic field applied from the flux transformer. These sensors can detect the magnetic fields generated at the scalp, at the femtotesla scale ($10^{-15}$). The recording must be conducted in a magnetically shielded room, to exclude external noise.
2.2.2 MEG data pre-processing

MEG has some advantages and disadvantages when compared with fMRI. Rather than measuring a vascular marker of metabolism, MEG directly measures electromagnetic signals and, like electroencephalography (EEG), is sensitive to real-time neural activity on the millisecond time-scale. However, whilst the voltage fluctuations measured with EEG are susceptible to distortions by the skull and scalp, the magnetic fields measured with MEG are less prone to these distortions (Okada et al., 1999). Furthermore, when decoding from neural activity, as in brain-computer interfaces, accuracy is often higher when MEG signals are used in comparison to EEG signals (Halme & Parkkonen, 2018). This may be due to MEG having a higher SNR than EEG, or MEG having access to neural sources that are not detectable with EEG. Nevertheless, MEG is relatively insensitive to subcortical neural activity, and even though cortical sources can be reconstructed with a forward model, the spatial resolution of these images is limited when compared to fMRI. Thus, MEG is useful for assessing how neural computations evolve over short timescales, whilst fMRI is useful for assessing how they are regionally distributed. In order to make meaningful inferences from MEG data, several pre-processing steps are required to remove artefacts from the signal.

The MEG experiment presented in this thesis used an event-related design where the evoked response field (ERF) to a stimulus is of interest. The ERF is the waveform that results from an external event. In order to obtain the ERF, the MEG data must be split up into small time windows, say a few seconds long. Each time window, or epoch, reflects the ERF in response to a specific stimulus on a specific trial. The epochs that correspond to a specific condition can then be averaged together to produce a mean ERF that shrinks inter-trial noise.
Figure 2-2: Examples of artefact correction in MEG

This figure shows three examples of MEG artefact correction using ICA. Each row is a different subject. The left column shows artefacts in raw MEG data. The column on the right shows artefacts after having performed ICA and component rejection. Each image is a map of the surface of the scalp, with the anterior portion towards the top of the page. Each MEG sensor is attributed an artefact percentage score. This is the percentage of trials for which the sensor's recorded field magnitude exceeded four standard deviations of the pre-stimulus baseline for a duration of twenty ms or longer. The values for the sensors have been smoothed and contour lines added to aid visualisation of the sensors that have high artefact scores. Most of these sensors are anterolateral, consistent with artefacts due to eyeblinks or eye movements.

Despite the use of a magnetically shielded room, MEG sensors will detect magnetic flux from non-neural dipoles. By filtering the data to exclude frequencies that are unlikely to reflect neural activity, some of these artefacts can be easily removed. However, physiological artefacts, such as from the heart and eyes, are likely to remain. Whenever the eyeballs or eyelids move, the electrical current in the eyeballs is modulated, producing a magnetic field that is often much larger than those produced by post-synaptic potentials in the brain (Carl et al., 2012). By performing independent component analysis (ICA) on the MEG data, artefactual components can often be easily identified.
(Vigario et al., 1998). The components can be identified in a number of ways. Firstly, they may contain field strengths that are much larger than what would be expected from a neural dipole. Secondly, their spatial topographies may reflect what would be expected from ocular dipoles. Thirdly, the magnetic field timecourse may reflect the typical waveform of the electrocardiogram. By manually identifying these components and then reconstructing the data from the remaining components, the effects of physiological artefacts can often be significantly reduced (Figure 2-2) without having to exclude trials or sensors from the analysis.

2.3 Computational modelling of brain and behaviour

Two types of modelling techniques are used, in this thesis, to describe neuroimaging and behavioural data (Figure 2-3). The first, ‘theory-driven’, approach relies on researcher-defined hypotheses about psychological processes. These hypotheses are generative models, in that they specify the process by which the observed data were generated, and can be used to simulate synthetic data. The second, ‘data-driven’, approach uses discriminative models to exploit statistical relationships between observed data and target variables, but has no way of reproducing the observed data. The two approaches are linked in that the target variables to be predicted by the discriminative model come from the generative model.

In both cases, a model is trained on some data, $X$, to make predictions about a variable of interest, $Y$. In other words, the model is used to estimate a posterior probability $\mathbb{P}(Y|X)$. In this thesis, the theory-driven approach trains a generative model on behavioural data to make predictions about subject-specific parameters. The generative model is used to compute the likelihood of the observed data, given the model parameters $\mathbb{P}(X|Y)$, with or without a prior $\mathbb{P}(Y)$, in order to estimate the posterior. The data-driven approach makes use a discriminative classifier, trained on neuroimaging data, to make predictions about psychological variables from the generative model. This discriminative model estimates $\mathbb{P}(Y|X)$ directly.
As shown in Figure 2-3, both approaches can be combined in a virtuous cycle. In this neurocomputational workspace, the theory-driven approach uses a hypothesis about cognition to estimate some unobservable, latent variables. These variables provide input to the discriminative model, which can be used to validate the theory behind the generative model and inform the development of future models.

**Figure 2-3: Neurocomputational workspace**

The figure illustrates how computational models can be applied to data from a single subject to make inferences about unobservable cognitive processes. Behavioural and neuroimaging data are collected simultaneously. A theory-driven, generative model (\(\mathcal{M}\)) formalises a psychological theory in terms of free parameters (\(\theta_\mathcal{M}\)), which are static, and psychological variables (\(Y\)), which are dynamic. The model is fit to behavioural data, such as choices or reaction times, by optimising the free parameters. When a suitable fit is found, the model is used to simulate psychological variables. A data-driven, discriminative model is then used to detect statistical relationships between neuroimaging data and the psychological variables from the generative model, by optimising new parameters (\(\theta_d\)), such as classifier weights. The performance of this data-driven model can be used as a test for convergent validity of the theoretical model, and to inform future theoretical models.
2.3.1 Theory-driven modelling

Formal, quantitative models can be devised as hypotheses to explain both behaviour and brain activity. A computational model is a hypothesis about the algorithms a subject uses when engaged in a specific psychological task. This ‘model-based’ approach to cognitive neuroscience thus has the advantage of testing hypotheses about the unobservable mental states that contribute to the observable behaviour. These unobservable mental states can be described as ‘decision variables’, in that they are numerical quantities, used by a hypothetical algorithm in the brain, to solve some computational problem.

Probabilistic models can be fit to behavioural choice and reaction time data using maximum likelihood estimation (MLE) or maximum a posteriori (MAP) estimation. MAP with a uniform prior over model parameters is equivalent to MLE.

\[
\theta_{\text{MAP}} = \arg\max_{\theta} \sum \log P(x_i|\theta)P(\theta)
\]

\[
\theta_{\text{MLE}} = \arg\max_{\theta} \sum \log P(x_i|\theta)
\]

Equation 2-1

Here, \(x_i\) denotes a single observation from the behavioural data and \(\theta\) denotes the free parameters of the model. Computing the joint likelihood \(\prod_i P(x_1, x_2, \ldots, x_n|\theta)\) would require multiplying many probabilities and so can lead to arithmetic underflow and numerical imprecision. Therefore, the logarithm of the likelihood is maximised instead. As the logarithm of a product is equivalent to the sum of the logarithm of the factors, it is the sum of the log-probabilities that is maximised per se, shown in Equation 2-1.

In this thesis, models are used to explain behaviour and describe the decision variables used in several different decision making tasks. Rescorla-Wagner (RW) models are used to explain how subjects update their beliefs about an environment and attribute those beliefs to different agents. A drift
diffusion model (DDM) is used to explain how subjects accumulate sensory evidence to reach a decision criterion. Delay discounting models are used to explain how subjects evaluate future rewards. The details of these models are described in subsequent chapters.

Model variables, such as prediction errors from a RW model, can then be regressed against BOLD or MEG signal. Typically, in a model-based neuroimaging analysis (Daw, 2011, Wilson & Niv, 2015), events in the design matrix are parametrically modulated by a variable from the model. Brain regions where the signal covaries with this regressor can be identified.

By identifying common spatiotemporal patterns of variable encoding at the group level, inferences can be made about how these algorithms are neurally implemented. However, beyond functional localisation, inferences can also be made about the algorithms themselves. The presence of any brain regions that show activity covarying with a model-based regressor speaks in favour of the cognitive model. The experiments in this thesis applied data-driven modelling to neuroimaging data to make inferences about algorithms in the brain.

2.3.2 Data-driven modelling

By treating different MRI voxels or MEG sensors as independent variables in a single statistical model, relationships between neural activity patterns and representational contents can be uncovered. Rather than conducting a separate analysis at each voxel or sensor, the interactions between multiple data features (sensors or voxels) can be exploited for predictive power (Figure 2-4). This approach has been previously described as a decoding or pattern-based analysis (Mahmoudi et al., 2012, Haxby et al., 2014, Haynes, 2015). As mental representations are thought to be encoded by large populations of neurons, rather than individual cells (Pouget et al., 2000), a pattern-based analysis is likely to be more sensitive in detecting neural correlates of representations (Norman et al., 2006). In the specific case of fMRI analysis, pattern-based techniques are also less
sensitive to co-registration errors and subject-specific idiosyncrasies in neurovascular coupling, because each statistical test is conducted over a group of voxels (Hall et al., 2016).

Typically, decoding analyses use supervised algorithms, meaning that the training data are labelled. In this sense, the modelling is not purely ‘data-driven’ per se, because the labels come from the researcher, not the data. The use of the term ‘data-driven’ in this context, refers to the fact that the relationship between the data and the labels is not specified by the researcher. Whilst the generative model fit to behaviour uses a researcher-defined process by which the parameters generate behaviour, the discriminative models used here exploit unknown relationships between the variables.

The choice of labelling depends on the kinds of representations that are of interest to the researcher. In this thesis, labels are derived from the generative behavioural model. By training a discriminative model on relationships between the latent variable and neural activity patterns, its discriminative accuracy can be measured on some out-of-sample, unlabelled data. If the model can attribute labels to this test data better than chance, then something can be said about the representational content of the neural activity pattern.
Figure 2-4: Pattern-based analysis of neuroimaging data

This figure uses an extreme example to demonstrate the utility of incorporating multiple data features into a single statistical model. A subject observes different colours on a screen whilst undergoing an MRI scan. Each point on the scatter plot represents a different trial. The trials have been categorised into ‘purple’ trials, when the subject saw a purple image on the screen, and ‘orange’ trials, when the subject saw an orange image on the screen. The researcher is interested to know whether, and where in the brain, information about colour is encoded. The BOLD signal from a single voxel, feature A, has been plotted for each trial along the x-axis (translucent dots). The BOLD signal from a different voxel, feature B, has been plotted for each trial along the y-axis (translucent dots). In neither of these 1-dimensional spaces is there a reliable distinction between purple trials and orange trials. In other words, there is no line, perpendicular to the x-axis or y-axis, that can robustly separate the two classes. However, if the data is projected into a 2-dimensional space (opaque dots), a clear separation is seen. Here a 1-dimensional line defines the class boundaries in a 2-dimensional space. In a typical pattern-based neuroimaging analysis, many more data features are used, and a multi-dimensional hyperplane is fit to the data to define the class boundaries.
Chapter 3: Agent-specific learning signals

The work presented in this chapter has previously been published (Ereira et al., 2018) in the following research article:


3.1 Abstract

Humans have a remarkable ability to simulate the minds of Others. How the brain distinguishes between mental states attributed to Self and mental states attributed to someone else is unknown. Here we investigated how fundamental neural learning signals are selectively attributed to different agents. Specifically, we asked whether learning signals are encoded in agent-specific neural patterns, or whether a Self-Other distinction depends on encoding agent identity separately from this learning signal. To examine this, we tasked subjects to learn continuously two models of the same environment, where one was selectively attributed to Self and the other was selectively attributed to another agent. Combining computational modelling with magnetoencephalography (MEG) enabled the tracking of neural representations of prediction errors (PEs) and beliefs attributed to Self, and of simulated PEs and beliefs attributed to another agent. We found that the representational pattern of a PE reliably predicts the identity of the agent to whom the signal is attributed, consistent with a neural Self-Other distinction implemented via agent-specific learning signals. Strikingly, subjects exhibiting a weaker neural Self-Other distinction also had a reduced
behavioural capacity for Self-Other distinction, and displayed more marked subclinical psychopathological traits. The neural Self-Other distinction was also modulated by social context, evidenced in a significantly reduced decoding of agent identity in a non-social control task. Thus, we show that Self-Other distinction is realised through an encoding of agent identity intrinsic to fundamental learning signals. The observation that the fidelity of this encoding predicts psychopathological traits is of interest as a potential neurocomputational psychiatric biomarker.

3.2 Introduction

Social interactions are underpinned by an ability to infer the mental states of Self and others, referred to as mentalising (Fonagy, 1991). The discovery of mirror neurons in the macaque premotor cortex (di Pellegrino et al., 1992) introduced the notion that in mentalising, the primate brain might directly simulate another agent’s cognitive process. More recently, functional magnetic resonance imaging (fMRI) studies (Burke et al., 2010, Cooper et al., 2012, Suzuki et al., 2012, Monfardini et al., 2013, Apps et al., 2015, Lockwood et al., 2015), intracranial recordings (Hill et al., 2016) in humans, as well as single-cell recordings in monkeys (Chang et al., 2013), have shown that when a subject observes another agent interact with its environment, the subject’s brain encodes not only the other agent’s motor activity, but also their reward prediction errors (RPEs). In other words, subjects appear to simulate the reinforcement learning processes of other agents.

These simulated learning signals localise to specific cortical regions, such as the anterior cingulate gyrus (Chang et al., 2013, Apps et al., 2013, Hill et al., 2016). A functional segregation of learning signals can allow the brain to encode information about whether learning is arising out of the individual’s own behavioural interactions with the environment, or whether learning is taking place vicariously through observing the behaviour of another agent. In a similar vein, it has been suggested that the medial
prefrontal cortex (mPFC) supports a functional axis that encodes whether behaviour is executed or imagined (Nicolle et al., 2012, Garvert et al., 2015).

For simulation to be useful in social interactions, the brain must discriminate signals attributed to Self from simulated signals attributed to other agents (Leslie, 1987, Steinbeis, 2016, Lamm et al., 2016, Lamm et al., 2017). It has been argued that autism spectrum disorder may involve an impairment in this Self-Other distinction ability (Baron-Cohen et al., 1985, Bird et al., 2010, Palmer et al., 2015, Sevgi et al., 2015). Similar impairments have also been reported in conditions such as schizophrenia (Liepelt et al., 2012, van der Weiden et al., 2015), psychopathy (Lamm et al., 2016) and borderline personality disorder (Beeney et al., 2015). An aberrant Self-Other distinction may also play a role in the social dysfunction seen in psychopathologies including depression (Weightman et al., 2014, Ladegaard et al., 2016) and addiction (Preller et al., 2014, Tomei et al., 2017, Quednow, 2017).

A prefrontal coding scheme that discriminates between instrumental and observational learning, or executed and imagined behaviour, could provide a useful heuristic for a Self-Other distinction, but would be insufficient for discriminating amongst signals attributed to different agents as a general-purpose computation. For instance, the false belief task (Wimmer & Perner, 1983), a standard test of mentalising ability, requires that subjects make inferences about an environment, and then selectively attribute one belief-state to Self and a different belief-state to another agent for whom the environment is only partially observable. These belief-states are not informed by the behaviour of the subject or the other agent, but arise through passively observing the environment. In this case, neural coding schemes that discriminate between executed behaviour and observed or imagined behaviour, cannot facilitate a Self-Other distinction, and a more fundamental computation for selectively attributing signals to different agents is required.

Thus, an open question for simulation theory is how Self-Other distinction is achieved (Lamm et al., 2017). If inferring another agent’s mental state requires the brain to simulate that agent’s computations, how are the
Chapter 3: Agent-specific learning signals

outputs of those computations identified as ‘belonging to Other’? One possibility is that variables for learning and decision making are encoded in distinct neural activity patterns, depending on the agent to whom these signals are attributed. Such architecture would entail an encoding of agent identity intrinsic to representations of these low-level signals. A second possibility is that a learning signal is always encoded in an agent-independent pattern. In this case the learning signal and the identity of the agent to whom the signal is attributed would need to be encoded in two separate activity patterns.

Here we test whether learning signals are encoded in agent-specific activity patterns, and thus whether Self-Other distinction requires agent identity to be encoded separately from a low-level learning signal. We used a novel paradigm inspired by false belief tasks, in which subjects learned about a fluctuating state in the environment. In so doing they were also required to intermittently switch their frame of reference between Self and Other. The two agents (Self and Other) received different information such that their belief trajectories were uncorrelated, enabling us to measure Self-attributed and Other-attributed learning signals independently. Unlike previous paradigms eliciting simulated signals, subjects did not observe the agent’s behaviour and there was no reinforcement of learning by either reward or punishment. Learning for Self and learning for Other thus recruited the same input channels, and were identically salient and identically motivating. We term this paradigm a ‘probabilistic false belief task’.

The task design rules out a potential confound of simulated reward learning (Burke et al., 2010, Cooper et al., 2012, Suzuki et al., 2012, Monfardini et al., 2013, Apps et al., 2015, Lockwood et al., 2015) wherein Self-attributed reward related decision signals (such as RPEs) pertain to rewards expected to be received by the participant, while Other-attributed reward related decision signals do not. We measured the neural encoding of learning signals using magnetoencephalography (MEG) in order to measure whether the representations of these signals are agent-specific and also how agent-specificity evolves over the time course of a single trial.
Chapter 3: Agent-specific learning signals

3.3 Methods

3.3.1 Participants

41 healthy adults (23 female) aged 18-42, participated in the experiment. They were recruited from the UCL Institute of Cognitive Neuroscience subject pool. All participants had normal or corrected-to-normal vision and had no history of psychiatric or neurological disorders. One participant was excluded from the analysis due to excessive head movements in the scanner and a further two were excluded due to technical faults with MEG data acquisition, leaving 38 subjects (21 female) for the analysis, with a mean age of 26.6 (SD 6.9). All participants provided written informed consent, which was approved by the Research Ethics Committee at University College London, under ethics number 9929/002.

3.3.2 Behavioural paradigm

During MEG scanning, subjects observed a sequence of samples from a Bernoulli distribution, with a drifting Bernoulli parameter \( P \). This is the probability, on each trial, of seeing one of two possible outcomes. Another participant, who sat outside the scanner in a different room, played the exact same game. This other subject was able to observe some of the samples seen by the scanned participant (shared trials) but not all of them (privileged trials). Additionally, the non-scanned participant was occasionally presented with misleading samples (decoy trials). Therefore, the non-scanned participant sampled evidence that induced a false belief about \( P \) (\( P_{fb} \)). These three types of trial (privileged, shared and decoy) were balanced in frequency and distributed evenly throughout the task in a pseudorandom order.

In a version of the game we refer to as the social version (SV) privileged and decoy trials were signalled to the scanned participant, who thus had access to information about both \( P \) and \( P_{fb} \). On Self probe trials, the scanned participant was required to report their estimate of \( P \), by positioning
an arrow along a virtual continuous scale that ranged from a probability of zero (certain to see one outcome) to one (certain to see the other outcome). On Other probe trials, the scanned participant had to put themselves in the shoes of the non-scanned participant and report their estimate of $P_{fb}$. Crucially, the information that the scanned subject used to compute $P_{fb}$ was sampled at the same rate as the information used to compute $P$. We refer to the subject’s belief about $P$ as $B$, and we refer to their belief about $P_{fb}$ as $B_{fb}$. The structure of sampling trials and probe trials is outlined in Figure 3-1.

All subjects also played a non-social version (NSV) of this game, which did not involve another participant. Here the scanned participant had to keep track of the belief-state of a fictional computer that received limited and misleading information, and stored a false estimate of $P$ ($P_{fb}$). On Other probe trials in NSV the scanned participant was asked to imagine themselves in a counterfactual situation, wherein they acted using the false information provided by the computer.

Thus, in SV participants switched their frame of reference between Self and Other, whilst in NSV participants switched their frame of reference between Self and a counterfactual Self. The only structural differences between SV and NSV pertained to the cover stories, the images used for stimuli and the wording of the Other probe trials.
Chapter 3: Agent-specific learning signals

Figure 3-1: Overview of behavioural paradigm

A) An example of a sampling trial and probe trial. Sampling trials consisted of a cue (privileged, shared or decoy) followed by a Bernoulli outcome (heads or tails), followed by a variable ITI. After 4-9 sampling trials, subjects were probed with either a Self or an Other probe trial, in which they had to report their own belief or the other subject’s belief about the current probability of seeing one outcome over the other, in this case a yellow sun-shade signifying sunshine and pink umbrella signifying rain, by moving an arrow along a continuous scale ranging from zero (certain for one outcome) to one (certain for the other outcome) B) The set of images used as cues for SV and NSV. C) The set of images used as outcomes for cover story 1 and cover story 2. D) An example pair of uncorrelated random walks used to generate a full trial sequence. Samples were drawn from these generative Bernoulli distributions to produce two sequences of ‘coin flips’, one for Self (blue) and one for other (red). E) An exemplar sequence of 6 sampling trials in SV with cover story 1. This schematic demonstrates how different information can be used to infer P and P_fb. In privileged trials (blue squares and arrows) only the subject sees the information. In shared trials (magenta squares and arrows) both agents receive the information. In decoy trials (red squares and arrows) both agents receive information but the scanned subject knows that it is misleading whilst the other agent does not. The three trial types were balanced in frequency and occurred in pseudorandom order.
3.3.2.1 Cover stories and task details

Every subject played two versions of this game while inside the MEG scanner, the SV and NSV, one after the other. We created two cover stories that could interchangeably be applied to either SV or NSV allowing for four possible games to be played (SV1, SV2, NSV1 or NSV2). The cover stories were designed to be immersive and to make the underlying structure of the task, particularly the drifting Bernoulli parameter (P), as intuitive as possible. Each subject played two games with different cover stories to make the SV and NSV games feel as different as possible.

Each subject was allocated to one of four groups. These groups were defined by the order in which SV and NSV were played and the cover stories that were applied to both of them. Group 1 played SV1 → NSV2. Group 2 played NSV2 → SV1. Group 3 played SV2 → NSV1. Group 4 played NSV1 → SV2. This 2x2 factorial design meant that game order and cover story mappings were counterbalanced across subjects. For each subject, 2 pre-generated trial sequences were selected, one for their first game and one for their second game.

In cover story 1, subjects played the role of a ‘shop assistant’ working in a shop on a tropical island selling only pink umbrellas and yellow sun-shades (i.e. heads or tails). Subjects were briefed with the following details: On every trial a ‘customer’ will come to the shop and buy an umbrella or a sun-shade. The weather on the island is unknown but it is always changing and this is reflected in the items that the customers choose to buy. It is the job of the ‘shop assistant’ to observe and remember the sequence of sales to infer the gradual changes in the weather in order to make predictions about what the next customers will buy.

In cover story 2, subjects played the role of an assistant in a shop in the centre of a city selling only red cans of cherry-cola and blue cans of diet-cola. Subjects were briefed with the following details: Outside the shop there are large digital adverts, which are always showing images of these products. The adverts may be biased towards one of the products but they are always changing. The assistant cannot see the adverts. On every trial a
customer will come to the shop and buy one of the drinks, as determined by which product is currently favoured by the adverts outside. It is the job of the ‘shop assistant’ to observe and remember the sequence of sales to infer the gradual changes in the adverts in order to make predictions about what the next customers will buy. In both cover stories subjects were instructed that the hidden states (weather or adverts) change constantly and slowly and that they must consider every sale in order to keep track of the fluctuations. These environmental fluctuations provided cover stories to justify the random drifting of P.

The differences between SV and NSV apply to both cover stories. In SV there was an accompanying ‘manager’. The manager represented a real person. This was another participant, outside the scanner in a different room. Subjects fully understood the perspective of this other person because they had previously participated in that role (see section 3.3.2.2 below for details). Subjects were told that the manager spends some time in a backroom and therefore does not get to observe all of the sales. Therefore, some of the sales contain privileged information that only the ‘shop assistant’ can see. However, on some trials the manager comes out of the backroom and does get to see the sale. These were the trials with shared information. Finally, subjects were told that the manager is watching security camera footage from the backroom to keep track of the sales, but that the manager is unaware that this is actually last week’s footage so the information is misleading. The ‘shop assistant’ is provided with a video-link of what the manager is watching on the security camera footage so they can see all the misleading information that the manager is receiving (decoy trials). Subjects had to try and keep track of the manager’s beliefs.

In the NSV there was a ‘hi-tech cash register’ instead of a manager. Instead of a real person’s beliefs, subjects had to keep track of a computer’s belief state. Subjects were told that the cash register keeps a record of every sale that’s been entered to it and computes a prediction of what the next customer will buy. Subjects were also told that some customers buy with coupons instead of cash, and that these sales will not be entered into the
Chapter 3: Agent-specific learning signals

cash register. These were the privileged trials. Subjects were told that some customers pay with cash and in these instances the cash register will be used and it will update its prediction. These were the shared trials. Finally, subjects were told that the cash register has an internet connection to a partner shop and receives updates from the sales happening there. However, the partner shop is far away from this shop, on another island with different weather for cover story 1, or in another city with different digital adverts for cover story 2. When these sales occur, the cash register is updated with misleading information and the subject can also see this information. These were the decoy trials.

All four games had identical designs and trial structures (Figure 3-1A). A sampling trial was where the subject samples the environment on behalf of themselves (privileged), on behalf of the manager/cash register (decoy) or on behalf of themselves and the manager/cash register at the same time (shared). Sampling trials started with cue image (Figure 3-1B) on the screen to indicate whether it was going to be a privileged, decoy or shared trial. The cue was presented in the centre of the screen with a grey background for 1100 ms. Then the cue disappeared and an outcome was presented (Figure 3-1C) at the centre of the screen for 900 ms. The outcome represented what the current customer had chosen to buy (i.e. the outcome of the Bernoulli trial: heads or tails). Then the outcome disappeared and was followed by an inter-trial interval (ITI) with a central fixation cross that lasted 750-1250 ms.

After 4-9 sampling trials there was a probe trial (Figure 3-1A) where subjects reported their estimate of either P or Pfb. Subjects were told that the two products of their shop were kept in two separate boxes. On probe trials a horizontal scale appeared on the screen with one of the boxes on the left and the other box on the right. The positions of the two boxes were randomly generated on every probe trial. A probe trial could be a Self trial (report P) or an Other trial (report Pfb). On Self trials, subjects were probed with the question ‘Which box would YOU reach into now?’ as if anticipating what the next customer would buy. Subjects had 7 seconds to give their response by moving a virtual arrow left or right along the scale and then pressing an
‘enter’ button once they were happy with the position. The arrow was initially invisible but appeared in a random location along the scale as soon as subjects pressed left or right. This was to avoid any systematic biases that the starting location of the arrow might have induced. Other trials were different between SV and NSV. In SV Other trials, subjects were asked ‘Which box would the MANAGER reach into now?’ and subjects had to respond by putting themselves into the shoes of the manager and reporting their estimate of $P_{fb}$. In NSV this question was rephrased as ‘Which box would you reach into now IF you used the readout on the cash register’. Subjects never received any feedback from their choices on probe trials, so they never knew how accurate their responses were. Probe trials were randomly selected to be ‘Self’ or ‘Other’ so subjects never knew which question was going to come next and they had to try and keep track of $P$ and $P_{fb}$ at all times. However, subjects were never probed with the same question more than 3 times in a row.

Subjects were incentivised to perform well because their final payment at the end of the experiment depended on their accuracy scores on probe trials. We calculated scores based on how much subjects’ responses deviated from the ‘optimal’ response on any given probe trial. In the SV Other trials, the optimal response was whatever the other participant selected for that trial as the ‘manager’. For all other probe trials, the optimal response was taken from the random walks $P$ or $P_{fb}$. Once again, participants never got to see these scores nor received any kind of feedback until they were paid at the end of the experiment.

Before playing the games, participants were carefully instructed and well-practiced to make sure they understood that after a probe trial, the hidden state of the environment would continue from the state it was in before the probe trial. Specifically, they were instructed that after a probe trial, they should not treat the upcoming information as independent of what they have already seen, but rather treat the entire stream of sampling trials as a continuous sequence while the hidden state of the environment (weather or adverts) changed gradually.
3.3.2.2 A subject’s eye view of the experiment

Deception was never used in this experiment; the set up was exactly as it was described to participants. All participants came to the lab on two occasions, T1 and T2. T2 occurred no more than 4 days after T1. Both sessions will be described with the fictional participants Sally and Anne.

Sally and Anne were both in group 1, which meant that while they were in the scanner they played SV1 followed by NSV2, but they never played SV2 or NSV1. Anne arrived at the lab at time T1 for a behavioural session with no scanning. Anne was instructed how to play a simplified version of SV1 called simple-SV1. In simple-SV1, Anne played the role of the ‘manager’. The game proceeded exactly as SV1 (described above) but with the following differences: 1) There was no mention of a ‘shop assistant’ 2) Privileged trials were excluded 3) Anne was instructed to ignore the cue images and told that they were not relevant to the game 4) There was only one type of probe trial, which asked ‘Which box would YOU reach into now?’. The presence of cue images was justified with instructions that said the images simply showed whether Anne (as the ‘manager’) was seeing the sale directly or via security camera footage from the backroom, but that this was irrelevant for Anne’s task.

After this, Anne was instructed how to play a simplified version of NSV2 called simple-NSV2. In this version of the game, Anne played the role of a ‘shop assistant using the information from a hi-tech cash register’. The same differences applied as in simple-SV1. The presence of cue images was justified with instructions that said the images simply showed whether the cash register was getting information from a sale in this shop or in a partner shop next door, but that this was irrelevant for Anne’s task. We did not analyse the behavioural data from simple-SV1 or simple-NSV2.

Finally, Anne was given the following set of personality questionnaires to complete: Balanced Inventory of Desirable Responding (BIDR), Empathy Quotient (EQ), Interpersonal Reactivity Index (IRI), Community Assessment of Psychic Experience (CAPE), Inventory of Callous-Unemotional Traits (ICU), Beck Depression Inventory (BDI). Anne, and all other participants,
filled out the BIDR first in order to provide us with a measure of response bias. The subsequent 5 questionnaires were completed in a random order.

The next day Anne returned for session T2. She was then informed that there was a social element to the experiment and that another participant, Sally, was also there, but that Sally was doing a T1 session. I.e. Sally was doing what Anne had done the previous day. We walked Anne past a testing room so that she would briefly see Sally to convince her that there was indeed another participant. Anne was then taken to a different testing room where she played a short working memory (n-back) task and then learned how to play SV1 and NSV2 with extensive practice. We explained to Anne that in SV1 she would be trying to predict Sally’s choices as the ‘manager’ and that in NSV2 she would be trying to make choices as if she were using the predictions of a computer (cash register).

For SV1 Anne was told that the security footage that Sally was observing was actually uninformative, and for NSV2 Anne was told that the partner shop down the road was closed and so by default the cash register was now receiving misleading information from another partner shop in a completely different place (different island or different city). Finally, Anne was taken to the MEG scanner where she played SV1 followed by NSV2.

For various logistical reasons Sally and Anne did not play their games as the ‘manager’ and ‘shop assistant’ simultaneously, which we explained to Anne. Sally’s responses as the ‘manager’ were saved to a network drive and they were used to calculate Anne’s score on Other probe trials in SV1. The same pre-generated trial sequence was given to Sally and Anne except that privileged trials were excluded from Sally’s simple-SV1 game, which consisted of only 272 sampling trials, while Anne’s SV1 game consisted of the full 444 sampling trials.

Anne had to play simple-SV1 the previous day for two reasons. Firstly, it provided ‘manager’ responses for a previous participant who was playing as the ‘shop assistant’ in the scanner (just like Sally provided ‘manager’ responses while Anne was in the scanner). Secondly, it was to make it more intuitive for Anne to put herself in Sally’s shoes and to understand exactly
what information was and wasn’t available to Sally. The reason why Anne had to play simple-NSV2 was simply so that the images and cover stories for the two games that Anne played at T2 (SV1 and NSV2) were equally familiar.

### 3.3.2.3 Pipeline for generating trial sequences

We followed a stringent pipeline to minimise the correlation between our variables of interest. First, we generated two random walks to represent the time courses of $P$ and $P_{fb}$. One walk started with a value randomly selected from a uniform distribution bound between 0.1 and 0.3 and the other walk started with a value randomly selected from a uniform distribution bound between 0.7 and 0.9. The walks proceeded with step sizes of 0.025. The sign of this step was random in most instances, but because $P$ and $P_{fb}$ are probabilities, the two walks were bound between 0 and 1. To achieve this, the walks were always reflected by these boundaries, which sometimes required a reversal of the randomly selected step sign. The walks were terminated after 444 steps. This resulted in two pseudorandom walks, each with 444 data points. If these two datasets had a non-significant Pearson correlation coefficient they were saved. This process was repeated iteratively until 300 pairs of uncorrelated pseudorandom walks had been generated. Each pair of walks was then used to generate a trial sequence.

To generate the trial sequence, we first generated a sequence of 444 trial types (privileged, shared or decoy). This sequence consisted of 37 concatenated blocks of 12 trials, where each block was a random sequence of 4 of each trial type. Thus, trial type frequencies were balanced throughout the experiment. We then used this sequence of trial types along with one of the pairs of random walks to simulate a sequence of Bernoulli trials. One of the random walks represented $P$ and the other represented $P_{fb}$. For a trial $t_i$, if the trial type was privileged we drew from a Bernoulli distribution with $P$ equal to the $i$th data point in random walk $P$. If the trial type was decoy we drew from a Bernoulli distribution with $P$ equal to the $i$th data point in random walk $P_{fb}$. If the trial type was shared we drew from a Bernoulli distribution with $P$ equal to 0.5. This resulted in one sequence of 444 heads and tails. The
complete trial sequence consisted of a sequence of Bernoulli outcomes and a corresponding sequence of trial types.

We then simulated an agent that observed this trial sequence to generate trial-by-trial estimates of our variables of interest. In order to do this, we used a 2-parameter Rescorla-Wagner (RW) model, which was the simplest RW model in our model space (see section 3.3.3). We selected parameters by taking the mean of parameter estimates across 18 subjects in a separate behavioural pilot study; we used a learning rate \( \alpha \) of 0.1. We tested for correlations between \( B \) and \( B_{fb} \) and between \( |PE^s| \) and \( |PE^o| \). If these correlations were non-significant we saved the trial sequence and moved on to the next pair of random walks. If a significant Pearson coefficient was discovered the process was attempted again with a different pseudorandom sequence of trial types. If, after 300 iterations, no trial sequence was generated that provided uncorrelated variables we moved on to the next pair of random walks. At the end of this process we ended up with 158 suitable trial sequences.

### 3.3.2.4 Task implementation

All tasks were implemented in MATLAB (Mathworks, MA, USA) using Cogent (Wellcome Trust Centre for Neuroimaging, University College London). All images used were edited in Inkscape and processed in MATLAB to ensure equal luminance (average luminance per pixel) with each other and with the plain grey background.

### 3.3.3 Computational models of learning

Our primary hypothesis was that subjects would use prediction errors (PEs) to update their beliefs and that they would also simulate the other person’s PEs to solve the belief inference problem. In order to test this hypothesis, we developed a series of computational models to try and explain subjects’ choice behaviours. A summary of all models is shown in Table 3-1 on page 109. All the models described here were fit to choice
behaviour from the SV and NSV tasks using identical procedures. We tested 3 different groups of models.

The models in group A (Models 1 and 2) assumed that subjects did not use PEs to update their beliefs on each trial, but rather used an averaging technique. Model 1 predicted a response on each probe trial by taking the average of information sampled since the last probe trial. This model predicted choices on Self probe trials by averaging information from privileged and shared trials and predicted choices on Other probe trials by averaging information from shared and decoy trials. Model 2 was the same as model 1 but instead of taking the average over the sampling trials since the last probe trial, this model took the average over the last 10 sampling trials.

The models in group B (Models 3 to 19) assumed that subjects used PEs to form trial-by-trial belief updates. All of these models used two different PE signals, a PE\textsubscript{s} and PE\textsubscript{o} (Equation 3-1) and assumed two update equations on each sampling trial, one for updating the beliefs of the Self and one for updating the beliefs of the Other (Equation 3-2). All of the group B models assumed that a non-zero PE\textsubscript{s} was generated on privileged and shared trials and that a non-zero PE\textsubscript{o} was generated on privileged and decoy trials. However, PE\textsubscript{s} was equal to 0 on decoy trials and PE\textsubscript{o} was equal to 0 on privileged trials (Equation 3-1).

\[
P_{E_s}^{(t)} = \begin{cases} 
0 & (Decoy) \\
Outcome(t) - B_{(t-1)} & (Otherwise) 
\end{cases} \\

P_{E_o}^{(t)} = \begin{cases} 
0 & (Privileged) \\
Outcome(t) - B_{fb(t-1)} & (Otherwise) 
\end{cases} 
\]

*Equation 3-1*

where B\textsubscript{t} is the subject’s belief about P on trial t and B\textsubscript{fb}(t) is subject’s belief about P\textsubscript{fb} on trial t. All group B models assumed that these estimates were both initialised at 0.5. Outcomes were coded as zero or one. The simplest models used the following pair of update equations:
where $\alpha$ is the learning rate. This is a free parameter that was fit to each subject and remained constant throughout the task. It should be noted that when $\text{PE}^s$ and $\text{PE}^o$ are zero (Equation 3-1) $B$ and $B_{fb}$ remain stationary.

To account for any information forgotten since the last update, some models included an additional memory-decay parameter $\delta$, which governs a decay of $B$ and/or $B_{fb}$ back to the initial value of 0.5:

\[
B(t) = B(t-1) + \alpha \text{PE}^s(t) + \delta(0.5 - B(t-1))
\]

\[
B_{fb}(t) = B_{fb}(t-1) + \alpha \text{PE}^o(t) + \delta(0.5 - B_{fb}(t-1))
\]

Equation 3-3

Finally, some models allowed for the possibility that the $\text{PE}^s$ might be used to erroneously update $B_{fb}$ or that $\text{PE}^o$ might be used to erroneously update $B$. Some models assumed a degree of this PE leakage with the parameter $\lambda$:

\[
B(t) = B(t-1) + \alpha \text{PE}^s(t) + \lambda \text{PE}^o(t)
\]

\[
B_{fb}(t) = B_{fb}(t-1) + \alpha \text{PE}^o(t) + \lambda \text{PE}^s(t)
\]

Equation 3-4

These three free parameters $\alpha$, $\delta$ and $\lambda$ could take any value between zero and one. Different models can be generated that incorporate various combinations of these parameters. $\alpha$ can be shared between the two update equations or alternatively, each equation can have its own $\alpha$ with different
values. The same can be said for $\delta$ and $\lambda$, which can also be excluded from either equation entirely. For example, some models included a $\lambda$ parameter for one update equation, but not the other, to allow for a unidirectional PE leak. Group B comprised a total of 17 models, allowing us to explore different ways these parameters could be used together. The differences between these models are summarised in Table 3-1 on page 109.

The final group of models, group C (Models 20 to 21), had the same update equations as in Equation 3-3 (incorporating a shared $\alpha$ and shared $\delta$) but these models tested the possibility that subjects did not selectively update their beliefs depending on the cues. Model 20 assumed that a non-zero $\text{PE}^s$ was generated on every sampling trial despite the fact that the information on decoy trials should not have been relevant for updating Self. Model 21 assumed that a non-zero $\text{PE}^o$ was generated on every sampling trial despite the fact that the information on privileged trials should not have been relevant for updating Other.
Chapter 3: Agent-specific learning signals

<table>
<thead>
<tr>
<th>Model Number</th>
<th>Model group</th>
<th>α</th>
<th>τ</th>
<th>δ</th>
<th>λ</th>
<th># Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>B</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>B</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1 (shared)</td>
<td>7</td>
</tr>
<tr>
<td>13</td>
<td>B</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1 (PE₀ updates B)</td>
<td>7</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1 (PEₙ updates B₀)</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>B</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>B</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1 (shared)</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>B</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1 (PE₀ updates B)</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>B</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1 (PEₙ updates B₀)</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>B</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>C</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>C</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3-1: Summary of all models fit to behavioural data

21 models were evaluated in total. The table illustrates the differences between these models by identifying the groups A-C in three different colours, and displaying which parameters were included in each model. In the α, τ and δ columns a ‘1’ indicates a parameter shared between the update equations for the two agents. A ‘2’ indicates that two different values of that parameter were fit for each of the two update equations. In the λ column, a ‘2’ indicates that two different leak parameters were fit, one for each update equation, allowing for an asymmetrical bidirectional leak. ‘1 (shared)’ indicates that one leak parameter was shared between both equations to allow for a symmetrical bidirectional leak. Alternatively, one parameter was included in only one equation to allow for a unidirectional leak. The direction of any unidirectional leak is described in the table.
3.3.4 Model-based behavioural data analysis

21 models were fit to the choice behaviour (on probe trials) of each subject with MATLAB’s non-linear optimisation function fmincon (Mathworks, MA, USA). We used this function to find the optimal model parameters for each subject as defined by the minimum negative log-likelihood of the subject’s choices conditioned on a set of estimated parameter values. To start the optimisation, every parameter was initialised with a value randomly drawn from a uniform distribution bound by the relevant upper and lower bounds for that parameter. The optimisation procedure was iterated at least 20 times for each model fit, with different initial parameter values each time, to avoid local minima. We selected the iteration with the best fitting optimised parameters and discarded the rest.

In order to obtain likelihood values, we derived a choice likelihood function (CLF) for each trial, which specified the likelihood of any choice that could have been made by the subject on that trial. The subject’s choice could take any value along a continuous scale. Because subjects were technically reporting a probability, this scale was bound between zero and one and we used a beta distribution to approximate this CLF. Beta distributions are conventionally parameterised by two shape parameters, $\alpha$ and $\beta$ (equation 5). It should be noted that this is not the same $\alpha$ as the learning rate used in the learning models.

$$f(x; \alpha, \beta) = \frac{x^{(\alpha-1)}(1-x)^{(\beta-1)}}{B(\alpha, \beta)}$$

Equation 3-5

where $x$ is a choice and $f(x; \alpha, \beta)$ is the likelihood of that choice given the shape parameters of the beta distribution. $B$ denotes the beta function, a normalisation constant to ensure the total probability integrates to one.

We wanted to parameterise the CLF with more meaningful parameters: 1) The most likely choice on that trial and 2) The variability or temperature of the subject’s choices. We assumed that the most likely choice
Chapter 3: Agent-specific learning signals

on a Self probe trial was the current estimate $B(t)$ and the most likely choice on an Other probe trial was the current estimate $B_{fb}(t)$ and the mode was assigned to one of these variables:

$$
Mode \left\{ \begin{array}{ll}
B(t) & \text{Self probe} \\
B_{fb}(t) & \text{Other probe}
\end{array} \right.
$$

Equation 3-6

The variance of the beta distribution was assigned to the value of a free parameter called $\tau$. This parameter was fit to each subject separately and remained constant throughout the task. It was bound between 0.0001 and 0.08. $\tau$ is a temperature parameter that captures how noisy each subject was in his or her mapping from belief to action. Like the previous parameters discussed, a model could have a shared $\tau$ parameter for Self and Other, or two separate parameters that can vary independently.

$$
\sigma^2 \leftarrow \tau
$$

Equation 3-7

In order to draw the CLF on each trial we derived the beta distribution shape parameters, $\alpha$ and $\beta$, which can both be expressed in terms of mode (Equation 3-8) and variance (Equation 3-9). These two equations could then be solved simultaneously to obtain the shape parameters and then the Beta distribution itself (Equation 3-5).

$$
\frac{\alpha - 1}{\alpha + \beta - 2} = Mode
$$

Equation 3-8
Chapter 3: Agent-specific learning signals

\[
\frac{\alpha \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \sigma^2
\]

Equation 3-9

The real choice that was made by the subject could then be read off the CLF and the corresponding likelihood would contribute to the joint likelihood of all choices made conditioned on the current model parameter estimates. We computed the Bayesian Information Criterion (BIC) for each model and each subject separately (Equation 3-10), and compared the mean BIC value across subjects for each model to assess relative model evidence. It should be noted that the CLF is a probability density function, and so the likelihoods were often larger than one. Therefore, the log-likelihoods were positive and their corresponding BIC values were negative.

\[
BIC = k\ln(N) - 2\ln(\hat{L})
\]

Equation 3-10

where \( N \) is the number of data points that the model was fit to (in this case number of probe trials), \( k \) is the number of free parameters in the model, and \( \hat{L} \) is the maximised value of the likelihood function of the model.

Using the BIC as an approximation for log model evidence, we also compared the winning model and the second-best model using a hierarchical Bayesian model to estimate the posterior probability that any randomly chosen subject in the sample had data generated by one of those models and not the other. This random-effects Bayesian model selection approach enables us to estimate the parameters \( \alpha \) of a Dirichlet distribution of the probabilities \( r \) of the models being compared. These probabilities inform a multinomial distribution over the model space. The hierarchical model is inverted using variational Bayesian approximation (Stephan et al., 2009).
Then from the Dirichlet parameters, one can compute the expected multinomial parameters $\langle r_k \rangle$ for each model $k$ as follows:

$$\langle r_k \rangle = \frac{\alpha_k}{(\alpha_1 + \ldots + \alpha_k)}$$  \hspace{1cm} \text{Equation 3-11}$$

One can also compute an exceedance probability $\varphi_k$, i.e. the belief that a particular model $k$ is more likely than any other model tested, given the group data $y$:

$$\varphi_k = p(r_k > 0.5 | y; \alpha)$$  \hspace{1cm} \text{Equation 3-12}$$

### 3.3.5 Model-free behavioural data analysis

For each subject, on each probe trial we computed the absolute difference between the subject’s report and the true underlying probability from random walk $P$ or random walk $P_{fb}$. We then computed this difference again for a simulated agent that positioned the arrow at random locations along the response scale. We then subtracted the real difference from the simulated difference on each probe trial and took the mean across probe trials. The resulting value describes how much better than a chance subject, the real subject performed. We took the mean across subjects to assess group level performance.

### 3.3.6 MEG data acquisition and pre-processing

MEG was recorded continuously at 600 samples/second using a whole-head 275-channel axial gradiometer system (CTF Omega, VSM MedTech), while participants sat upright inside the scanner. 2 gradiometers (ML042 and MRC12) were out of service throughout data collection; we pre-processed and analysed data from the remaining 273 channels. We recorded four runs of data for each subject (two runs for SV and two runs for NSV).
Participants made responses on four buttons with a button box using the fingers they found the most comfortable. The buttons had the following functions: Move arrow left, move arrow right, move arrow faster, enter choice.

All MEG pre-processing was carried out using the FieldTrip data analysis toolbox (Oostenveld et al., 2011) on each run of data independently. First, we epoched the data around the relevant triggers and scanned the data for any SQUID jump artefacts. In the whole experiment, one jump artefact was found, and the relevant trial was excluded from the analysis. Then we down-sampled the data from 600 Hz to 100 Hz, and filtered the data using a bandpass of 0.5-150 Hz and a stopband of 48-52 Hz to remove line noise. We then ran an independent component analysis (ICA) using the built-in ft_componentanalysis function in FieldTrip and manually inspected the components for obvious eye artefacts and cardiac ECG artefacts. The relevant components were removed and the data was reconstructed. All analyses were conducted on the epoched, filtered, resampled and cleaned data in units of femtotesla.

3.3.7 Mass-univariate linear regression

All MEG analyses were performed on data time locked to the onset of the outcome of the sampling trial. This is the information that subjects would require to generate a PE and update their beliefs. In order to regress PE against recorded brain activity, we performed a mass-univariate regression analysis where we took the absolute PE magnitude as our regressor to ensure we were not capturing brain activity that correlated with the visual attributes of, or other associations with, the image that represented the outcome. We had to consider the fact that PE and PE were coded as zero on some trials. Therefore, we conducted two separate regressions. One regression model for PE excluded decoy trials and another regression model for PE excluded privileged trials. Therefore, in both regression models, the regressor would contain no systematic pattern of zero-valued
elements. If we hadn’t excluded the zero-valued trials, the regression would have been confounded by trial type and cue images (PE\textsuperscript{s} was 0 only after a particular cue image on a decoy trial and PE\textsuperscript{o} was zero only after a particular cue image on a privileged trial).

We regressed PE magnitude against the ERF at each sensor and each time point, to produce a spatiotemporal map of unsigned regression weights. At each sensor we subtracted the median pre-stimulus value. We then upsampled the data to create a 95x95 2D pixel map of these baseline-corrected effect sizes. Including the time dimension, we ended up with a 3D image for each subject. At the group level we performed a one-sided Wilcoxon signed rank test at each pixel in these 3D images to ask whether the group median was significantly greater than 0. We used a non-parametric test here because baseline-corrected unsigned regression weights are not normally distributed. Here we are testing the null hypothesis that the effect size (unsigned regression weight) is no larger than the pre-stimulus effect size.

We identified points with activations above a cluster-forming threshold (\(p < 0.001\)). We then identified clusters of contiguous supra-threshold points in this 3D image, which could extend through space and time. We made cluster level inference by repeating this analysis 300 times, using permuted trial sequences, to generate null distributions of cluster-extent, from which we derived significance thresholds (\(p = 0.05\) FWE corrected).

### 3.3.8 ‘Pseudotrial’ data construction

We wanted to test whether we could distinguish a neural pattern encoding a PE\textsuperscript{s} from a pattern encoding a PE\textsuperscript{o} and thus determine whether a Self-Other distinction can be achieved on the basis of these signals.

A typical way to identify the neural pattern encoding a PE is to regress the magnitude of the PE (derived from our learning model) against the brain activity, across trials. This would yield a single beta estimate at each sensor, capturing the slope of the relationship between PE and brain activity at that
sensor. However, in order to use powerful multivariable methods like support vector machine (SVM) classification to look for differences in the spatial patterns of PE\textsuperscript{s} and PE\textsuperscript{o}, it was necessary to obtain multiple samples of each pattern. One way to achieve this is to divide the data into multiple partitions (without replacement) and repeat the analysis in each partition to obtain multiple independent samples of the spatial pattern for each type of PE. This is the approach we opted for, using the smallest possible partitions: pairs of trials (Figure 3-6A).

To maximise power without introducing bias, we randomly partitioned trials into pairs under the constraint that each pair contained one trial above the median PE and one trial below the median PE (Figure 3-2). Thus, the difference in brain activity between the two trials within a pair corresponded to a representation of PE. We performed this random partitioning independently for PE\textsuperscript{s} and PE\textsuperscript{o}. This resulted in two sets of difference images, corresponding to neural representations of PE\textsuperscript{s} and PE\textsuperscript{o}. Finally, we could then apply multivariable methods to classify whether each difference image was a representation of PE\textsuperscript{s} or PE\textsuperscript{o}.

It should be noted that this method differs slightly from typical pattern-based neuroimaging analyses described in, for example (Haynes, 2015). Usually, such an analysis looks for a neural representation of some variable. This is achieved by training a classification or regression model to distinguish patterns of neural activity corresponding to different values of that variable. Above-chance accuracy of the model indicates that the brain activity contains information about the variable. However, in our case we were interested in a difference in the representation of a variable between two conditions. Because the representation itself is defined by a difference in neural activity between a large PE and a small PE, we were looking for a difference of differences. Thus, it was necessary to train classifiers on patterns of subtracted activity rather than activity patterns from individual trials.

We started with N (444) trials in total. First, we partitioned all privileged and shared trials (2N/3 trials) by median split on |PE\textsuperscript{s}|. We then
randomly sampled two trials, one from either side of this partition, and subtracted the ERF on the low $|PE^s|$ trial from the ERF on the high $|PE^s|$ trial, at every sensor and time point. For ease of reference, we call this subtracted image a ‘pseudotrial’. We continued randomly sampling pairs of trials without replacement to obtain a total of $N/3$ pseudotrials. Each of these pseudotrials describes the difference in activity between a trial with a high $|PE^s|$ and a trial with a low $|PE^s|$. The brain activity in the difference image thus constituted a representation of $|PE^s|$.

Second, we partitioned all decoy and shared trials ($2N/3$ trials) by median split on $|PE^o|$. We carried out the same procedure as for $|PE^s|$, resulting in a second set of $N/3$ pseudotrials, each of which constituted a representation of $|PE^o|$. At each time point we trained a classifier to distinguish $|PE^s|$ pseudotrials from $|PE^o|$ pseudotrials.

![Figure 3-2: Illustration of how Pseudotrial MEG data were constructed](image)

For each variable of interest, in this case the Self-attributed PE signal, we conducted a median split across trials (horizontal dashed line). Note that only trials where the variable is non-zero are included, so here all decoy trials are excluded. Pairs of trials are randomly sampled, with one trial above the median and one trial below the median. Here three example pairs are selected: Pink, brown and purple. For each pair, the MEG signal in the low-PE trial is subtracted from the MEG signal in the high-PE trial to generate a whole-brain contrast image. Each contrast image is a noisy representation of the variable of interest. A whole stack of these contrast images is generated. For the classification analysis, this process is repeated for the Other-attributed variable and we attempt to classify the Self-images from the Other-images.
3.3.9 Classification analyses for agent decoding

We used a non-linear support vector machine, LIBSVM for MATLAB (Chang & Lin, 2011), with a radial basis function (RBF) kernel, which was trained on pseudotrial data with 273 features, one for each sensor. Cross-validation was performed with repeated random subsampling. We performed 200 iterations, and on each iteration two pseudotrials from each class (Self and Other) were randomly sampled and constituted a testing set while the remainder constituted a training set. On each fold of cross-validation, the training set and testing set were both de-meaned and scaled with respect to the means and standard deviations of the data in each feature (MEG sensor) in the training set. Accuracy was measured as the percentage of correct predictions, averaged over the 200 folds of cross-validation. This value indicated how much information was available in the \( |PE| \) signals to discriminate between PE\(^s\) and PE\(^o\). We performed this analysis at each point in peristimulus time to generate a time course of classification accuracies. This procedure was conducted once for the SV and again for the NSV.

We tested all classifiers on a range of hyperparameter combinations (regularisation constant \( C \) and RBF parameter \( \gamma \)). We optimised each classifier by selecting the combination of hyperparameters that produced the highest classification accuracy averaged across subjects. The time course of classification accuracies was then smoothed with a moving average filter (moving average span of 10 data points).

For statistical inference, we adopted a permutation-based method to determine whether any classification accuracy was significantly better than chance. This procedure has been recommended for making inferences on information-based neural measures such as classification accuracy (Allefeld et al., 2016).

We computed significance thresholds by running 150 simulations of these exact analyses, each time with data generated from permuting the trial order. For each permutation we obtained a maximal classification accuracy and used these values to generate a null distribution. If the real classification accuracy exceeded the 95\(^{th}\) percentile of the null distribution, it was deemed
significantly above chance. This method corrects for multiple comparisons in the time domain. We also used this method to determine whether the difference in classification accuracy between SV and NSV, at each time point, was significant. For this we generated a null distribution of maximal (SV–NSV) classification accuracies and another null distribution of maximal (NSV–SV) classification accuracies.

3.3.10 Analysis of questionnaire data

We used scores from the BIDR questionnaire to determine whether any subjects were likely to have a large response bias. This questionnaire allocates a point whenever a subject gives an extreme response (6 or 7 on a 7-point Likert scale) to a question, indicating that they might be answering in such a way as to preserve their reputation. No subject had a score more than 2.5 standard deviations greater than the sample mean so we had no reason to believe that any subject had an unusually large response bias.

For each subscale of our five questionnaires of interest we set up a regression model with gender and age as predictor variables and the subscale score as a dependent variable. We then took the residuals from these regression models as age- and gender-controlled scores for each subscale. We then z-scored each of these nine age- and gender- controlled subscales and entered them into a principal components analysis (PCA). We investigated the principal component that explained the most variance in the data (PC1).

When conducting the correlational analysis at each time point between the PC1 score and the neural agent decoding value we computed a significance threshold with a permutation-based null distribution. For each of the 150 simulations of the classification analysis we simulated the correlation analysis to generate a time course of –ln(p) values for each of the three types of pseudotrial and then concatenated these three time courses together. For each permutation we took the maximal value of this triple-length time course, resulting a null distribution of maximal –ln(p) values. The
95th percentile of this distribution represented a significance threshold correcting for multiple comparisons across time and across the three types of pseudotrial.

3.4 Results

3.4.1 Behaviour

For each subject, we assessed behavioural accuracy independently for the two versions of the task (SV and NSV) as well as on the two types of probe trial (Self and Other). This resulted in four conditions overall, where accuracy was defined relative to chance performance (see section 3.3.5). Where an accuracy of zero is equivalent to chance level performance, mean accuracies (with standard deviations) for SV were 0.11 (0.04) in Self probe trials and 0.11 (0.04) in Other probe trials. For NSV, these were 0.11 (0.04) for Self probe trials and 0.10 (0.05) for Other probe trials (Figure 3-3).

The group performed significantly better than chance in all four conditions as assessed with 4 separate one-sample t-tests on the mean accuracies per subject [p < 0.0001 in all four conditions]. There were no differences in accuracy between the two probe trial types, or between the two versions of the game [ANOVA: main effect of probe trial type: F(1, 148) = 1.54, p = 0.22. Main effect of game version: F(1, 148) = 0, p = 0.96. Interaction: F(1, 148) = 0.06, p = 0.81]. Thus, all four conditions were similar in difficulty.
Chapter 3: Agent-specific learning signals

Figure 3-3: Subject performance in behavioural task

Performance of all subjects, relative to chance. Accuracy is measured as distance, in pixels, along the virtual scale. Higher numbers indicate higher accuracy relative to a player that positions the arrow randomly. Chance performance is zero, which would indicate that a random player’s deviation from ground truth is no larger than the subject’s deviation from ground truth. Horizontal bars indicate mean and standard error of the mean. There were no significant differences in performance between the different types of probe trials or the different games (SV and NSV).

3.4.2 Modelling simulated belief updates with simulated PEs

We fit 21 models to the probe trial behaviour of each subject, separately for the SV and NSV games. There were three principal groups of model (detailed in section 3.3.3). Group A models assumed that subjects’ beliefs were constructed from an average over recently sampled information.

Group B models were based on an assumption of Rescorla-Wagner updating (Rescorla & Wagner, 1972), where the models derive PEs on each trial from the difference between the actual and expected outcomes. PE^Self updated the beliefs of Self, while PE^Other was a simulation of the other agent’s PE, for updating the beliefs of Other in SV, or counterfactual Self in NSV. A subset of Group B models also included leak parameters that allowed PE
Chapter 3: Agent-specific learning signals

signals to erroneously update the wrong agent’s belief, thus capturing an inability to maintain separate belief updates for the two agents. All group B models also assumed that the PE\textsuperscript{S} had a value of zero on decoy trials whilst PE\textsuperscript{O} had a value of zero on privileged trials. Qualitative differences between group A and group B models are shown in supplementary Figure 3-9.

Group C models were like group B models except they did not make this assumption; instead they allowed PE\textsuperscript{S} and PE\textsuperscript{O} to update the beliefs of Self and Other respectively in all three trial types.

We compared models separately for SV and NSV using the Bayesian Information Criterion (BIC). For both the SV and NSV, model 8 had the lowest mean BIC value (Figure 3-4A). This model incorporated two separate PE signals and included four free parameters: a learning rate ($\alpha$) regulated the update of the beliefs of the two agents, a memory decay parameter ($\delta$) controlled the rate of ‘forgetting’ for the beliefs of the two agents, and two temperature parameters ($\tau_s$ and $\tau_o$) governed choice stochasticity on Self probe trials and Other probe trials respectively. This model generated synthetic choice data qualitatively similar to subjects’ real choice data (Figure 3-4B). Parameter recovery is shown in supplementary Figure 3-10.

Noting large intersubject variability in BIC values, we also employed a random effects Bayesian model selection (Stephan et al., 2009) to compare the winning model with the second best model (supplementary Figure 3-11) and found, for both SV and NSV, an exceedance probability in excess of 0.99. This is the probability that the winning model better explains a randomly chosen subject’s data.

We also assessed the correlation between parameter estimates fit to SV and parameter estimates fit to NSV (Figure 3-4C). For each model we obtained a correlation coefficient for each parameter and then took the mean of those coefficients as a summary statistic for the between game consistency of the model. Because parameter values were not normally distributed, we computed the non-parametric Spearman’s rank correlation coefficient. We found that model 8 also had the highest between game
consistency. Thus, this model captured consistent dispositions in subjects’ choice behaviour across the two games.

After fitting models to the behavioural data, we then had parameter estimates for each model and each subject. We used model 8 along with each subject’s parameters for this model to generate trial-by-trial estimates of latent PEs and beliefs, which we then used in subsequent analyses on the MEG data. Note that PEs and belief values generated by other models were very similar and consequently our findings were not sensitive to the selection of a particular model.

**Figure 3-4: Behavioural model comparison**

A) Results from a Bayesian model comparison for SV (left) and NSV (right). The bar representing the winning model in both versions of the game is highlighted in bold (model 8). Error bars show standard error of the mean (SEM). The winning model in both cases is a four-parameter model that incorporates two types of PE signal, each one attributed to a different agent. B) Generative performance of the winning model (model 8) compared to the generative performance of a less successful model (model 1). In the top panel we used
model 8 to simulate choice data for one subject. In the bottom panel we used model 1 to simulate choice data for the same subject. To select which subject to use for this display, we computed the median BIC score of the model 8 fits to SV data and selected the subject whose fit was closest to this value. Thus, the subject can be considered an ‘average’ subject in terms of goodness-of-fit of the winning model. The simulated data (red) is qualitatively similar to the subject’s real choice data. C) Consistency of each model between the two games. Each bar shows, for a different model, how correlated the subjects’ parameter estimates fit to SV data are to subjects’ parameter estimates fit to NSV data. The model with the highest Spearman coefficient (averaged across parameters) is highlighted in bold. This was again model 8, the very same model with the best average predictive performance for SV and NSV independently. Note that model 1 and model 2 do not have error bars because these models both contained a single parameter, and so only one correlation coefficient was computed for both these models. Error bars show SEM.

3.4.3 Neural representations of PEs and simulated PEs

We next asked whether $|PE^S|$ and $|PE^o|$ were encoded in the MEG signal recorded during task performance using a mass-univariate analysis (Figure 3-5). For each subject, we fit two separate linear regression models at each sensor and each peristimulus time point, time locked to the onset of the trial outcome. We obtained trial-by-trial estimates of $|PE^S|$ and $|PE^o|$ using the winning model’s estimated free parameters fit to the choice data. The first model regressed $|PE^S|$ against the event-related field (ERF) on privileged and shared sampling trials (i.e. trials where $|PE^S|$ was non-zero). The second regression model regressed $|PE^o|$ against ERFs on decoy and shared trials (i.e. trials where $|PE^o|$ was non-zero). This resulted in four statistical maps over sensors and time, two for the SV game and two for the NSV game.

We converted each of these maps into a 3D image (two spatial dimensions and one temporal dimension) of baseline-corrected effect sizes (see section 3.3.7 for details). To make group level inferences we conducted a one-sided Wilcoxon signed rank test at each pixel to determine whether the group median was significantly greater than zero. We thresholded the resulting 3D image with a cluster-forming threshold ($p < 0.001$) and identified
clusters of contiguous supra-threshold pixels, which could extend through space and time.

**Figure 3-5: Regressing PE magnitude against MEG ERF**

Group level statistical maps projected onto scalp surface (frontal sensors towards top of page) after regressing $|PE^s|$ and $|PE^o|$ against the ERF for SV and NSV, time locked to the onset of the outcome. All four regressions yielded significantly large clusters that extended through multiple sensors and time points. The maps highlight the spatial extent of these clusters (dotted lines) at specific points in peristimulus time. The plots underneath show the temporal waveform (smoothed with a moving average filter with span of ten timepoints) of the regression effect size from averaging over all spatial regions within the cluster. This has been averaged across subjects. The shaded regions show SEM.

We determined whether any clusters were significantly larger than chance with a non-parametric permutation test to generate null distributions of cluster extent. In each of the four regression models we found clusters significantly larger than chance at a 0.05 family wise error (FWE) level. In SV the clusters extended through parietal and occipital sensors whilst in NSV the clusters extended through frontal and parietal sensors. For SV PE$^s$ the largest cluster extended from 330 ms to 390 ms and comprised 2628 pixels (threshold 612). For SV PE$^o$ the largest cluster extended from 340 ms to 420 ms and comprised 2032 pixels (threshold 624). For NSV PE$^s$ the largest cluster extended from 370 ms to 440 ms and comprised 1621 pixels (threshold 554). For NSV PE$^o$ the largest cluster extended from 310 ms to
Chapter 3: Agent-specific learning signals

370 ms and comprised 847 pixels (threshold 569). Despite finding significant clusters at the group level, we also noted large intersubject differences in these spatiotemporal patterns (supplementary Figure 3-12).

3.4.4 Decoding agent identity from learning signals

In order to test whether Self-Other distinction can be achieved on the basis of these PE signals, we conducted a classification analysis on ‘pseudotrials’ labelled as Self or Other. The process by which we generated these pseudotrials is described in section 3.3.8.

We tested classifiers in cross-validation, yielding a time course of classification accuracies (CAs) and compared the SV with the NSV tasks. The absolute difference in CA underlying reliable effects was in some cases as small as one percent. In observing this we note that effect sizes cannot be inferred from absolute CAs (Allefeld et al., 2016, Jamalabadi et al., 2016, Hebart & Baker, 2017). Therefore, for statistical inference we used a permutation-based approach, as described in section 3.3.9. This procedure has been recommended for making inferences on ‘information-based’ neural measures such as CA (Allefeld et al., 2016). This allowed us to make statistical inferences without making assumptions about how CAs (or CA differences) are distributed, whilst also correcting for multiple comparisons across time points, at a 0.05 FWE level.
Figure 3-6: Decoding agent identity from neural patterns of learning signals

A) Generation of pseudotrial data. These plots show the distributions of, and relationship between, trial-by-trial variables for the two agents (Self on x-axis and Other on y-axis). The left panel is for the $|PE|$ variable, the middle panel is for the ‘signed belief’ variable and the right panel is for the ‘unsigned belief’ variable. These are the typical estimates for one subject and each dot corresponds to a sampling trial. In all three plots, the vertical blue line shows the median split used to generate pseudotrials that were labelled as Self while the horizontal red line shows the median split used to generated pseudotrials that were labelled as Other. Note that in the left panel, the data form three clusters. The cluster of points on the x-axis correspond to privileged trials, the cluster of points on the y-axis correspond to decoy trials and the cluster of points in the middle correspond to shared trials. In order to generate the $|PE|$ pseudotrials, the median splits were only performed on data where the $|PE|$ was non-zero. B) Time course of group mean classification accuracies in SV (red) and NSV (blue) for the $|PE|$ pseudotrials (left), ‘signed belief’ pseudotrials (middle) and ‘unsigned belief’ pseudotrials (right). Classifiers trained on all three types of pseudotrial could predict agent identity in SV ($p < 0.05$ FWE corrected). However, in NSV the classifiers only showed a small, late decoding effect for the ‘signed belief’ pseudotrials. In all three analyses there were multiple time points where classification accuracy was significantly higher in SV than in NSV ($p < 0.05$ FWE corrected). Red stars indicate significant above-chance classification for SV. Blue stars indicate significant above-chance classification for NSV. Black stars indicate significant above-chance difference between SV classification and NSV classification. Shaded regions show standard error of the mean across subjects. The
Chapter 3: Agent-specific learning signals

MEG sensors that contributed to decoding accuracy are shown in supplementary Figure 3-14.

We found that Self and Other could be classified significantly above chance level from the spatial patterns of activity that represented $|PE^a|$ and $|PE^o|$ approximately 300 ms after stimulus onset (Figure 3-6B). However, CA did not exceed chance level when we conducted this same analysis on NSV data. Moreover, at approximately 300 ms there was a significant difference between CA in SV and CA in NSV. Thus, distinct spatial activity patterns for $|PE^a|$ and $|PE^o|$ were evident in SV but not in NSV. This implies information about Self and Other is intrinsic to the representations of low level learning signals, whilst information about Self and counterfactual Self is not.

To test the robustness of this finding we performed two further variants of the analysis, by constructing pseudotrials from subjects’ trial-by-trial ‘signed beliefs’ (B and $B_{fb}$) and ‘unsigned beliefs’ ($|B - 0.5|$ and $|B_{fb} - 0.5|$). The former are the subject’s trial-by-trial estimates of the underlying Bernoulli parameter from the perspective of each agent. The latter are the absolute distances of these estimates from 0.5, which represents an equal probability of either outcome. The ‘unsigned belief’ is thus a measure of confidence in what the next outcome will be.

It should be noted that here we can use all trials to generate pseudotrials. We found that classifiers trained on pseudotrial data, generated from either of these latent variables, could predict agent identity (Self or Other) significantly above chance in the SV game. However, in the NSV game the classifiers could only predict agent identity (Self or counterfactual Self) for pseudotrials generated from ‘signed beliefs’, and in this instance the signal was weaker and occurred later in time than was the case for the SV game (Figure 3-6B). Furthermore, we found that CAs for SV were significantly larger than CAs for NSV at multiple time points for both of these pseudotrials. Finally, for comparison, in a separate analysis classifying between the visual stimuli, we obtained similar decoding accuracies in SV and NSV (supplementary Figure 3-13).
3.4.5 Neural decoding of agent identity predicts Self-Other distinction

An important question is whether the neural distinction in learning signals is related to a behavioural measure of Self-Other distinction. A subset of our behavioural models (models 11-19) included a leak ($\lambda$) parameter that governed the extent to which PE$^s$ was erroneously used to update $B_{hb}$ and/or PE$^o$ was erroneously used to update $B$, thus indexing an inability to discriminate between two different agents’ learning processes. We estimated $\lambda$ values by selecting the best fitting $\lambda$-containing model for each individual subject. If the best model contained two $\lambda$ parameters we took the mean of the two values. We derived two estimates of $\lambda$ for each subject, one for the SV and one for the NSV.

We then computed, for each subject, a metric describing overall neural Self-Other distinction (SOD). In order to do this, we took the maximal CA from each of the time courses from the three types of pseudotrial (Figure 3-6B) and summed these three numbers. This provided one number for neural agent decoding in SV and another number for neural agent decoding in NSV.

Because $\lambda$ in SV and $\lambda$ in NSV were strongly correlated across subjects, we examined the difference between SV and NSV. Due to the non-normally distributed parameter estimates, we computed a non-parametric Spearman’s rank correlation coefficient. We found a strong negative correlation (Figure 3-7) between the neural decoding contrast [SV − NSV] and the estimated $\lambda$ contrast [SV − NSV]: Spearman’s rho: $-0.43$, $p < 0.01$.

We also tested the accuracy of a linear regression model that used neural decoding contrasts to predict the estimated $\lambda$ contrasts. Here we used cross-validation with random subsampling (train on half, test on half) and recorded the correlation between predicted and observed values on every fold. The median Pearson coefficient across 10,000 folds was 0.31, which was significantly greater than chance as determined by a non-parametric permutation test ($p = 0.039$).
Figure 3-7: Relationship between neural SOD and behavioural SOD

Each dot is a subject. The x-axis shows the contrast (SV – NSV) of maximum classification accuracies, pooling data from all three types of pseudotrial. The y-axis shows the contrast (SV – NSV) of estimated $\lambda$ parameters, which quantify misattributing belief updates to the wrong agent’s belief. The negative correlation implies that the subjects for whom we obtained more accurate agent decoding in SV than NSV, also behaviourally better discriminated between agents in SV than NSV. The inset violin plot shows the distribution of prediction accuracies (correlation coefficients) of a linear regression model across 10, 000 folds of cross-validation, where neural SOD (SV – NSV) predicted estimated $\lambda$ (SV – NSV). The horizontal black bar indicates the median of this distribution.

3.4.6 Neural decoding of agent identity predicts subclinical traits

Finally, we asked whether neural agent decoding relates to intersubject differences in subclinical psychopathological traits. All subjects filled out five questionnaires of interest: Beck Depression Inventory (BDI), Empathy Quotient (EQ), Interpersonal Reactivity Index (IRI), Inventory of Callous-Unemotional traits (ICU) and the Community Assessment of Psychic Experience (CAPE). These questionnaires were specifically chosen to assess the presence of psychopathological traits previously proposed to
Chapter 3: Agent-specific learning signals

relate to a dysfunctional Self-Other distinction or more general social
cognitive deficits (Liepelt et al., 2012, Sevgi et al., 2015, Lamm et al., 2016,
Ladegaard et al., 2016). These questionnaires assessed autistic (EQ),
SCHIZOTYPAL (CAPE), antisocial (ICU) and depressive (BDI) traits as well as
general capacities for empathy and sympathy (EQ, IRI). We also obtained
measures of response bias using an additional questionnaire, the Balanced
Inventory of Desirable Responding (Li & Bagger, 2007). None of the subjects
were considered to have an unacceptably high response bias (see section
3.3.10 for details).

We performed dimensionality reduction on age- and gender-
controlled personality questionnaire data (see section 3.3.10 for details)
using a principal components analysis (PCA). Including all subscales of the
five questionnaires of interest gave nine dimensions in total (Figure 3-8A).

The principal component (PC1) that explained the most variance in
the data (32%) loaded negatively with both subscales of the CAPE
questionnaire (schizotypy), BDI (depression), ICU (antisocial behaviour) and
one subscale of the IRI (personal distress in social situations) and loaded
positively with EQ and other subscales of the IRI (Figure 3-8A). Thus, PC1
negatively captured psychopathological features in our personality data in a
nonspecific manner.

We projected the personality data into the space of this principal
component to obtain a score for each subject. First, we correlated the neural
SOD metric, as described in the previous section, with the PC1 scores.
When using the contrast of [SV – NSV] this yielded a significant correlation: r
= 0.39, p = 0.017 (Figure 3-8B). We also tested the accuracy of a linear
regression model that used neural decoding contrasts to predict the PC1
scores, using the same method as described for Figure 3-7. The median
Pearson coefficient across 10,000 folds was 0.34, which was significantly
greater than chance as determined by a non-parametric permutation test (p =
0.04). Thus, subjects for whom we obtained higher classification accuracies
in SV than NSV scored lower on PC1. In other words, subjects for whom it
was easier to neurally decode Self from Other, than to decode Self from
counterfactual Self, scored higher on a nonspecific anti-psychopathological component.

When looking at SV and NSV neural SOD metrics separately, we found significant positive correlation for SV \((r = 0.43, p < 0.01)\) but no significant correlation for NSV \((r = 0.01, p = 0.94)\).

We then investigated the temporal evolution of this relationship for each of the three types of pseudotrial. At each peristimulus time sample we correlated the subjects’ PC1 scores with \([CA(SV) – CA(NSV)]\) to generate a time course of Pearson coefficients (Figure 3-8C). Using permutation-based thresholding, to correct for multiple comparisons across time and across the three types of pseudotrial, we found a significant positive correlation \((p < 0.05 \text{ FWE})\) approximately 110 ms after stimulus onset (Figure 3-8C) when using the ‘signed belief’ pseudotrials. This falls within the window of significant Self-Other distinction in signed beliefs \((100 – 340 \text{ ms})\) as shown in Figure 3-6B.
Chapter 3: Agent-specific learning signals

Figure 3-8: Relationship between neural agent decoding and subclinical traits

A) Component coefficients (loadings) for personality data in the first principal component (PC1). This component loads negatively on psychopathological traits, including both subscales of the CAPE questionnaire (schizotypy), the BDI (depression), the personal distress subscale of the IRI and the ICU (unemotional and callous traits). B) Correlation between neural SOD metric $[SV - NSV]$ with PC1 scores. Each dot is subject. The inset violin plot shows the distribution of prediction accuracies (correlation coefficients) of a linear regression model across 10,000 folds of cross-validation, where neural SOD $[SV - NSV]$ predicted PC1 scores. The horizontal black bar indicates the median of this distribution. C) Pearson coefficient time course (red) from correlating subjects’ PC1 scores with $[SV$ classification accuracy $- NSV$ classification accuracy] for the classifiers trained on 'signed belief' pseudotrials. A time course of corresponding $-\ln(p)$ values is shown (blue) to indicate the significance of the correlation at each time point. The horizontal black line shows the permutation-based corrected threshold for statistical significance in terms of $-\ln(p)$.

3.5 Discussion

We show that a representation of a learning signal (PE or belief) is encoded with a different neural spatial pattern when the signal is attributed to Self as compared to when it is attributed to another agent. Intersubject variability in this difference correlated between subjects with a behavioural measure of Self-Other distinction, and with subclinical psychopathological
traits. This suggests that Self-Other distinction is realised by an encoding of agent identity that is intrinsic to low level learning signals, and the fidelity with which this occurs is an important dimension of variation between individuals.

In our experiment subjects had to solve two simultaneous computational problems. The first problem was predicting what the next outcome would be. The second problem was identifying whether this belief-state about the next outcome should be attributed to one agent or another, a computation that requires a Self-Other distinction. We found a spatial segregation between Self-attributed and Other-attributed learning signals. This means that the neural representations of beliefs and PEs in this task also contained information about the agent to whom these signals belong, and consequently the neural resources that compute the next outcome also inevitably contribute to computing a Self-Other distinction. It is of interest therefore that the degree of spatial segregation was correlated with a behavioural measure of Self-Other distinction derived from our learning models.

Previous work has shown that neuronal populations in the macaque anterior cingulate cortex preferentially encode simulated RPEs (Chang et al., 2013) and the future decisions (Haroush & Williams, 2015) of another monkey. Conversely, human fMRI data has identified common activations in the mPFC that represent RPEs (Suzuki et al., 2012) or subjective preferences (Garvert et al., 2015) for both Self and Other in an agent independent manner. Likewise, mirror neurons recorded from the macaque premotor cortex are also agent independent (di Pellegrino et al., 1992). A Self-Other distinction in the affective domain has been reported in terms of dissociable networks for experienced versus vicarious pain (Krishnan et al., 2016, Lopez-Sola et al., 2017), though other reports suggest that these are both are subserved by the same structures (Lamm et al., 2011, Rütgen et al., 2015a).

The above accounts are conflicting with respect to whether Self- and Other- attributed signals have common or distinct neural activations. One possible reason for this is that previous studies eliciting simulated signals
Chapter 3: Agent-specific learning signals

were not indexing a Self-Other contrast per se, but rather a contrast of executed behaviour versus observed behaviour, where the subject receives feedback from the observee’s behaviour. In these cases, learning or decision variables are discriminated not by virtue of the agent to whom they are attributed, but instead by virtue of distinct input modalities and cognitive demands required for instrumental learning and observational learning respectively. As these factors are heavily task-specific, for instance dependent on the way in which the subject accrues information about the other agent’s behaviour, it is unsurprising that observational learning paradigms have produced inconsistent accounts of the neural encoding of agent identity.

In the present study, we devised a Self-Other contrast per se by allowing subjects to observe what the other agent observed but not the other agent’s behaviour. By requiring subjects to switch between attributing a signal to Self and attributing a signal to Other, with fixed sensory input modalities and cognitive demands, we could show that learning signals do contain information about the identity of the agent to whom they are attributed. Consistent with this are our findings that the neural Self-Other distinction is modulated by individual personality traits, as well as by the precise contextual relationship between the agents in question, as assessed with social and non-social versions of the paradigm. Our findings do not rule out a possibility that the brain uses additional mechanisms to distinguish Self from Other, for instance with an explicit encoding of agent identity that is separate from low-level learning signals. However, our results support the theory that agent-specific learning signals are sufficient for the brain to achieve a Self-Other distinction during mentalising.

Our results support the idea that the brain updates simulated beliefs of another agent using PEs calculated within the frame of reference of that agent. Previous work on Other-referenced processing has shown that humans and other primates simulate another agent’s experience of unexpected reward (Burke et al., 2010, Suzuki et al., 2012, Chang et al., 2013). Our work extends these findings to the domain of updating beliefs.
Chapter 3: Agent-specific learning signals

about non-reward related quantities. A simulated sensory prediction error such as what we observed, combined with information about a preceding state, could also be used to simulate how another agent learns transition models of complex environments with multiple states, a requisite for goal-directed behaviour (Glascher et al., 2010, Daw et al., 2011, Dolan & Dayan, 2013, Smittenaar et al., 2013, Smittenaar et al., 2014).

Differences between neural representations of signals attributed to Self and Other, which we attribute to agent identity, might relate to some other features of the task frame (Hunt et al., 2013, Marti et al., 2015, Peters et al., 2016). However, if this were true, we would expect representations to be similarly distinct in NSV as in SV, since both versions shared a shift in task frame (between Self and Other in SV, and between Self and a counterfactual Self in NSV). The finding of significantly less distinct representations in NSV supports our conclusion that features intrinsic to agent identity are fundamental to the distinct representations observed in SV. Although representations were overall less distinct in NSV, there was nevertheless a detectable difference in ‘signed belief’ representation between Self and counterfactual Self, occurring approximately 550 ms after stimulus onset. It will be important in future work to clarify why this long latency separation occurs in a non-social context.

The differences we found between SV and NSV do not necessarily mean that subjects were not engaging in social computations in NSV. Despite evidence for a so called Theory of Mind network (Saxe & Wexler, 2005, Amodio & Frith, 2006, Jenkins et al., 2008) recruited during mentalising, there is evidence that the brain might also rely on domain general computations for social cognition (Behrens et al., 2008, Heyes, 2012). It has been suggested that ‘social’ computations like mental state inference and ‘non-social’ computations, such as mental time travel and metacognition, are underpinned by the same general capacity for meta-representation (Carruthers, 2009, Frith, 2012, Ladegaard et al., 2016, Chambon et al., 2016). If Self-Other distinction is a special case of a domain
general computation, future work should seek to understand why this computation is executed differently in social and non-social contexts.

We observed substantial intersubject heterogeneity in the spatiotemporal pattern of PE signals in our task. Although anatomical inferences are limited for data acquired in sensor space (Baillet et al., 2001, Troebinger et al., 2014), the heterogeneity would suggest a diversity of cortical regions encoding PEs. fMRI studies, employing both learning and non-learning paradigms, have reported unsigned sensory PE activity or activity corresponding to unexpected neutral stimuli in a range of cortical and subcortical regions, including the anterior insula and inferior frontal gyrus (Weilnhammer et al., 2017), primary sensory cortices (den Ouden et al., 2009, Horga et al., 2014, Lee & Noppeney, 2014), superior temporal sulcus (Arnal et al., 2009), hippocampus (Kumaran & Maguire, 2006), cerebellum (Schlerf et al., 2012), striatum (Zink et al., 2003, den Ouden et al., 2009, den Ouden et al., 2010) and midbrain (Bunzeck & Duzel, 2006).

With regards to timing, previous studies in a non-social context using electroencephalography (EEG) (San Martin, 2012, Walsh & Anderson, 2012, Sambrook & Goslin, 2014) and MEG (Talmi et al., 2012) have identified signed PE signals 200-350 ms after stimulus onset. However, unsigned PE signals are less well characterised and appear to be encoded across a much broader time window, ranging from 145 ms to 640 ms after stimulus onset (Sambrook & Goslin, 2014). Finally, previous false belief experiments using EEG (Liu et al., 2004, Ferguson et al., 2015) and MEG (Mossad et al., 2016) have reported latencies ranging from 100 ms to 800 ms after stimulus onset, at which time signals have differentiated true beliefs from false beliefs. Here we show that in a social setting, these PE signals are agent-specific as soon as they are detectable, approximately 300 ms after stimulus onset. Conversely, in the NSV task, these PE signals were not agent-specific at any time point.

The contrast in agent decoding accuracy between the SV and NSV correlated with subjects’ behavioural ability to differentiate between agents and with the first principal component of subclinical personality traits.
Chapter 3: Agent-specific learning signals

Specifically, subjects for whom Self and Other brain representations were more distinct than Self and counterfactual Self scored higher on this principal component. An inability to differentiate between Self and Other is a feature of psychopathology (Baron-Cohen et al., 1985, Bird et al., 2010, Liepelt et al., 2012, Sevgi et al., 2015, Palmer et al., 2015, van der Weiden et al., 2015, Beeney et al., 2015, Lamm et al., 2016). Our measures of agent decoding might be useful as a sensitive gauge of a Self-Other distinction in the context of phenotypic markers for psychopathology.

Computational phenotyping in psychiatry has recently been mooted (Montague et al., 2012) and posits individualised diagnostic and therapeutic tools in mental health based on computational models of behaviour and brain function. Recent efforts to develop computational models of Theory of Mind (Hill et al., 2017a, Baker et al., 2017) do not address how representations are attributed to different agents. Here we present the foundations for a model of Theory of Mind that specifically addresses computations that contribute to a Self-Other distinction, a quantifiable characteristic necessary in both social and non-social contexts.
3.6 Supplementary material

**Figure 3-9: Qualitative differences between group A and group B models**

A) The red line shows the latent beliefs of a simulated subject who takes the average of the last 10 trials (same as model 2 in Group A). The blue line shows the latent beliefs of another simulated subject who uses Rescorla-Wagner updating on each trial, with a learning rate of 0.1. The Group A model simulation has beliefs that change with large steps whilst the Group B model simulation has beliefs that change in smaller gradations. B) Simulated choice behaviour of the two simulated agents in panel A using a temperature parameter of 0.001. The Group A model simulation is more likely to overshoot and use the extremes of the scale. C) Choice behaviour of two real subjects. The red line shows a subject who displayed the strongest evidence (relative BIC) for Group A models. The blue line shows a subject who displayed weak evidence (relative BIC) for Group A models. The
Chapter 3: Agent-specific learning signals

behavioural pattern mirrors that shown in the simulations in panel B, with subject 1 using the extremes of the scale more often.

Figure 3-10: Parameter recovery

To further test the identifiability and construct validity of the winning model and its parameters, we ran a parameter recovery test. We used model 8 to simulate choice data for each subject and then refitted the model to the simulated data. We assessed the degree of parameter recovery by computing the correlation between true parameter estimates and refitted parameter estimates for each of the four parameters and then taking the average of these correlation coefficients. Due to the non-normal distribution of the parameter estimates, we computed non-parametric Spearman’s rank correlation coefficients. The figure shows two scatter plots displaying parameter estimates fit to data simulated by model 8 against parameter estimates fit to empirical data. Each dot represents a parameter estimate for one subject. SV is on the left and NSV is on the right. The parameters fit to the simulated data are highly correlated with the parameters fit to the real data, demonstrating successful parameter recovery. Parameter values are z-scored for display purposes.
Chapter 3: Agent-specific learning signals

To further quantify the difference between the best and second-best models, we used a hierarchical Bayesian model to estimate the posterior probability that one model, and not the other, generated a randomly chosen subject's data. For SV (left) and NSV (right) the exceedance probability (probability that model 8 is more likely than model 4) was at least 0.99. $\alpha_1$ and $\alpha_2$ are the Dirichlet parameter estimates that define the probability density function. $r_1$ is the expected likelihood that model 8 (rather than model 4) generated the data for any randomly chosen subject. $r_2$ is the expected likelihood that model 4 (rather than model 8) generated the data for any randomly chosen subject. EP stands for exceedance probability.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{posterior_distributions.png}
\caption{Posterior Dirichlet distributions from Bayesian model selection}
\end{figure}
Figure 3-12: Patterns of PE encoding in two exemplar subjects

Statistical maps (regression effect size) plotted over scalp for two subjects at same time points as in Figure 3-5. There is substantial intersubject heterogeneity in the spatiotemporal patterns of the signals.

Figure 3-13: Decoding the visual stimuli

To make sure the MEG data were clean and appropriate for conducting decoding analyses, we tested to see if we could decode the visual stimuli used for cues and outcomes. For SV (left) and NSV (right) we trained a linear support vector machine on the MEG ERFs in response to cues for privileged trials and decoy trials (green) and another linear support vector machine on the two outcome stimuli (magenta). We only used data from 94 occipital sensors as here we are exploiting visual information. We trained and tested these classifiers at every time point during a trial. The early and late vertical black bars indicate the onset of the cue and the outcome respectively. For both SV and NSV, we could decode the cue image immediately after cue-onset and we could decode both the cue and the outcome immediately after outcome onset. The levels of decoding were similar for both
Chapter 3: Agent-specific learning signals

SV and NSV, indicating that visual discrimination of the stimuli was similar in both games. Shaded regions indicate standard error of the mean across 38 subjects.

Figure 3-14: Spatial topography of neural Self-Other distinction

Group level statistical maps of z-scored classification accuracies (frontal sensors towards top of the page). To determine which sensors contributed to neural Self-Other distinction, we repeated the pseudotrial analysis 3000 times, each time using an independent random subsample of 10 MEG sensors. For each sensor we found all the samples that included that sensor and calculated the average classification accuracy of those samples. This produced a spatial map of classification accuracies, which we averaged across subjects. This method is described in more detail in (Kurth-Nelson et al., 2015). On average, left posterior frontal sensors were more implicated for decoding PEs whilst occipital sensors were more implicated in decoding beliefs.
Chapter 4: Meta-representational associative learning

This work presented in this chapter is in preparation as a research article. I would like to acknowledge Tobias Hauser for providing guidance on the fMRI analysis and Rani Moran for providing guidance on drift-diffusion modelling.

4.1 Abstract

Humans show variability in their ability to distinguish between the mental states of different agents. Self-Other distinction may depend on the identity of the Other in question, or on the general ability of the individual to construct distinct agent-specific models. Here, we propose a learning mechanism to account for this variability. The hypothesis is that Self- and Other-attributed learning signals can become associated, effectively reducing the burden of representing two distinct models. We tested this hypothesis by asking healthy adults to track the beliefs of two different social agents. With one agent, there was a strong contingency between Self- and Other-attributed belief updates. With the other agent, there was a weak contingency between Self- and Other-attributed belief updates. In a testing phase, subjects were re-exposed to the two agents. Even though there was no difference between the agents in testing, subjects were more proficient at Self-Other distinction in the low-contingency condition than the high-contingency condition. This behavioural effect was associated with neural plasticity in Self- and Other-attributed prediction error representations in sensory cortical regions, such that neural Self-Other distinction was reduced.
in the high-contingency condition with respect to the low-contingency condition. This learning transferred to a different cognitive task that assessed perspective-taking on the same social agents, suggesting that the training induced generalised changes in Self-Other distinction. Finally, we found that the microstructure of ventromedial prefrontal white matter was associated with the degree of functional plasticity induced by the training, and also with subclinical traits of socio-cognitive dysfunction. The results suggest that the vmPFC is involved in deploying meta-representational structural knowledge about Self-Other relations, in a context-dependent manner.

4.2 Introduction

The findings presented in Chapter 3 raise the question of why there is variability in Self-Other distinction. We observed two types of systematic variability. The first type was a within-subject, context-dependent variability, whereby neural Self-Other distinction was greater in a social condition than a non-social condition. This finding indicates that the identity of the other agent, another individual or counterfactual Self, is a determinant of Self-Other distinction. The second type was a between-subject variability, associated with differences in task-related behaviour and subclinical psychopathological traits.

Here we propose a hypothetical mechanism for explaining the degree to which mental states can be selectively attributed to Self and Other. The hypothesis states that neural representations of Self- and Other-attributed mental states are subject to associative learning mechanisms, and that this is sufficient for variability in Self-Other distinction to emerge. This kind of learning can be described as ‘meta-representational associative learning’. Meta-representational associative learning differs from basic associative learning in that the animal learns associations between representations of subjective mental states, rather than representations of objective quantities. Formally, we can describe this as learning an association between a change in one representation and a change in another representation. For instance,
an animal learns that tone A predicts reward B with some unknown probability. Through repeated exposures, uncertainty in the transition probability from A to B will be resolved, and the representation of A will change. The animal has learned a simple model of the environment, $M_1$. The animal might learn that another tone X predicts a punishment Y with some other unknown probability. The animal has learned another simple model of the environment, $M_2$.

\[
M_1 \rightarrow p(A) \propto p(B) \\
M_2 \rightarrow p(X) \propto p(Y)
\]

**Equation 4-1: Two representational associations (models)**

$P(n)$ denotes the probability of encountering a state $n$ at some moment in time. The proportionality between two probabilities indicates that encounters of one state are associated with encounters of another state. The fact that the representations constitute mere subjective models, rather than objective quantities of the environment, is represented in the prediction error signals, $PE_{AB}$ that scales the change in the representation of A, and $PE_{XY}$ that scales the change in the representation of X. Meta-representational associative learning refers to learning an association between these two PE signals, such that a change in $M_1$ becomes associated with a change in $M_2$.

\[
\mu \rightarrow \frac{dM_1}{dt} \propto \frac{dM_2}{dt}
\]

**Equation 4-2: One meta-representational association (meta-model)**

In (Shea, 2012) it is argued that a reward prediction error signal satisfies the conditions that philosophers claim a meta-representation should satisfy (Millikan, 1984, Shea, 2007, Proust, 2007). These include containing information about the correctness of another representation, and this information being used for some downstream function. Thus, a neural signal that merely correlates with a quantity in the environment is not necessarily a
representation’ of that quantity. Similarly, a neural signal that correlates with a neural representation is not necessarily a ‘meta-representation’. Furthermore, in Chapter 3 we showed that sensory prediction error signals can embody a conjunctive representation of sensory information and agent identity, meaning they contain all the information necessary for a propositional attitude like ‘I am surprised that I saw pink umbrella’. The meta-representational associative learning hypothesis provides a mechanism by which such signals can become more or less agent-specific, and hence more or less able to support propositional attitudes.

The hypothesis predicts that if there is a contingency between updates to a Self-attributed model and updates to an Other-attributed model, then the representations of these two update signals will become more similar to each other. Consequently, the content of the models themselves will become correlated. This would mean that when an individual knowingly shares experiences with another agent, the Self- and Other-attributed mental states would become more similar, even when the mental states pertain to unfamiliar experiences or environments. This may facilitate both informational and normative types of social conformity (Sistrunk, 1973, Kaplan & Miller, 1987, Campbell & Fairey, 1989, Abrams et al., 1990, Wood et al., 1994, Toelch & Dolan, 2015), which can promote efficient model updating and social cohesion respectively.

This is consistent with humans conforming their beliefs and values to those they interact with (Campbell-Meiklejohn et al., 2010, Zaki et al., 2011, Edelson et al., 2011, Garvert et al., 2015, Moutoussis et al., 2016) and conforming more to peers and ingroup members than outgroup members (Abrams et al., 1990, Cialdini & Goldstein, 2004, Stallen et al., 2012, De Dreu & Kret, 2016). However, in such studies of conformity, the subject typically learns about the stationary belief or preference of another agent, as if it were a trait. This means that the low-level learning, about the Other, and the meta-representational learning process, about the relation between the Self model and the Other model, are indistinguishable.
Chapter 4: Meta-representational associative learning

The meta-representational associative learning hypothesis is also consistent with ingroup and outgroup effects in social psychology. Humans are less likely to attribute secondary emotions to outgroup members than ingroup members (Leyens et al., 2000, Leyens et al., 2001). Humans also regularly dehumanise outgroup members by considering socially distant Others as being very similar to each other (Haslam, 2006, Waytz et al., 2010), and are less able to differentiate mental states of socially distant Others (Thornton et al., 2019a). If an individual has associated their own model of the world with socially close Others, then when uncertainty is resolved in their Self-attributed model, it will also be resolved in the models they attribute to close Others. The models of socially distant Others, however, will be represented with a higher degree of uncertainty.

Meta-representational learning could explain how humans acquire the kinds of priors that influenced the behaviour of our subjects in Chapter 3. A subtle change in cover story, and the perceived identity of an Other (counterfactual Self or another person) was sufficient to induce detectable changes in a neural measure of Self-Other distinction. Furthermore, an inability to engage in this kind of learning might limit the flexibility in Self-Other distinction that would be useful for engaging in fluent social interactions. Thus, variability in this learning mechanism could be a predictor of the kind of socio-cognitive dysfunctions seen in schizophrenia (Liepelt et al., 2012, van der Weiden et al., 2015), autism spectrum disorder (ASD) (Baron-Cohen et al., 1985, Bird et al., 2010, Palmer et al., 2015, Sevgi et al., 2015) or borderline personality disorder (BPD) (Beeney et al., 2015).

Here, we reused the probabilistic false belief task described in Chapter 3 in which low-level learning processes, for Self and Other, are divorced from the meta-representational learning about the relationship between Self and Other. We tested for the predicted effects of meta-representational learning by employing two training conditions. In the first condition, subjects played with an agent where there was a strong temporal contingency between belief updates attributed to Self and belief updates attributed to Other. In the second condition, subjects played with an agent
where there was a weak contingency between these belief updates. In neither condition was it optimal to use one agent’s belief updates on behalf of the other, thus if Self- and Other-attributed belief updates were to become associated in the first condition, it would actually impair task performance.

We tested for a learning effect by re-exposing subjects to the two agents. Firstly, they played the false belief task with the two agents again, but this time with no systematic difference between the two tasks. Thus, any difference in behaviour between the two conditions, at this test phase, would indicate that subjects learned something about relations to the other agents in training. Secondly, subjects played a separate visual perspective-taking task with the same two agents. This was a visual target-detection task rather than a learning task, and was designed as an assay for generalised meta-representational learning. It enabled us to test whether the training in the false belief task was general enough to induce changes in the represented relations between Self- and Other-attributed mental states, in a manner independent from the peculiarities of the training environment.

We tested for plasticity in the representations of Self- and Other-attributed prediction errors by conducting the test phase of the false belief task with concurrent functional magnetic resonance imaging (fMRI). This enabled us to compare neural Self-Other distinction for the two conditions. We predicted that neural measures of Self-Other distinction would be reduced with the agent for whom there was a strong contingency of Self- and Other-attributed updates, when compared with the agent for whom there was a weak contingency.

Finally, we performed a multi-parameter mapping (MPM) MRI sequence, which provided us with subject-specific high-resolution anatomical maps of magnetisation transfer (MT), an in vivo marker of myelin density (Stanisz et al., 1999, Stuber et al., 2014, Marques et al., 2017, Allen et al., 2017, Ziegler et al., 2019). Childhood social experience is an important determinant of white matter structure (Eluvathingal et al., 2006) and false belief understanding in infancy has been associated with the development of local white matter structure in ‘social brain’ regions such as medial prefrontal
cortex (mPFC) and temporo-parietal junction (TPJ) (Grosse Wiesmann et al., 2017). Furthermore, recent studies in mice have found that social isolation causes prefrontal hypomyelination in both juveniles (Makinodan et al., 2012) and adults (Liu et al., 2012). However, myelinating oligodendrocytes can be re-activated to increase prefrontal myelination and restore normal social behaviour, either through social re-integration (Liu et al., 2012, Makinodan et al., 2017) or pharmacological manipulation (Liu et al., 2016). Here we explored whether myeloarchitectural variability is associated with our computationally-derived measures of Self-Other distinction, or with subclinical traits of socio-cognitive dysfunction. We expected that MT levels in mPFC and TPJ would be associated with these measures.

4.3 Methods

4.3.1 Participants

47 healthy adults (26 female) aged 19-54, participated in the experiment. They were recruited from the UCL Institute of Cognitive Neuroscience subject pool. All participants had normal or corrected-to-normal vision and had no history of psychiatric or neurological disorders. No participants had taken part in the experiment described in Chapter 3. All participants provided written informed consent, which was approved by the Research Ethics Committee at University College London, under ethics number 4446/003.

Six participants did not complete the scanning part of the experiment, either because they were uncomfortable during scanning, or because they had metallic implants in their body, making them ineligible for scanning. One participant was excluded from all analyses as it was evident, on debriefing, that they did not understand all of the tasks. This left 46 subjects (25 female) with a mean age of 26.5 (SD 7.8) who were included in the analysis of the visual perspective-taking task, and a subgroup of 40 subjects (22 female) with a mean age of 26.8 (SD 8.1) who were included in all other analyses.
4.3.2 Modified false belief task

Subjects were exposed to the probabilistic false belief task that was introduced in Chapter 3. However, they did not play a non-social and social version. Rather, they played two social versions, each time with a different social partner. A cartoon avatar that represented the other agent was presented on screen on every trial (Figure 4-1), regardless of whether it was ‘privileged’, ‘shared’, or ‘decoy’. Participants played once, where an avatar called ‘Fred’ was the manager, and once where an avatar called ‘Maria’ was the manager. In one of these games there was a very high proportion of shared trials compared to privileged and decoy trials. Thus, there was a strong contingency between the onsets of Self-attributed PEs and Other-attributed PEs. The other agent in this task will hereafter be described as the Hi-Share agent. For the other game, there was a very low proportion of shared trials compared to privileged and decoy trials. Thus, there was a weak contingency between the onsets of Self-attributed PEs and Other-attributed PEs. The other agent in this task will hereafter be described as the Lo-Share agent. Subjects were not told about this manipulation. They were simply told that they would play two games, each with a different participant as the manager.

With the Hi-Share agent, there were 112 privileged trials, 112 decoy trials, and 224 shared trials (448 total). With the Lo-Share agent, there were 196 privileged trials, 196 decoy trials, and 56 shared trials (448 total). The trial sequences were generated in the same way as described in section 3.3.2.3 but instead of using blocks of 12 trials with 4 of each trial type, the Hi-Share task was constructed using blocks of 16 trials, composed of 4 privileged trials, 4 decoy trials and 8 shared trials. The Lo-Share task was constructed using blocks of 16 trials, composed of 7 privileged trials, 7 decoy trials and 2 shared trials. For both conditions, trial sequences were selected that produced no correlation between trial-by-trial Self beliefs and Other beliefs (assuming a leak parameter of zero).
The mapping between condition (Lo-Share or Hi-Share) and avatar (Fred or Maria) was counterbalanced across subjects, as was the order in which the two games were played. One game was presented with the umbrellas cover story described in Chapter 3, whilst the other game was presented with the drinks cover story. The mapping between cover story and condition was also counterbalanced across subjects.

The timing of the task was modified to make it suitable for fMRI scanning. The cue and outcome were now presented simultaneously for 1500 ms, followed by a variable ITI with a fixation cross on screen for 1000-1500 ms (Figure 4-1).

Subjects played both versions of the task back to back, outside of the scanner, in a testing room. This was a training phase. About 24 hours later, they played both tasks again, inside the MRI scanner. This was a testing phase. In this testing phase, the trial design for Lo-Share and Hi-Share was identical. In both cases, the three trial types were presented in a 1:1:1 ratio with 148 of each trial type, as in Chapter 3. During the testing phase, the only difference between the Lo-Share and Hi-Share tasks were the stimuli used to represent the other agent (Fred or Maria) and the stimuli used to represent outcomes (umbrellas or drinks), both of which were counterbalanced between subjects. Thus, any difference in behaviour or BOLD signal between the Lo-Share and Hi-Share tasks, during the testing phase, could only be attributed to learning that took place during the training phase the previous day.
Figure 4-1: Modified false belief task

The task we used was very similar to the task introduced in Chapter 3. The task was slightly modified in that the cue and outcome stimuli were presented simultaneously. Furthermore, the avatar that represented the manager was on screen during every sampling trial.

Fred and Maria represented two real people, who were other participants that had taken part in the experiment. Thus, every participant had to play a simplified version of the task, as the manager, four times. They played twice with one cover story and twice with the other cover story. This simplified version of the task is described in section 3.3.2.2. This way, every participant could be represented as ‘Fred' in training and testing for one future participant, and also as ‘Maria', in training and testing, for a different future participant. This set up made it clear to participants that the experiment was genuinely social in nature. The behavioural data from the simplified tasks were not analysed.

4.3.3 Visual perspective-taking task

In order to test whether a training effect could generalise beyond the false belief task, we developed a transfer paradigm (Figure 4-2). Whilst the probabilistic false belief task required participants to discriminate between the learning processes of Self and Other, this transfer task required
participants to discriminate between the visual perspectives of Self and Other. We reasoned that if our training procedure could induce learning about the relationship between mental states attributed to Self and mental states attributed to Other, then the ability to discriminate between the visual perspectives of Self and that same Other should be affected.

This visual perspective-taking task was adapted from a task developed by Samson and colleagues (Samson et al., 2010). In the original task, participants observed an avatar facing one wall of a room. Dots were visible on the wall faced by the avatar. On some trials there were additional dots on the wall behind the avatar. Thus, on ‘congruent’ trials, the subject and the avatar could see the same number of dots, whilst on ‘incongruent’ trials the subject could see more dots than the avatar could see.

In the original task, on ‘Self’ trials participants were first presented with the word ‘SELF’, followed by a number. Participants were then shown the room with the avatar. They had to respond with a ‘yes’ key or a ‘no’ key as quickly as possible to confirm whether or not the number of dots they could see in the room was the same as the number previously shown. ‘Other’ trials proceeded in the same way, except that participants were first presented with the word ‘HE’ or ‘SHE’ and they had to respond from the perspective of the avatar. Thus the ‘yes’ and ‘no’ responses pertained to whether the number of dots that the avatar could see was consistent with the number previously shown.

The main finding was that subjects made more errors, and responded slower, on incongruent trials relative to congruent trials. The effect on Other trials has been described as ‘egocentric’ interference, in that the subject’s own personal perspective interferes with the processing of the Other’s perspective. The effect on Self trials has been described as ‘altercentric’ interference. Altercentric interference supposedly describes the automatic and involuntary computation of another agent’s perspective, which interferes with judgement of one’s own visual experience (Samson et al., 2010).

This finding, that another agent’s irrelevant perspective interferes with visual processing, has been replicated by several investigators (Ramsey et
al., 2013, Schurz et al., 2015, Ferguson et al., 2017, Drayton et al., 2018, Marshall et al., 2018). However, similar results have been observed when the avatar is replaced with an arrow (Santiesteban et al., 2014). Thus, some have argued that the incongruency effect does not reflect automatic mentalising per se, but is merely driven by a directional cue that biases attention towards one side of the room, interfering with visual processing of the whole room.

It should be noted that this ‘directional cueing’ explanation can only explain the incongruency effect on Self trials and may be an alternative to the ‘altercentric interference’ explanation. It cannot explain the ‘egocentric interference’ effect seen on Other trials, where if anything, the directional cue should facilitate performance because the participant only needs to attend to that side of the room. Some researchers have suggested that as soon as participants see the room, the saliency of the dots makes them ‘pop-out’ against the background, allowing for subitisation. On Other trials, this quick subitisation could lead to a response that is incorrect because only some of the dots are relevant to the trial (Santiesteban et al., 2017). In summary, the key finding from the dots task can be explained by appealing to domain-general cognitive processes that are not necessarily mentalistic.

Our visual perspective-taking task was structured similarly to the dots task. However, instead of counting dots, subjects had to count the number of patterns that matched a target pattern (Figure 4-2). On every trial a target pattern was shown, along with a target number. Then the room was shown with an avatar facing one wall. On every trial, there were two patterns on the wall that the avatar could see and two patterns on the wall behind the avatar. These latter two patterns were not visible to the avatar. Some of the patterns matched the target pattern, and some of them were distractor patterns, which looked like the target pattern but were rotated 60 degrees clockwise. The orientation of the target pattern changed randomly on every trial, preventing

\[12\] Subitisation is the ability to instantaneously recognise the number of objects in a scene without explicitly counting.
subjects from learning to anticipate a specific target pattern over the course of the task.

![Image](image.jpg)

**Figure 4-2: Visual perspective-taking task**

A single trial of the task that was adapted from (Samson et al., 2010) is shown. First, the perspective was displayed (YOU, HE, SHE or ARROW). Then a target pattern and number were shown. Then the room was shown, along with one of the two avatars or an arrow (all three possibilities are displayed in the figure, but in the task only one of these would be shown to the participant). Subjects had three seconds to respond with either a ‘yes’ or a ‘no’ key. In this example, the participant can see three target patterns and one distractor pattern. The avatar can see one target pattern and one distractor pattern. The correct response here is ‘no’. If the perspective cue had said ‘SHE’ then the correct response would have been ‘yes’.

This task has several advantages over the original dots task. Firstly, in the dots task, congruent trials always have dots on one wall, whilst on incongruent trials, there are dots on both walls. There is an inherent confound between congruency and the spatial distribution of objects in the scene. In the current task however, there are always four patterns on the walls. On Self trials, the participant is required to compare each of the four patterns with the target pattern held in working memory. On Other trials, the participant is required to compare the two patterns, on the wall seen by the avatar, with the target pattern. However, there is no systematic difference in spatial distribution between congruent and incongruent trials.

Secondly, in the original task congruent trials where the correct response is ‘no’ are trivially easy. This is because the target number
corresponds to a number of dots that is inconsistent with the perspectives of both Self and Other. Researchers conventionally only analyse trials where the correct answer is 'yes', leaving half of the trials unanalysed. In the current task, none of the trials are trivially easy because a comparison between presented patterns and target pattern is always required.

Finally, the current task is more difficult than the original dots task. In the original dots task, participants often perform close to ceiling, limiting the sensitivity of analyses of error rates. Conventionally, response time (RT) is taken as the dependent measure of interest. Because the current task is more difficult, RT and error rate can both be analysed with adequate sensitivity. This also makes the task more amenable to drift-diffusion modelling, a classic approach for modelling binary decisions under time pressure.

As the training procedure in the false belief task was designed to induce learning about relations between Self and specific Others, our visual perspective-taking task required perspective-taking on multiple different agents. Rather than seeing the same avatar in the room on every trial, subjects saw three different avatars, throughout the task. These avatars were Fred and Maria, representing Lo-Share and Hi-Share from the false belief task, and an arrow.

Each trial started with a fixation cross that was presented for 500-1000 ms, followed by the perspective that the participant was required to take ('YOU', ‘HE’, ‘SHE’ or ‘ARROW’) for 500 ms. When the cue said ‘HE’ or ‘SHE’, subjects had to adopt the perspective of Fred and Maria respectively. When the cue said ‘ARROW’, subjects simply had to report whether the number of target patterns that an arrow pointed to was consistent or inconsistent with a target number. After the perspective was displayed, the target pattern and target number were displayed for 750 ms. Finally, the room, along with an avatar, was presented and subjects had up to three seconds to respond with a ‘yes’ or ‘no’ key (L and K keys on the keyboard, counterbalanced across subjects).
The task consisted of 384 trials in total and had a factorial design (Figure 4-3). Perspective (Self or Other), condition (congruent or incongruent), response (yes or no), avatar on screen (Lo-Share, Hi-Share or arrow) and avatar gaze (left or right) were balanced in a $2 \times 2 \times 2 \times 3 \times 2$ design, with 8 trials in each cell of the design matrix. These 384 trials were presented in a randomised order. Subjects had a short rest every 94 trials.

<table>
<thead>
<tr>
<th># trials</th>
<th>Perspective</th>
<th>Condition</th>
<th>Response</th>
<th>Avatar on screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL (384)</td>
<td>SELF (192)</td>
<td>Cong (96)</td>
<td>Yes (48)</td>
<td>LoShare (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HiShare (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrow (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No (48)</td>
<td></td>
<td>LoShare (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HiShare (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrow (16)</td>
</tr>
<tr>
<td></td>
<td>OTHER (192)</td>
<td>Cong (96)</td>
<td>Yes (48)</td>
<td>LoShare (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HiShare (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrow (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No (48)</td>
<td></td>
<td>LoShare (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HiShare (16)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrow (16)</td>
</tr>
</tbody>
</table>

**Figure 4-3: Factorial design of visual perspective-taking task**

The task consisted of 384 trials in total, presented in a randomised order. For visualisation, the final factor in the task design (direction of avatar eye gaze) is not shown in this table. We expected to see an interaction between condition (congruent or incongruent) and avatar on screen (Lo-Share, Hi-Share, Arrow) on response time and accuracy.

Subjects played the visual perspective-taking task twice, before and after training with the false belief task. These baseline and follow-up measurements could be used to see whether there was any evidence of transfer from the training on the false belief task. The arrow trials provided a
control condition to obtain a measure of any non-specific changes in performance from baseline to transfer.

In order to help participants imagine the avatars’ perspectives in this task, they first completed a simplified version of the visual perspective-taking task, as if they were the avatar in the middle of the room. In this simplified version, subjects never had to take another agent’s perspective. They simply saw a target number and pattern, and then saw one wall with two patterns and had to respond with ‘yes’ or ‘no’. They only saw two patterns because in the full version of the task, the avatar in the room also only sees one wall with two patterns. The behavioural data from this simplified task were not analysed.

4.3.4 A subject’s eye view of the experiment

Deception was never used in this experiment; the set up was exactly as it was described to participants. All participants came to the lab on three consecutive days (Figure 4–4). On day one, subjects completed a computerised intertemporal choice task (described further in Chapter 5). Then subjects completed the simplified version of the visual perspective-taking task. Following this, subjects were introduced to ‘Fred’ and ‘Maria’ as avatars that represented two real previous participants. They played the full version of the visual perspective-taking task (baseline) and then completed some personality questionnaires on paper. Finally, subjects returned to the computer and were introduced to the cover stories used in the probabilistic false belief task. They played four short simplified versions of the task, as the manager, two with each cover story.
Figure 4-4: Three-day training procedure

This figure shows the timeline of behavioural tasks analysed in this chapter. On day one, subjects played the visual perspective-taking task for a baseline measurement. On day two, they completed the false belief tasks with the Hi-Share and Lo-Share agents and then played the visual perspective-taking task again to measure any transfer effect. On day three, they went into the MRI scanner and completed the false belief tasks with Hi-Share and Lo-Share agents but this time, with balanced trial designs. In this example, ‘Maria’ is the Lo-Share agent and ‘Fred’ is the Hi-Share agent, but the avatar-condition mapping was counterbalanced across subjects.

Twenty-four hours later subjects returned for day two of the experiment. They received a full briefing about the social nature of the experiment. They were told that the four short games played on day one would provide choice data for the next two participants. They played the Hi-Share version and the Lo-Share version of the false belief task, without being explicitly instructed about these two conditions. Subjects then completed the remaining personality questionnaires, and finally, played the visual perspective-taking task again (transfer phase). The questionnaires included the BDI, CAPE, IRI, EQ and ICU (as in Chapter 3), the Borderline Scale of the Personality Assessment Inventory (PAI-BOR) (Morey, 1991) and the Borderline Personality Questionnaire (BPQ) (Poreh et al., 2006). These two additional questionnaires are designed to assess for traits of BPD.

Twenty-four hours later subjects returned for day three of the experiment. They went inside the MRI scanner and played the Lo-Share and
Hi-Share games (testing phase) and then underwent structural brain scans (see section 4.3.6) while watching a movie. Following the scans, subjects were debriefed and reimbursed for participation. Final payment was determined by performance on all of the tasks over the three days. Depending on the choices made in the intertemporal choice task, some of the reimbursement may have been delayed by up to twelve weeks.

4.3.5 Model fitting

Models were fit to the probabilistic false belief task data using the same method as described in Chapter 3. All models assumed that subjects used parallel Rescorla-Wagner style belief updating for Self and Other, like the Group B models described in Chapter 3. 72 models were fit in total, covering an exhaustive list of different combinations of learning rate, memory decay, leak and temperature parameters. We also explored the possibility that subjects might use different learning rates on Shared trials, compared to privileged and decoy trials. Therefore, a model could include up to four different learning rates: \( \alpha_{s1} \) for Self on privileged trials, \( \alpha_{s2} \) for Self on shared trials, \( \alpha_{o1} \) for Other on decoy trials and \( \alpha_{o2} \) for Other on shared trials. A summary of all of the models is shown in supplementary Table 4-1.

A drift-diffusion model was fit to the visual perspective-taking task data using the fast-dm toolbox (Voss & Voss, 2007) using MLE. All trials where subjects responded very slowly (> 3000 ms) or quickly (< 500 ms) were excluded from the analysis. In the fast-dm toolbox, the diffusion coefficient (within-trial drift variability) is fixed a priori to a value of 1.

4.3.6 Scanning protocol

Scanning took place in a 3T whole body MRI scanner (Magnetom Prisma system from Siemens Healthcare, Erlangen, Germany) with a body coil for transmission and a 64-channel receive head coil. We first ran a localiser scan. Then we collected the functional data with four EPI scanning sequences (two runs for the Lo-Share task and two runs for the Hi-Share
This was a 2D EPI sequence, previously optimised for regions near the OFC and amygdala, as per the recommendations in (Weiskopf et al., 2006). Each volume comprised 40 slices with a resolution of 3 mm isotropic, with a TR of 2.8 s, TE of 30 ms, slice tilt of –30°, and Z-shim of –0.4. Each scan comprised about 280 volumes, and lasted about 12 minutes, depending on how long the subject took to finish the run. Heart rate was monitored using a Nonin 8600FO pulse-oximeter and respiration rate was monitored using a Siemens breathing belt during scanning.

Following the functional scans, a field mapping sequence was used to measure inhomogeneity of the $B_0$ field. This was a double-echo FLASH sequence with a short TE of 10 ms and a long TE of 12.46 ms.

Lastly an MPM protocol was applied for microstructural imaging. First, calibration data were acquired to map and correct for inhomogeneities in the $B_1$ transmit field using a 3D spin echo/stimulated echo EPI method (Lutti et al., 2012). Then three 3D multi-echo FLASH acquisitions were made, with predominantly $T_1$, PD and MT weighting respectively. The flip angle was 6° for the PD-weighted and MT-weighted images, and 21° for the $T_1$-weighted images. MT-weighting was achieved through the application of a Gaussian RF pulse 2 kHz off-resonance with 4 ms duration and a nominal flip angle of 220°. The data were acquired with whole-brain coverage at an isotropic resolution of 0.8 mm. Gradient echoes were acquired with alternating readout gradient polarity at eight equidistant echo times ranging from 2.3 to 18.4 ms in steps of 2.3 ms. Only six echoes were acquired for the MT-weighted acquisition in order to maintain a TR of 25 ms of all volumes. Each multi-echo FLASH acquisition took approximately 7 minutes.

Prior to each FLASH acquisition, two additional low resolution (8 mm isotropic) volumes were acquired, one with the 64-channel head and neck array coil and the other with the body coil. A single echo, with a TE of 2.2 ms, was acquired in each case using a 6° flip angle and a TR of 6 ms. The acquisition time of each of these calibration volumes was 5.9 s. These ‘sensitivity maps’ can be used to correct the position-specific modulation of receive sensitivity field (Papp et al., 2016, Tabelow et al., 2019).
4.3.7 qMRI data preprocessing

Quantitative MT maps were created and then spatially processed using the hMRI toolbox in SPM12 (Tabelow et al., 2019). Spatial processing involves three steps: segmentation, diffeomorphic deformation and tissue-weighted smoothing. The segmentation step uses a unified segmentation algorithm (Ashburner & Friston, 2005), which takes an MT map for each subject along with a series of tissue probability maps. Each map is converted into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) tissue class images. The diffeomorphic deformation step uses the DARTEL toolbox (Ashburner, 2007) to iteratively align tissue class images from all of the subjects to their own average before normalising the images to MNI space.

Finally, we smoothed the maps with a Gaussian kernel of full width at half maximum (FWHM) 6 mm isotropic. As described in (Draganski et al., 2011), this smoothing step should account for the partial volume contribution of the tissue density in each voxel in native space. The resulting tissue-weighted smoothed map only includes those voxels with an a priori probability of being considered in the relevant tissue class (GM, WM or CSF) above 5% and an original tissue density above 5%.

4.3.8 fMRI data preprocessing

The first six volumes of each functional run were discarded. Slice-timing correction was applied. Motion correction was carried out using the ‘realign and unwarp’ toolbox within SPM12. Images were co-registered to the first volume acquired for each subject. The motion-corrected images were then unwarped using the field map. The functional images were co-registered to the respective subject’s MT map, normalised into MNI space and then smoothed with a Gaussian kernel of FWHM 8 mm isotropic.

Physiological data were converted into 18 nuisance regressors with the PhysIO Toolbox (Kasper et al., 2017). These were entered into first-level GLMs along with 6 motion parameters estimated at the realignment stage.
4.3.9 Mass-univariate analysis

Two separate GLMs were estimated, one for localising PE\textsuperscript{s} and one for localising PE\textsuperscript{o}. The Self GLM modelled the onsets of privileged and shared trials, parametrically modulated by the absolute PE\textsuperscript{s} on those trials. The Other GLM modelled the onsets of shared and decoy trials, parametrically modulated by the absolute PE\textsuperscript{o} on those trials. Temporal and dispersion derivatives were also included. The onsets of probe trials were also included in both GLMs, as were 24 nuisance regressors, describing motion and physiological noise. Both GLMs modelled four runs of functional data, two runs for the Hi-Share task and two runs for the Lo-Share task.

First-level maps were entered into a t-test at the second level. Significantly large clusters ($p_{\text{FWE}} < 0.05$) were identified at the group-level using a cluster-forming threshold of $p < 0.001$, in a whole-brain analysis.

For all mass-univariate and decoding analyses, model-based regressors were generated using the group median parameters from the complex model (M68).

4.3.10 Searchlight analysis

In rapid event-related designs, such as the current design, there is a large overlap in BOLD signal across adjacent trials. In order to obtain trial-specific activation patterns, we used the ‘Least Squares - Separate’ approach described in (Mumford et al., 2012). For every trial, a new GLM was estimated with unsmoothed images as input. One regressor represented the onset of the trial of interest, and one regressor represented the onsets of all other trials. Temporal and dispersion derivatives of these two regressors were also included. 24 nuisance regressors, describing motion and physiological noise, were included. A beta map was produced that represented the contribution of the trial of interest to the whole functional run. Decoding analyses were performed on these beta maps.

We ran two whole-brain searchlight procedures, one for the PE\textsuperscript{s} and one for the PE\textsuperscript{o}, using the ‘Decoding Toolbox’ (Hebart et al., 2014). For each
analysis we trained a LASSO (least absolute shrinkage and selection operator) linear regression model on three runs of functional data, to predict absolute PE values from the BOLD signal across voxels within spherical searchlights of radius four voxels.

The regression models were fit using the ‘Penalized toolbox’ (McIlhagga, 2016). The regression model was then tested on the fourth, held-out run of functional data. Performance was quantified as the Fisher Z-transformed correlation between the model's predicted PE magnitudes and the actual PE magnitudes. The transformed correlation coefficients for each of the four test sets were averaged to produce a mean cross-validated decoding accuracy, which was attributed to the voxel at the centre of the searchlight. This was repeated for each subject with a range of $L_1$ penalty parameters ($10^{-5}$ to $10^{-3}$ in increments of $2.5 \times 10^{-5}$). The whole-brain accuracy maps were then smoothed with a Gaussian kernel of FWHM 8 mm isotropic.

We extracted decoding accuracies from voxels masked by the co-ordinates of significantly large clusters ($p_{FWE} < 0.05$, cluster-level) identified in the two respective mass-univariate analyses. For each subject, an optimal penalty parameter was selected for the PE$^s$ analysis, by identifying which penalty produced the highest median decoding accuracy across masked voxels in the PE$^o$ analysis. Concurrently, optimal penalty parameters were selected for the PE$^o$ analysis by identifying which penalties produced the highest median decoding accuracies across masked voxels in the PE$^s$ analysis. By optimising the hyperparameters for one analysis on a different analysis, we mitigated the risk of overfitting.

Once penalty parameters were selected for each subject, first-level decoding accuracy maps were entered into a t-test at the second level. Significantly large clusters ($p_{FWE} < 0.05$) were identified at the group-level using a cluster-forming threshold of $p < 0.001$, in a whole-brain analysis. These clusters from the PE$^s$ and PE$^o$ searchlight analyses were combined to form a single Self-Other multi-cluster mask. This mask was used to select voxels for the subsequent decoding analyses.
4.3.11 Pseudotrial analysis

We first conducted a pseudotrial analysis, as described in Chapter 3 (Figure 3-2), to measure the distinctiveness of $\text{PE}^2$ representations and $\text{PE}^2_o$ representations. The BOLD signal was extracted from all of the voxels in the Self-Other multi-cluster mask. Self-pseudotrials and Other-pseudotrials were generated by subtracting signal in low PE-magnitude trials from signal in high PE-magnitude trials, categorised with a median split.

We used a two-step decoding approach. The first step was a feature extraction step, using PCA. The second step involved training a LASSO logistic regression model to classify pseudotrials as being Self or Other.

This approach required tuning of two hyperparameters, the $L_1$ penalty and the percentage variance to be explained by the principal components used to train the classifier. We used nested cross-validation to optimise these two hyperparameters. This used a grid-search, sampling over a range of $L_1$ values ($10^{-5}$ to $10^{-3}$ in increments of $2.5 \times 10^{-5}$) and a range of variance-explained percentages (90%, 92.5%, 95%, 97.5%). Two pseudotrials from each class were randomly sampled to constitute a hold-out set. The remainder constituted a training set. For each possible pair of hyperparameter values, 40 inner folds of cross-validation were performed, by randomly sampling two pseudotrials of each class from the training set. For each hyperparameter combination, a performance measure was quantified as the median cross-entropy across folds between predicted class probability and actual class label. Hyperparameters were selected that produced the minimum median cross entropy.

\[
\text{Cross Entropy} = -(y \ln(p) + (1 - y) \ln(1 - p))
\]

Equation 4-3

Here, $y$ denotes the true class label (Self or Other) and $p$ denotes the probability that the classifier assigns to the observed pseudotrial being a Self pseudotrial. Therefore, this performance measure incorporates not just the
accuracy of the predictions of the logistic regression model, but also the confidence of the predictions. Cross-entropy is lowest when confident, correct predictions are made, and highest when confident incorrect predictions are made.

Finally, the classifier, with the optimised hyperparameters, was applied to the hold-out set, and cross-entropy was measured. This whole procedure was repeated for 40 outer folds of cross-validation and performance was quantified as the median cross-entropy across the 40 folds. This analysis was conducted twice, once for the Hi-Share task data and once for the Lo-Share task data.

4.3.12 Cross-decoding analysis

This analysis followed the same pipeline as described for the pseudotrial analysis, but instead of training a logistic regression model, a linear regression model was trained with labels of PE^s magnitudes, and tested to predict PE^o magnitudes, and vice versa. There were also only eight outer folds and one inner-fold of cross-validation, due to the natural split in train and test sets. For tuning hyperparameters, the Fisher Z-transformed correlation between predicted PE magnitudes and true PE labels was maximised. Final performance was quantified as the mean Fisher Z-transformed correlation across the eight outer folds of cross-validation. This analysis was conducted twice, once for the Hi-Share task data and once for the Lo-Share task data.

4.3.13 Analysis of questionnaire data

For each subscale of our questionnaires we set up a regression model with gender and age as predictor variables and the subscale score as a dependent variable. We then took the residuals from these regression models as age- and gender-controlled scores for each subscale. We then z-scored each of these age- and gender-controlled subscales and entered
them into a PCA. We investigated the principal component that explained the most variance in the data.

4.4 Results

4.4.1 Self-Other leak emerges in training and persists

Performance on the probabilistic false belief task was defined as the correlation between subjects’ reports and the generative probabilities through which outcomes were sampled (Figure 4-5). We divided probe trials into Self and Other trials and performed a two-way repeated measures ANOVA with condition (Lo-Share or Hi-Share) and probe trial type (Self or Other) as factors.

In training, there was a main effect of condition on performance, with subjects performing worse with the Hi-Share agent \(F(1, 39) = 12.64, p = 0.001\). The same main effect was seen in the test session \(F(1,39) = 6.76, p = 0.013\). In the test session, there was also an interaction between condition and probe trial type on performance, whereby the impairment in performance with the Hi-Share agent was only seen on Self trials and not on Other trials \(F(1,39) = 5.58, p = 0.023\).

Each subject generated four behavioural datasets: Lo-Share and Hi-Share in training, and Lo-Share and Hi-Share in testing. Behavioural models were fit, per subject, to these four datasets independently. 72 models were fit in total, which tested different numbers of learning rates, temperatures, memory decay parameters, and leak parameters. BIC was summed across subjects to determine a winning model for each of the four datasets (Figure 4-6A).
In training, subjects performed better with the Lo-Share agent than the Hi-Share agent ($p = 0.001$). This was also true in testing ($p = 0.013$). There was also an interaction effect at testing; performance was impaired with the Hi-Share agent, specifically on Self trials ($p = 0.023$). White circles indicate predictions from the winning models for each dataset. Error bars indicate SEM across subjects. The female avatar represents the Lo-Share condition. The male avatar represents the Hi-Share condition. ** denotes $p < 0.01$. * denotes $p < 0.05$.

For the Lo-Share condition, evidence favoured a simple model (M4) in training, which included three parameters: a learning rate, a decision temperature parameter and a memory decay parameter. In testing, a slightly more complex model (M13) was favoured, which differed in that it included two separate decision temperature parameters, one for Self trials and one for Other trials. Neither of these models included a leak parameter.

For the Hi-Share condition, evidence favoured a complex model (M65) in training, which included seven parameters. The model included four learning rates, two for Self updates (one for privileged trials and one for
shared trials) and two for Other updates (one for decoy trials and one for shared trials). In addition to the four learning rates, the model included two temperature parameters and one leak parameter. In testing, a slightly more complex model (M68) was favoured, which differed in that it also included a memory decay parameter. Both of these models included a bidirectional leak parameter, implying that in the Hi-Share condition, subjects were more likely to erroneously update their own belief with the Other’s PE, and vice versa.

By using subjects’ parameter estimates to simulate synthetic data, we could refit the models to the synthetic data and quantify parameter identifiability. The parameters that generated the synthetic data were highly correlated with the recovered parameters, indicating that they were identifiable. The confusion matrix generated by correlating every generative parameter with every recovered parameter for the most complex winning model (M68) is shown in Figure 4-6B.

We hypothesised that the two-way interaction between condition (Hi-Share or Lo-Share) and probe trial type (Self or Other) on performance, shown in Figure 4-5, was due to the leak between belief updates for Self and Other. Despite the fact that the winning model for the Hi-Share agent in testing only contained a single bidirectional leak parameter, it contained multiple learning rates. The learning rate for updating the belief of the Other on decoy trials was significantly higher than the learning rate for updating the belief of the Self on privileged trials \([t(39) = -2.27, p = 0.029]\). For each subject we calculated the proportion of belief updating that was contributed to by the leak parameter, for Self and Other independently. For each subject, this was simply the ratio of the leak parameter to the mean of the learning rates \((\lambda : \alpha)\). We call this term a ‘leak factor’. Indeed, the difference in leak factor between Self and Other, in the testing task with the Hi-Share agent, was strongly negatively correlated with the difference in performance between Self trials and Other trials [Spearman’s rho = –0.45, p = 0.004] (Figure 4-6C). This implies that the contribution of leak parameters to learning is a strong determinant of performance in this task.
Figure 4-6: Model fits and parameter estimates

A) Parameter estimates for the winning model of each of the four datasets are shown. Insets show BIC summed over subjects for each of the four models with the winning model highlighted in yellow. B) Parameter recovery for most complex winning model (Hi-Share on test day). C) The difference in leak factor for Self and Other is correlated with the difference in performance on Self and Other trials in testing with the Hi-Share agent.

To further examine the impact of leak on belief updates, we computed the correlation between trial-by-trial Self-attributed beliefs and Other-attributed beliefs derived from the winning models. We divided each of the four behavioural datasets into eight non-overlapping trial bins of approximately 56 trials each and then derived a Self-Other correlation for
each bin (Figure 4-7). Self-Other correlation was higher with the Hi-Share agent than with the Lo-Share agent in all eight bins, in both training and testing. In a two-way repeated measures ANOVA with bin number (1-8) and condition (Lo-Share or Hi-Share) as factors, there was a main effect of condition on Self-Other correlation, in both training \([F(1,39) = 21.7, p < 0.001]\) and testing \([F(1,39) = 3.8, p < 0.001]\). This result was not sensitive to the number of bins that trials were divided into (Figure 4-12). Crucially for the fMRI analysis, the trial-by-trial PE magnitudes were not correlated between Self and Other, in either the Lo-Share task or the Hi-Share task (supplementary Figure 4-13).

![Graph](image)

**Figure 4-7: Correlations between Self- and Other-attributed beliefs**

As a result of leak parameters in the Hi-Share task, there was a higher correlation between trial-by-trial Self-attributed and Other-attributed beliefs in the Hi-Share task than the Lo-Share task. This was the case in each of eight trial bins in both Training and Testing. The result of this analysis was not sensitive to the choice of the number of bins (see supplementary Figure 4-12). Error bars show SEM.

These behavioural findings show that, in testing, subjects were better at distinguishing between their own beliefs and the beliefs of the Lo-Share agent, than they were at distinguishing between their own beliefs and the beliefs of the Hi-Share agent. Because there was no systematic difference between these tasks in testing, this effect could only be due to learning that
took place in the training phase. This learning appeared to be implicit, as on debriefing none of the subjects reported noticing a difference between the Hi-Share and Lo-Share tasks in training. Our visual perspective-taking task was designed to test how domain-general this learning was.

**4.4.2 Training transfers to a different cognitive domain**

Subjects performed well in the visual perspective-taking task, achieving an average accuracy of 89% at baseline (SD 10.9%) and 91% at transfer phase (SD 9.8%). In order to make use of both RT and accuracy data, we fit a drift-diffusion model (DDM). The DDM assumes that subjects accumulate evidence on each trial by sampling sensory evidence from the visual scene, until reaching one of two possible decision thresholds. The model is tuned by several parameters, including non-decision time, boundary separation, starting point and drift rate.

Responses were faster on correct trials than incorrect trials [baseline: \( t(45) = -10.35, p < 0.001 \), transfer: \( t(45) = -7.93, p < 0.001 \)]. Higher drift rates produce faster and more accurate responses. In order to allow the model to generate different RT distributions for correct and incorrect trials, we allowed for between-trial drift rate variability (Ratcliff & Rouder, 1998, Ratcliff & McKoon, 2008). This was achieved by randomly sampling the drift rate on each trial from one of twelve possible Gaussian distributions. These distributions accounted for different trial types, namely the three avatars (Hi-Share, Lo-Share, arrow), two perspectives (Self, Other) and two conditions (congruent, incongruent), giving twelve distributions in total. These distributions had separate mean parameters, but shared a subject-specific variance parameter.

The model was fit twice to each subject, once for the baseline dataset and once for the transfer dataset. In total, 1.4% of all trials were excluded from the analysis due to responses that were too slow (> 3000 ms) or too fast (< 500 ms).
We predicted an avatar-specific change in the incongruency effect from baseline to transfer. Specifically, we predicted that at transfer phase, there would be a stronger incongruency effect with the Hi-Share agent and a weaker incongruency effect with the Lo-Share agent, reflecting differential Self-Other distinction capacities. We did not have a prediction about Self and Other trials, but in this task, RT was significantly lower on Other trials than on Self trials \[ \text{baseline: } t(45) = 14.24, \ p < 0.001, \text{ transfer: } t(45) = 13.02, \ p < 0.001 \]. This is because on Other trials, subjects only needed to attend to the two patterns that the avatar was facing, whilst on Self trials subjects needed to attend to all four patterns. Different mean drift rates were fit for Self and Other trials to accommodate this bias.

The arrow trials were included to obtain a measure of within-task training effects that were not related to the false belief task. By subtracting any changes in performance on arrow trials from changes in performance on the Hi-Share and Lo-Share trials, we could quantify training effects over and above those induced merely by repeated exposure to the visual perspective-taking task.

We first checked that there was indeed an incongruency effect by comparing the mean drift rates for congruent trials with the mean drift rates for incongruent trials, averaging over the differential drift rates for perspective and avatar (Figure. 4-8A). There was a strong incongruency effect at both baseline and transfer with lower mean drift rates on incongruent trials \[ \text{baseline: } t(45) = 7.6, \ p < 0.001, \text{ transfer: } t(45) = 6.7, \ p < 0.001 \]. The task therefore replicated the main finding from the original dots task of (Samson et al., 2010).

We then explored variability across all twelve mean drift rate parameters. For each subject, for each of the twelve drift rate parameters, we subtracted the estimated parameter at baseline from the estimated parameter at transfer to obtain a change score. All twelve drift rates were higher at transfer than at baseline, reflecting a general improvement in performance. The change scores were then normalised by subtracting from them the change scores for arrow trials. A positive normalised change score
indicated that the relevant drift rate increased more than the increase seen on arrow trials. A negative normalised change score indicated that the relevant drift rate increased less than the increase seen on arrow trials.

We performed a three-way repeated measures ANOVA on the normalised change scores with avatar (Hi-Share or Lo-Share), perspective (Self or Other) and condition (congruent and incongruent) as factors. We expected to see a two-way interaction between condition and avatar but this was not significant \( F(1, 45) = 0.28, p = 0.6 \). Instead, we found a significant main effect of avatar, with higher normalised change scores for Lo-Share trials than Hi-Share trials \( F(1, 45) = 7.4, p = 0.009 \). This main effect is visualised in Figure 4-8B. Higher drift rates indicate both faster \textit{and} more accurate responses. Therefore, this finding means that false belief task training improved subjects’ performance on the perspective-taking task on Lo-Share trials more than on Hi-Share trials.

**Figure 4-8: Analysis of drift rates in visual perspective-taking task**

A) Drift rate (summed over perspective and avatar parameters) was lower on incongruent trials compared to congruent trials. This effect was found at both baseline and transfer. Scatter points represent individual subjects B) Normalised change scores for Lo-Share and Hi-Share trials, averaged over perspective. A score of zero indicates that the change in drift rate from baseline to transfer is the same as the change seen on arrow trials. There is a significant main effect of avatar, with reduced change scores for the Hi-Share agent. This is seen across both congruent and incongruent trials, yielding no significant interaction between congruency and avatar on drift rate change. Error bars denote SEM. *** denotes \( p < 0.001 \). ** denotes \( p < 0.01 \).
Whilst we expected to see a transfer effect only on incongruent trials, when Self-Other distinction is required, the effect was seen on both congruent and incongruent trials. Nevertheless, this effect demonstrates that the Hi- versus Lo-share manipulation in the false belief task caused lasting plasticity which transferred to the visual perspective-taking task. This suggests that, in training, subjects may have learned something about the relations between Self-attributed and Other-attributed mental states that is general with respect to the type of mental state in question. These mental states were learning signals in the training task and visual perspectives in the transfer task.

4.4.3 Functional localisation of prediction errors

In order to assess whether the training had induced plasticity in the representations of Self- and Other-attributed PEs, we first had to localise these signals in the brain. $|PE^s|$ and $|PE^o|$ were not correlated with each other (Figure 4-13). The following functional localisation analyses were conducted across the Hi-Share and Lo-Share tasks, and did not model them as separate conditions.

We conducted a whole-brain mass-univariate analysis, regressing unsigned $PE^s$ and $PE^o$ against BOLD signal. Significant clusters for $PE^s$ were found in the bilateral extrastriate visual cortex as well as the right intraparietal sulcus extending to the right supramarginal gyrus (rSMG) and precentral gyrus. Significant clusters for $PE^o$ were found in bilateral primary and extrastriate visual cortex, as well as the rSMG and precentral gyrus. Clusters are visualised in Figure 4-9A and peak co-ordinates are reported in supplementary Table 4-2.

We then conducted a whole-brain searchlight multi-voxel pattern analysis (MVPA), optimising model parameters with masks of clusters discovered in the preceding mass-univariate analysis (see section 4.3.10 for details). Significant clusters for $PE^s$ were found in right extrastriate visual cortex extending to the right posterior cingulate, right intraparietal sulcus and
right supplementary motor cortex. Significant clusters for PE\(^o\) were found in left extrastriate visual cortex extending to the cerebellum. Clusters are visualised in Figure 4-9B and peak co-ordinates are reported in supplementary Table 4-2.

**Figure 4-9: Group-level clusters of PE-related activity**

A) Results from a mass-univariate regression analysis. B) Results from a multivoxel searchlight pattern analysis. In both A) and B), the overlaid maps show significantly large clusters (\(p_{\text{FWE}} < 0.05\)) of voxels at second level, with BOLD signal co-varying with either unsigned PE\(^s\) (blue) or unsigned PE\(^o\) (yellow). Cluster-forming threshold \(p < 0.001\).
4.4.4 Plasticity of neural Self-Other distinction

The clusters derived from the searchlight analysis were used to select voxels for the subsequent analyses on experience-dependent plasticity. Here we explored whether the similarity between PE$_s$ and PE$_o$ representations was different between the Hi-Share and Lo-Share conditions. These subsequent analyses are not biased by the fact that they are limited to voxels that were found to be significant in the previous analyses. The previous analyses searched for voxels where fluctuations in BOLD signal correlated with PE$_s$ or PE$_o$, in two independent analyses. In the previous analyses, there were no statistical tests of the relationships between these two signals. However, the subsequent analyses explore whether, across all these voxels where PE$_s$ or PE$_o$ are encoded, patterns of BOLD variation can predict whether the subject is experiencing a PE$_s$ or a PE$_o$. Additionally, whilst these subsequent analyses are performed for Hi-Share and Lo-Share independently, the previous analysis averaged over both of these tasks. As the subsequent hypotheses are specifically concerned with a difference between these two tasks, the subsequent analyses cannot be biased by the selection of voxels from the previous analyses.

We first conducted a ‘pseudotrial’ analysis (see Chapter 3 and section 4.3.11) to obtain a measure of dissimilarity of Self- and Other-attributed signals. A higher decoding accuracy for classifying PE$_s$ and PE$_o$ signals would indicate more distinct Self-Other representations. We ran this analysis twice, once for the Hi-Share condition and once for the Lo-Share condition. In both cases we trained classifiers on the same set of features, derived from the searchlight analysis.

The training phase was designed to strengthen associations between PE$_s$ and PE$_o$ in the Hi-Share condition and weaken them in the Lo-Share condition. Accordingly, we expected to see a higher classification accuracy in the Lo-Share condition than in the Hi-Share condition. Indeed, we found a significant difference in classification accuracy in the predicted direction, with better Self-Other classification for the Lo-Share condition than the Hi-Share condition [$t(39) = 2.1$, $p = 0.042$, two-tailed] (Figure 4-10A).
This analysis provides a statistical test for distinctiveness of representations of PE\textsuperscript{s} and PE\textsuperscript{o}. However, the null classification result for the Hi-Share condition, could simply mean that the data for this condition were noisier. For instance, subjects may simply have paid less attention in this task, which would explain the impairment in performance and also the reduced decoding accuracy from the BOLD signal. To determine whether this was the case, we conducted a ‘cross-decoding’ analysis which tested for the logical inverse of the pseudotrial analysis.

The cross-decoding analysis provided a statistical test for the similarity of representations of PE\textsuperscript{s} and PE\textsuperscript{o}, by training a decoder on PE\textsuperscript{s} magnitudes and testing its ability to make predictions about PE\textsuperscript{o} magnitudes from BOLD signal, and vice versa. The cross-validation scheme that we used is shown in detail in supplementary Figure 4-14. If the regression model can generalise well between the two types of signals, it indicates that the representations are similar. We expected the cross-decoding performance to be better with the Hi-Share agent than with the Lo-Share agent. Indeed, we found a significant difference in cross-decoding accuracy in the predicted direction, with better decoding performance for the Hi-Share condition than the Lo-Share condition \([t(39) = -2.8, p = 0.009, \text{two-tailed}]\) (Figure 4-10B).

This result provides evidence against the hypothesis that the data in the Hi-Share task were simply noisier. It also shows that the representations of PE\textsuperscript{s} and PE\textsuperscript{o} were significantly more similar with the Hi-Share agent than with the Lo-Share agent.

As we found a stronger between-condition effect with the cross-decoding measure, we treated this as our primary fMRI measure of neural Self-Other distinction. To further validate it, we tested whether the measure was associated with our behavioural measure of Self-Other distinction. Indeed, the difference in cross-decodability between the Hi-Share and Lo-Share tasks was positively associated the \(\lambda:\alpha\) leak factor \([r = 0.44, p = 0.004]\) (Figure 4-10C). This is a behavioural measure, which captures the extent to which misattributed PE signals contribute to belief updates (see section 4.4.1). We also found that the extent to which the more complex
model (M68) better fit the Hi-Share condition than the Lo-Share condition, in training and testing, correlated with the difference in cross-decodability between conditions \([r = 0.34, p = 0.03]\).

In summary, these results show that the learning that took place in the training phase changed the way Self- and Other-attributed learning signals were represented with respect to each other. The results are consistent with the hypothesis that exposure to a strong contingency between two types of PE signals, as in the Hi-Share training, can strengthen the association between these signals, whilst exposure to a weak contingency can weaken the association. The relationship between the strength of this association and a behavioural index of Self-Other distinction is consistent with the hypothesis that this kind of associative learning could be sufficient for learning about relationships between the mental states of different agents.

Figure 4-10: Evidence of plasticity of Self- and Other-attributed signals

A) Classifiers were more accurate at identifying whether a PE signal was attributed to Self or Other with the Lo-Share agent than with the Hi-Share agent, indicating that the representations were more distinct in the Lo-Share condition than in the Hi-Share condition. The y-axis shows transformed cross-entropy between predicted and true class labels, such that zero indicates chance decoding. B) Decoders were better at generalising between Self and Other with the Hi-Share agent than with the Lo-Share agent, indicating that the representations were more similar in the Hi-Share condition than in the Lo-Share condition. The y-axis shows Fisher-transformed correlation between predicted and true class labels. C) Each point is a different subject. The difference in cross-decodability between the Lo-Share and Hi-Share conditions is positively correlated with the \(\lambda : \alpha\) leak factor derived from the
behavioural data, indicating that the cross-decodability measure is an index of Self-Other distinction ability. Error bars show SEM. ** denotes \( p < 0.01 \). * denotes \( p < 0.05 \).

### 4.4.5 Prefrontal microstructure and meta-representational knowledge

The previous results showed that our training procedure induced plasticity in Self- and Other-attributed PE representations in sensory processing regions (e.g. extrastriate visual cortex). However, we expected that other brain regions might be involved in representing the different contexts of the Hi-Share and Lo-Share tasks, shaping the extent to which the evoked learning signals were distinct or overlapping.

To this end, we exploited the between-subject variability in Self-Other cross-decodability. Using an MPM scanning protocol, we derived subject-specific magnetisation transfer (MT) maps. MT can be used as a biophysical marker of myelin density. By regressing subjects’ cross-decoding effect from the BOLD signal against white matter MT, we could look for regions where differences in white matter microstructure were associated with our measure of functional plasticity. We conducted a whole-brain, mass-univariate regression at the second level. The cross-decoding effect (Hi-Share – Lo-Share) was the independent variable, and white matter MT was the dependent variable within each voxel. Age, gender and intra-cranial volume were included as covariates. Significantly large clusters (\( p_{FWE} < 0.05 \)) were identified using a cluster-forming threshold of \( p < 0.001 \).

When testing for a negative contrast, there were no significant voxels. When testing for a positive contrast, we found one significant cluster of white matter adjacent to the right vmPFC [844 voxels, \( p_{FWE} < 0.001 \), peak co-ordinates: \( x = 12.8, y = 59.2, z = -18.4 \)] (Figure 4-11A). Subjects with greater MT in this cluster showed a more profound difference in cross-decodability between the two conditions. This suggests that subjects with a higher myelin density in fibres connecting to the vmPFC had neural Self-Other distinctions that were more sensitive to the social context. MT in this cluster was also
positively associated with the leak factor from the Hi-Share context, but this association was not significant \( r = 0.28, p = 0.075 \).

We predicted that inter-subject variability in microstructure of this region might also be important for socio-cognitive function outside the laboratory. We ran a PCA on all of the subscales of a set of personality questionnaires, chosen to assess for traits of subclinical socio-cognitive dysfunction (Figure 4-11B). We tested whether MT in the right vmPFC cluster was associated with subjects’ scores in the first principal component. This component explained 32% of the variance in the questionnaire data and loaded positively with all subscales of two BPD trait questionnaires (PAI-BOR and BPQ) apart from one subscale measuring self-harm, all subscales of a schizotypy questionnaire (CAPE), depression symptoms (BDI), and a subscale of the IRI that measures personal distress in social situations. The component did not load strongly or positively with empathy (EQ or other subscales of IRI) or antisocial behaviour (ICU).
Figure 4-11: vmPFC microstructure is associated with plasticity and subclinical traits

A) In a whole-brain second-level analysis, one significantly large cluster of white matter MT was found to correlate with the difference in cross-decodability between Hi-Share and Lo-Share ($p_{FWE} < 0.001$). This cluster was located adjacent to the right vmPFC. The inset scatter plot visualises the relationship between the functional effect (x-axis) and the measure of right vmPFC microstructure (y-axis) averaged across the voxels within the cluster. The units on the y-axis are the residuals after regressing out age, gender and intra-cranial volume from MT.

B) Component coefficients (loadings) for questionnaire data in the first principal component. This component loaded positively on psychopathological traits, including all subscales of the CAPE questionnaire (schizotypy), the BDI (depression), the personal distress subscale of the IRI and all subscales of the PAI-BOR and BPQ (BPD) except for the self-harm subscale of the BPQ. Subscales are labelled as follows: Identity (ID), Affective Instability (AI), Negative Relationships (NR), Self Harm (SH), Positive (P), Negative (N), Depressive (D), Impulsivity (IM), Abandonment (Ab), Relationships (R), Self Image (SI), Emptiness (E), Anger (An), Psychotic (P), Fantasising (FS), Perspective Taking (PT), Empathic Concern (EC), Personal Distress (PD).

C) The average MT within the cluster shown in part (A) is negatively correlated with subjects’ scores in this principal component, suggesting that subjects with a low myelin density in this region have more of these subclinical psychopathological traits. The units on the y-axis are the residuals after from regressing out age, gender and intra-cranial volume from MT.
Indeed, we found that MT, adjusted for covariates, in this right vmPFC cluster was significantly negatively correlated with subjects’ scores in this principal component [Spearman’s rho = –0.46, p = 0.003] (Figure 4-11C). This result suggests that subjects with a lower myelin density in white matter adjacent to the right vmPFC had more subclinical traits of these conditions. Thus, the same cluster in which microstructure is associated with the impact of social context on Self-Other distinction in our false belief task is also related to a more ecological measure of general social functioning. Subjects’ principal component scores were also negatively correlated with the cross-decoding training effect but this was not a significant correlation [r = –0.14, p = 0.4].

4.5 Discussion

We show that Self-Other distinction is susceptible to experience-dependent plasticity in line with meta-representational associative learning. When tracking the mental states of an agent for whom there had been strong contingency between Self- and Other-attributed learning signals, representations of prediction errors become similar. When tracking the mental states of an agent for whom there had been a weak contingency between Self- and Other-attributed learning signals, representations of prediction errors become distinct.

This kind of learning may drive the formation of priors that look like biases in various social tasks. For instance, in Chapter 3 (Ereira et al., 2018), it was shown that a subtle manipulation of social context was sufficient to induce a change in measured Self-Other distinction. Furthermore, neural representations underpinning reflection on Self and similar Others are more similar than representations underpinning reflection on Self and dissimilar Others (Mitchell et al., 2006, Jenkins et al., 2008). It has recently been shown that a neural representation of another person is simply a representation of the mental states that the Other habitually experiences (Thornton et al.,
Thus, these findings are consistent with associated mental states for Self and socially proximal Others.

This model predicts that the degree of similarity in prediction error signals should be higher between Self and friends than Self and strangers, because Self- and Other-attributed meta-representations will have been associated for the case of the friend, but not for the stranger. Indeed, close relationships have been described as an incorporation of an Other into the Self (Aron et al., 1991, Aron et al., 2004, Smith & Mackie, 2016). An interesting future direction will be to explore whether this kind of training can manipulate the perceived familiarity of other agents. We would predict high-contingency training with outgroup members to increase perceived familiarity and increase the vividness with which mentalisation can be performed.

The higher-order learning induced by our training was general enough to affect behaviour in a different cognitive task. Both tasks involved the same agents, but whilst the training task invoked representations of numerical probabilities, the transfer task invoked representations of visual perspectives. Consistent with meta-representational learning, the ability to represent the mental states of Self and Other was changed through training, in a manner that was independent of the form of representational content. Notably the effect of transfer was not what we expected. The visual perspective-taking task is designed such that Self-Other distinction is required on incongruent trials but not required on congruent trials. Thus, we only expected to see an agent-specific effect on incongruent trials. Instead, performance was impaired on all Hi-Share trials and improved on all Lo-Share trials.

On congruent trials in the visual perspective-taking task, using a shared model for both Self and Other should not impair performance. However, sensory sampling on behalf of Self and Other simultaneously, could carry a cognitive load that slows down evidence accumulation, regardless of whether the samples for Self and Other are the same or different. This may be akin to the kinds of performance deficits seen under dual-task demands (Otto et al., 2013, Economides et al., 2015). This
explanation might predict that the agent-dependent impairment would become specific to incongruent trials if the time pressure were removed.

Consistent with previous findings (Santiesteban et al., 2014, Catmur et al., 2016, Santiesteban et al., 2017), we observed an incongruency effect in the visual perspective-taking task when an arrow was on the screen, rather than an actual avatar. Thus, the incongruency is likely to be due, at least in part, to mere spatial biases induced by the stimuli. However, the fact that we observed an agent-specific transfer effect, provides evidence in favour of the argument that subjects represent other agents’ mental states in perspective-taking tasks (Hamilton et al., 2009, Samson et al., 2010).

In localising the PE signals, we found both distinct and overlapping brain regions for Self and Other, consistent with previous experiments on simulated learning (Suzuki et al., 2012, Lockwood et al., 2016). The regions identified were consistent with previous work examining unsigned prediction error signals. Extrastriate visual cortex has previously been shown to encode visual sensory PEs, such as when stimulus colour unexpectedly changes (Egner et al., 2010, de Gardelle et al., 2013, Jiang et al., 2013, 2016, Stefanics et al., 2019). The intraparietal sulcus has been associated with the encoding of state prediction errors during navigation of a probabilistic environment (Glascher et al., 2010). Of note, the signals were also found in the rSMG, which has previously been described as having a role in Self-Other distinction in the affective domain (Silani et al., 2013, Steinbeis, 2016).

In line with our prediction, we found that myelin-related MT, in white matter adjacent to vmPFC, was associated with the degree functional plasticity induced by training. Specifically, subjects with higher MT in this region showed a larger difference in Self-Other representational similarity between the two conditions. This suggests that the vmPFC plays a role in either acquiring, or in deploying knowledge about social context and Self-Other relations. This is consistent with previous research showing that mPFC microstructure in rodents (Liu et al., 2012, Makinodan et al., 2012, Liu et al., 2016, Makinodan et al., 2017), and macrostructure in humans (Grosse Wiesmann et al., 2017), is related to social behaviour and social cognition.
Chapter 4: Meta-representational associative learning

Myelination is a developmental process, extending into adulthood (Ziegler et al., 2019), which increases the speed of axonal electrical conduction, and thus information transmission between different neuronal populations (Bunge, 1968, McDougall et al., 2018). Therefore, myelin density is a determinant of functional integration between different brain regions. From this perspective, our findings may index the fidelity with which the vmPFC influences sensory processing. In light of the known connectivity profile of the vmPFC, this would likely be through multi-synaptic projections, for instance via the mediodorsal thalamus (Ongur & Price, 2000, Mitchell, 2015).

However, myeloarchitecture is not just a determinant of neural function, but also a consequence. Neuronal activity can promote myelination by stimulating the proliferation of oligodendrocytes (Gibson et al., 2014), providing a non-synaptic form of functional plasticity (Fields, 2015). Motor learning, for instance, is associated with rapid oligodendrogenesis and myelination (McKenzie et al., 2014, Xiao et al., 2016). Therefore, our findings may index the extent to which prior experiences have shaped the development of the vmPFC. While it is unclear whether vmPFC myelination is more of a cause or consequence of sensory plasticity, our data support this region’s role in meta-representational associative learning.

The human vmPFC is involved in self-referential processing (D'Argembeau, 2013). It is more active during self-reflection (Johnson et al., 2002, Mitchell et al., 2006) and it has recently been shown to track the strength with objects are associated with Self (Lockwood et al., 2018). The broader mPFC is associated with representing mental states of both Self and Other (Amodio & Frith, 2006). However, Other-attributed mental states can indeed be represented in the vmPFC, depending on the precise framing of the cognitive task, and the extent of Self-Other distinction (Mitchell et al., 2006, Nicolle et al., 2012, Garvert et al., 2015).

We note that in our experiment, meta-representations were not encoded in the vmPFC. This is not surprising because, whereas previous studies investigating the simulation of Others’ mental states have
investigated value-based decision variables (Nicolle et al., 2012, Suzuki et al., 2012, Garvert et al., 2015), the mental states being represented in our task were sensory surprise signals. The vmPFC is known to represent state and action values, (Knutson et al., 2005, Plassmann et al., 2007). Thus, in studies where other agents’ preferences are simulated, one should expect the contents of the mental states to be encoded in the vmPFC. In the present study however, the contents of the mental states, and the meta-representational structure, were encoded in visual sensory processing regions. The vmPFC in this case, may map these learnt structures onto contextual cues, allowing them to be flexibly deployed when the relevant cue, such as a face, is re-encountered. This is consistent with prior accounts that describe the mPFC as modulating behaviour to suit the social context (Wang & Hamilton, 2012, 2015).

From this perspective, the structure of vmPFC may reflect the bandwidth of meta-representational learning capacity. A more richly innervated vmPFC may facilitate the association of meta-representational structures with different social contextual cues. A failure to use contextual cues to deploy such structured knowledge might manifest itself in socio-cognitive dysfunctions, such as those indexed with our personality questionnaires. Indeed, individuals, with lower vmPFC MT exhibited more severe subclinical traits of BPD, ASD and schizotypy, conditions characterised by impaired everyday social functioning. This symptom measure did not correlate with our functional measure of meta-representational learning, so our data does not support the hypothesis that this form of learning mediates the relationship between vmPFC microstructure and psychopathology. This null result may reflect our small sample size, and that white matter MT is a less noisy measure than our functional measure, which accumulates noise in behaviour, model fitting and BOLD signal.

Meta-representational associative learning may be useful for acquiring efficient representations of the environment, beyond the social domain. Humans and primates can infer and learn about many statistical properties of
their environments, such as covariance structure (Acuna & Schrater, 2010, Wunderlich et al., 2011), volatility (Behrens et al., 2007, Iglesias et al., 2013, Diaconescu et al., 2014, Lawson et al., 2017, Diaconescu et al., 2017), variance (Tobler et al., 2009) and skewness (Symmonds et al., 2010). This kind of rich structural knowledge may help animals to infer latent features of their environment (Kemp et al., 2010, Gershman, 2017, Tomov et al., 2018) and learn relations between different environments.

Relational learning can support efficient representations of the world (Solway et al., 2014, Radulescu et al., 2019). By representing environments in terms of abstract ‘concepts’ (Goodman et al., 2008), ‘task sets’ (Collins & Frank, 2013) or ‘cognitive maps’ (Tolman, 1948, Stachenfeld et al., 2017, Behrens et al., 2018), animals can rapidly generalise the structure learned in one environment to a new environment (Harlow, 1949, Wang et al., 2018, Baram et al., 2019, Liu et al., 2019). If meta-representations, like prediction errors, can become associated, then learning about one model or environment would enable an animal to learn about a different, but related model or environment. The vmPFC has also been associated with mapping latent, contextual states of the environment, in non-social situations (Wilson et al., 2014, Constantinescu et al., 2016, Schuck et al., 2016). Thus, agent identity may be just one example of a latent environmental state, that shapes learning and behaviour to suit the current context.

In summary, our results support a computational mechanism that enables updates to one model to influence change in another model. This may facilitate generalisation of knowledge structure to new situations, and in the special case of social cognition, the generalisation of one agent’s mental states to make inferences on another. We show that contingent updating of two models strengthens this association in a task-independent manner, even in a situation where generalisation hinders task performance. The ability to acquire and deploy this structural knowledge may be important for representing multiple agents’ mental states in an efficient representational format.
### 4.6 Supplementary material

<table>
<thead>
<tr>
<th>ID</th>
<th>$\alpha$</th>
<th>$\tau$</th>
<th>$\delta$</th>
<th>$\lambda$</th>
<th>Total</th>
<th>BIC LO1</th>
<th>BIC HI1</th>
<th>BIC LO2</th>
<th>BIC HI2</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>-2956</td>
<td>-2871</td>
<td>-2706</td>
<td>-2912</td>
</tr>
<tr>
<td>M2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>-2948</td>
<td>-2819</td>
<td>-2649</td>
<td>-2877</td>
</tr>
<tr>
<td>M3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>-2908</td>
<td>-2757</td>
<td>-2609</td>
<td>-2845</td>
</tr>
<tr>
<td>M4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td><strong>-3044</strong></td>
<td><strong>-2778</strong></td>
<td><strong>-2727</strong></td>
<td><strong>-2928</strong></td>
</tr>
<tr>
<td>M5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>-2983</td>
<td>-2710</td>
<td>-2655</td>
<td>-2876</td>
</tr>
<tr>
<td>M6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>-2901</td>
<td>-2610</td>
<td>-2613</td>
<td>-2853</td>
</tr>
<tr>
<td>M7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>-2940</td>
<td>-2695</td>
<td>-2632</td>
<td>-2860</td>
</tr>
<tr>
<td>M8</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>-2900</td>
<td>-2678</td>
<td>-2588</td>
<td>-2813</td>
</tr>
<tr>
<td>M9</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>-2802</td>
<td>-2571</td>
<td>-2530</td>
<td>-2731</td>
</tr>
<tr>
<td>M10</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>-2950</td>
<td>-3048</td>
<td>-2869</td>
<td>-2980</td>
</tr>
<tr>
<td>M11</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>-2915</td>
<td>-3001</td>
<td>-2809</td>
<td>-2958</td>
</tr>
<tr>
<td>M12</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>-2910</td>
<td>-2951</td>
<td>-2771</td>
<td>-2906</td>
</tr>
<tr>
<td>M13</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>-3036</td>
<td>-2956</td>
<td><strong>-2879</strong></td>
<td>-3004</td>
</tr>
<tr>
<td>M14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>-2960</td>
<td>-2857</td>
<td>-2815</td>
<td>-2971</td>
</tr>
<tr>
<td>M15</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>-2956</td>
<td>-2803</td>
<td>-2736</td>
<td>-2900</td>
</tr>
<tr>
<td>M16</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>-2930</td>
<td>-2850</td>
<td>-2759</td>
<td>-2916</td>
</tr>
<tr>
<td>M17</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>-2888</td>
<td>-2791</td>
<td>-2702</td>
<td>-2898</td>
</tr>
<tr>
<td>M18</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>-2785</td>
<td>-2716</td>
<td>-2636</td>
<td>-2807</td>
</tr>
<tr>
<td>M19</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>-2923</td>
<td>-2898</td>
<td>-2688</td>
<td>-2873</td>
</tr>
<tr>
<td>M20</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>-2925</td>
<td>-2871</td>
<td>-2621</td>
<td>-2856</td>
</tr>
<tr>
<td>M21</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>-2905</td>
<td>-2817</td>
<td>-2590</td>
<td>-2795</td>
</tr>
<tr>
<td>M22</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>-2983</td>
<td>-2788</td>
<td>-2626</td>
<td>-2860</td>
</tr>
<tr>
<td>M23</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>-2931</td>
<td>-2734</td>
<td>-2557</td>
<td>-2833</td>
</tr>
<tr>
<td>M24</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>-2876</td>
<td>-2662</td>
<td>-2560</td>
<td>-2779</td>
</tr>
<tr>
<td>M25</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>-2854</td>
<td>-2674</td>
<td>-2519</td>
<td>-2779</td>
</tr>
<tr>
<td>M26</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>-2852</td>
<td>-2671</td>
<td>-2533</td>
<td>-2758</td>
</tr>
<tr>
<td>M27</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>-2762</td>
<td>-2566</td>
<td>-2432</td>
<td>-2679</td>
</tr>
<tr>
<td>M28</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>-2903</td>
<td>-3049</td>
<td>-2810</td>
<td>-2932</td>
</tr>
<tr>
<td>M29</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>-2984</td>
<td>-2997</td>
<td>-2742</td>
<td>-2886</td>
</tr>
<tr>
<td>M30</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>-2926</td>
<td>-2935</td>
<td>-2680</td>
<td>-2831</td>
</tr>
</tbody>
</table>
### Chapter 4: Meta-representational associative learning

<table>
<thead>
<tr>
<th>M31</th>
<th>2</th>
<th>2</th>
<th>1</th>
<th>0</th>
<th>5</th>
<th>-2975</th>
<th>-2940</th>
<th>-2798</th>
<th>-2955</th>
</tr>
</thead>
<tbody>
<tr>
<td>M32</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>-3008</td>
<td>-2866</td>
<td>-2721</td>
<td>-2910</td>
</tr>
<tr>
<td>M33</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>-2861</td>
<td>-2812</td>
<td>-2646</td>
<td>-2844</td>
</tr>
<tr>
<td>M34</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>-2847</td>
<td>-2828</td>
<td>-2688</td>
<td>-2843</td>
</tr>
<tr>
<td>M35</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>-2829</td>
<td>-2770</td>
<td>-2634</td>
<td>-2821</td>
</tr>
<tr>
<td>M36</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>-2752</td>
<td>-2693</td>
<td>-2564</td>
<td>-2747</td>
</tr>
<tr>
<td>M37</td>
<td>2*</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>-2793</td>
<td>-2817</td>
<td>-2558</td>
<td>-2712</td>
</tr>
<tr>
<td>M38</td>
<td>2*</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>-2842</td>
<td>-2880</td>
<td>-2561</td>
<td>-2726</td>
</tr>
<tr>
<td>M39</td>
<td>2*</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>-2788</td>
<td>-2791</td>
<td>-2571</td>
<td>-2758</td>
</tr>
<tr>
<td>M40</td>
<td>2*</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>-2869</td>
<td>-2732</td>
<td>-2557</td>
<td>-2757</td>
</tr>
<tr>
<td>M41</td>
<td>2*</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>-2875</td>
<td>-2773</td>
<td>-2554</td>
<td>-2781</td>
</tr>
<tr>
<td>M42</td>
<td>2*</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>-2812</td>
<td>-2672</td>
<td>-2559</td>
<td>-2810</td>
</tr>
<tr>
<td>M43</td>
<td>2*</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>-2793</td>
<td>-2685</td>
<td>-2521</td>
<td>-2735</td>
</tr>
<tr>
<td>M44</td>
<td>2*</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>-2817</td>
<td>-2729</td>
<td>-2555</td>
<td>-2771</td>
</tr>
<tr>
<td>M45</td>
<td>2*</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>-2749</td>
<td>-2633</td>
<td>-2583</td>
<td>-2777</td>
</tr>
<tr>
<td>M46</td>
<td>2*</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>-2787</td>
<td>-2923</td>
<td>-2728</td>
<td>-2800</td>
</tr>
<tr>
<td>M47</td>
<td>2*</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>-2837</td>
<td>-3020</td>
<td>-2717</td>
<td>-2777</td>
</tr>
<tr>
<td>M48</td>
<td>2*</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>-2809</td>
<td>-2924</td>
<td>-2698</td>
<td>-2815</td>
</tr>
<tr>
<td>M49</td>
<td>2*</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>-2860</td>
<td>-2846</td>
<td>-2724</td>
<td>-2857</td>
</tr>
<tr>
<td>M50</td>
<td>2*</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>-2835</td>
<td>-2922</td>
<td>-2734</td>
<td>-2859</td>
</tr>
<tr>
<td>M51</td>
<td>2*</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>-2849</td>
<td>-2824</td>
<td>-2696</td>
<td>-2868</td>
</tr>
<tr>
<td>M52</td>
<td>2*</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>-2769</td>
<td>-2734</td>
<td>-2642</td>
<td>-2795</td>
</tr>
<tr>
<td>M53</td>
<td>2*</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>-2843</td>
<td>-2866</td>
<td>-2659</td>
<td>-2828</td>
</tr>
<tr>
<td>M54</td>
<td>2*</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>-2790</td>
<td>-2762</td>
<td>-2623</td>
<td>-2834</td>
</tr>
<tr>
<td>M55</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>-2875</td>
<td>-2961</td>
<td>-2706</td>
<td>-2812</td>
</tr>
<tr>
<td>M56</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>-2921</td>
<td>-2995</td>
<td>-2728</td>
<td>-2868</td>
</tr>
<tr>
<td>M57</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>-2880</td>
<td>-2927</td>
<td>-2731</td>
<td>-2840</td>
</tr>
<tr>
<td>M58</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>-2916</td>
<td>-2864</td>
<td>-2665</td>
<td>-2827</td>
</tr>
<tr>
<td>M59</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>-2917</td>
<td>-2893</td>
<td>-2647</td>
<td>-2926</td>
</tr>
<tr>
<td>M60</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>-2891</td>
<td>-2807</td>
<td>-2654</td>
<td>-2900</td>
</tr>
<tr>
<td>M61</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>-2786</td>
<td>-2760</td>
<td>-2589</td>
<td>-2769</td>
</tr>
<tr>
<td>M62</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>-2862</td>
<td>-2783</td>
<td>-2640</td>
<td>-2846</td>
</tr>
<tr>
<td>M63</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>-2789</td>
<td>-2678</td>
<td>-2581</td>
<td>-2819</td>
</tr>
<tr>
<td>M64</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>-2832</td>
<td>-3058</td>
<td>-2821</td>
<td>-2870</td>
</tr>
<tr>
<td>M65</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>-2952</td>
<td>-3078</td>
<td>-2842</td>
<td>-2945</td>
</tr>
<tr>
<td>M66</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>-2897</td>
<td>-3034</td>
<td>-2793</td>
<td>-2871</td>
</tr>
<tr>
<td>M67</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>-2878</td>
<td>-2955</td>
<td>-2794</td>
<td>-2922</td>
</tr>
<tr>
<td>M68</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>-2981</td>
<td>-2999</td>
<td>-2802</td>
<td>-3018</td>
</tr>
</tbody>
</table>
Chapter 4: Meta-representational associative learning

|  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| M69 | 4 | 2 | 1 | 2 | 9 | -2932 | -2914 | -2755 | -2975 |
| M70 | 4 | 2 | 2 | 0 | 8 | -2767 | -2850 | -2702 | -2824 |
| M71 | 4 | 2 | 2 | 1 | 9 | -2913 | -2929 | -2757 | -2950 |
| M72 | 4 | 2 | 2 | 2 | 10 | -2830 | -2817 | -2684 | -2906 |

Table 4-1: List of models fit to false belief task

72 models were fit to each of the four behavioural datasets. The number of parameters in each model is shown. In the learning rate column, a ‘2’ indicates that there was a Self learning rate and an Other learning rate. A ‘2’ indicates that there was a learning rate for shared trials and a learning rate for privileged/decoy trials. A ‘4’ indicates that there was a learning rate for Self on privileged trials, Other on decoy trials, Self on shared trials and Other on shared trials. For all other parameters, a ‘1’ indicates a parameter that was shared for Self and Other, and a ‘2’ indicates separate parameters for Self and Other. The BIC columns show the summed BIC across all subjects for each of the four datasets. The lowest BIC for each dataset is highlighted in yellow.

Figure 4-12: Main effect on Self-Other correlation with different bin sizes

The analysis shown in Figure 4-7 was repeated with different numbers of trial bins. The p value for the main effect of condition (Lo-Share or Hi-Share) on Self-Other correlation in testing is reported at each number of bins used in the analysis. The correlations are significantly higher with the Hi-Share agent than the Lo-Share agent for all bin sizes, showing evidence of correlations at both short and long timescales with respect to the duration of the experiment. Where only one bin was used (i.e. correlation across all trials in the task), the p value reflects the result of a t-test rather than an ANOVA.
Figure 4-13: Regressors for fMRI analyses are not correlated

For each subject we correlated unsigned Self- and Other-attributed PE on shared trials in the testing session. In both Lo-Share and Hi-Share tasks, no subject showed a significant correlation.
Chapter 4: Meta-representational associative learning

<table>
<thead>
<tr>
<th>Method</th>
<th>Contrast</th>
<th>$p_{FWE}$ (cluster level)</th>
<th>Cluster size</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>PE$p^a$</td>
<td>&lt; 0.001</td>
<td>1835</td>
<td>29</td>
<td>-54</td>
<td>44</td>
</tr>
<tr>
<td>GLM</td>
<td>PE$p^b$</td>
<td>&lt; 0.001</td>
<td>981</td>
<td>47</td>
<td>-60</td>
<td>-6</td>
</tr>
<tr>
<td>GLM</td>
<td>PE$p^c$</td>
<td>&lt; 0.001</td>
<td>802</td>
<td>-44</td>
<td>-66</td>
<td>-5</td>
</tr>
<tr>
<td>GLM</td>
<td>PE$p^d$</td>
<td>&lt;0.001</td>
<td>2365</td>
<td>-9</td>
<td>-93</td>
<td>2</td>
</tr>
<tr>
<td>GLM</td>
<td>PE$p^e$</td>
<td>&lt;0.001</td>
<td>1486</td>
<td>36</td>
<td>-77</td>
<td>32</td>
</tr>
<tr>
<td>GLM</td>
<td>PE$p^f$</td>
<td>&lt;0.001</td>
<td>723</td>
<td>35</td>
<td>-39</td>
<td>41</td>
</tr>
<tr>
<td>GLM</td>
<td>PE$p^g$</td>
<td>0.001</td>
<td>541</td>
<td>-29</td>
<td>-84</td>
<td>26</td>
</tr>
<tr>
<td>Searchlight</td>
<td>PE$p^h$</td>
<td>0.018</td>
<td>1632</td>
<td>41</td>
<td>-65</td>
<td>24</td>
</tr>
<tr>
<td>Searchlight</td>
<td>PE$p^i$</td>
<td>0.044</td>
<td>1230</td>
<td>29</td>
<td>-53</td>
<td>48</td>
</tr>
<tr>
<td>Searchlight</td>
<td>PE$p^j$</td>
<td>0.045</td>
<td>1223</td>
<td>18</td>
<td>-21</td>
<td>54</td>
</tr>
<tr>
<td>Searchlight</td>
<td>PE$p^k$</td>
<td>0.02</td>
<td>1586</td>
<td>-18</td>
<td>-92</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 4-2: Summary of significant clusters

Peak co-ordinates of all significantly large ($p_{FWE} < 0.05$) clusters identified using a cluster-forming threshold of $p < 0.001$. Co-ordinates are reported in MNI space. Cluster sizes are reported in number of voxels.
Chapter 4: Meta-representational associative learning

Figure 4-14: Cross-validation scheme for cross-decoding

An illustration of how cross-validation was performed for the cross-decoding analysis. For each condition (Hi-Share or Lo-Share) there were two runs of fMRI data. Each run was split into a validation half and a test half. Decoders were trained on the PE\textsuperscript{*} or PE\textsuperscript{o} data for one a whole run, validated on half of a run for the alternate signal, and tested on the other half of that same run. Eight folds of cross-validation were carried out in total.
Chapter 5: Self-Other distinction in intertemporal choice

The work presented in this chapter is in preparation as a research article. I would like to acknowledge Giles Story for developing and sharing MATLAB code for hierarchical fitting of discounting models.

5.1 Abstract

Construal-level theory suggests that social distance and temporal distance can be represented using the same cognitive machinery. Here we test whether there is an association between subjects’ propensities to traverse these two psychological dimensions. We operationalise social distance with behavioural and neural measures of Self-Other distinction, the ability to selectively attribute belief updates to Self and Other. We operationalise temporal distance using an intertemporal choice task, wherein subjects must choose between an immediate small reward and a delayed large reward. We find that the extent to which an individual discounts the value of future rewards is negatively associated with the extent to which they engage in Self-Other distinction. This is consistent with the notion that future Self is represented like another agent, and that a common mechanism may underlie how the Self-attributed mental states are represented with respect to mental states attributed to Others and future Selves. Finally, we show that the microstructure of a sub-region of the vmPFC, which we previously showed to be related to knowledge about Self-Other distinction, is associated with temporal discounting behaviour. The results are consistent with the view that the vmPFC is involved in deploying meta-representational structural knowledge for attributing mental states to different agents, both within and outside the social domain.
5.2 Introduction

Mental time-travel is the ability to represent one’s future and past mental states. Like Theory of Mind, it is a form of mentalisation and requires meta-representation (Suddendorf & Corballis, 1997, Redshaw, 2014). There are consistent patterns of brain activation across different tasks that require subjects to represent states displaced from the Self, in the here and now (Buckner & Carroll, 2007). Construal-level theory (Trope & Liberman, 2010) exploits this observation to argue for a common neurocomputational architecture, which enables humans to traverse ‘psychological distances’, such as space, time and social distance.

Consistent with this theory, autobiographical memory and Theory of Mind are abilities that emerge at a similar stage in childhood development (Perner et al., 2007). Furthermore, people with autism spectrum disorder (ASD) are not only impaired at representing other agents’ mental states, but also at engaging in autobiographical memory (Adler et al., 2010) and episodic future thinking (Terrett et al., 2013).

Intertemporal choice paradigms are particularly popular tools for investigating mental time travel. These tasks tend to elicit a typical behaviour whereby subjects discount the value of future rewards as a function of temporal delay (Ainslie, 1974, Thaler, 1981). The ‘multiple selves’ theory considers intertemporal choice to reflect a conflict between the goals of the immediate Self, and an abstract future Self, which is represented like a different agent (Parfit, 1971, Ainslie, 1989). This hypothetical relationship between social and temporal distance exemplifies the construal-level theory.

Recent neuroimaging findings are consistent with the use of a common mechanism in traversing social distance and temporal distance. Activity in the medial prefrontal cortex (mPFC) is associated with whether a subject is thinking about Self or Other, and also with whether a subject is thinking about Self or future Self, a physiological marker predictive of
temporal discounting propensity (Ersner-Hershfield et al., 2009, Mitchell et al., 2010).

Furthermore, mPFC activity is predictive of social discounting propensity, whereby the value of a hypothetical reward for another agent is discounted as a function of social distance (Hill et al., 2017b). Social discounting and temporal discounting are both enhanced by transcranial stimulation of the temporo-parietal junction (TPJ) (Soutchek et al., 2016) and people who show more activity in this region during a false-belief task also show less temporal discounting in an intertemporal choice task (O'Connell et al., 2018). Finally, pre-schoolers who discount future rewards more are also less likely to share preferred toys with socially distant Others (Garon et al., 2011).

The putative relationship between temporal and social discounting may be evidence in favour of common mechanism for representing multiple types of psychological distance. However, as the decision variables involved in both tasks are so similar, an association between them may simply reflect the common use of value-based decision making systems. Furthermore, social distance is usually difficult to quantify, and relies on subjective self-reports from participants. Tasks that make use of friend versus stranger conditions cannot guarantee that they are manipulating social distance in the same way that temporal distance can be so carefully manipulated in an intertemporal choice task.

If meta-representational learning is a general feature of cognition, then one would expect that people who learn associations between the mental states of Self and Other, also form associations between the mental states of Self and future Self. Here we tested whether our formal measures of Self-Other distinction, derived from behaviour and functional imaging (see Chapter 4) are indeed related to intertemporal choice behaviour. We hypothesised that subjects who represent the mental states of Self and Other more distinctly, also represent the mental states of Self and future Self more distinctly. Thus, we predicted that subjects who showed more ‘leakage’ between Self and Others would also represent the mental states of Self and
future Self more similarly, and would show less discounting for future rewards.

In Chapter 4 we observed that the microstructure of white matter adjacent to the vmPFC was associated with the capacity to either acquire or deploy meta-representational structural knowledge. This region is known to be involved in representing rewards for future Self (Bechara et al., 2000, Kable & Glimcher, 2007, Nicolle et al., 2012, Garvert et al., 2015). Therefore, we were also interested to test whether microstructure in the same region of prefrontal white matter was related to both Self-Other distinction and intertemporal choice.

5.3 Methods

5.3.1 Intertemporal choice task

For participant details see section 4.3.1. All subjects completed a computerised intertemporal choice task on day one of a three-day experiment. The task consisted of sixty binary forced choice questions. Each question presented the subject with a choice between an immediate reward and a delayed reward. Immediate rewards were monetary gifts ranging from £1 to £9, which would be given to the participant at the end of the experiment. Delayed rewards were monetary gifts ranging from £2 to £10, which would be given to the participant after some temporal delay after the end of the experiment.

The immediate reward was always smaller than the delayed reward, such that subjects were always choosing between a ‘smaller sooner’ outcome and a ‘larger later’ outcome. The set of possible delays for the ‘larger later’ outcome comprised: one day, one week, two weeks, four weeks, six weeks, eight weeks and twelve weeks. An example question is ‘Would you rather have £3 now, or £7 in one week?’ All subjects were presented with the same sixty questions.
Chapter 5: Self-Other distinction in intertemporal choice

Subjects were instructed to make choices that were consistent with their own personal preferences. They were told that one of their sixty choices would be chosen at random at the end of experiment, and that their choice on that trial would be honoured.

5.3.2 Delay discounting models

We fit two different discounting models to subjects’ intertemporal choice behaviour. Both models assumed a hyperbolic discount function, in line with previous findings from investigations on human discounting behaviour (Rachlin et al., 1991, Green & Myerson, 1996). The first model was a simple hyperbolic discounting model that took the following form:

$$V_{\text{later}} = \frac{V}{1 + kD}$$

Equation 5-1

Here $V_{\text{later}}$ denotes the value of the later option after passing it through the discount function. $V$ denotes the raw, undiscounted value of the option. $D$ denotes the delay of the reward, in days, and $k$ is a free parameter that modulates the extent to which delayed rewards are discounted. The second model was a generalised hyperbolic discounting model that took the following form:

$$V_{\text{later}} = \frac{V}{1 + kD^g}$$

Equation 5-2

This model contains one additional free parameter, $g$, which allows additional flexibility in the shape of the discounting curve.

In both models, the difference between the values of the two options was passed through a softmax function to account for probabilistic choice behaviour.
\[ P_{\text{later}} = \frac{1}{1 + e^{\beta(V_{\text{sooner}} - V_{\text{later}})}} \]

**Equation 5-3**

\( P_{\text{later}} \) denotes the probability of the subject choosing the ‘larger later’ option on a specific trial. \( \beta \) is a free parameter that governs choice stochasticity. A high \( \beta \) generates choice behaviour that is more strongly determined by the values of the two options. A low \( \beta \) generates stochastic choice behaviour, which is less sensitive to the values of the two options.

Thus, the generalised hyperbolic model contained three free parameters. The hyperbolic model is a simpler, nested model with only two free parameters, equivalent to the generalised hyperbolic model with the \( g \) parameter fixed to one.

### 5.3.3 Model fitting and model comparison

Models were fit with unconstrained parameter values in log space. Thus, in native space, parameter values had a lower bound of zero and no upper bound.

We optimised the maximum a posteriori (MAP) of observed data, given a likelihood and an empirical group level prior over model parameters. The empirical prior effectively regularises the model, preventing parameter estimates taking on extreme values. The hyperparameters of this Gaussian prior (mean and variance) were estimated by maximising the likelihood of all the data from all subjects. As applied to the binary choice data of an intertemporal choice task, this procedure is similar to fitting a mixed effects logistic regression model.

To optimise the hyperparameters of the prior distribution we used a technique developed by Quentin Huys and colleagues (Huys et al., 2011a), which has been adapted for fitting delay discounting models (Story, 2015). The technique makes use of an expectation-maximisation (E-M) algorithm, that iterates between E-steps, where posterior parameter distributions are
Chapter 5: Self-Other distinction in intertemporal choice

estimated for each subject, and M-steps, where the empirical prior is updated. The algorithm iterates between these two steps until convergence.

At every E-step, a MAP estimate is computed for each subject by minimising the negative log posterior probabilities with the fminunc function in MATLAB (Mathworks, MA, USA). The variance on this MAP parameter is computed using a Laplace approximation, which assumes that the posterior distribution is simply a Gaussian around the MAP estimate.

These subject-specific means and variances are then used for updating the hyperparameters of the prior on the M-step. Here, the mean and variance of the Gaussian prior are updated. The mean is simply set to the mean of all subjects’ MAP estimates. The variance update is more complex, and incorporates both subject-level estimation error and between-subject variability:

\[
\text{Prior Variance} = \frac{1}{N} \sum_{j}^{N} [(m_j - \mu_j)^2 + \sigma^2_j]
\]

Equation 5-4

The variance of the prior distribution incorporates the individual variance (\(\sigma^2\)) of the posterior for each subject \(j\), as well as the deviation of the mean \((m)\) of each subject from the prior mean \((\mu)\). \(N\) denotes the total number of subjects. The resulting prior distribution is then used for computing new subject-specific MAP estimates in the subsequent E-step.

To compare the goodness-of-fit for each model, we computed an ‘integrated BIC’ score (iBIC) for each model. Where the conventional BIC (Equation 3-10) approximates the model evidence, \(p(data|model)\), for each subject, the iBIC approximates the evidence for the full hierarchical model at the group level:

\[
iBIC = h\log(|\mathcal{A}|) - 2\log(\mathcal{A}|\hat{\theta}^{ML})
\]

Equation 5-5
Chapter 5: Self-Other distinction in intertemporal choice

This equation takes the same form as that of the BIC. \( h \) denotes the number of hyperparameters fitted, i.e. a mean and variance for each model parameter. \( \mathcal{A} \) denotes all of the choice data from all of the subjects, with \(|\mathcal{A}|\) denoting the total number of data points. \( \log(\mathcal{A}|\hat{\theta}^{ML}) \) is the log likelihood of all the data given the optimised prior hyperparameters \( \hat{\theta}^{ML} \). This is equivalent to the sum of subject-specific model evidences after integrating out the subject-specific parameters.

\[
\log(\mathcal{A}|\hat{\theta}^{ML}) = \sum_{j}^{N} \log \int d\theta \, p(y_j|\theta)p(\theta|\hat{\theta}^{ML}) \\
\approx \sum_{j}^{N} \log \frac{1}{K} \sum_{k=1}^{K} p(y_j|\theta^k)
\]

Equation 5-6

\( y_j \) is the data for a single subject \( j \), \( p(\theta|\hat{\theta}^{ML}) \) is the prior distribution of model parameters \( \theta \), set by the optimised hyperparameters \( \hat{\theta}^{ML} \). The integral is approximated by summing over \( K \) samples, drawn from the empirical prior \( p(\theta|\hat{\theta}^{ML}) \). In this implementation, \( K \) was set to 2000. The model with the lowest iBIC was considered the best model for explaining subjects’ choice behaviour.

5.3.4 Logistic regression analysis

We conducted a leave-one-subject-out logistic regression analysis to see if we could predict whether a subject was a high or low discounter (via a median split), using various subject-specific variables related to the false belief task described in Chapter 4.

The first logistic regression model included two predictor variables: the leak parameter estimated for the Hi-Share task in training and the leak parameter estimated for the Hi-Share task in testing. The second, third and fourth logistic regression models were control analyses to see if any of the other parameters from the false belief task were predictive of discounting
propensity. The second logistic regression model had two predictor variables: the mean learning rate for the Hi-Share task in training and the mean learning rate for the Hi-share task in testing. The third logistic regression model included two different predictor variables: the mean temperature parameter for the Hi-Share task in training and the mean temperature parameter for the Hi-Share task in testing. The fourth logistic regression model included only one predictor variable: the memory decay parameter for the Hi-Share task in testing (note that the winning model for the Hi-Share task in training did not include a memory decay parameter). The fifth and final logistic regression model was the same as the first, but included two additional predictor variables: the cross-decoding accuracy from the fMRI analysis on the Hi-Share task, and the cross-decoding accuracy from the fMRI analysis on the Lo-Share task. This fifth logistic regression model therefore included four predictor variables.

In each of these logistic regression analyses, for each left-out subject, we computed the cross-entropy between the model’s estimated probability of the subject being a high discounter, and the true label for that subject. Each logistic regression model’s performance was quantified as the median cross-entropy across folds.

Statistical inference was made by generating permutation-based null distributions. For each statistical test, the analysis was repeated 5000 times, each time randomly permuting the class labels. The performance of the logistic regression model was considered significantly better than chance if the cross-entropy was below the 5th percentile of the null distribution of cross-entropies.

This technique also allowed us to quantify whether the prediction accuracy achieved using one variable was significantly better or worse than the prediction accuracy achieved using other variables. To make these statistical inferences we constructed null distributions of cross-entropy differences. The performance of a logistic regression model was considered significantly better than another model, if the difference in cross-entropy was
greater than the 95th percentile of the null distribution of cross-entropy differences.

5.4 Results

5.4.1 Common computation of social and temporal distance

We found that subjects’ choice behaviour reflected a hyperbolic discounting curve, which was best explained by the generalised hyperbolic discounting model (Figure 5-1). The iBIC was 1891 for the hyperbolic model and 1861 for the generalised hyperbolic model. This model contains two parameters that govern the impact of temporal delay on subjective value. Both parameters, k and g, generate steeper discounting curves at higher values. Thus, to quantify discounting propensity for each subject we simply summed these two parameters, in log space, to produce a single subject-specific discount factor: \( \log(g) + \log(k) \).

To see whether this model-derived construct had face validity, we tested whether it was correlated with a model-free measure of discounting behaviour, area under the discounting curve (Figure 5-1). A steeper discounting curve has a smaller area beneath it, and thus the discount factor should be negatively correlated with the area under the curve. Indeed, these two measures were negatively correlated across subjects \( r = -0.72, p < 0.001 \) (supplementary Figure 5-4).
Our hypothesis was that a common mechanism underlies representations of temporal distance and social distance, meaning that subjects who discount future rewards more will show a larger Self-Other distinction. We therefore predicted that discount factor would be negatively correlated with the Self-Other leak parameter fit to behaviour in the false belief task (see Chapter 4). For each subject, we summed the leak parameters from the training and testing sessions of the Hi-Share task. Indeed, this measure was significantly negatively correlated with discount factor [Spearman’s rho = –0.34, p = 0.03] (supplementary Figure 5-5).

To test the robustness of this relationship, we asked whether we could predict a subject’s discounting propensity from their leak parameter. We classified subjects as high and low discounters, using a median split. We then trained a logistic regression model to predict whether a subject was a
high or low discounter, using the two leak parameters from the train and test tasks. The performance of the model was assessed with leave-one-subject-out cross-validation (Figure 5-2).

This model performed significantly better than chance on left-out subjects \([p = 0.005]\). In order to test whether this predictive accuracy was specific to the leak parameters, we conducted the same analysis but using the other parameters from the learning model fit to the false belief task data (Hi-Share task). None of the logistic regression models that used these other parameters as predictors were able to predict discounting propensity better than chance \([\text{learning rate: } p = 0.4, \text{ memory decay: } p = 0.34, \text{ temperature: } p = 0.18]\). Furthermore, the model trained with the leak parameters performed significantly better than these other three models \([\text{learning rate: } p = 0.002, \text{ memory decay: } p = 0.004, \text{ temperature: } p = 0.001]\). All p-values were derived from permutation-based null distributions.

These results suggest a relationship between a behavioural measure of Self-Other distinction, and discounting behaviour. We wanted to see whether this association would become even stronger if we made use of our neural measures of Self-Other distinction. We constructed another logistic regression model which included the leak parameters as predictors, and also included the cross-decodability fMRI measures from the test session, which index a neural Self-Other distinction. Specifically, a high cross-decodability reflects a poor Self-Other distinction.

This model could predict discounting behaviour significantly better than chance, but this is a trivial result since we simply added more predictors to a model that could already predict discounting behaviour. To ensure that this model was performing better than the simpler model, over and above what would be expected from simply adding additional features, we tested the improvement in prediction accuracy against a permutation-based null distribution where the additional predictors were shuffled. Indeed, the improvement in prediction accuracy was significantly better than chance \([p = 0.016]\).
Prediction accuracies (cross-entropy) for five different logistic regression models trained to predict whether a subject was a high or low discounter. Diamonds represent median cross-entropy across folds of cross-validation (leave-one-subject-out). The only behavioural parameter that could predict discounting behaviour better than chance was the leak parameter \((p = 0.005)\). The prediction accuracy of this logistic regression model was significantly better than the accuracies of the models that used learning rate \((p = 0.002)\), memory decay \((p = 0.004)\) or temperature parameters \((p = 0.001)\). When neural measures of Self-Other distinction were added as predictors to the logistic regression model that used a leak parameter, prediction accuracy was significantly improved \((p = 0.016)\). *** denotes \(p < 0.001\). ** denotes \(p < 0.01\). * denotes \(p < 0.05\). Error bars show interquartile range across folds of cross-validation. P-values estimated from permutation-based null distributions of cross-entropy or cross-entropy difference. Vertical bars show chance level prediction accuracy for each logistic regression model.

Finally, we tested for a simple correlation between discount factor and a composite Self-Other distinction score, which was the sum of the behavioural and neural measures (i.e. sum of leak parameter and cross-decodability measures). As expected, the composite score and the discount factor were significantly negatively correlated \([\text{Spearman's rho} = -0.48, p = \ldots] \)
0.002]. This correlation was stronger than the correlation between the
discount factor and the summed leak parameters on their own (supplementary Figure 5-5). These findings show that a subject’s temporal
discounting behaviour can be predicted above chance, from behavioural and
neural measures of Self-Other distinction.

5.4.2 vmPFC microstructure is associated with discounting
behaviour

Given the association between Self-Other distinction and temporal
discounting, we wanted to test whether the same microstructural variability
that was associated with meta-representational learning (section 4.4.5) was
associated with discount factor. We conducted a second-level regression
analysis to see whether myelin-related white matter MT was associated with
discount factor, controlling for age, sex and intracranial volume. We restricted
the analysis to a region of interest, which was the right vmPFC cluster
discovered in section 4.4.5. This cluster comprised 844 voxels.

Indeed, within this region, there was a subcluster of 201 voxels where
MT was negatively associated with discount factor (Figure 5-3) [peak co-
ordinates: x = 12.8, y = 54.4, z = −15.2, \( p_{FWE} = 0.022 \), voxel-level, small-
volume corrected]. This means that subjects with more myelin-related MT in
this region had lower discount factors. When testing for a positive contrast,
there were no significant voxels.

We also conducted an exploratory whole-brain analysis with a
negative contrast, looking for regions where white matter myelin-related MT
was negatively associated with discount factor (Figure 5-3). This yielded one
significantly large cluster, of 174 voxels. This cluster was also adjacent to the
right vmPFC, but was more ventral than the previous cluster [peak co-
ordinates: x = 15.2, y = 38.4, z = −24, \( p_{FWE} = 0.043 \), cluster-level]
Figure 5-3: Ventromedial prefrontal white matter MT is associated with discount factor

The blue overlay shows a subcluster of 201 voxels of vmPFC white matter (within a ROI defined from Chapter 4) that was negatively associated with discount factor \([p_{FWE} = 0.022, \text{voxel-level, small-volume corrected}]. \) The red overlay shows a cluster of 174 voxels of vmPFC white matter discovered in a whole-brain analysis that was negatively associated with discount factor \([p_{FWE} = 0.043, \text{cluster-level}]. \)

5.5 Discussion

Our findings satisfied our prediction that behavioural and neural measures of Self-Other distinction are related to discounting behaviour in an intertemporal choice task. Subjects who discounted future rewards more steeply also represented other agents’ mental states more distinctly from their own mental states and were better able to distinguish the beliefs of Self and Other.

In intertemporal choice tasks, people with better cognitive function across a range of tasks tend to discount future rewards less than those with cognitive impairments (Willner et al., 2010, Boyle et al., 2012). Despite this, the association we found here predicts a positive relationship between
performance on the false belief task and discounting propensity. In the false belief task, optimal performance requires segregation of the belief updates for Self and Other. It is striking then, that a negative association between Self-Other leak and discounting propensity is detectable, in spite of a probable opposing effect driven by general cognitive performance.

Our findings are consistent with a common mechanism underlying the attribution of mental states to future Self and to Others. Cast in terms of meta-representational learning, some individuals may simply be more likely to learn associations between meta-representations than others. An individual who is more likely to form these associations would be more likely to generalise between models attributed to Self and Other, and also more likely to generalise between models attributed to Self and future self.

These results may also help to clarify some of the findings presented in Chapter 3 (Ereira et al., 2018). If a general mechanism for meta-representational learning can explain how mental states are attributed to both future Self and Others, it may also explain how mental states are attributed to counterfactual Selves, of which the future Self is a special case. Thus, the difference in neural Self-Other distinction that we observed between a ‘social’ and ‘non-social task’ may not reflect two qualitatively distinct cognitive mechanisms, but may reflect a spectrum of associative strengths. Through meta-representational associative learning, subjects may acquire a shared model for attributing mental states to Self and counterfactual Self, which simply hasn’t been learned for the Other in the ‘social’ condition. From this perspective, the terminology ‘social’ and ‘non-social’ may be rather misleading.

We also found that myeloarchitecture of the same region of prefrontal white matter, which was associated with acquiring or deploying meta-representational knowledge, was associated with discounting behaviour. The role of the vmPFC in representing the value of future rewards is well established (Bechara et al., 2000, Kable & Glimcher, 2007, Nicolle et al., 2012, Garvert et al., 2015). Structural brain imaging has shown that discounting behaviour is associated with frontostriatal tract integrity (Peper et
al., 2013, Achterberg et al., 2016), and impulsivity is associated with myelin-related MT in the mPFC (Ziegler et al., 2019). Our finding that vmPFC MT is negatively associated with discount factor is consistent with these findings.

Our finding that the structure of this same region is involved in learning about relations between the mental states of Self and Other may help in understanding its role. The vmPFC is strongly implicated in both self-referential processing (D'Argembeau, 2013) and social cognition (Mitchell et al., 2006). In addition to representing the value of delayed rewards for Self it has also been shown to represent the subjective value of rewards for Others during altruistic behaviour (Hare et al., 2010, Cutler & Campbell-Meiklejohn, 2019), consistent with the vmPFC playing a role in traversing both social and temporal distance.

A recent meta-analysis found that the vmPFC was the largest area of overlap between regions active during metacognitive tasks and Theory of Mind tasks (Vaccaro & Fleming, 2018), both of which invoke meta-representations. Furthermore, a recent study found that patients with vmPFC lesions were no more detailed in their descriptions of future events pertaining to Self than of events pertaining to Other, unlike controls, who report richer descriptions of events pertaining to future Self (Verfaellie et al., 2019).

These findings are consistent with the vmPFC having a role in acquiring or deploying meta-representational structural knowledge, which is necessary for the attribution of mental states to different agents, such as Self, past Self, future Self and Other.
5.6 Supplementary material

Figure 5-4: Validating the discount factor
To ensure that our discount factor (log(k) + log(g)) was a reasonable measure of
discounting behaviour, we tested whether it was correlated with a model-
free measure of discounting behaviour, area under the discounting curve. Indeed, these two measures were
strongly negatively correlated.

Figure 5-5: Correlations between Self-Other distinction and discount
factors
A) The leak parameter, a behavioural measure of Self-Other distinction, was weakly
negatively correlated with discount factor. B) A composite measure (sum) of leak parameter
and cross-decodability from the fMRI analysis was strongly negatively correlated with
discount factor. Thus, when both behavioural and neural measures of Self-Other distinction
are used, there is a stronger association with discount factor.
Chapter 6: General discussion

6.1 Summary of findings

The overall aim of this thesis was to find out how the human brain selectively attributes mental states to different agents such that Self-Other distinction is achieved. In Chapter 3 I showed that it is possible to classify a sensory prediction error (PE) as being Self-attributed or Other-attributed, on the basis of the whole-brain topography of that signal's activity pattern. This provided evidence that these learning signals can conjunctively represent both the error of a model, and the agent to whom that model belongs. I showed that the extent to which these neural signals are agent-specific predicts how well people can distinguish the belief updates of Self and Other, and also the severity of subclinical psychopathological traits, demonstrating that this measure was functionally meaningful.

I showed that the extent to which these signals were agent-specific could be significantly modulated by a subtle change in context. When the other agent was framed as a counterfactual version of Self rather than another person, the signals were no longer encoded in topographically distinct activity patterns. This meant that, despite not requiring any social interaction, the social version of this task induced learning signals that were represented differently compared to the non-social version.

In Chapter 4 I showed that the agent-specificity of these learning signals could be altered through training. This plasticity was consistent with a form of learning whereby exposure to a strong contingency between Self- and Other-attributed sensory PEs causes the encodings of these signals to become less distinct. This reduction in neural Self-Other distinction was also associated with a reduction in the ability to distinguish between the belief updates of Self and Other, demonstrating that this plasticity was functionally meaningful. These results highlighted the possibility that the social and non-social contexts in Chapter 3 might not engage fundamentally different
computational processes, but rather reflect priors about Self-Other distinction learned in advance of the experiment.

Chapter 4 also showed that the microstructure of ventromedial prefrontal white matter was associated with this plasticity in PE representations. Subjects who had a higher myelin-related magnetisation transfer (MT) in this region were more susceptible to the training effect, suggesting that the vmPFC may play a role in relating meta-representational structural knowledge to contextual cues, such as faces.

Finally, in Chapter 5 I showed evidence for an association between intertemporal choice behaviour and Self-Other distinction. Subjects who were less able to distinguish their own beliefs from the beliefs of Others, also discounted future rewards more. This is consistent with the idea that people represent their future Self as either a distant or close Other. Given that, in Chapter 4 it was shown how Self-Other distinction is a variable that can be learned through experience, these results suggest that this form of learning may support the acquisition of meta-representational structural knowledge, for attributing mental states to Self and Other in both social and non-social settings.

6.2 Interpretations and implications

The learning signals detected in the probabilistic false belief task were encoded in agent-specific activity patterns during the social context. That these signals contain information about both the environment and the agent to whom the signal is attributed, suggests that this signal contains sufficient information for supporting a propositional attitude like ‘I am surprised that I saw a pink umbrella’.

Despite the fact that the same signal provides a conjunctive representation of both PE and agent identity, these two components are represented differently. PE magnitude is often encoded through the firing rate of a specific population of neurons (Schultz et al., 1993, Takahashi et al., 2017). This is consistent with the sensor-specific evoked field changes and
voxel-specific BOLD changes that correlated with PE magnitude in Chapter 3 and Chapter 4 respectively. The agent identity component however, was encoded in the relative spatial topographies of different PE signals.

We can describe these two components as being represented at different levels of abstraction. PE magnitude is encoded at a low level of abstraction, whilst agent identity is encoded at a higher level of abstraction. At the low level, information about PE magnitude can be decoded from one instance of a PE signal. At the higher level, information about agent identity can be decoded from multiple instances of PE signals, each one encoded in a different, but possibly overlapping, neuronal population. At this higher level, information about PE magnitude has been abstracted away. The expression of multiple PE signals, each pertaining to the same environmental state, in different neuronal populations, provides a level of abstraction at which agent identity can be represented.

If mental state attribution is indeed computed at this level of abstraction, then Self and Other can only be represented in terms of distance from each other. In other words, the Self-attributed signal is only meaningful with respect to an Other-attributed signal, and vice versa. This is consistent with theoretical work that considers introspection and Theory of Mind as relying on the same underlying process (Carruthers, 2009).

Considering mental states as being encoded at this higher level of abstraction may help to tie together the Theory Theory (TT) and Simulation Theory (ST) of mentalising. As discussed in Chapter 1, TT is grounded in simple learning rules, which an observer can use to make predictions about another agent’s behaviour. ST, on the other hand, is grounded in attributing a simulated mental state to an Other, a meta-representational process. Empirical evidence exists for both TT (Ruffman, 1996b, Saxe, 2005a) and ST (di Pellegrino et al., 1992, Fogassi et al., 2005, Senju et al., 2011). Thus, a comprehensive model of mentalising ought to accommodate the representational flavour of TT with the meta-representational flavour of ST.

Tamir and Thornton’s model, introduced in Chapter 1 (Figure 1-10), grounds behaviour reading and mind reading in a single predictive
processing framework (Tamir & Thornton, 2018), but does not make a computational distinction between TT and ST. The model describes a superficial layer for representing the observed behavioural states of the observee, in addition to a deep layer for representing their hidden mental states, which are latent causes of the observed behaviour. The observer can exploit a pre-learned probabilistic transition structure, whereby deep states can be used, not only to make predictions about subsequent deep states, but can also be used to make predictions about subsequent shallow states. For instance, if Sally observes Anne laughing, Sally may exploit a pre-learned model to infer that a hidden state of Anne’s belief, ‘Sally loves playing marbles with me’, causing the observed behaviour in Anne. Sally can leverage that hidden state to make predictions about what Anne might next covertly believe, and also what she might next overtly do.

This framework can be enriched by considering the deep states as being encoded at the higher level of abstraction, described above, and the shallow states as being encoded at the lower level. In this sense, the deep states in Sally’s model are meta-representations, which are selectively attributable to different agents, either Sally or Anne. The shallow states, however, are mere representations, Sally’s beliefs about the world that she takes for granted, which cannot be selectively attributed to different agents. Describing this framework in TT/ST parlance, the transition structure is analogous to the set of rules in the ‘theory’, whilst the deep meta-representational states, are ‘simulations’. TT describes the construction of the predictive model through associative learning, whilst ST describes the fact that some states in this model are meta-representational.

The distinction between these two kinds of mental states, representational and meta-representational, is similar that between phenomenologically transparent states and opaque states (Moore, 1903, Metzinger, 2008, Limanowski & Friston, 2018). Here, transparency and opacity refer to the ‘window’ through which an observer sees the world. Some mental states are opaque, in that the subject can ‘see the window’, or rather, attend to the process by which that mental state is constructed. Other
mental states are transparent, in that the subject can only ‘see through the window’, or rather, has no access to the way in which their mind has constructed the mental state. The shallow states, such as the sensory representation of Anne’s laughing, are transparent to Sally. These constitute Sally’s beliefs about the world, which are attributable only to herself and are taken to be true. The deep states, on the other hand, are opaque because here Sally is simulating Anne’s beliefs, and representing the fact that these are refutable and subjective mental states, constructed from the Anne’s prior experiences.

Taking this idea forward, an important question is why some mental states should be opaque and others transparent. The results presented in Chapter 4 suggest that these two labels might describe a continuous dimension, rather than a dichotomy. Our findings showed that the agent-specificity of a learning signal can be increased or decreased as a function of experience. If two instances of a PE signal become associated following temporal contingency, then their representations change to become more similar. Thus, they will encode less information at the higher level of abstraction. As these signals condense into a unitary signal, they can only encode information about PE magnitude at the lower level of abstraction. The mental state has thus shifted to be less opaque and more transparent. Rather than supporting a propositional attitude such as ‘I am surprised that I saw a pink umbrella’ or ‘Anne is surprised that she saw a pink umbrella’, the signal can now only support an agent-independent ‘surprise at seeing a pink umbrella’. Thus, in a predictive model used to make inferences about Others, the probabilistic transition structure may be acquired through simple associative learning, whilst the opacity of each state may be determined through meta-representational associative learning.

These findings have significant implications for the ontogenetic development of meta-representational skills. It is conventionally thought that children acquire Theory of Mind at about age four (Wimmer & Perner, 1983, Wellman et al., 2001, Rakoczy, 2017). This delay could reflect the time it takes to acquire meta-representational structural knowledge, and learn that
some mental states are agent-specific. Consistent with the view that Theory of Mind is something that children learn, a study of 1,116 pairs of five-year-old twins indicated a much stronger contribution of the environment on mind reading abilities than the genetic contribution (Hughes et al., 2005).

In Chapter 5 it was shown that people who distinguish between the mental states of Self and Other more, also discount rewards for their future Self more steeply. Furthermore, it was shown that the microstructure of ventromedial prefrontal white matter is associated with both discounting behaviour and meta-representational associative learning. These findings suggest that the kind of learning through which representations can transition between transparency and opacity may be important for the emergence of a coherent sense of Self. This is consistent with Axel Cleeremans’ Radical Plasticity Thesis (Cleeremans, 2011), which proposes that the brain learns to form meta-representations through experience-dependent plasticity. The theory suggests that the brain can learn not only about the environment, but also about its own representations, and thus learns to have consciousness.

Contrary to the idea that humans learn to meta-represent, the dominant view of mind reading is that it is a cognitive instinct (Heyes, 2018). Meta-representational abilities appear to have evolved late in phylogenetic history, and may be accessible to a small number of species, and perhaps even only humans (Suddendorf & Corballis, 2007). Some research suggests that very young human infants can represent other agents’ mental states (Kovács et al., 2010, Scott et al., 2010, Király et al., 2018). Many researchers have synthesised the comparative and developmental psychological literatures to conclude that mentalising is not learned over ontogenetic timescales, but rather is an adaptation honed through generations of evolution. Thus, suggesting that humans are born without a sense of Self, or any meta-representations, and simply acquire them through learning is a strong claim.

However, the hypothesis that Self and Other are acquired through meta-representational associative learning does not preclude the possibility of a ‘Self instinct’. The ability to form these associations, build meta-
Chapter 6: General discussion

representational structural knowledge, and represent agent identity at a higher level of abstraction, may depend on a neurobiology that has only evolved in certain species. From this perspective, the ability to engage in this kind of learning could be considered an instinct that humans are born with.

6.3 Limitations

The work presented in this thesis prioritised tight experimental control over ecological validity. Social neuroscientists have long acknowledged that humans constantly interact with other agents, and there have been recent calls to embrace this with naturalistic and interactive paradigms (Schilbach et al., 2013, Redcay & Schilbach, 2019). Indeed, social systems emerge through the co-ordinated behaviour of social groups, and it may seem farfetched to make inferences about social cognition from behaviour and neural activity in a non-interactive, individualistic setting.

The probabilistic false-belief task introduced in this thesis is, in essence, just a dual learning task. Subjects need to infer and keep track of two variables. The cover story frames the variables as the belief states of two different agents, but the computations performed in this task are not social per se. Subjects never get to observe another agent’s behaviour, nor do they even get feedback on how well they are predicting the behaviour of the other agent. This task was designed to induce the most basic computations required for social cognition, that is inferring the presence of another agent’s belief. The same can be said for classic false belief tasks used to test for Theory of Mind in children, infants and animals (Baron-Cohen et al., 1985, Kovács et al., 2010, Krupenye et al., 2016). Typically, these tasks do not involve any observational learning or social interaction, but simply tracking what another agent can or cannot observe, and test the ability to make predictions about the other agent’s behaviour based on mental state inference.

The probabilistic false belief task imposes strict restrictions on the streams of information available to Self and Other, in a manner which does
not emulate real-life social situations. Thus, whether the computations we index are general to mentalising, and not merely specific to this paradigm, is a valid concern. However, our measurements from the task are sensitive to the social context and also correlate with subclinical psychopathological traits (see Chapter 3). Furthermore, training on the task impacts behaviour in a different paradigm (see Chapter 4). These findings go some way to validate the measurements derived from this task as computations relevant for social cognition more generally. Additionally, the ability to represent another agent’s representations is necessary for more complex social cognition, such as observational learning or intention inference. Thus, the work in this thesis explored a specific computational process, which, whilst not evidently social per se, provides a substrate for more specifically social processes.

A further limitation, which stems from the paradigm being under tight control, is that it is very demanding for participants. Keeping track of two fluctuating variables imposes heavy working memory demands. This means that subjects have to exert a lot of effort to play the task, which is fatiguing. Though the participants who took part in these experiments could do the task sufficiently well, it may not appropriate to use in certain patient groups. General cognitive impairments, particularly related to working memory and attention, are frequently seen in both schizophrenia (Heinrichs & Zakzanis, 1998, Elvevag & Goldberg, 2000, Lencz et al., 2006) and autism spectrum disorder (ASD) (Bennetto et al., 1996, Russell et al., 1996, Corbett et al., 2009).

Though the paradigm might be a good measure of Self-Other distinction, a purported transdiagnostic dimension of variability in mental health disorders (Lamm et al., 2016), it may be too cognitively demanding for some patient groups. If performance is at chance then the task simply has no sensitivity to detect impairments in Self-Other distinction per se. In Chapter 3 it was shown that the neural measures of Self-Other distinction can predict variability in subclinical psychopathological traits. This shows promise for potential simplified versions of the paradigm, which could be used in patient
groups, but it is important to note that no clinical groups were studied as part of this thesis.

Caution is required when making inferences about any mental health disorders from this small sample of healthy participants. The emerging field of computational psychiatry calls for a dimensional approach to understanding mental illnesses, which seeks to ‘carve the brain’s computational nature at its joints’, rather than cluster conditions based on symptomatology (Huys, 2018). Recent efforts to relate variability in computational phenotypes to transdiagnostic dimensions of subclinical traits have been met with success (Gillan et al., 2016, Rouault et al., 2018). However, whether or not variability across the general population reflects the same variability between diagnosed and undiagnosed individuals, is an empirical question. It is likely that some symptoms, more than others, reflect the extremes of a continuous dimension. Discovering whether a dimension, such as Self-Other distinction ability, is clinically meaningful, will ultimately require testing in clinical groups.

6.4 Future directions

To build on this research, it will be important to develop adapted paradigms that can be used in patient groups. The current version of the probabilistic false belief task has already been included in a cognitive battery as part of a drug development trial for borderline personality disorder (BPD). People with BPD have been described as having an impairment in mentalising (Fonagy & Bateman, 2008) and an inability to distinguish Self from Other (Beeney et al., 2015). People diagnosed with BPD do not have generalised cognitive impairments to the same extent that patients with schizophrenia or ASD have, thus this group may be a population in which these hypotheses can be rigorously tested in a computational framework, without having to develop new paradigms. The primary purpose of including the task in the cognitive battery is to assess differences between patients
and controls. However, the effect of pharmacological manipulation on Self-Other distinction is also of interest.

The impact of psychedelic drugs on Self-Other distinction will be of particular interest for future studies. Psychedelics, such as psilocybin and lysergic acid diethylamide (LSD), can induce profound changes to the subjective sense of Self, often described as 'ego-dissolution' (Carhart-Harris et al., 2016, Tagliazucchi et al., 2016). These acute effects can predict long term changes in personality (Lebedev et al., 2016) and psychosocial functioning (Smigielski et al., 2019). Psychedelics have also been shown to modulate neural activity in so-called ‘social’ brain regions (Preller et al., 2018) and increase emotional empathy without increasing cognitive empathy (Pokorny et al., 2017). This may reflect a reduction in Self-Other distinction, which would produce a heightened sensitivity to Other’s emotional states, whilst impairing the ability to infer Others’ beliefs. Psychedelics may then be a useful tool for modulating Self-Other distinction and gaining further insight into its physiological underpinnings.

Psychedelics are known to enact their consciousness-altering and social effects via agonism at the 5-HT$_{2A}$ receptor (Vollenweider et al., 1998, Preller et al., 2018), which appears to result in gross changes in functional integration between sensory and executive cortical regions (Lebedev et al., 2015, Carhart-Harris et al., 2016, Tagliazucchi et al., 2016, Lebedev et al., 2016). If these functional effects can predict behavioural effects in Self-Other distinction tasks in a specific manner, this may help to ground accounts of psychedelic action and its purported therapeutic effects (Ross et al., 2016, Mithoefer et al., 2018), within a computationally-inspired framework.

An important theoretical development to this thesis will be investigating the adaptive benefit of Self-Other leakage. Our findings suggested that people can learn to conflate the computations of Self and Other, in an environment where this impairs performance. However, other environments may favour a degree of Self-Other leakage. In interactive settings, where a group of agents share a common goal, performance of the whole group might be improved if agents erroneously attribute each Other’s
belief states to each other. This spreading of belief states through the group might facilitate behavioural co-ordination. Even if false beliefs spread through the group, they may be corrected more quickly if agents can observe each other acting on false beliefs and then revising them. A recent study has shown that people are willing to forego monetary rewards to conform to the group, providing evidence that social conformity is intrinsically rewarding (Mistry & Liljeholm, 2019). This is consistent with humans having a bias towards adopting the beliefs of Others. However, as has been discussed at length in this thesis, the ability to identify a belief state as belonging to Self or Other is crucial for accurate mentalising. Recent work in the artificial intelligence community has seen the development of multi-agent simulations, where each agent has its own deep reinforcement learning architecture. (Rabinowitz et al., 2018, Raileaunu et al., 2018). These simulations are useful for modelling different aspects of social interactions, including observational learning and even Theory of Mind. Such simulations may be helpful for determining the optimal trade-off between Self-Other distinction and leakage in different social environments, while further empirical studies can be used to examine whether and how people can learn to shift this balance in a context-dependent manner.

6.5 Concluding remarks

This thesis was motivated by a general question:

‘How does the brain process information about the world, such that social systems emerge?’

From this general question, a more specific question was raised, of how the brain distinguishes between signals that are attributed to Self and simulated signals of other agents. The experimental work presented here shows evidence that Self-Other distinction is a fundamental computation that the human brain performs. By expressing learning signals in agent-specific activity patterns, basic learning signals do contain information about agent
identity, and the extent to which these signals contain that information, is itself a variable that people learn through experience. The results suggest that Self-Other distinction can be achieved through the kinds of learning mechanisms, which have been observed throughout the brain and across many different species.

This thesis has not demonstrated that Self-Other distinction is necessary, or sufficient, for the emergence of social systems. The focus on Self-Other distinction was borne out of intuition, and the full hypothesis remains to be tested. These findings about Self-Other distinction sketch out the structure of a biological system which is assumed to function at the level of social systems. This structure provides a substrate for future experimental manipulations, which can start to directly test this assumption. The neural and computational variables that have been described in this thesis should be assessed in their ability to make predictions about the emergence of social systems. If they can, then they may contribute to the development from cell to society.
References


References


Bechara, A., Damasio, A. R., Damasio, H. & Anderson, S. W. 1994, 'Insensitivity to future consequences following damage to human prefrontal cortex', *Cognition*, vol. 50, no. 1-3, pp. 7-15.


References


Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J. & Tomasello, M. 2017, 'Great apes distinguish true from false beliefs in an interactive helping task', *PLoS One*, vol. 12, no. 4, p. e0173793.


References


References


References


236
References


Gershman, S. J. 2019, 'What does the free energy principle tell us about the brain?', *arXiv*.


References


References


References


References


Huys, Q. J., Moutoussis, M. & Williams, J. 2011b, 'Are computational models of any use to psychiatry?', *Neural Netw.*, vol. 24, no. 6, pp. 544-51.


References


References


References


References


Li, L., Li, K. K. & Li, J. 2019, 'Private but not social information validity modulates social conformity bias', *Hum Brain Mapp.*

Lieberman, M. D. & Cunningham, W. A. 2009, 'Type I and Type II error concerns in fMRI research: re-balancing the scale', *Soc Cogn Affect Neurosci.*, vol. 4, no. 4, pp. 423-8.


References


Marques, J. P., Khabipova, D. & Gruetter, R. 2017, 'Studying cyto and myeloarchitecture of the human cortex at ultra-high field with quantitative imaging: R1, R2(*) and magnetic susceptibility', *Neuroimage*, vol. 147, pp. 152-63.

References


Mitchell, A. S. 2015, 'The mediodorsal thalamus as a higher order thalamic relay nucleus important for learning and decision-making', Neuroscience and Biobehavioral Reviews, vol. 54, pp. 76-88.


References


References


Parfit, D. 1971, 'Personal Identity', *Philosophical Review*, vol. 80, no. 1, pp. 3-27.


References


Peters, U. 2013, 'Interpretive sensory-access theory and conscious intentions', *Philosophical Psychology*, vol. 27, no. 4, pp. 583-95.


References


References

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T. & Reissland, N. 2017, 'The Human Fetus Preferentially Engages with Face-like Visual Stimuli', *Curr Biol*, vol. 27, no. 12, pp. 1825-+


Ruffman, T. 1996b, 'Do Children Understand the Mind by Means of Simulation or a Theory? Evidence From Their Understanding of Inference', *Mind & Language*, vol. 11, no. 4.


Saxe, R. 2009, 'The neural evidence for simulation is weaker than I think you think it is', *Philosophical Studies*, vol. 144, no. 3, pp. 447-56.

Saxe, R. & Kanwisher, N. 2003, 'People thinking about thinking people - The role of the temporoparietal junction in "theory of mind"', *Neuroimage*, vol. 19, no. 4, pp. 1835-42.


References


References


References


References


References


Thornton, M. A., Weaverdyck, M. E., Mildner, J. N. & Tamir, D. I. 2019a, 'People represent their own mental states more distinctly than those of others', *Nature Communications*, vol. 10.


References


References


Whiten, A. 2013, 'Humans are not alone in computing how others see the world', Anim Behav, vol. 86, no. 2, pp. 213-21.


References


