Author:
Tyrer, Ashley

Title:
Modelling of Behaviour and Neuroimaging in Decision Making and Memory Through Application of Variational Bayesian Methods, In Health and Disease

Author:
Tyrer, Ashley

Title:
Modelling of Behaviour and Neuroimaging in Decision Making and Memory Through Application of Variational Bayesian Methods, In Health and Disease

University of Bristol

Doctoral Thesis

Modelling of Behaviour and Neuroimaging in Decision Making and Memory Through Application of Variational Bayesian Methods, In Health and Disease

Ashley Tyrer

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Engineering

School of Engineering Mathematics

June 2021

Word Count: 46,201

Abstract

Computational Psychiatry is a rapidly emerging field, which combines traditional neuroscience with formal computational methods to investigate the transfer of information in neural circuits.

In this thesis, I aimed to combine analyses of behavioural and neuroimaging data with computational models of cognition and biological brain circuits, by applying Bayesian computational modelling techniques: specifically Active Inference and Dynamic Causal Modelling, in both health and disease.

I optimised a behavioural study with the aim of inverting a probabilistic Markov decision task for individual phenotyping. I observed a range of behavioural profiles across cohorts of healthy volunteers, and revealed distinct exploratory and exploitative behaviours.

I then applied human behavioural data to an Active Inference modelling framework, in which I inverted generative models to estimate subject-specific parameters encoding key mechanisms underlying behaviour and reward. I found that model inversion was successful in the accurate retrieval of these parameters within-subject, and that these parameters could predict coarse behavioural metrics on the group level.

By combining these techniques, I conducted a functional MRI study in which healthy participants performed the optimised behavioural task, then underwent a drug manipulation to induce selective noradrenaline reuptake inhibition. By constructing generative models for these participants, I found significant associations between neural activity in the locus coeruleus and anterior cingulate cortex, and model parameters estimated through inversion.

Finally, to examine biological circuits in neurogenerative disease, I analysed electroencephalography data collected from patients with Alzheimer’s disease and healthy older controls. This revealed left-lateralized memory circuit dropout in deeper memory tasks, with potential right-hemisphere compensation in simpler visual memory recall.

Taken together, these studies demonstrate the application of computational modelling in the study of problems in psychiatry and neuroscience to link mechanism to behaviour. The studies provide evidence that emerging Bayesian frameworks in computational psychiatry provide robust and mechanistically interpretable phenotypes.

Acknowledgements

I would like to thank all those who have offered me help and support throughout my PhD and during the process of writing this thesis.

Firstly, I would like to thank my supervisors, Rosalyn Moran and Iain Gilchrist, for all of their guidance, help, and patience over the last four and a half years during my PhD. This includes assistance with research and also encouragement to make the most of opportunities that were available to me during my studies; I have learned so much from their supervision which has made me grow so much as a researcher.

I would also like to thank Richard Apps, Conor Houghton and all those behind the Wellcome Trust Neural Dynamics PhD Programme, without whom I would not have had the incredible opportunity to study at the University of Bristol.

Furthermore, I would like to thank the members of the Bristol Computational Neuroscience Unit for introducing me to computational neuroscience and giving me the opportunity to learn about others' research.

I would also like to thank my family for their continued support and advice throughout my PhD.

Finally, I would like to personally thank my fiancé Thomas, for his unconditional love and support (including tech support!), and for the kindness and helpfulness he has shown me since we met.

Declaration of Authorship

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: [redacted] DATE: 18/06/2021

Contents

List of Figures

List of Tables and Boxes

List of Abbreviations

AAL	Anatomical Automatic Labelling
Aβ	Amyloid-beta
AC	Anterior Commissure
ACC	Anterior Cingulate Cortex
ACE	Addenbrooke's Cognitive Examination
ACh	Acetylcholine
AD	Alzheimer's Disease
ADHD	Attention Deficit Hyperactivity Disorder
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
ANOVA	Analysis Of Variance
BDNF	Brain-Derived Neurotrophic Factor
BMA	Bayesian Model Average
BMS	Bayesian Model Selection
BOLD	Blood Oxygen Level Dependent
cAMP	cyclic Adenosine Monophosphate
COVID-19	Coronavirus Disease 2019
CR	Cognitive Reserve
DA	Dopamine
DC	Direct-Coupled
DCM	Dynamic Causal Modelling
d.f.	Degrees of Freedom
DSM	Diagnostic and Statistical Manual of Mental Disorders
EEfRT	Effort Expenditure for Rewards Task
EEG	Electroencephalography
Ep	Effect size
ERP	Event Related Potential
FLAIR	Fluid-Attenuated Inversion Recovery
fMRI	functional Magnetic Resonance Imaging
FOV	Field Of View
FWE	Family Wise Error
FWHM	Full-Width at Half Maximum

GABA	Gamma-aminobutyric acid
GLM	General Linear Model
HMM	Hidden Markov Model
IFG	Inferior Frontal Gyrus
ITG	Inferior Temporal Gyrus
KL	Kullback–Leibler
LC	Locus Coeruleus
L-DOPA	L-3,4-dihydroxyphenylalanine
LFP	Local Field Potential
LOAD	Late-Onset Alzheimer's Disease
MANOVA	Multivariate Analysis Of Variance
MAP	Maximum *a posteriori*
MCI	Mild Cognitive Impairment
MDD	Major Depressive Disorder
MDP	Markov Decision Process
MEG	Magnetoencephalography
MMSE	Mini Mental State Examination
MNI	Montreal Neurological Institute
MTL	Medial Temporal Lobe
NA	Noradrenaline
NET	Norepinephrine Transporter
NMDA	*N*-methyl-ᴅ-aspartate
OCG	Occipital Gyrus
OCP	Occipital Pole
OFC	Orbitofrontal Cortex
PA	Pernicious Anaemia
PANSS	Positive and Negative Syndrome Scale
PC	Posterior Commissure
PEB	Parametric Empirical Bayes
PET	Positron Emission Tomography
POMDP	Partially Observable Markov Decision Process
Pp	Posterior probability
PTSD	Post-Traumatic Stress Disorder

RL	Reinforcement Learning
RPE	Reward Prediction Error
RT	Reaction Time
SAPE	State-Action Prediction Error
SARSA	State-Action-Reward-State-Action
SD	Standard Deviation
SEM	Standard Error of Mean
SNP	Single Nucleotide Polymorphism
SNRI	Selective Noradrenaline Reuptake Inhibitor
SPM	Statistical Parametric Mapping
SSE	Summed Square Error
TD	Temporal Difference
TE	Time to Echo
TPM	Tissue Probability Map
TR	Repetition Time
VB	Variational Bayes
VLPFC	Ventro-Lateral Pre-Frontal Cortex
VOI	Volume Of Interest
VTA	Ventral Tegmental Area
WFU	Wake Forest University
WM	White Matter

Chapter One

General Introduction

1.1 Overview

Computational Neuroscience and Psychiatry are rapidly-growing fields, in which traditional neuroscience, psychology and the biology of the brain are combined with formal computational approaches to investigate the transfer of information in neural circuits. This combination of computation with neural biology and aspects of psychology contribute powerful methodologies that can steer treatments and diagnoses in clinical settings (Redish and Gordon, 2016).

In order to effectively characterise, diagnose, and eventually treat a vast range of neurological and psychological disorders, we must gain an insight into the underlying neuronal mechanisms. Just as with the disease of any organ other than the brain, diagnosis begins with an investigation into the symptoms expressed by the patient. Gaining access to such mechanisms, however, is where the main issue lies. Direct observation of human neuronal activity is currently unattainable, as this would involve extremely invasive and potentially dangerous procedures deemed unethical. Therefore we must look to alternative, indirect measures of brain activity. Neuroimaging methods such as electroencephalography (EEG), magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) are frequently used in both research and clinical settings to offer a glimpse into neural activity across the whole brain, but these methods are very indirect and each have their imperfections and compromises. EEG, for example, involves a wearable headset which uses scalp surface electrodes to measure the electrical signals emitted by cortical neurons at synaptic timescales of milliseconds. However, this method offers extremely limited access, if any, to deeper brain structures such as the midbrain or hippocampus. FMRI, on the other hand, provides a visualisation of structures across a greater brain volume with very high spatial resolution, but this comes at the price of reduced temporal resolution and an even more indirect measure of activity, namely blood oxygen level dependent (BOLD) signals as opposed to electrical activity.

Furthermore, MRI as a modality is relatively restrictive, as the scanner can be an uncomfortable environment, particularly for those suffering from psychiatric or neurodegenerative diseases.

The vast catalogue of animal studies has offered huge contributions to our current knowledge about the brain on a molecular and cellular level (Aston-Jones *et al.*, 1994; Aston-Jones *et al.*, 1997; Schultz *et al.*, 1997; Berridge and Waterhouse, 2003). However, these are also highly imperfect, particularly in the study of neurological disease; most neurological diseases and conditions display vastly different phenotypes in animal models compared with the human conditions, therefore we cannot reliably replicate disease presentations or treatments using animal models of disease.

When studying human participants, behaviour can be directly measured, but this alone gives very little useful information about the underlying biological mechanisms of both health and disease; in such studies, one can only record discrete sets of actions or decisions, or timing information related to these actions. Written psychological batteries pose similar issues, in that they mainly rely on self-report of behaviour and symptoms, which in itself is unreliable.

An alternative option is to build and apply computational models to either behavioural or neuroimaging data, and make inferences about the mechanisms underlying the causes of phenotypic differences. One of the main roles of computational modelling in neuroscience is to bridge this gap between behaviour, disease phenotype and indirect neuroimaging data, and the underlying neuronal mechanisms which result in such observable phenotypes. Computational methods may therefore be used to individually characterise elements of human behaviour and disease phenotype.

Even so, computational modelling comes with its own challenges and compromises. The brain itself is the most complex organ in the body, and yet highly complex computational models have proven to be problematic. One key compromise that computational neuroscientists must address is that of model accuracy and model complexity. Occam's razor, for example, which states that "entities should not be multiplied without necessity", is frequently applied in many computational models to avoid over-complex models, which can lead to model overfitting. An informative model must be complex enough so that enough components of the neural responses are captured, but must also be simple enough to avoid overparameterization, and so that the model can be applied to multiple datasets. Model overparameterization can result in overfitting, and the model therefore is useless when applied to neural data to which it has not been trained. It is vital to remember, when building and fitting computational models, that "All models are wrong, but some are useful" (Box, 1976): excessive complexity does not ensure increased accuracy. Every useful model starts with a theory describing what one is aiming to find out, and what the model aims to achieve. Strong theories, and good ideas about your hypotheses (and generally experience) form the vital ingredients necessary for building useful computational models.

Over the past decade, many different branches of computational psychiatry have developed, with a similar goal in mind. Individualised medicine is set to become increasingly developed and applied in healthcare settings, as many diseases present in greatly different ways across a population. Specifically, psychiatric conditions such as depression, anxiety or attention deficit hyperactivity disorder (ADHD), can express contrasting phenotypes in different patients. Currently, the method of diagnosis for such conditions is the Diagnostic and Statistical Manual of Mental Disorders (DSM), which describes categories of signs and symptoms for a large catalogue of psychiatric disorders and diseases, and the symptoms experienced by the patient are assessed in line with these discretely-defined disorders prior to receiving diagnosis (American Psychiatric Association, 2013). Once diagnosed, a patient may be prescribed with a *go-to* initial medication, treatment or therapy, and so begins a trial-and-error method of treatment. This is highly unsuitable for many patients, and this problem of a one-size-fits-all treatment is being increasingly recognised. This is because, while the DSM has provided a huge contribution to the understanding of psychosis, it fails to take into account the underlying neurobiological processes associated with psychiatric conditions that have nevertheless been studied extensively over recent years.

Computational psychiatry could play an extremely important role here, in that the symptoms experienced by the patient could be assessed on an individual level, and, using models developed on a range of behavioural and phenotypic profiles, an individualised treatment programme may be recommended. Such models may enable us to *infer* the putative causes of a disease, *given* the observable signs, symptoms, and measurable behaviour and/or neural signals, and therefore provide the patient with a more accurate, individualised treatment programme. Such a method of diagnosis and treatment would save the patient a great deal of time and potential trauma through the process of the currently-used trial-and-error method of psychiatric treatment.

To achieve this, one may make probabilistic statements about such underlying neuronal mechanisms or latent variables (a latent variable is an unknown or unobservable parameter/state, which can be inferred based on known or directly observable parameters/states), through the application of Bayesian inference. Bayesian inference is both a statistical method and a modelling framework, which can be used for modelling both behavioural and neurological datasets.

In contrast with frequentist methods, which only describe the probability of any given effect for a current experiment and do not update beliefs in light of new information, Bayesian methodologies apply prior and posterior information to infer the probabilities of outcomes. Using probability theory, with Bayesian inference one can infer the state of latent variables relevant to our investigation, given (often noisy) observed data and previously-acquired prior knowledge.

1.2 Bayesian Inference

Bayes' Theorem, as initially proposed by Thomas Bayes (Bayes, 1763), describes the posterior probability of an event (or disease), as a result of two conditions (or phenotypes) that may be associated with this event (or disease). Bayes Theorem is as follows:

$$p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E)} \qquad \textit{Eq. 1.1}$$

Here, we are looking to find out the probability p of any given hypothesis, H, given some evidence, E. The parameters H and E refer to true or false statements, which could represent the presence or absence of a disease. For example: in the context of clinical diagnosis, one may wish to calculate the probability that a patient is suffering from Pernicious anaemia (PA), given that the patient has produced a significant blood test result, i.e. the blood test has detected low levels of intrinsic factor in the blood. Bayes Theorem is particularly useful due to its ability to estimate the probability that H is true, given that we already know that E is true. $p(E|H)$ is known as the *likelihood*; in this example, the likelihood would be the probability of a patient producing a significant blood test result given that we already know that they suffer from PA, i.e. the true positive test rate. $p(H)$ represents our prior knowledge of our hypothesis, in this case, the probability that any individual is suffering from PA – the incidence of PA in the general population. This is known as a *prior probability distribution*, or simply the *prior*. Finally, $p(E)$ represents the *evidence*, here, the probability of a significant test result. Once we have observed the evidence, $p(E)$, using Bayes Theorem we can then estimate $p(H|E)$, known as the *posterior probability distribution*, or the *posterior,* which forms a compromise between the prior distribution and observed evidence (**Figure 1.1**). Bayes Theorem can also be extended to the following form:

$$p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E|H) \cdot p(H) + p(E|\neg H) \cdot p(\neg H)} \qquad \textit{Eq. 1.2}$$

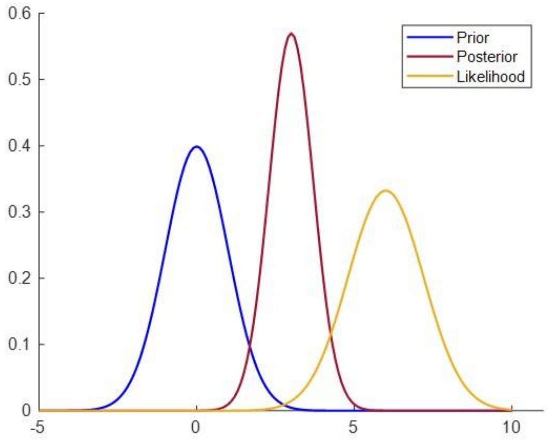
where $p(E|\neg H)$ denotes the probability of receiving a false positive blood test result (receiving a significant test result when the patient does not have PA).

An important benefit of using Bayesian inference is that as we acquire information over time, we can continue to update our statements about our hypotheses, given new observations or evidence. Through repeated observation, we may continue to acquire evidence regarding the latent variables and make better predictions in the future.

A further benefit of using Bayesian inference is the ability to compare models of the same data to evaluate which model best fits the data we are analysing. This is can be done using a generic approach known as Bayesian Model Selection (BMS). BMS is used to assess the 'goodness of fit' of any given models, i.e. how well the model explains the data, or which model has the highest 'model evidence' (highest probability of the data, given the model). One drawback of BMS, however, is that the data for which the competing models are compared must be identical; BMS cannot be used to compare models that are applied to different datasets.

Over recent years, a range of computational tools have been developed to tackle the issues outlined above, some of which can be applied to behavioural and neuroimaging data. In this chapter, I will outline a selection of these tools, which range from biophysical models which act upon the microscale molecular level of neuronal dynamics, to models of Bayesian inference, which model macroscale, system-level and behavioural-level data. Examples of Bayesian modelling methods include Active Inference and Dynamic Causal Modelling (DCM), both of which are discussed below. Whereas Active Inference is used to measure the algorithmic content of brains, DCM is used to measure brain data directly.

Figure 1.1



$$p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E)}$$

$$posterior = \frac{likelihood \cdot prior}{evidence}$$

Figure 1.1 Bayesian Inference and Posterior Probability Distributions. In order to estimate the posterior probability distribution (maroon), one must apply the prior probability distribution (blue), i.e. one’s prior knowledge or previous experiences of the hypothesis in question, and the likelihood (orange) to Bayes’ Theorem, along with the evidence.

1.3 Model-Free and Model-Based Reinforcement Learning

Reinforcement learning (RL) is a computational approach to the learning and application of optimal action in order to maximise reward. In RL, an animal or artificial learning agent evaluates actions based on the rewards that are expected to result. The main aim of a RL agent is to select an action or set of actions within a dynamic environment, with the goal of maximising reward (Sutton and Barto, 2018). At the core of RL and such decision-making is Thorndike's 'Law of Effect' (1911), which states that any action which is followed by a positive response or reinforcement is more likely to be repeated by the animal in future decisions (Thorndike, 1912). RL consists of two main categories of algorithm: *model-free* and *model-based* learning (Daw *et al.*, 2005). In model-free learning, an animal or artificial agent learns the values of particular actions directly and evaluates reward mappings retrospectively, without building a model of the environment. The values of situations or states (i.e. specific locations within the task or environment) are learned via trial and error, and subsequent actions are selected solely to maximise the reward earned at the next location. Model-based learning, however, acts by building a learned 'internal model' of the environment, and uses this model to assess actions. This approach, in contrast to model-free learning, acts prospectively, using information acquired from previous experience to determine the availability of future reward possibilities (Daw *et al.*, 2005; Gläscher *et al.*, 2010; Daw *et al.*, 2011).

An example of a model-free RL application is temporal difference (TD) learning, in which an animal can learn the dynamics of the task environment directly through experience, without applying an internal model. In a landmark study, Schultz *et al.* (1997) presented dopaminergic firing patterns in the monkey brain, in response to reward and conditioned stimuli (Schultz *et al.*, 1997). Schultz *et al*. used a TD learning model in combination with primate recordings to demonstrate that dopaminergic activity in the ventral tegmental area (VTA) occurs immediately following reward prior to training, and after training, a similar dopaminergic firing pattern occurs immediately following the conditioned stimulus rather than after the presentation of reward. Conversely, when the conditioned stimulus is presented to the monkey to indicate forthcoming reward but the reward does not follow, this was shown to result in a depression of dopaminergic activity at the very point when the reward would normally have been presented. This suggested that this pattern of dopaminergic firing in the VTA represented a *reward prediction error* (RPE).

Another example of a model-free reinforcement learning approach is Q-learning: a TD control algorithm initially developed by Watkins in 1989 (Watkins, 1989). This method, where Q stands for ‘quality’, is an off-policy values-based learning algorithm in which the value function is updated, commonly using standard Bellman equations which require the current state (*s*) and action (*a*) as inputs. ‘Off-policy’ refers to the agent’s learning of the optimal policy, which occurs independently of the agent’s action selection. Q*(*s*,*a*) represents the expected value, or cumulative discounted reward, of an agent taking action *a* when occupying state *s*. In Q-learning, the agent’s main aim is to maximise its expected cumulative discounted reward by exploring the environment to learn the value of the optimal policy. To do this, the agent continually updates its internal value of Q(*s*,*a*), without learning the transition model, i.e. the probability of reaching any particular state after taking action *a* at state *s*. Q(*s*,*a*) is the agent’s most up-to-date estimate of Q*(*s*,*a*). The agent then selects future actions based on its updated value of Q. If the agent were to continue to perform every action and experience every state within the task space, the agent’s estimate of Q(s,a) would eventually converge towards Q*(*s*,*a*) for all available states and actions (Watkins and Dayan, 1992).

In later studies, a Bayesian approach to Q-learning was proposed. This approach introduces the specification of prior probability distributions over Q values as opposed to point estimates, which allows the agent to update these priors based on past experiences of the environment and therefore develop more informed estimates of Q. Dearden *et al*. (1998) examined four different Q-learning algorithms, including Bayesian Q-learning, in three different task domains, and found that the Bayesian Q-learning algorithm consistently outperformed the other conventional RL algorithms tested (Dearden *et al.*, 1998). This method, however, proved to be more computationally expensive due to this Q-learning algorithm using more prior information to update Q values and for the agent to select actions (Dearden *et al.*, 1998). A more recent study used a variation of deep Q-learning in the context of Atari 2600 computer games, and was able to successfully improve upon previous RL attempts for six out of seven games, also achieving scores higher than the current human expert on three games (Mnih *et al.*, 2013).

A notable RL study examined the interaction between model-free and model-based learning mechanisms in human decision-making, using fMRI (Daw *et al.*, 2011). Previous studies had suggested that model-free and model-based evaluation employed different networks of brain activity, specifically ventral striatum activation during generation of model-free prediction errors (McClure *et al.*, 2003; O'Doherty *et al.*, 2003), and medial prefrontal cortex in model-based learning (Hampton *et al.*, 2006). Daw *et al*. used a two-stage decision-making paradigm to examine human action choices combined with BOLD signalling, and found that the human behaviour reflects aspects of both model-free and model-based learning hallmarks (**Figure 1.2**), in addition to the recruitment of both brain networks during decision making, combining both model-free and model-based evaluation methods in both brain networks, indicating that humans employ both strategies simultaneously during decision making.

Figure 1.2

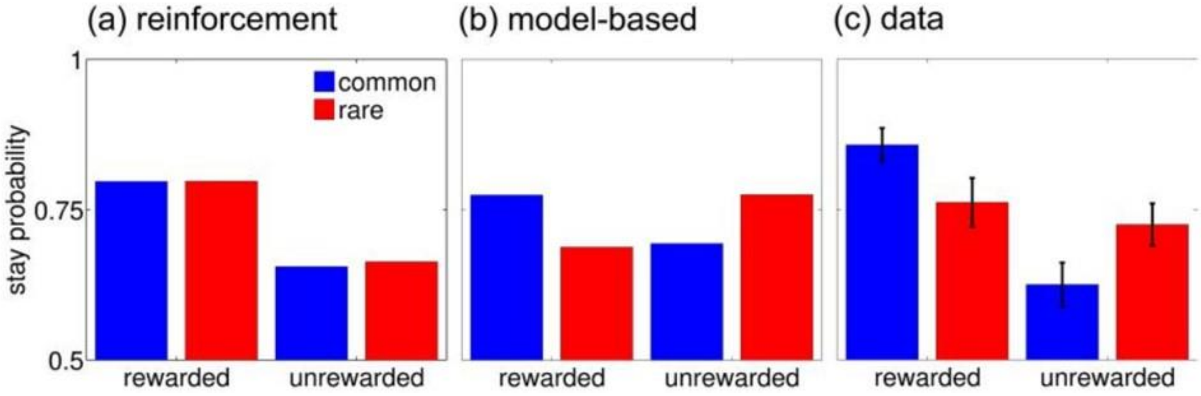


Figure 1.2 Model-Free Versus Model-Based Learning. Taken from Daw *et al*. (2011). Human behaviour in a two-stage decision-making task reflected aspects of both model-free reinforcement and model-based learning. In panel (**A**), the model-free learner would repeat any first action, if selecting that action on the previous trial resulted in reward, regardless of whether this reward was associated with a common or rare transition probability. On the other hand, in (**B**), a model-based learner would predict interactions between reward and transition probability. (**C**) Daw *et al*. found in their paradigm that human participants exhibited the signatures of both approaches (Daw *et al.*, 2011).

1.4 Reversal Learning, Noradrenaline, and Uncertainty

Reversal learning has been broadly studied as a learning process in which animals or humans must adapt their behaviour in response to changes in the environment structure, or in stimulus-reward mappings. For example, in a Go/No-go task, there may be a 90% chance of an animal receiving a food reward if a particular tone is heard, and a 90% of no reward if a different tone is heard. Prior to a contextual reversal, the animal may believe that the stimulus-reward contingencies in the task are relatively stable. However, a reversal in cue meanings, in that the tone previously associated with reward switches to become associated with no reward and vice versa, would contravene the animal's beliefs about the environment. Thus, the reversal may be referred to as an *unexpected* uncertainty (Yu and Dayan, 2005).

Yu and Dayan (2005) postulated that two distinct forms of uncertainty, *expected* and *unexpected* uncertainty, are neurobiologically represented by the neurotransmitters, acetylcholine (ACh) and noradrenaline (NA). Yu and Dayan suggested that ACh signals *expected* uncertainty; uncertainty about predictable unreliability in a known environment. For example, a probabilistic decision-making task in which the probability of receiving reward as a result of taking a specific set of actions varies, but in a consistent and predictable way, is considered *expected* uncertainty; the task contains an element of known consistent unreliability or lack of precision. *Unexpected* uncertainty, on the other hand, refers to an unpredictable, unsignalled alteration in context or task structure that generates observations that are largely unexpected, such as the reversal example described above. In response to such a large change in the environment without warning, the animal must adapt to the changes by rebuilding its internal model and beliefs about the environment. Yu and Dayan postulated that NA encodes this *unexpected* uncertainty.

A more recent study by Parr and Friston builds on this, by generating a simulation of epistemic foraging in a volatile environment, applying the Free Energy Principle to their simulations (Parr and Friston, 2017). In their model, they introduce parameters which encode precisions over state transitions and attentional gain. These parameters were introduced to signal NA and ACh, respectively, thus representing these two distinct forms of uncertainty.

NA, a catecholamine neuromodulator, is crucial for arousal, memory modulation, decision-making and executive function (Sara, 2009). The major noradrenergic nucleus in the brain is the Locus Coeruleus (LC) in the pons, which possesses extensive projections throughout the brain and is responsible for the vast majority of NA signalling across the brain, from the neocortex to the spinal cord (Ramos and Arnsten, 2007). An important aspect of LC-NA function is the two distinct firing patterns exhibited by the LC: phasic and tonic firing. In phasic firing of the LC, phasic bursts of activity are initiated in response to task-relevant salient stimuli and decision-related outcomes, which act as a catalyst for short-term behavioural changes and task performance optimisation. Conversely, the tonic firing mode of the LC is strongly linked to global behavioural flexibility and arousal, and is also associated with exploration, which occurs in response to reduced utility in a task context (Aston-Jones and Cohen, 2005). The LC applies both modes of activity in order to optimise behavioural performance.

Optimal behavioural performance may be achieved through an intermediate level of arousal, in combination with task-relevant phasic LC activity. This optimal firing pattern can be described in relation to the classic Yerkes-Dodson inverted-U shaped curve (**Figure 1.3**). The Yerkes-Dodson inverted-U was originally composed in 1908 to describe the empirical relationship between task performance, generally in difficult cognitive tasks, and levels of arousal (Yerkes and Dodson, 1908). They proposed the theory that performance improves with mental or physiological arousal, up to a certain point. Once arousal exceeds this threshold, performance level would start to decrease, due to divided attention and anxiety. Different tasks may even require different levels of arousal in order to achieve optimal performance, therefore different cognitive paradigms may be described by different inverted-U shaped functions, shifted with respect to the task at hand.

Figure 1.3

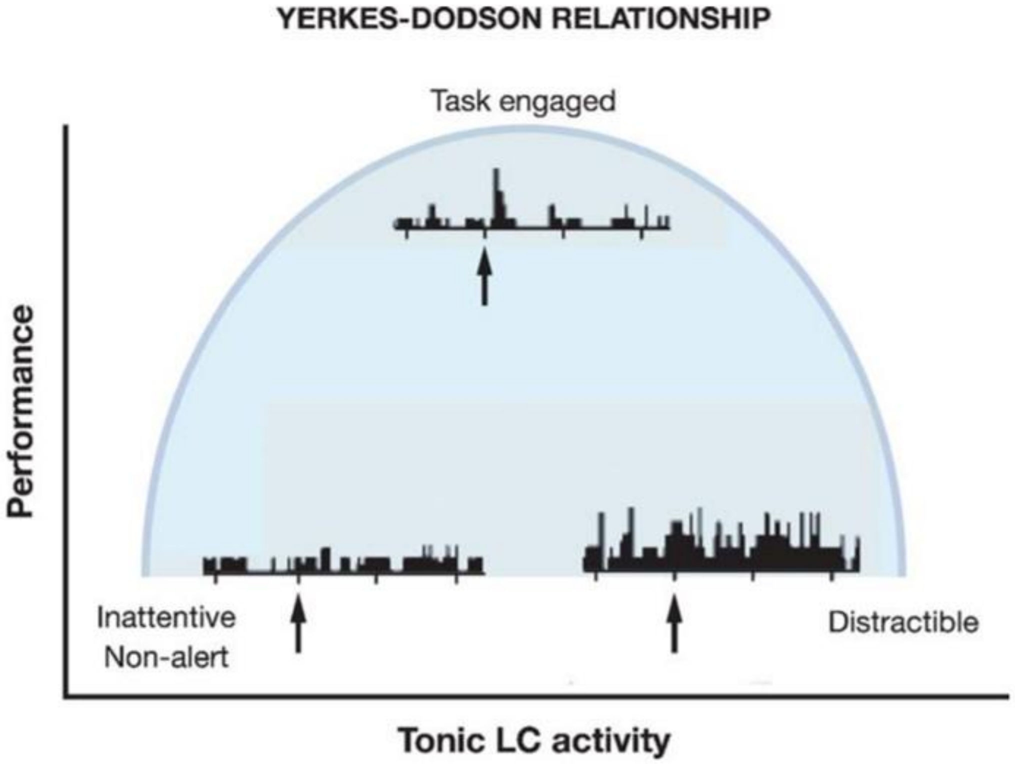


Figure 1.3 Yerkes-Dodson Inverted-U Relationship. Taken from (Aston-Jones and Cohen, 2005). The relationship between tonic LC activity and task performance reflects that of the classic Yerkes-Dodson inverted-U, in that a moderate/intermediate level of tonic LC firing yields optimal performance, with prominent phasic bursts of LC activity in response to task-related stimuli, i.e. during phasic LC firing mode. If tonic LC firing levels are too low, animals may be inattentive, non-alert and drowsy. In contrast, if tonic LC firing levels are too high, animals may become anxious or distractable, and unable to focus on the task at hand, and therefore display poor performance.

This curve can also be applied to the firing patterns of the LC. At one end of the curve, very low levels of tonic firing in the absence of phasic activity results in lack of sufficient arousal, and ultimately drowsiness. At the other far end of the curve, excessive tonic activity can drown out the effects of task-related phasic firing, leading to heightened distractibility and levels of arousal that are too high to lend themselves to optimal behavioural performance. A moderate level of tonic activity, combined with large phasic spikes in response to task-related stimuli or events, is suggested to promote optimal performance, and therefore this phasic-tonic trade-off must be modulated to maintain high levels of cognitive functioning.

A modulatory role for adrenoceptors has also been widely studied in the relationship between NA release and neuronal transmission. Adrenoceptors are G protein-coupled membrane-bound adrenergic receptors, which are subdivided into three main receptor subtypes: alpha-1 adrenoceptor (α_1), alpha-2 adrenoceptors (α_2) and beta adrenoceptors (β). Many studies have examined the modulatory effect of adrenoceptors through the use of pharmacological manipulation. For example, a study by Winder-Rhodes *et al*. investigated the involvement of α_1 adrenoceptors in the cognitive effects of modafinil, a wake-promoting medication which weakly inhibits dopamine (DA) reuptake, in human participants (Winder-Rhodes *et al.*, 2010). By applying modafinil and prazosin, an α_1 adrenoceptor antagonist which also has high affinity for α_2 adrenoceptors, in a randomised control trial, they found that performance in cognitive tasks which examined executive function and working memory were enhanced by modafinil, but this cognitive enhancement was subsequently blocked by prazosin. This highlights the importance of modulation of NA and adrenoceptors in arousal and attention.

The LC also receives strong cortical projections, particularly from the anterior cingulate cortex (ACC) and the orbitofrontal cortex (OFC). The OFC has been demonstrated to show significant activations during reversal learning, in both animal studies and in human neuroimaging (Rolls, 1999; Amodeo *et al.*, 2017). The OFC, in addition to the ACC, has major projections to the LC, and has been reported to play key roles in the evaluation of reward, in both reward anticipation and reward delivery, and goal direction (O'Doherty *et al.*, 2001; Aston-Jones and Cohen, 2005; Valentin *et al.*, 2007). A reversal learning study by Rygula *et al.* (2010) examined activations in the OFC, alongside the ventrolateral prefrontal cortex (VLPFC) in marmosets, in the context of a serial reversal learning paradigm (Rygula *et al.*, 2010). In animals with lesions in the OFC, there was a deficit in reversal performance, in that OFC-lesioned monkeys displayed slower improvement of reversal learning performance to reach pre-surgery levels compared with those that had undergone VLPFC lesions only.

A key aspect of this study is that the monkeys examined here were previously trained in reversal learning paradigms pre-surgery, and therefore the reversals that occurred in the task were an expected, rather than unexpected uncertainty. Another study applied this principle: Costa *et al*. (2015) formed a Bayesian analysis method to examine reversal learning in rhesus monkeys under conditions of dopaminergic agonism (L-DOPA), or antagonism (haloperidol) (Costa *et al.*, 2015). The monkeys had, again, experienced extensive reversal learning training and the reversal was therefore an expected uncertainty in the task. It was found in this study that administration of haloperidol resulted in a greater reliance of the animals on their prior beliefs about the occurrence of a reversal, while administration of either drug manipulation resulting in increased performance across the task, including abrupt changes in choice behaviour in response to the reversals rather than gradual behavioural changes (Costa *et al.*, 2015). Another study combined expected and unexpected uncertainty in a reversal learning task, which also gave negative feedback on rare trials to promote perseverative behaviours (Cools *et al.*, 2002). Perseveration describes the persistence of a behaviour, even in the absence of reward. This study used event-related fMRI to examine the importance of a ventral frontostriatal network in reversal learning, and found that their probabilistic reversal learning paradigm employed these regions, in line with previous studies (Zald and Pardo, 1997; O'Doherty *et al.*, 2001).

1.5 Active Inference

Active Inference is a behavioural theory, based on the premise that all self-organising organisms (animals, humans, or artificial agents) always act to minimise variational free energy. Free energy mathematically describes the difference between the agent's generative model of the world and the true or 'real' state of the world. Under the free energy principle, any agent or self-organising system must always act to minimise its free energy in order to maintain equilibrium with its (dynamic) environment, thus minimising surprise. Surprise is defined as the negative log-probability of any outcome, and free energy provides a bound on log-evidence for any model (Friston *et al.*, 2007; Friston, 2010). In contrast to RL techniques, an agent is not predominantly driven to maximise reward; under Active Inference, the agent instead aims to stay in or navigate to states which minimise uncertainty. It does this by constructing a generative (internal) model of the world, as a Partially Observable Markov Decision Process (POMDP) – it aims to represent the true state of the world, and infers information about the world by making predictions about consequences of actions, then subsequently updating those predictions according to real-world outcomes, and whether the sensory information received was as expected. This continued updating of the agent's generative model should, if the environment is relatively stable, reduce its uncertainty about the environment, and thus, lower free energy. An agent can actively minimise free energy by either selecting actions that it believes will result in unsurprising states, or make adjustments to its generative model according to new sensory information (Friston *et al.*, 2016).

In Active Inference, an agent may make an observation, i.e. obtain some sensory information acquired from the environment in a particular state. Here, a state is defined as a location or set of contextual features relevant to the agent for its selection of choice behaviour. Using this sensory information, the agent will then update its inferences over state probabilities for each available policy (policies: sets of multiple actions selected by the agent to reach a desired outcome) for each time point. Based on this information, the agent then calculates the past free energy and expected future free energy for each policy, and updates its precision and policy probabilities. All of these computations aim to minimise free energy, and so the agent uses these probabilities to calculate a Bayesian Model Average (BMA) over states, then selects its next action based on the current BMA.

A vast number of studies have been conducted to apply Active Inference to simulate the behaviour of artificial agents in various decision-making or foraging contexts (Mirza *et al.*, 2016; Cullen *et al.*, 2018; Kaplan and Friston, 2018; Mirza *et al.*, 2019; Sales *et al.*, 2019). A recent study employed an OpenAI Gym paradigm to directly compare Active Inference with the off-policy RL method of Q-learning, and also to a Bayesian RL agent (Sajid *et al.*, 2021). Sajid *et al*. demonstrated that the Active Inference agents were able to conduct epistemic exploration of the task (or foraging) in a Bayes-optimal fashion, which under Active Inference appears to emerge naturally, and they do not significantly rely on an explicit reward signal, in contrast to model-free RL. Both the Active Inference agent and Bayesian RL agent were able to engage in information-seeking behaviour following the removal of explicit reward signals (Sajid *et al.*, 2021).

1.6 Dynamic Causal Modelling

Initially introduced by Karl Friston in 2003 as an fMRI analysis method (Friston *et al.*, 2003), DCM is a generic Bayesian approach for the inference of (hidden) neuronal states, from recordings of brain activity (here, I use the term DCM to refer to both DCM as a computational method, and the specific dynamic causal models themselves) (Stephan *et al.*, 2010). Since its inception, DCM has been extensively developed and can now be applied to multiple imaging modalities, such as MEG, EEG, and local field potential (LFP) recordings (Kiebel *et al.*, 2006; Garrido *et al.*, 2008; Kiebel *et al.*, 2009; Moran *et al.*, 2009). A key benefit of using DCM is that DCMs aim for neurophysiological plausibility. Alternative methods for analysing patterns of activity in neuroimaging data on a large scale, such as Hidden Markov Models (HMMs) which can be used effectively to model network dynamics over the whole brain and how these fluctuate over time (Vidaurre *et al.*, 2017), do not specifically tap into biophysical dynamics at the meso-scale or micro-scale. DCM employs dynamic (linear or non-linear) differential equations, such as Morris-Lecar type non-linear differential equations to describe synaptic dynamics in conductance-based neural mass models (Moran *et al.*, 2013), which enable the investigation of biophysical parameters on a per-subject basis.

A further benefit of DCM is that it estimates effective connectivity. There are three major branches of connectivity in functional imaging analyses: structural connectivity, functional connectivity, and effective connectivity. Structural connectivity describes the physical connections between brain regions or sources of activity, i.e. the interconnection of regions by white matter tracts, and may be investigated using diffusion weighted imaging. Functional connectivity, on the other hand, describes the statistical covariation or correlation of activity between discrete sources in the brain, obtained through fMRI (Greicius *et al.*, 2009; Uddin, 2013). The third form, effective connectivity, is defined as the causal influences between neuronal populations. Effective connectivity takes one step further than functional connectivity, as it describes the contextual influence that one region has over another region, such that connections can be interrogated in a context-dependent or task-dependent fashion. This is particularly beneficial for task-based imaging studies, in which the effect of one task condition compared to an alternative task condition, or a comparison between a patient group and a healthy control group, may be a vital aspect of data interpretation. Therefore, rather than being limited to simply asking questions about the strengths of activity in particular sources, with DCM one can investigate specific hypotheses about the activations between brain regions in a predefined network, relevant to one's specific task, or even in the context of a neurodegenerative disease.

Through DCM, one can then invert the generative model of sources of brain activity according to a Variational Bayesian scheme, to examine the likelihood of parameters in the model, given the data and model. This inversion approximates the posterior probability: $p(\theta|y,m)$. As detailed above, this represents the probability p of the parameters θ, given the data y and the model m (**Figure 1.4**).

A further key benefit of using DCM is that Bayesian inversion yields an approximation to the log model evidence, which can be used to compare alternative models of the same data to make statistical inferences about which model best represents the data, i.e. has the highest model evidence, through BMS described above. However, in comparison to alternative conventional neuroimaging analysis methods, DCM is relatively complex, and requires some understanding of model selection, model inversion, and Bayesian statistics.

Figure 1.4

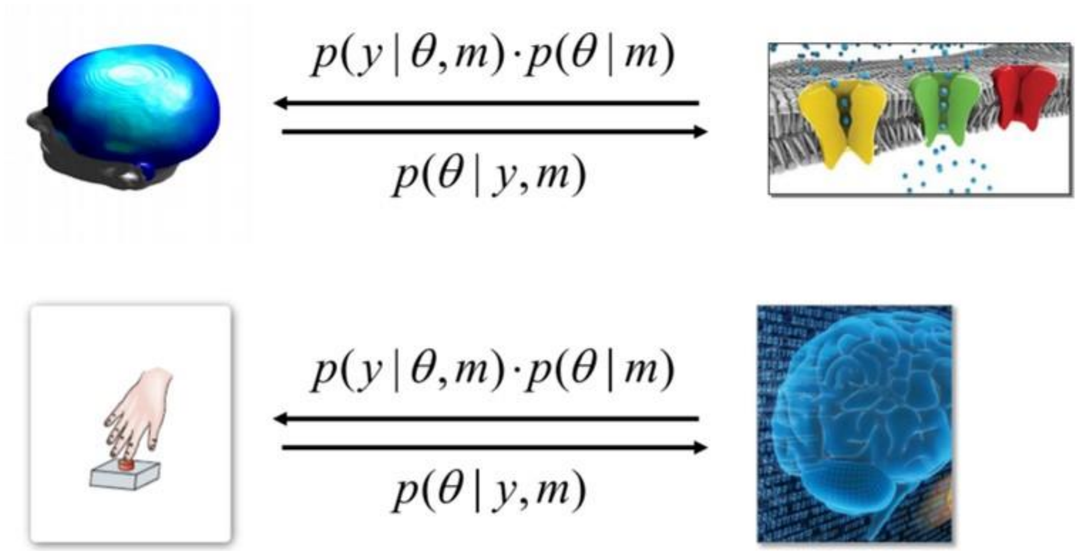

$p(y \mid \theta, m) \cdot p(\theta \mid m)$
$p(\theta \mid y, m)$
$p(y \mid \theta, m) \cdot p(\theta \mid m)$
$p(\theta \mid y, m)$

Figure 1.4 Mapping from neural data to synaptic dynamics using DCM. Adapted from (Stephan, 2017). DCMs can be used to make inferences about the underlying neuronal mechanisms which explain a behavioural phenotype or disease state. In the inverse model, $p(\theta|y, m)$, one can infer the probability of model parameters θ from neuroimaging data or sets of actions from a behavioural paradigm, given the data y and the model m, which may describe microscale synaptic dynamics. In the forward model, one can estimate the data y that may be produced as a result of some given model m and set of parameters θ.

Recent work by Brodersen *et al.* demonstrates how DCM can be used in psychiatric diagnostics, in patients suffering from schizophrenia (Brodersen *et al.*, 2014). In this study, DCM was combined with generative embedding to characterise patients with schizophrenia and healthy controls, based on task-based fMRI data. It was found that when exclusively analysing patient data, three distinct subgroups could be characterised, which mirrored clinical subgroups as defined by cognitive assessment of negative symptoms (via the Positive and Negative Syndrome Scale (PANSS)). Earlier work by the same group examined this method of combining DCM with generative embedding, looking specifically at speech processing networks to classify aphasic patients following a stroke, and healthy controls (Brodersen *et al.*, 2011). Such modelling approaches may be invaluable in the characterisation of psychiatric spectrum diseases such as schizophrenia, to advance further in more specific diagnoses than those obtained through solely symptom-based diagnosis.

Similar statistical classification efforts in MRI have been implemented to predict the likelihood that individuals suffering from Mild Cognitive Impairment (MCI) will go on to develop Alzheimer's disease within a particular time frame (Davatzikos *et al.*, 2008; Lehmann *et al.*, 2012). Alzheimer's disease is the leading cause of dementia in aged adults (Zhang *et al.*, 2016), and despite extensive research into the neurobiological mechanisms of this disease since its discovery over 100 years ago, an effective treatment or cure is still out of reach. The main histopathological signatures of Alzheimer's disease consist of extracellular amyloid-beta (Aβ) aggregates and intracellular hyperphosphorylated tau neurofibrillary tangles (Buckner *et al.*, 2005), with depositions of Aβ appearing to be broadly distributed across the brain, in contrast to tau pathology, which originates in the entorhinal cortex, then progresses outwards to other brain structures as the disease progresses (Marks *et al.*, 2017; Pasquini *et al.*, 2019). Alzheimer's disease is therefore a prime candidate for investigation using computational modelling, in combination with prior knowledge of molecular mechanisms involving tau and Aβ and neuroimaging data; potentially the application of DCM to infer how effective connectivity between disease-relevant brain regions alter as a result of neurodegenerative disease, or how synaptic dynamics differ between patients suffering from Alzheimer's disease and healthy aged controls.

Overall, at the top level (macroscale), a patient may have to undergo EEG recording while completing a behavioural task, perhaps for diagnostic purposes. The recording electrodes record the scalp-level activity transmitted from neuronal sources, and using DCM, one can interpolate between sensors or electrodes in order to predict summed local circuit currents. Finally, at the lowest level (meso/microscale), synaptic activity can be described using dynamic equations such as those described above, which can be used to infer the biophysical parameters of synaptic currents using neurobiologically-interpretable mathematical models.

1.7 Summary

In summary, computational methods and models such as those described above may provide highly valuable insights into latent neuronal dynamics which underlie observable and measurable behaviour or neural signals. A wide range of recent work has demonstrated the capacity of computational models to provide inferences about hidden states and mechanisms in the brain, ranging from those that can assign subject-specific parameters to characteristics of behaviour, to those that are able to classify subgroups of patients suffering from psychiatric disease or brain injury, both in comparison to healthy controls and within patient groups themselves.

In this thesis, I aim to contribute to this bridging of the gap between human behaviour and neuroimaging data, and the underlying neural dynamics which results in various behavioural phenotypes, through the use of Bayesian computational modelling techniques; namely, Active Inference and DCM, in both health and disease.

To examine behaviours in response to *expected* and *unexpected* uncertainty, I ran four behavioural studies examining optimal decision-making and exploration/exploitation behaviours (*Chapter Two*), also with the goal of fine-tuning a probabilistic Markov decision task to be used in subsequent experiments. By conducting these studies I aimed to optimise a probabilistic decision-making paradigm, with the addition of a task reversal. I also aimed to investigate task performance based on broad behavioural metrics on an individual level, and consider whether the behaviours exhibited by participants could be explained by *model-free* or *model-based* learning methods.

Using the finalised task structure, I constructed an Active Inference model of this task, and generated simulated task behaviour (*Chapter Three*). In this study, I aimed to estimate subject-specific parameters signalling precision over reward (how sensitive a participant is to rewarding outcomes) and internal model volatility (model flexibility, i.e. how much a participant relies on their prior beliefs about the world) through the inversion of a generative model. I also aimed to examine how these parameter estimates might predict broad behavioural measures of task performance on the group level. I used this pipeline to optimise model parameters to produce highly-rewarding 'optimal' behaviour in the task, then conducted a model inversion of simulated data to produce conditional estimates of three model parameters which signalled precision over rewarding states and internal model flexibility. I then used this model inversion scheme to invert the behavioural data of human participants from the main behavioural study to generate subject-specific conditional Maximum *a posteriori* (MAP) estimates of the three model parameters described above, in order to phenotype individuals based on their model volatility/flexibility and precision over reward, and examine how these parameters may describe broad measures of behavioural performance.

Building upon this line of investigation, I conducted an fMRI study using the same probabilistic decision-making paradigm, with the addition of a pharmacological manipulation – specifically selective NA reuptake inhibition using reboxetine, to delve into the neurobiological mechanisms underlying model flexibility, reward, and decision-making (*Chapter Four*). In this study, I aimed to investigate how selective NA reuptake inhibition influenced belief updating in the decision-making paradigm and how this might affect participants' responses to a contextual reversal.

These investigations were all conducted from the perspective of neuronal dynamics and decision-making in the (young) healthy brain. However, in order to really gain insights into hidden neuronal mechanisms underlying vital processes in the brain, we must also consider these processes in the context of disease, and examine aberrant pathways as a result of brain dysfunction, particularly in neurodegeneration. Also, thus far, I have used computational models to gain insights into cognition by modelling *behaviour*. It is also vital to consider the modelling of *biological circuits* of cognition in order to understand the mechanisms underlying pathology. Therefore, finally I conducted analysis of task-based EEG data of Alzheimer’s disease patients and healthy aged controls, using DCM (*Chapter Five*). Data was collected during the execution of visual priming and recognition tasks; a paradigm that taps into visual memory networks which are compromised in Alzheimer’s disease and other forms of dementia.

In this thesis, I utilise multiple neuroimaging modalities and combine these with multiple computational methods, to offer valuable insights into neural dynamics in both healthy participants and Alzheimer’s disease patients.

Chapter Two

Examining Unexpected Uncertainty in Probabilistic Decision Making: A New Model for Reversal Learning

2.1 Introduction

Reversal learning is a behavioural process that has been widely studied, which involves the inhibition of previously rewarded actions, and the relearning of stimulus-reward contingencies. Reversals in the context of a decision-making task refer to the change or switch of stimulus-reward or state-reward mappings during a task, which requires the animal or human participant to learn an opposite, previously irrelevant state-reward mapping. Task reversals, in contrast to standard probabilistic cueing, occur without prior warning and require the participant to shift their attention to alternative cues in the task environment to relocate the rewarding stimulus or state. Such reversals are an example of *unexpected* uncertainty, whereas probabilistic cueing represents a form of *expected* uncertainty (Yu and Dayan, 2005). Yu and Dayan postulated that these distinct forms of uncertainty may be biologically represented by the neuromodulators ACh and NA, where ACh signals *expected* uncertainty, and NA encodes *unexpected* uncertainty. Parr and Friston examined this proposal further, through simulations of epistemic foraging in the context of Active Inference (Parr and Friston, 2017). The model described by Parr and Friston links these neuromodulators to precision over beliefs about state transitions (NA), and beliefs about outcomes (ACh), by modelling the effects of NA as an inverse volatility parameter and ACh as precision over attentional gain (Parr and Friston, 2017).

The primary source of NA in the brain is the LC. The LC fires in a phasic fashion; the LC responds to behaviourally-relevant stimuli, including highly unexpected stimuli, with high-frequency (10-15 Hz) bursts of activity (Aston-Jones *et al.*, 1994; Dayan and Yu, 2006; Kane *et al.*, 2017). The LC also fires in a tonic firing pattern (2-6 Hz) with high levels of spontaneous activity (Kane *et al.*, 2017), which is associated with behavioural flexibility and arousal, and is positively correlated with levels of alertness (Rajkowski *et al.*, 1994; Berridge and Waterhouse, 2003). The firing patterns of the LC have also been linked to specific behavioural characteristics, in that phasic firing of the LC has been known to be associated with exploitative behaviours due to its activation in response to task-relevant processes, whereas tonic LC firing is more strongly associated with exploratory behaviours, which may be triggered as a result of experiencing a task reversal (Aston-Jones and Cohen, 2005).

A recent study by Sales *et al*. probed both firing patterns in the LC, simulating the behaviour of a synthetic agent in an explore/exploit paradigm with contextual reversals in the location of rewards, also in the context of Active Inference (Sales *et al.*, 2019). After the artificial agent had acquired sufficient information about the task environment to build strong prior probabilities on reward availability in particular locations, the reversal occurred, after which increased tonic LC activity and the generation of state-action prediction errors (SAPEs) were observed. SAPEs occur as a result of a substantial difference between the expected and actual outcome of an action, and therefore the agent experiences a significant change in its beliefs about its past and future states (Sales *et al.*, 2019).

Cools *et al*. (2002) also utilised a reversal learning paradigm, using a probabilistic decision task to probe the importance of a ventral frontostriatal network in reversal learning (Cools *et al.*, 2002). This task combined reversal learning with a probabilistic structure, in which participants rarely received negative feedback in response to 'correct' responses, independent of the task reversals. Such a design was used to promote perseverative behaviours following reversals, meaning that participants may be inclined to persist with their previous strategies for obtaining reward, even after stimulus-reward contingencies changed.

In this chapter, I present four behavioural studies (three pilot studies and one main study) in which I examined SAPEs in a navigational decision-making task, using a behavioural paradigm similar to that used by Gläscher *et al*. (Gläscher *et al.*, 2010). Gläscher *et al*. used a probabilistic Markov decision task combined with computational models of RL and fMRI to identify neural signatures of prediction errors. They focused on teasing apart BOLD signals correlated with *model-free* reward prediction errors and *model-based* state prediction errors, and found trial-by-trial neural correlates of state prediction errors in lateral prefrontal cortex and posterior intraparietal sulcus, alongside correlates of reward prediction errors in the ventral striatum, as previously identified (McClure *et al.*, 2003; O'Doherty *et al.*, 2003).

Gläscher *et al*. also used three different computational models of model-free and model-based RL: the model-free SARSA (state-action-reward-state-action) learner, the model-based FORWARD learner, and a HYBRID model which combined both model-free and model-based learning methods. They hypothesized that participants would be able to gain knowledge about the transition probabilities in the task prior to receiving information about reward, therefore acquiring information through model-based learning. Their results reflected that of Daw *et al*., who observed in a human fMRI study that participants employed aspects of both model-free and model-based learning methods in a two-step choice task (Daw *et al.*, 2011).

Gläscher *et al.* aimed to emulate the design of classical 'latent learning' animal studies which involved pre-training in the task environment, followed by the introduction of rewards into the environment to test if participants were able to employ new strategies to seek rewards, using their previously acquired knowledge of the environment from the training session. This initial unrewarded training session would reveal evidence of state prediction errors only, as no reward information was given until the rewarded testing session, thus examining model-based RL. However, contrasting such 'latent learning' animal studies, their experiment was non-spatial since abstract fractal images were used as visual stimuli. Here, I utilised Gläscher *et al*.'s probabilistic Markov decision paradigm (**Figure 2.1A**) but with an added spatial element, in that the images used in my study depicted spatial locations in nature to emulate navigation through a real-world environment.

Figure 2.1

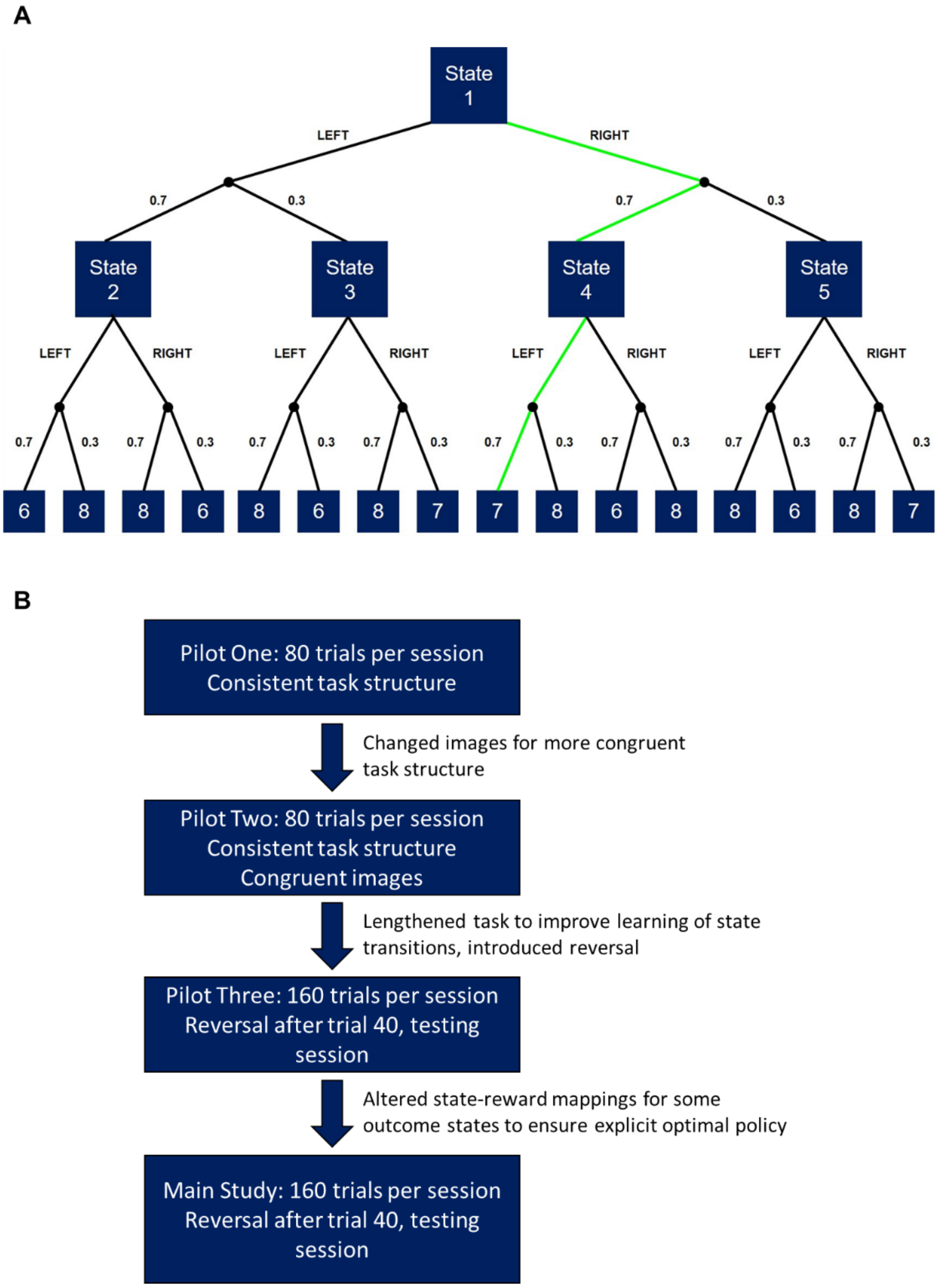

A
State 1
LEFT
RIGHT
0.7
0.3
0.7
0.3
State 2
State 3
State 4
State 5
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
0.7
0.3
0.7
0.3
0.7
0.3
0.7
0.3
0.7
0.3
0.7
0.3
0.7
0.3
0.7
0.3
6
8
8
6
8
6
8
7
7
8
6
8
8
6
8
7
B
Pilot One: 80 trials per session
Consistent task structure
Changed images for more congruent task structure
Pilot Two: 80 trials per session
Consistent task structure
Congruent images
Lengthened task to improve learning of state transitions, introduced reversal
Pilot Three: 160 trials per session
Reversal after trial 40, testing session
Altered state-reward mappings for some outcome states to ensure explicit optimal policy
Main Study: 160 trials per session
Reversal after trial 40, testing session

Figure 2.1 Task and Study Structures. (**A**) State structure of the decision-making task, with the optimal policy highlighted in green. A *policy* is defined as a set of multiple actions selected by the participant to reach a desired state or outcome. (**B**) Summary of full set of behavioural studies, outlining the changes made between studies.

This chapter consists of four main sections, each describing the methods and results of four different behavioural studies (three pilot studies and one main study, see **Figure 2.1B**). I made modifications to each study over time with the aim of increasing participants' ability to select optimal routes, and ran the main study once the task structure had been optimised (task structure and modifications detailed in **Figure 2.1B**). In these behavioural studies, I aimed to optimise the structure of a probabilistic decision-making paradigm to examine behavioural responses to expected and unexpected uncertainty, and determine whether participants' choice behaviour could be elucidated by *model-free* or *model-based* learning or, similarly to the findings of Daw *et al*. and Gläscher *et al.* detailed above, a combination of both mechanisms. I hypothesized that: 1) participants would be able to successfully learn optimal routes within the task structure, but that I would also observe a range of behavioural profiles across the cohort; 2) participants would display hallmarks of model-based learning approaches, suggesting that model-free learning theory cannot fully explain participants' behaviour, in line with previous findings; and 3) through various modifications made to the task, participants' learning of the task structure and optimal routes would improve.

Part of the rationale behind conducting these behavioural studies was also to inform a subsequent fMRI and pharmacology study, which investigated how the use of a selective NA reuptake inhibitor (SNRI) can influence belief updating and exploratory/exploitative behaviours in this decision-making task. The study also used fMRI to identify neural signatures of SAPEs in this spatial memory and decision-making task, and how such neural signatures may be influenced by NA manipulations (see *Chapter Four*).

2.2 Pilot Study One

2.2.1 Methods

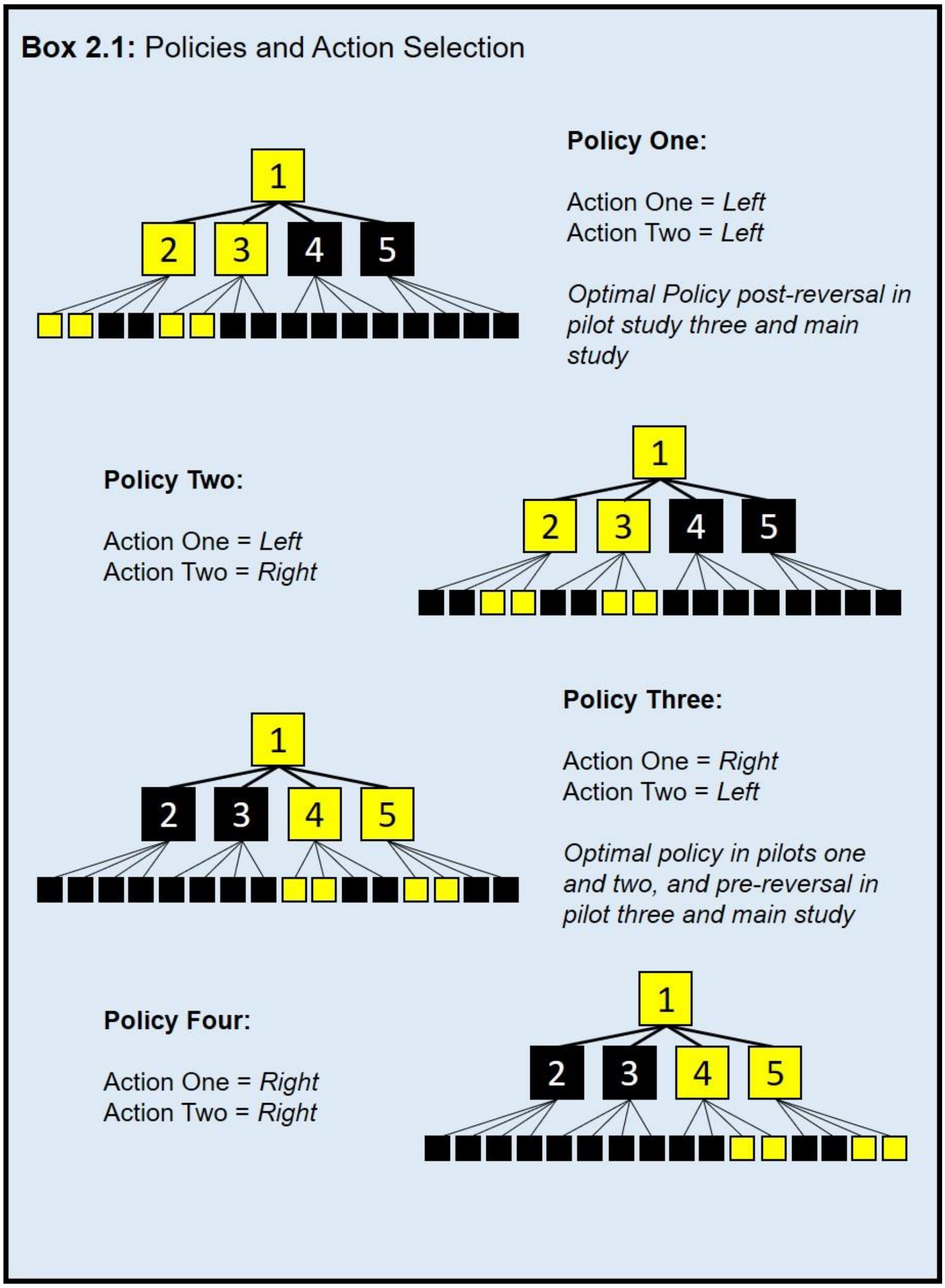
Participants

Twenty participants (11 females) were recruited in pilot one. All participants in pilot one were aged 18 or over (mean = 26.9 ± 4.08 SD), right-handed, had no current or history of neurological or psychiatric conditions; were not taking any anti-depressant medication and were a mix of both males and females. All participants were recruited via online, email, and poster adverts from the student and staff population of the University of Bristol, and the general public.

Behavioural Paradigm

I designed a Markov decision task using a probabilistic binary tree structure of state transitions based on the task used by Gläscher *et al*. (Gläscher *et al.*, 2010), with the modification of using images of spatial locations/scenes rather than fractals (**Figure 2.1A**). In the task, participants were asked to make two sequential choices, one choice in each of two successive decision states to reach the outcome (end) state and either receive a reward or be notified of the absence of reward. A set of multiple actions selected by the participant to reach a desired state or outcome is known as a *policy*, and in subsequent analyses the term *policy* will be used to denote each action set chosen by participants (see **Box 2.1** for outline of policies). Moving left then left is defined as policy one, moving left then right is policy two, right then left is policy three, and right then right policy four. Each state was represented by an image of the current scene/location in the virtual game environment, for example an image of a forest, with action choices defined by an upper-right-pointing orange arrow and an upper-left-pointing blue arrow, corresponding to right and left arrow keys, respectively (**Figure 2.2**). This indicated that during the first two states the participant had to choose either a left or right arrow key press. The initial state remained the same for every trial with the same two action choices; the initial state at t = 1 is denoted the ‘level one’ state.

Box 2.1



Box 2.1: Policies and Action Selection

Policy One:

Action One = *Left*
Action Two = *Left*

Optimal Policy post-reversal in pilot study three and main study

Policy Two:

Action One = *Left*
Action Two = *Right*

Policy Three:

Action One = *Right*
Action Two = *Left*

Optimal policy in pilots one and two, and pre-reversal in pilot three and main study

Policy Four:

Action One = *Right*
Action Two = *Right*

Box 2.1 Outline of policies within the task structure.

Figure 2.2

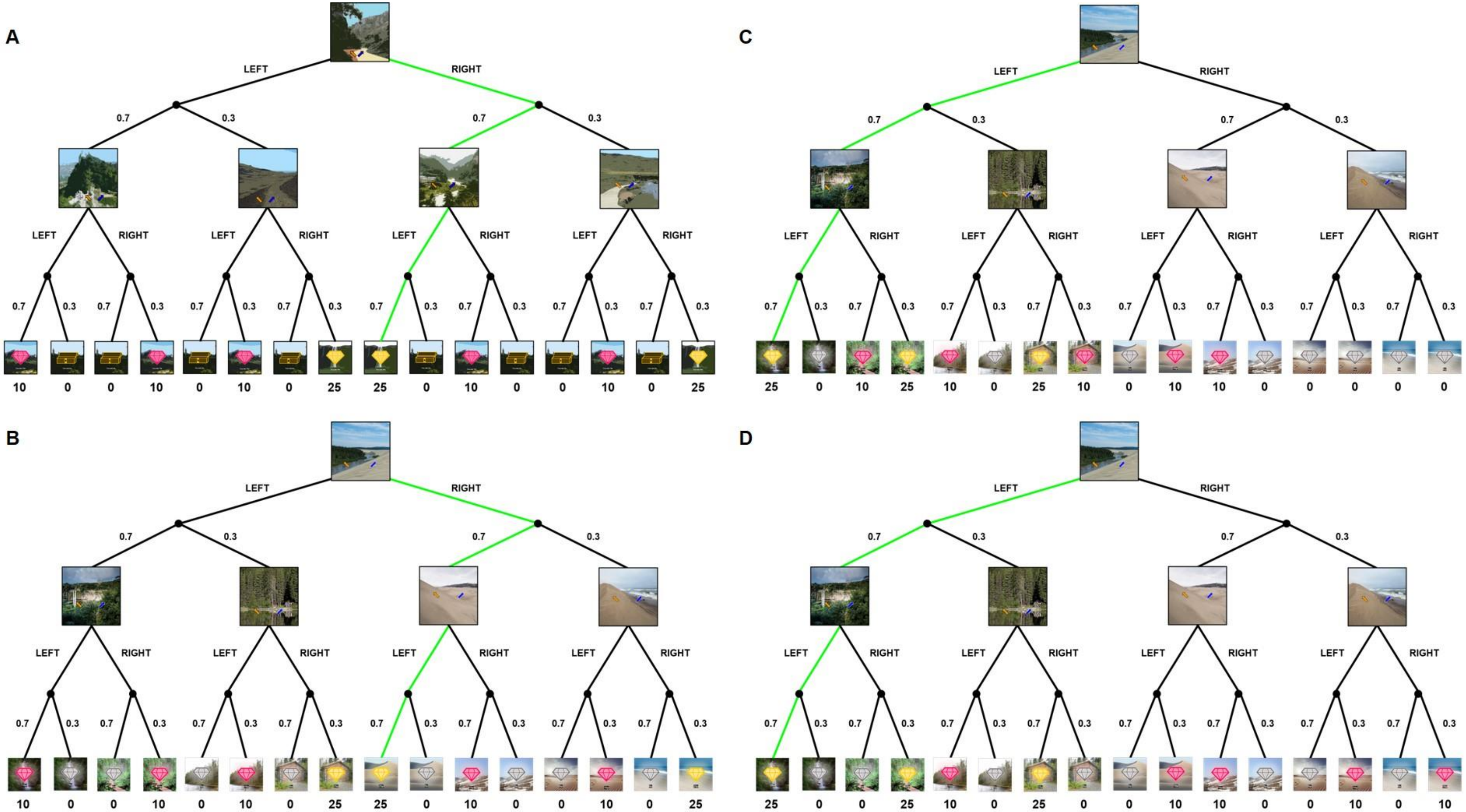

A
LEFT
RIGHT
0.7
0.3
0.7
0.3
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3
10 0 0 10 0 10 0 25 25 0 10 0 0 10 0 25
C
LEFT
RIGHT
0.7
0.3
0.7
0.3
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3
25 0 10 25 10 0 25 10 0 10 10 0 0 0 0 0
B
LEFT
RIGHT
0.7
0.3
0.7
0.3
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3
10 0 0 10 0 10 0 25 25 0 10 0 0 10 0 25
D
LEFT
RIGHT
0.7
0.3
0.7
0.3
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
LEFT
RIGHT
0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3 0.7 0.3
25 0 0 25 10 0 25 0 0 10 10 0 0 10 0 10

Figure 2.2 Task design and state structure. The task is a Markov decision task with a probabilistic binary decision tree structure. Each trial starts with the same state, scene one (state one). In the outcome states, participants will either receive low-level reward, a pink gem (10p), high-level reward, a gold gem (25p), or no reward, an empty treasure chest (0p). Probabilities are indicated on each branch (here, either 0.7 or 0.3). Optimal policies are highlighted in green. A *policy* is defined as a set of multiple actions selected by the participant to reach a desired state or outcome. (**A**) This displays an example of how the locations of images may be counterbalanced at levels two and three for a particular participant, with images used in pilot one. The numbers below the outcome states reflect the reward obtained upon reaching the state per trial in pence. (**B**) Task structure with new images used in pilot two, using the silver gem instead of the empty treasure chest to indicate a reward of 0p. No image counterbalancing occurred for pilot two. (**C**) Task structure after the reversal in pilot three, which occurs after trial 40 in the testing session. The new optimal policy has changed from policy three to policy one (49% chance of winning 25p), and the pre-reversal (former) optimal policy now only offers 0p or 10p instead of 25p. (**D**) Task structure after the revised reversal for the main behavioural study. Participants have a chance of no reward in every arm of the task, and the optimal policy is unambiguously policy one.

After choosing the first action, the participant moved into one of two different intermediary states with different transition probabilities (0.7 or 0.3, see **Figure 2.2**), each state with two different action choices; intermediary states at t = 2 are denoted 'level two' states. Images in the second scene were counterbalanced across participants (**Figure 2.2A**). After the second action choice, the participant moved into one of the outcome states with different levels of reward as described below, once again with different transition probabilities (0.7 or 0.3); outcome states at t = 3 are denoted 'level three' states. Scenes associated with rewards in the outcome states were also counterbalanced across participants, but the locations of the reward values themselves remained constant across participants (**Figure 2.2A**).

The inter-state interval was randomly sampled from a uniform distribution from 1.5-2.5 s. The inter-trial interval was randomly sampled from a uniform distribution from 5-7 s, and a fixation cross was displayed between each trial and between each state transition. For each state, participants had 4 s to choose an action and make a key press. If they failed to do so in this time, the current trial restarted, and the participant was presented with the restart screen, i.e. an image of a skull, for 4 s. The task consisted of two sessions in one experimental sitting, the training session and the testing session, each session consisting of 80 trials and lasting approx. 25 minutes.

Training Session:

In the training session (80 trials), all actions were predetermined: only one arrow appeared on the screen pointing either left or right, therefore choices of which direction to take were fixed. Participants did not receive any rewards at the outcome states during the training session and were not notified of the reward values associated with the objects in the outcome states. The trials in the training session were pseudorandomised but reflected exactly the underlying state transition probabilities, reflecting the paradigm used by Gläscher *et al.* (Gläscher *et al.*, 2010).

Testing Session:

In the testing session (80 trials), after a 15-minute break, participants were free to make their own action choices and were rewarded accordingly at the outcome states. During the break, participants were informed of the reward-object mappings in the outcome states by completing a short choice task. These rewards were reflected as real-world monetary rewards which were awarded once the task was completed, along with reimbursement for time.

Outcome States:

- 'Pink Gem' = low-level, medium probability reward (10p per trial)
- 'Gold Gem' = high-level, low probability reward (25p per trial)
- 'Empty Treasure Chest' = no reward, high probability (0p per trial)

Exclusion Criteria

The study by Gläscher *et al*. used a threshold for minimal learning and participants who scored below this threshold were excluded from all their subsequent analyses. In this study I have included all participants in the behavioural analyses, however I conducted similar tests to explore how many of my participants would meet the criteria set by Gläscher *et al*. This threshold was defined as the upper 95th percentile of total reward distribution obtained from a Monte-Carlo simulation of 10,000 randomly behaving agents (Gläscher *et al.*, 2010). I replicated this Monte-Carlo simulation (**Appendix Figure A.1A**) and obtained an upper 95th percentile of £7.40. For pilot one, eight participants (seven females) did not pass this threshold, contrasting with two participants in Gläscher *et al*.'s study whom did not meet the criterion for minimal learning.

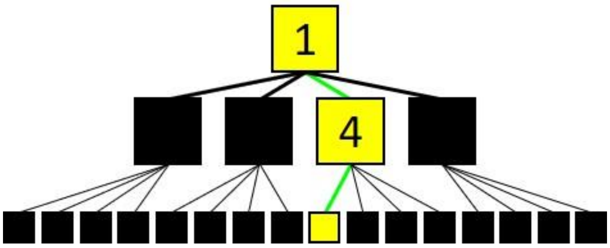
The expected level of reward based on the transition probabilities and reward values in the task for pilot one is £6.09. Using this lower value as an alternative minimum reward threshold, three participants (two females) did not exceed this threshold in pilot one. As these are exploratory behavioural pilot studies, the participants who did not pass these thresholds were not excluded.

Statistical Analyses

I used the two-sided binomial (sign) test to evaluate if participants chose the optimal policy significantly more than chance level, and chi-squared tests to evaluate if participants' policy selection across all available policies deviated significantly from chance level. I also used chi-squared tests to examine action selection at individual states, i.e. if participants preferred to move left or right from the initial state or any intermediary states. To examine reaction time (RT) differences I used paired *t*-tests.

2.2.2. Results

Policy frequency across trials in testing session



In the behavioural paradigm, there is a single optimal route (or *policy*) where participants are most likely to obtain the highest available reward, i.e. moving right from state one, reaching state four, then moving left which gives the participant a 70% chance of winning 25p after reaching state four (**Figure 2.2A,** also see **Box 2.1** and ***Insert***). Therefore, I expected participants to attempt this route more frequently than other possible routes throughout the testing session of the task if they had successfully learned the task structure during the fixed training session. Policy three is the optimal policy as the participant has the greatest chance of obtaining the high-level reward: 49% chance of 25p (**Box 2.1**). The mean frequency of policy three choices over all trials across participants is significantly different to chance level (p = 5.18 x 10^{-4}; sign test) (**Figure 2.3A**). Also, I conducted binomial sign tests per participant across trials; 14 out of 20 participants chose policy three significantly more than expected by chance (p < 0.05), where chance level is 25% of choices. Furthermore, I conducted chi-squared tests per participant to examine whether policy choice frequency was due to chance, regardless of which policy was preferred by the participant. Nineteen out of 20 participants reliably chose a particular policy significantly more than would be expected by chance (p < 0.05). This suggests that almost every participant (19/20) developed a particular strategy for navigating the behavioural task space, even if the chosen strategy was not optimal (16/20 chose the optimal policy). (***Insert:*** *initial state at level one numbered as state one. To take the optimal policy, policy three, the participant must move right to reach state four, then left to earn 25p (70% chance); optimal states in yellow. Green path lines indicate the optimal policy*).

Figure 2.3

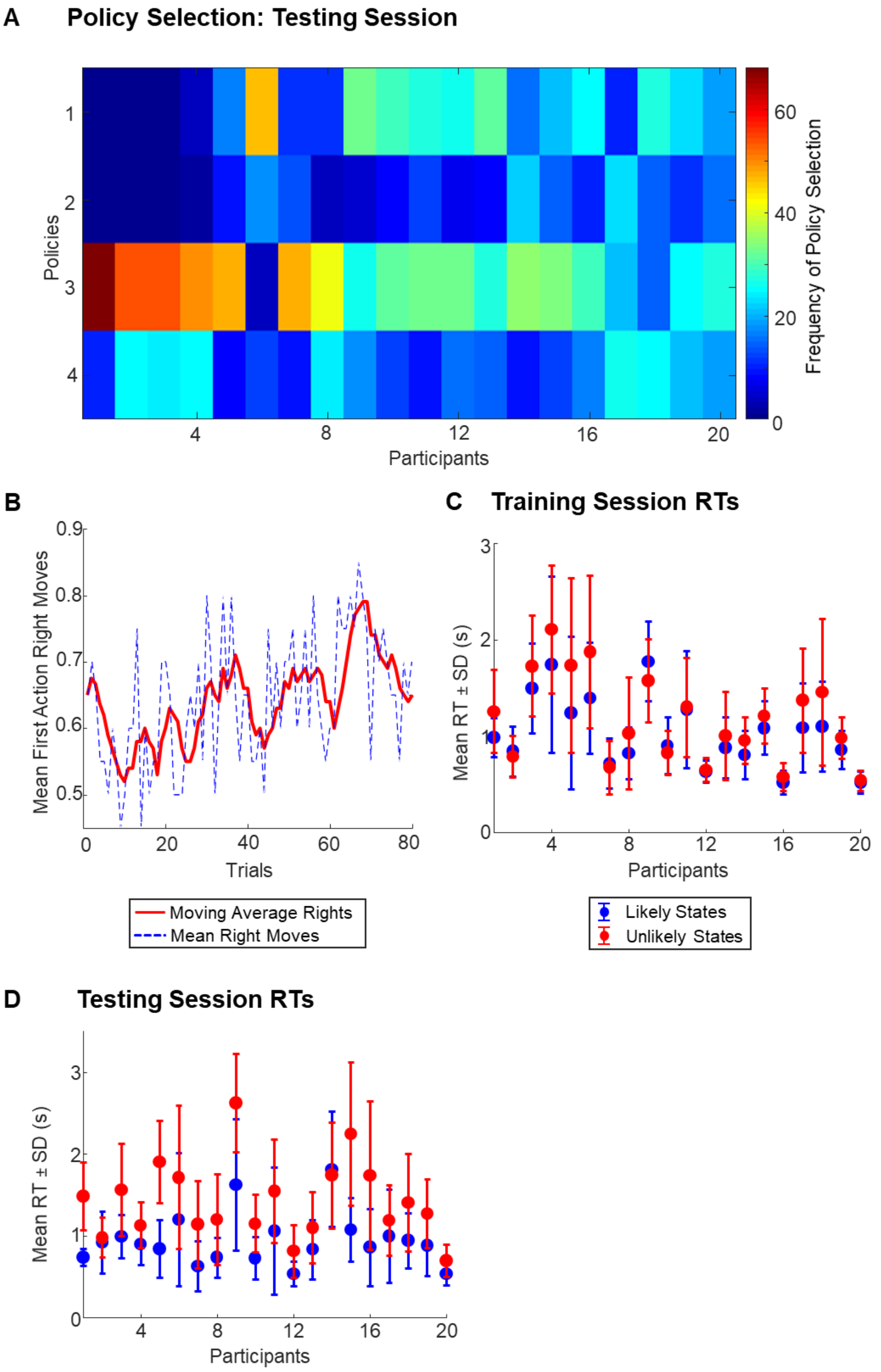
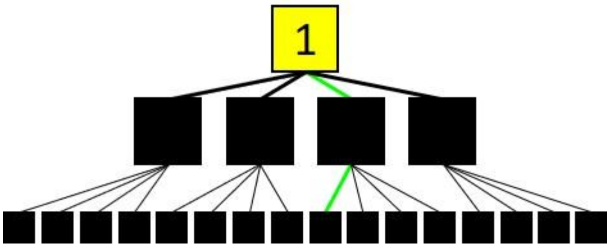


Figure 2.3 Behaviour in Pilot One. (**A**) Heatmap of policy frequency across participants in pilot one testing session, ordered according to their chi-squared test *p* value. Colour bar indicates frequency of policy selection across all trials in the testing session. Policy one: Left, left. Policy

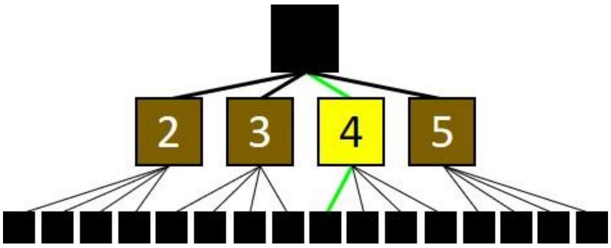
two: Left, right. Policy three: Right, left. Policy four: Right, right. Policy three is the optimum policy (very good likely outcome, 49% chance of 25p). Policies one and four are good policies (good likely outcome, 49% chance of 10p). Policy two is a bad policy (bad likely outcome, 49% chance of 0p). Mean of policy three choice frequency across the cohort was significantly different to chance level ($p = 5.18 \times 10^{-4}$, sign test). (**B**) Moving average across participants of right moves taken in the first action choice of each trial, i.e. at level one (red solid line), and the mean of right moves taken at $t = 1$ per trial (blue dashed line) in the testing session. Moving average used a window of five trials. The moving average of right actions significantly increased as trials progressed through time (*rho* = 0.656, $p = 4.07 \times 10^{-11}$). (**C**-**D**) RT across trials for actions in likely (blue) and unlikely (red) intermediary states in the training (**C**) and testing (**D**) sessions of pilot one, mean ± SD. Likely states had significantly faster RTs in both training and testing sessions compared to unlikely states across participants (training: $t(19) = -3.50$, $p = 0.00240$; testing: $t(19) = -6.49$, $p = 3.22 \times 10^{-6}$; paired *t*-tests). Participants are ordered by testing session performance (*p* values of policy selection chi-squared tests) in both training and testing RT plots, as actions in the training session were fixed. For actions at $t = 2$ from the unlikely level-two states three and five (30% chance of reaching these states from action at $t = 1$), only three participants showed a preference in state three, and four showed a preference in state five. This is expected, as neither of the most likely outcome states from states three and five offer any reward regardless of action choice, and the unlikely intermediary states are less likely to be learned by participants in the training session. SD = standard deviation.

Participants Made Optimal Action Choices Following Likely Intermediary States

In addition to policies, I examined individual action choices at each timepoint in the trial. For the first action choice in state one at $t = 1$, i.e. level one, participants did not show an overall preference for a left or right action, as only eight out of 20 participants chose one action over the other significantly above chance (where chance level is 40 left moves and 40 right moves). I also looked at how the number of right moves at $t = 1$ changed over the course of the testing session by calculating the moving average of right first moves over participants, using a window of five trials. The number of right moves at $t = 1$ increases significantly over time across participants ($rho = 0.656$, $p = 4.07 \times 10^{-11}$; Pearson's correlation) (**Figure 2.3B**). Therefore, as trials progressed participants exploited the optimal route more frequently to obtain higher levels of reward. (***Insert:*** *initial state at level one numbered as state one, in yellow. To take the optimal route from state one, the participant must move right. Green path lines indicate the optimal policy*).



However, for level two second action choices at $t = 2$ from both states two and four, the likely intermediary states (70% chance of reaching these states from action at $t = 1$), 11 out of 20 participants showed preferences for one action over the over (either left or right) in state two, and 15 out of 20 participants showed action preferences in state four significantly more than chance. Of the 11 participants who showed significant action preference in state two, all showed preference for moving left, which is the optimum action from state two with the highest chance of obtaining the most reward. Similarly, of the 15 participants who showed preferences in state four, 13 participants showed preference for moving left, which is the optimum action from state four and follows the optimal policy. (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state is state four, in yellow. Green path lines indicate the optimal policy*).



To summarise, most participants in pilot one selected policy three significantly more than chance, suggesting that the task structure was successfully learned by the majority of participants. However, the mean number of right moves at $t = 1$ continued to increase in the testing session, indicating that participants may not have learned this structure completely during the training session, and learning continues into the testing session (**Figure 2.3B**).

Reaction Times Were Significantly Faster in Likely Compared to Unlikely Intermediary States

In the both task sessions, I investigated the RTs of actions at $t = 2$ across trials and compared RTs of actions taken from level two states: likely intermediary states two and four and unlikely intermediary states three and five. I examined the RTs of second actions across trials in the training session, and found that RTs were significantly faster in likely versus unlikely states ($t(19) = -3.50$, $p = 0.00240$; paired *t*-test) (**Figure 2.3C**). Similarly in the testing session, participants responded significantly faster to likely states compared to unlikely states, and this RT difference was more pronounced in the testing session compared to the training session ($t(19) = -6.49$, $p = 3.22 \times 10^{-6}$; paired *t*-test) (**Figure 2.3D**). This suggests that participants successfully learned the transition probabilities between likely and unlikely intermediary states, and that this learning improves from the training session to the testing session.

2.3 Pilot Study Two

Based on my statistical findings in pilot one in addition to verbal self-report from participants, it appeared that although most participants were able to locate the optimal route in the task, participants struggled to make strong distinctions between the different images, and saw each image as a discrete location or state, as opposed to single points that occurred on the same continuous route in space. The main modification to the task that was made for pilot two, was therefore a change in the images to use a set of more congruent images, which appeared to be located at different points on the same physical route. I expected learning of the optimal policy to increase as a result of this change.

2.3.1 Methods

Participants

Nine participants (five females) were recruited and tested in pilot two. All participants were recruited with the same exclusion criteria and tested on the behavioural paradigm in the exact same way as in pilot one, and were aged 18 or over (mean = 23.6 ± 3.13 SD).

Behavioural Paradigm

In the second pilot study the same task structure as in pilot one was used with the same number of trials. However, I used new, more congruent spatial images to make the virtual path look more realistic, i.e. the new images would follow more closely from one to the next in a more realistic sequential environment. Using these new images, however, meant that I could no longer counterbalance images for level two and three states between participants as each image followed on from the previous image in a more realistic map, so could not be rearranged. All participants therefore experienced the same set of images in the same state locations (**Figure 2.2B**). I also changed the image of the treasure chest, indicating no reward, to a silver gem to increase consistency across reward-associated objects at the outcome states. All other aspects of the behavioural task structure, such as transition probabilities, reward locations and inter-state/trial intervals, remained the same as pilot one.

Outcome States for Pilot Two:

- ‘Pink Gem’ = low-level, medium probability reward (10p per trial)
- ‘Gold Gem’ = high-level, low probability reward (25p per trial)
- ‘Silver Gem’ = no reward, high probability (0p per trial)

Exclusion Criteria

Based on the Monte-Carlo simulation described above for pilot one, as the task structure itself did not change between pilot studies one and two, five participants in pilot two did not score greater than the threshold set by Gläscher *et al.* – the upper 95th percentile of this simulation was £7.40. However, using the expected level of reward as a threshold for minimal learning, which was calculated above to be £6.09, two participants in pilot two scored below this threshold.

Statistical Analyses

All statistical analyses performed for pilot one were also replicated exactly for pilot two. I also wanted to examine differences between reward earned and optimal policy selection frequency across pilot studies one and two, in order to determine whether the more congruent images enhanced reward earning and optimal policy exploitation in the task. I therefore used two-tailed two-sample Welch's *t*-tests, i.e. assuming unequal variances. Welch's *t*-tests were used due to the unequal variances between the first two pilot studies, and unequal group sizes.

2.3.2 Results

Policy Frequency Across Trials in Testing Session Showed Optimal Policy Preference

Consistent with pilot one, the optimal policy in pilot two was also policy three. All definitions of policies remain consistent with that in pilot one (**Box 2.1**). The mean frequency across all trials and participants of policy three choice was significantly above chance (p = 0.00290; sign test) (**Figure 2.4A**). On a per-participant basis, six out of nine participants chose policy three significantly more than chance across trials ($p < 0.05$, sign test), where chance level is 25% of choices. By conducting chi-squared tests per participant, I found that all participants chose one of the policies significantly more than chance. This concurs with pilot one, in that all participants learned the task to a level at which they could choose a preferred policy and use this to navigate the environment. Contrasting with pilot one in which the second most-preferred policy was policy one, the second most-preferred policy in pilot two was policy four. This may have been due to participants remembering that the desert-like level two images (states four and five) were more likely to be rewarding, as policy three passed through the desert locations and so participants were more likely to move right at the start of the trial, regardless of their second action. (***Insert:*** *initial state at level one numbered as state one. To take the optimal policy, policy three, the participant must move right to reach state four, then left to earn 25p (70% chance); optimal states in yellow. Green path lines indicate the optimal policy*).

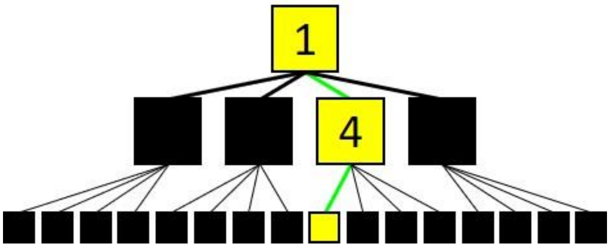


Figure 2.4

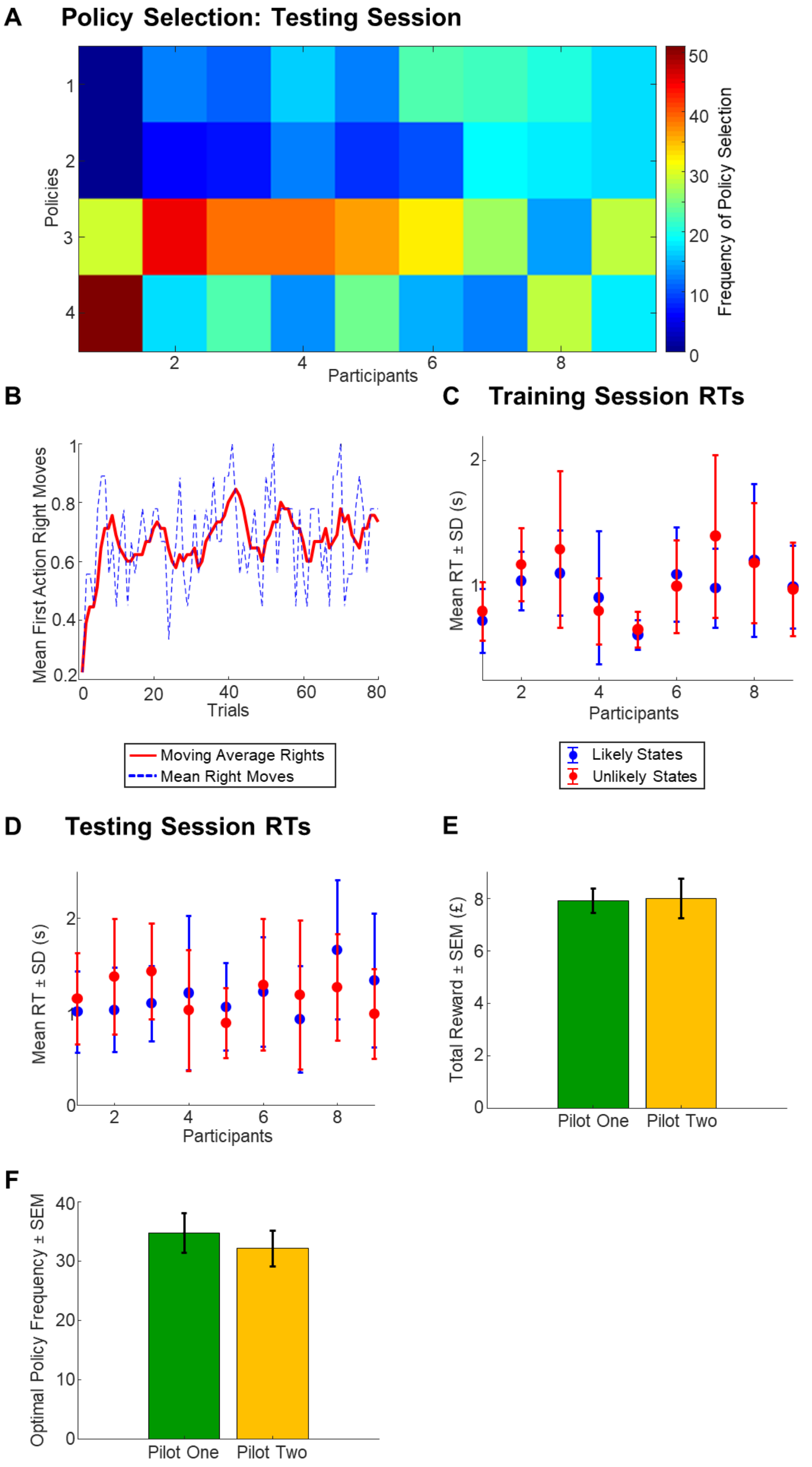
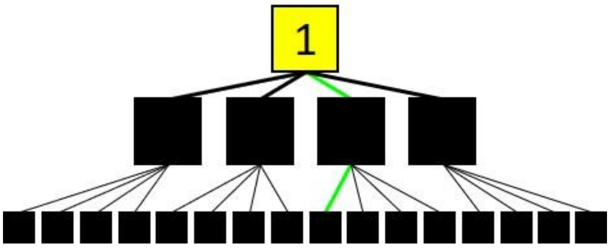


Figure 2.4 Behaviour in Pilot Two. (**A**) Heatmap of policy frequency across participants in pilot two testing session, ordered according to their chi-squared test *p* value. Colour bar indicates

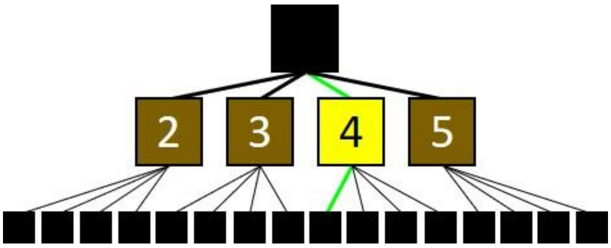
frequency of policy selection across all trials in the testing session. Policy one: Left, left. Policy two: Left, right. Policy three: Right, left. Policy four: Right, right. Policy three is the optimum policy (very good likely outcome, 49% chance of 25p). Policies one and four are good policies (good likely outcome, 49% chance of 10p). Policy two is a bad policy (bad likely outcome, 49% chance of 0p). Mean of policy three choice frequency across participants was significantly different to chance level (p = 0.00290, sign test). (**B**) Moving average across participants of right moves taken in the first action choice of each trial (red solid line), and the mean of right moves taken per trial (blue dashed line) in the testing session. Moving average used a window of five trials. The moving average of right actions significantly increased as trials progressed through time (rho = 0.476, $p = 7.93 \times 10^{-6}$). (**C**-**D**) RT across trials for actions in likely (blue) and unlikely (red) intermediary states in the training (**C**) and testing (**D**) sessions of pilot two, mean ± SD. There was no significant difference between RTs in likely versus unlikely states, in both training and testing sessions across participants (training: $t(8) = -1.24$, p = 0.250; testing: $t(8) = -0.0705$, p = 0.946; paired t-tests). Participants are ordered by testing session performance (p values of policy selection chi-squared tests) in both training and testing RT plots, as actions in the training session were fixed. (**E**-**F**) Comparison of pilot studies one and two, examining total reward earned (**E**) and optimal policy selection frequency in testing sessions (**F**). Pilot one = green; pilot two = orange. (**E**) Total reward earned per participant was not significantly different between pilots one and two ($t(14.2) = -0.102$, p = 0.920; Welch's two-sample t-test). (**F**) There was no significant difference in the optimal policy selection per participant between pilots one and two ($t(24.4) = 0.583$, p = 0.565; Welch's two-sample t-test). For actions at $t = 2$ from unlikely level two states, in state three, one participant moved right significantly more than chance. Five out of nine participants chose to move right significantly more than chance from state five, selecting policy four. SD = standard deviation; SEM = standard error of mean.

Participants Made Optimal Action Choices Following Intermediary States After Moving Right in State One, but Not Left

For the first action choice in state one at $t = 1$, five out of nine participants moved right first significantly more than chance, and no participants moved left first significantly more than chance. This was a slight improvement from pilot one and supports the finding above that policies three and four were preferred over policies one and two. Examining the moving average of first action right moves at $t = 1$ over 80 trials, I found a similar result to pilot one, in that there was a steady increase in the number of right first moves over the course of the testing session across participants, but the correlation was slightly weaker compared to pilot one (*rho* = 0.476, $p = 7.93 \times 10^{-6}$; Pearson's correlation) (**Figure 2.4B**). (***Insert:*** *initial state at level one numbered as state one, in yellow. To take the optimal route from state one, the participant must move right. Green path lines indicate the optimal policy*).



For second action choices at $t = 2$, participants learned the most optimal outcomes from states four and five but not states two and three (*see **Figure 2.4** legend for unlikely states*). From state two, only two out of nine participants moved left significantly more than chance (optimal action choices from these states). However, from state four, seven out of nine participants chose to move left significantly more than chance, therefore selecting policy three. The proportion of left to right moves from state two was similar to that from state four, however, because very few trials encountered state two due to reduced left actions from state one at $t = 1$ across participants this was not statistically significant. This suggests increased learning of the right arm of the task structure and reduced experience of the left arm compared to pilot one. (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state is state four, in yellow. Green path lines indicate the optimal policy*).



In pilot two, therefore, I saw a similarly high level of significant preference for policy three, with a slight improvement in the action selection of participants. Also, reduced learning appeared to occur in the testing session of pilot two, evidenced by the reduced increase in mean number of right moves at $t = 1$ compared with pilot one, suggesting that in pilot two participants may have learned the task structure quicker in the training session (**Figure 2.4B**).

No Significant Difference in RTs Following Likely Compared to Unlikely Intermediary States

As with pilot one, I examined the RTs of actions at $t = 2$ across trials in the training and testing sessions separately, comparing likely and unlikely intermediary states. In contrast to pilot one, there were no significant differences in RT between likely and unlikely states, neither in the testing nor training sessions, which was consistent across all participants (training: $t(8) = -1.24$, $p = 0.250$; testing: $t(8) = -0.0705$, $p = 0.946$; paired t-tests) (**Figure 2.4C-D**). This was unexpected, as learning of the optimal policy was more consistent in pilot two and learning of the right arm of the task was improved compared to pilot one. Also, the overall RT was similar to pilot one across participants, so this was not a result of a general slowing of responses across the whole task.

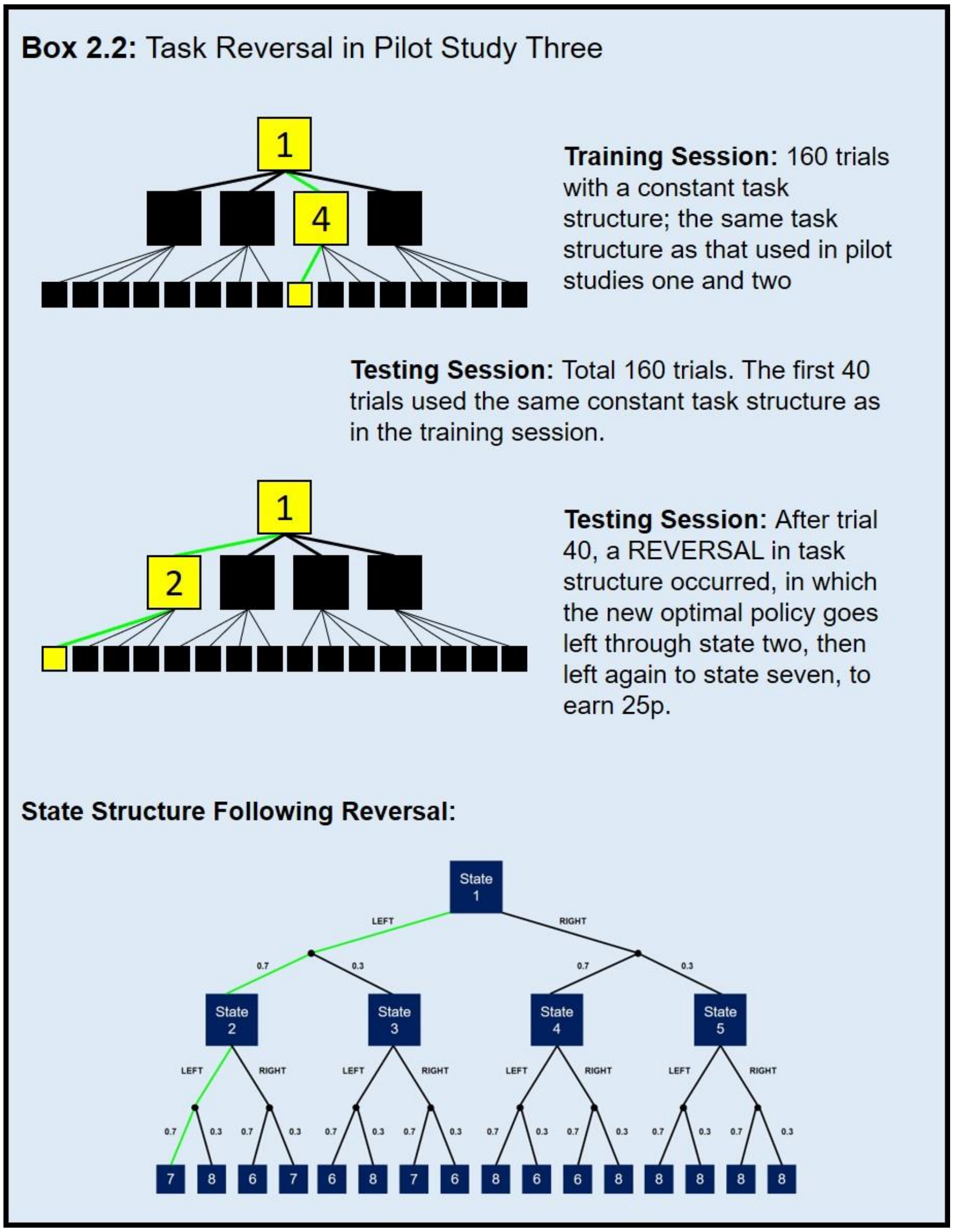
No Significant Improvement in Reward or Policy Choice From Pilot One to Pilot Two

I then investigated whether the alterations made to the task in pilot two produced a significant improvement from pilot one. There was no significant increase in the level of reward earned over the testing session across participants (mean reward, pilot one: £7.91 ± £2.04 SD; pilot two: £8.00 ± £2.26 SD; $t(14.2) = -0.102$, $p = 0.920$; Welch's two-sample t-test) (**Figure 2.4E**). Also, there was no significant difference in the frequency of optimal policy selection between pilot one and pilot two ($t(24.4) = 0.583$, $p = 0.565$; Welch's two-sample t-test) (**Figure 2.4F**). From this I concluded that although the new images seemed to improve the learning of the right (optimal) arm of the task structure, the effect this had on the overall performance of participants in the task is negligible.

2.4 Pilot Study Three

Based on my findings from pilot studies one and two, the more congruent images used in pilot two appeared to yield a slightly improved learning of the optimal policy, as in pilot one, 70% of participants selected the optimal policy greater than chance, whereas in pilot two, 77.8% of participants selection the optimal policy greater than chance. In pilot three, I aimed to investigate optimal task length and introduce unexpected uncertainty into the task. I increased the trial number from 80 to 160 per session, and after trial 40 in the testing session, introduced a reversal, as described below (**Box 2.2**). I also switched the fixed training session for a free-choice training session, with the aim of focusing on model-based SAPEs.

Box 2.2



Box 2.2: Task Reversal in Pilot Study Three

Training Session: 160 trials with a constant task structure; the same task structure as that used in pilot studies one and two

Testing Session: Total 160 trials. The first 40 trials used the same constant task structure as in the training session.

Testing Session: After trial 40, a REVERSAL in task structure occurred, in which the new optimal policy goes left through state two, then left again to state seven, to earn 25p.

State Structure Following Reversal:

Box 2.2 Outline of task reversal structure in pilot study three.

2.4.1 Methods

Participants

Seven participants (one female) were recruited and tested in pilot three. All participants were recruited using the same exclusion criteria and tested on the behavioural paradigm in the exact same way as in pilots one and two, and were aged 18 or over (mean = 22.6 ± 3.65 SD).

Behavioural Paradigm

For pilot three, the two sessions consisted of 160 trials each rather than 80 trials, and the break between sessions was extended from 15 minutes to one hour to emulate the necessary break that would be required in the fMRI study for drug manipulation. Each session lasted for approx. 45 minutes due to the increased trial numbers. Also, the training session was no longer predetermined. Rather than being directed by one arrow on the screen, the structure of the training and testing sessions in pilot three were the same as the testing sessions in pilots one and two, with two arrows presented on the screen and participants able to move freely in the environment and earn rewards from the start. Participants could also earn rewards in both sessions, so were informed of the reward-object mappings before starting the training session and therefore did not complete the choice task in the break between task sessions. The reward-object mappings remained the same as in pilot two. In the testing session only of pilot three, a reversal was introduced after trial 40, so participants had to learn the new task structure and adapt their strategies accordingly (**Figure 2.2C, Box 2.2**). This was the first behavioural study conducted here that included a task reversal, and the reversal only took place in the testing session. After this reversal, the optimal policy switched from policy three to policy one. Therefore, I analysed participants' behaviour in both the training and testing sessions and how they responded to the reversal.

Exclusion Criteria

I conducted a Monte-Carlo simulation for the 320-trial task structure with reversal used in pilot three and obtained a mean of £26.53 and upper 95th percentile of £29.25 (**Appendix Figure A.1B**). Two participants in pilot three did not score greater than this threshold of the upper 95th percentile, but only one participant scored below the expected level of reward of £26.52, calculated in the same way as that for pilots one and two above.

Statistical Analyses

All statistical analyses performed for the previous pilots were also replicated exactly for pilot three, but both the training and testing sessions were analysed in pilot three rather than solely the testing session, and the testing session was divided into pre-reversal and post-reversal trials.

2.4.2 Results

Training Session

Policy Frequency Across Trials in Training Session Divided 'Exploitative' and 'Exploratory' Behaviours in Participants

In the training session, four out of seven participants chose policy three significantly more than chance, and one participant chose policy three significantly less than chance (p = 3.07×10^{-12}; sign test, over all trials across participants) (**Figure 2.5A**). However, the four participants who preferred policy three showed an extremely strong preference, as these participants chose policy three for over 100 trials out of 160, which was not seen in other participants who preferred a different policy. By conducting chi-squared tests per participant, I found that six participants chose one of the policies significantly more than chance ($p < 0.05$, individual chi-squared tests). Contrasting with pilots one and two in which I saw a gradient of preference for policy three, here participants seemed to display characteristics of either one of two behavioural profiles: those that strongly preferred policy three and *exploited* this highly rewarding route greater than chance, and those that had a weaker preference for a different policy but tended to *explore* different routes throughout the task.

Figure 2.5

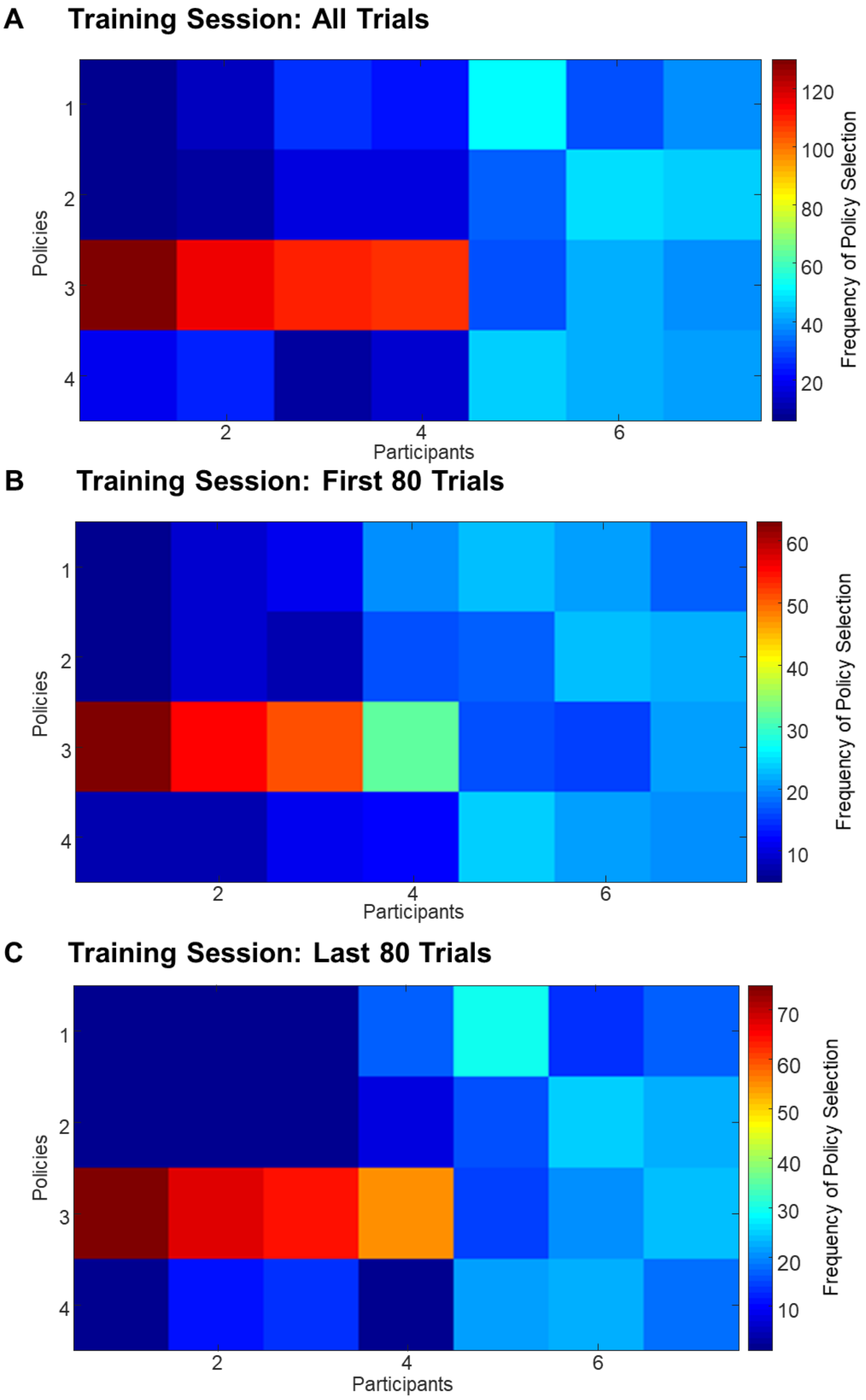
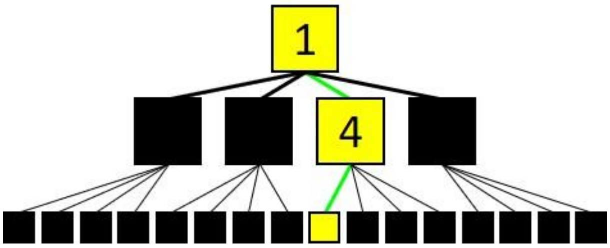


Figure 2.5 Behaviour in Pilot Three Training Session. (**A**) Heatmap of policy frequency over all 160 trials across participants in the training session of pilot three, ordered according to

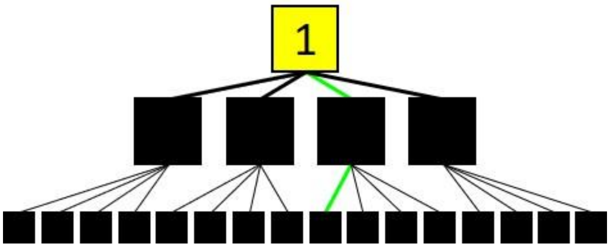
their chi-squared test *p* value. Colour bar indicates frequency of policy selection across all trials in the testing session. Mean of policy three choice frequency across participants was significantly different to chance level: (p = 3.07 x 10^{-12}, sign test). (**B**) Heatmap of policy frequency over the first 80 trials across participants in pilot three training session, ordered according to their chi-squared test *p* value. Colour bar indicates frequency of policy selection. Mean of policy three choice frequency across the cohort was significantly different to chance level: (p = 8.16 x 10^{-5}, sign test). (**C**) Heatmap of policy frequency over the last 80 trials across participants in the training session of pilot three, ordered according to their chi-squared test *p* value. Colour bar indicates frequency of policy selection. Mean of policy three choice frequency across participants was significantly different to chance level: (p = 2.64 x 10^{-9}, sign test). For actions at t = 2 for level two unlikely states: for state three no participants had a preferred direction in which to move; all were at chance level. For state five, despite the optimal action being to move right from this state, three participants moved left significantly more than chance and the remaining four were at chance. This may have been due to participants persisting with the optimal policy, as participants were less likely to reach state five and therefore unlikely to learn the optimal action following this state, and the three participants with a significant preference for moving left all displayed *exploitative* behaviours.

I then split the training session into the first and last 80 trials and examined the policy choice frequency in each 80-trial block, expecting the frequency of participants that chose the optimal policy to increase as the training session progressed. In the first 80 trials, four out of seven participants chose policy three significantly more than chance (p = 8.16 x 10^{-5}; sign test, across participants) (**Figure 2.5B**), and in the last 80 trials, the same four participants again chose policy three significantly more than chance (p = 2.64 x 10^{-9}; sign test, across participants) (**Figure 2.5C**). No participants chose any policy other than policy three significantly more than chance; during both the first and last 80 trials the remaining three participants remained at chance level. The significance across participants increased from the first 80 to the last 80 trials; a result of these four participants learning the optimal policy within the first 80 trials and then maintaining their optimal policy choice throughout the last 80 trials, choosing this policy more frequently than in the first 80, as expected (***Insert:*** *initial state at level one numbered as state one. To take the optimal policy, policy three, the participant must move right to reach state four, then left to earn 25p (70% chance); optimal states in yellow. Green path lines indicate the optimal policy*).



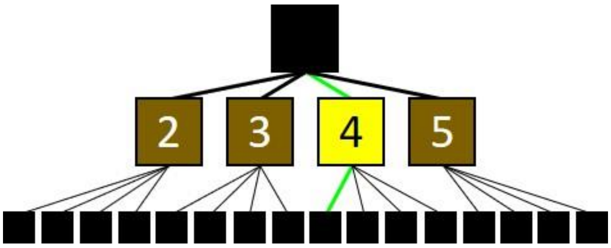
Participants Made Optimal Action Choices Following States One and Four in Training Session

I then examined action selection at each time point in the trial. Across all 160 trials of the training session, four out of seven participants moved right significantly more than chance from state one at t = 1; the same four participants who preferred policy three more than chance. The remaining three participants were at chance, once again reflecting their policy selections. When splitting the training session into the first and last 80 trials, only three of these more *exploitative* participants moved right significantly more than chance, and the remaining four participants were at chance. However, in the last 80 trials the same four participants moved right significantly more than chance, suggesting that one *exploitative* participant took slightly longer than the



other three *exploitative* participants to learn the optimal route, and the three *exploratory* participants remained at chance level; no participants moved left significantly more than chance at any point during the training session. When examining the moving average of first action right moves at $t = 1$ over all 160 trials, I again saw a gradual increase in participants moving right after state one with a moderate correlation between right moves and trials, very similar to pilots one and two (*rho* = 0.488, $p = 6.07 \times 10^{-11}$; Pearson's correlation) (**Figure 2.6A**). This further suggests that participants successfully learned the task structure over the course of the training session (***Insert:*** *initial state at level one numbered as state one, in yellow. To take the optimal route from state one, the participant must move right, pre-reversal. Green path lines indicate the pre-reversal optimal policy*).

For the level two second actions at $t = 2$, most participants (four out of seven, all *exploiters*) learned the optimal route and moved left from state four significantly more than chance, but did not learn the optimal actions following other intermediary states, indicating that participants only learned the optimal policy and not the full environmental structure. One participant moved right significantly more than chance from state four but did not have a significant policy preference, and two participants were at chance. From state two, only two participants moved left significantly more than chance and the rest were at chance level (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state pre-reversal is state four, in yellow. Green path lines indicate the pre-reversal optimal policy*).



Overall, in pilot three I observed two distinct behavioural profiles: one in which participants highly exploited the optimal policy but did not learn very much of the rest of the environment, and one in which participants were very exploratory and showed no preference for any particular policy or set of actions, i.e. chance-level policy selection. These more exploitative participants learned the optimal policy very quickly, as three out of four of the more exploitative participants had learned this policy in the first half of the training session, enabling the categorisation of participants into two groups, separating those displaying more *exploratory* and *exploitative* behavioural profiles.

Figure 2.6

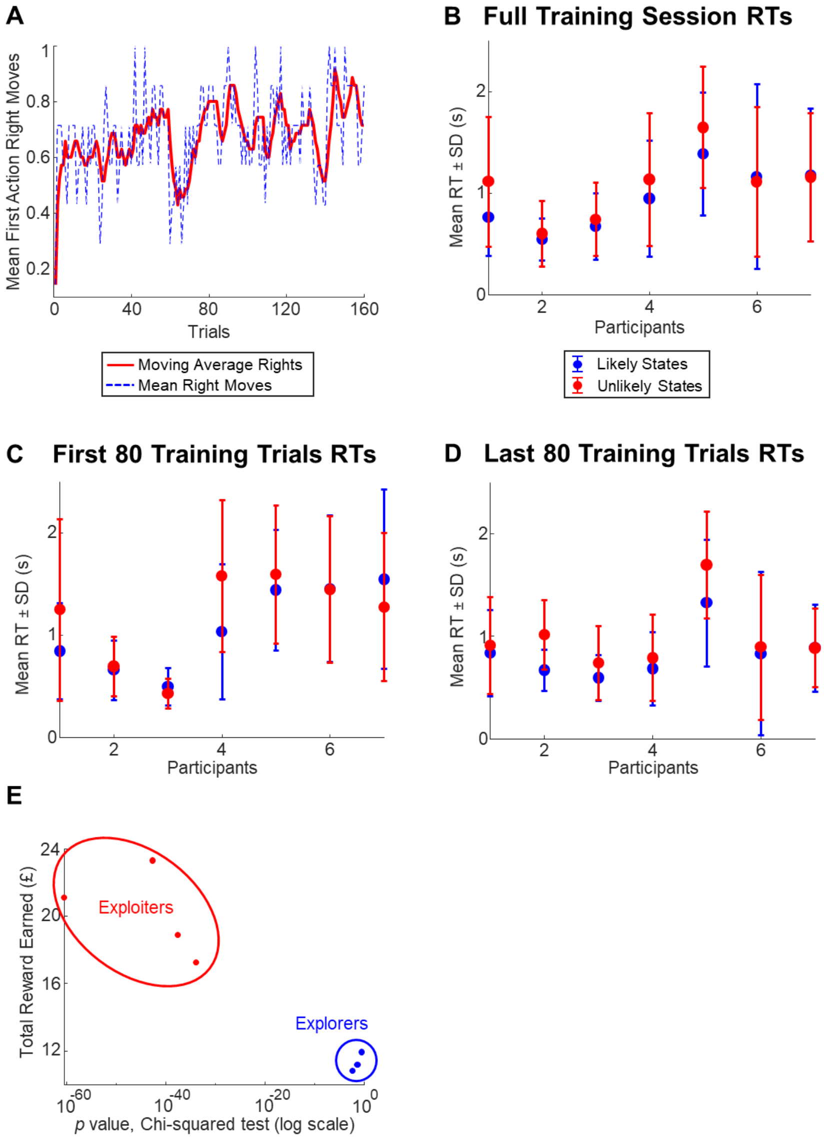


Figure 2.6 Actions in Pilot Three Training Session. (**A**) Moving average across the cohort of right moves taken in the first action choice of each trial (red solid line), and the mean of right moves taken per trial (blue dashed line) over all trials in the training session of pilot three. Moving average used a window of five trials. The moving average of right actions significantly increased as trials progressed through time (*rho* = 0.488, *p* = 6.07 x 10^{-11}). (**B**) RT across all 160 trials for actions in likely (blue) and unlikely (red) intermediary states in the training session of pilot three, mean ± SD. There was no significant difference between RTs over all 160 trials

in likely versus unlikely states across the cohort (*t*(6) = -2.1758, *p* = 0.0725; paired *t*-test). (**C**) RT across the first 80 trials for actions in likely and unlikely intermediary states, mean ± SD. There was no significant difference between RTs over the first 80 trials in likely versus unlikely states across the cohort (*t*(6) = -1.07, *p* = 0.327; paired *t*-test). (**D**) RT across the last 80 trials for actions in likely and unlikely intermediary states, mean ± SD. Likely states had significantly faster RTs than unlikely states in the last 80 trials across participants (*t*(6) = -2.90, *p* = 0.0273; paired *t*-test). Participants are ordered by performance (*p* values of policy selection chi-squared tests) in all RT plots. (**E**) Relationship between policy choice (chi-squared *p* value) and total reward earned across pilot three training session. Participants could be divided into two explicit groups: those who displayed *exploratory* behaviours (blue) with policy choice close to chance and lower overall reward, and those who displayed more *exploitative* behaviours (red) with high policy preference (low *p* value) and higher overall reward. SD = standard deviation.

Reaction Times Following Likely States Were Significantly Faster Than Following Unlikely States in the Last 80 Trials of the Training Session

I then looked at RTs following likely versus unlikely states over the whole 160 trials, and the first and last 80 trials separately. Overall, there was no significant difference between RTs following likely versus unlikely states ($t(6) = -2.18$, $p = 0.0725$; paired *t*-test) (**Figure 2.6B**). This was expected, as for the first half of the training session participants were unaware of which states are more or less likely. Similarly, there was no significant difference between RTs of likely and unlikely states during the first 80 trials ($t(6) = -1.07$, $p = 0.327$) (**Figure 2.6C**). However, RTs following likely states were significantly faster than that following unlikely states during the final 80 trials of the training session ($t(6) = -2.90$, $p = 0.0273$), showing that in the second half of the task participants had learned which states had the higher or lower probability of appearing (**Figure 2.6D**). This finding reflects the RT effect seen in pilot one, which was missing in pilot two.

Exploratory and Exploitative Behavioural Profiles Could Be Separated Using Total Reward and Policy Choice

When examining the policy choices of participants, four participants showed a very strong preference for policy three, showing *exploitation* behaviour, and three participants were at chance level, showing *exploration* behaviour. The more exploitative participants earned significantly more as a subgroup compared to the more exploratory participants ($t(5) = 5.57$, $p = 0.00260$; two-sample *t*-test), and when plotting each participants' chi-squared test *p* value against the total reward earned in the training session these subgroups were clearly very distinct (**Figure 2.6E**).

Testing Session

Policy Frequency Across Trials in Testing Session Showed Post-Reversal Learning

For the following analyses of the testing session, trials were divided into the first 40 trials before the reversal, and the last 120 trials post-reversal (**Figure 2.7**). Before the reversal, four out of seven participants chose policy three significantly more than chance: three exploiters and one explorer as labelled in the training session analysis above (p = 1.75 x 10^{-4}; sign test, over all trials across participants) (**Figure 2.7A**). By conducting chi-squared tests I saw a similar result; four participants preferred one policy significantly more than chance and the remaining three participants were at chance level for policy choice in the first 40 trials (p < 0.05, individual chi-squared tests).

Figure 2.7

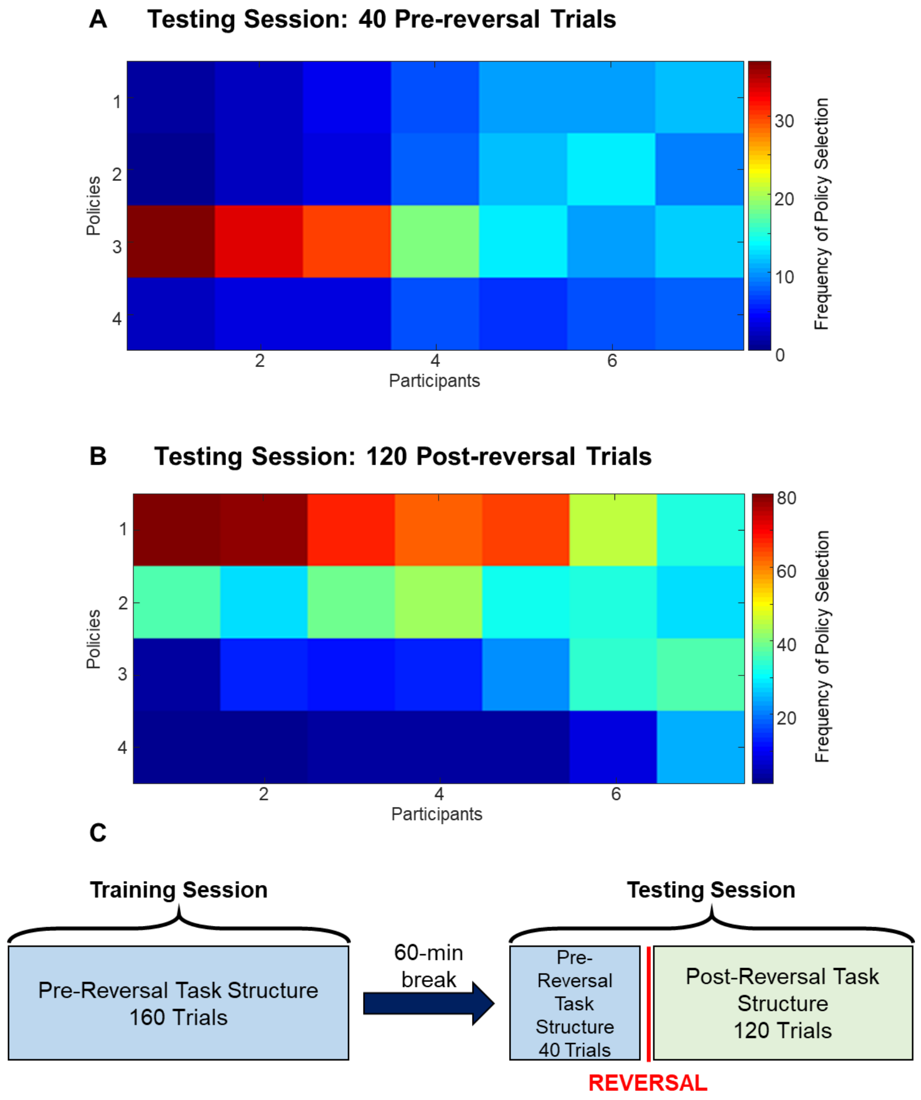
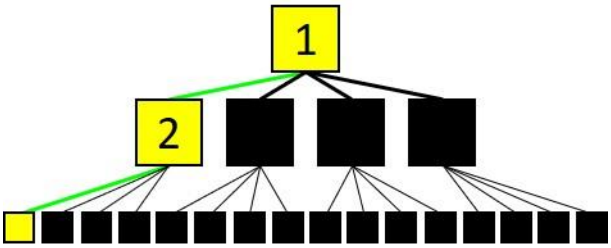


Figure 2.7 Behaviour in Pilot Three Testing Session. (**A**) Heatmap of policy frequency over the first 40 trials (pre-reversal) across participants in pilot three testing session, ordered according to their chi-squared test *p* value. Colour bar indicates frequency of policy selection. Mean of policy three choice frequency across participants was significantly different to chance level: (p = 1.75 x 10^{-4}, sign test). (**B**) Heatmap of policy frequency over the last 120 trials (post-reversal) across the cohort in pilot three testing session, ordered according to individual chi-squared test *p* values. Colour bar indicates frequency of policy selection. Mean of policy one choice frequency across the cohort was significantly different to chance level: (p = 1.11 x 10^{-}

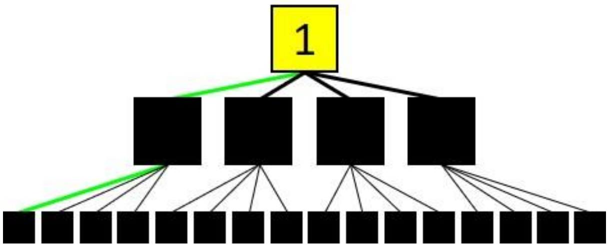
^{9}, sign test). (**C**) Reversal task structure. The training session lasted for 160 trials with a constant pre-reversal task structure; the same structure as that used in pilots one and two. After the 40th trial in the testing session, a reversal occurred. The remaining 120 trials in the testing session used the post-reversal task structure, as detailed in **Figure 2.2C**. For actions at t = 2 from level two unlikely states: from state three, only two participants moved differently to chance pre-reversal, which was expected due to the low probability of visiting this state and the even lower probability of receiving reward moving in either direction from state three. Post-reversal, four participants moved right from state three significantly greater than chance level (the four exploiting participants in the training session), utilising the new second most optimal policy, policy two. The remaining three participants were at chance level; no participants preferred the less optimal route of moving left from state three. From state five, only one participant moved differently to chance in the first 40 trials, indicating that very few participants learned the optimal route to take after reaching state five, as it is an unlikely level-two state. After the reversal, this lack of state five learning remained consistent; one participant moved from state five differently to chance, further suggesting that this state was not visited frequently enough for participants to learn the optimal action.

After the reversal, I expected participants to switch their preferred policy from three to one. During the last 120 trials, six participants showed significant preference for policy one (p = 1.11 x 10^{-9}; sign test, over all trials across participants), showing that most participants noticed that the reversal had occurred and changed their strategy accordingly, whether or not they had displayed exploitative or exploratory behaviours in the training session (**Figure 2.7B,** see also **Figure 2.7C** for structure of reversal onset). This was also reflected in the chi-squared tests, in which the same six participants showed policy choices significantly different to chance ($p < 0.05$, individual chi-squared tests); the participant that did not select a policy significantly different to chance had exhibited similarly exploratory behaviour in the training session (***Insert:*** *To take the post-reversal optimal route from state one, the participant must move left to reach state two, in yellow, then left to earn 25p (70% chance). Green path lines indicate the current optimal policy*).



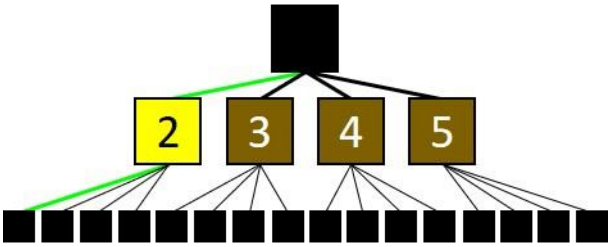
Participants Successfully Changed Actions at Each Timepoint to Follow Optimal Routes

Examining actions at timepoint t = 1, I expected participants to switch from moving right during the first 40 trials pre-reversal, to moving left in the last 120 trials post-reversal. From state one, three participants moved right significantly more than chance in the first 40 trials, and four participants were at chance level; no participants moved left significantly more than chance at t = 1. This was reduced compared to the training session as participants may have explored the environment initially after the one-hour break to ensure that the environment was the same as in the training session, and then selected the optimal policy after a brief exploration. After the reversal occurred, six participants moved left at t = 1 significantly more than chance, following the new optimal policy, policy one, further suggesting that participants successfully learned the post-reversal task structure. The moving average of right moves at t = 1 was high from trial 1-40, then very steeply dropped away after the reversal and remained low,



fluctuating around 20% of actions, showing that participants changed their actions very quickly after the reversal took place, despite participants being uninformed about the reversal in the testing session of the task (*rho* = -0.838, *p* < 0.001; Pearson's correlation) (**Figure 2.8A**). (***Insert:*** *initial state at level one numbered as state one, in yellow. To take the optimal route from state one after the reversal, the participant must move left. Green path lines indicate the post-reversal optimal policy*).

During the first 40 trials, no participants moved reliably differently to chance level from state two, likely due to very few visits to this state. However, after the reversal, six participants moved left from state two significantly more than chance, taking the new optimal route. In the right arm of the task, five out of seven participants moved left from state four pre-reversal, following the optimal policy, and this was reduced to four participants after the reversal. Despite the post-reversal optimal action changing to moving right from state four, this may have been due to participants persisting with the optimal policy shortly after the reversal took place, displaying perseverative behaviours. No participants moved right from state four significantly more than chance, suggesting that this new optimal action after the reversal was not learned by any participants (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state post-reversal is state two, in yellow. Green path lines indicate the post-reversal optimal policy*).



To summarise, participants who learned the optimal policy in the training session were able to recall the pre-reversal optimal policy after the lengthened break between task sessions. Most participants were also able to notice the reversal very quickly and adapt their policy selection accordingly, regardless of whether they had exhibited exploitative or exploratory behavioural profiles in the training session.

Reaction Times Following Likely States Remained Faster than Following Unlikely States Across the Testing Session

Similar to the training session, RTs after likely states were significantly faster than RTs after unlikely states across all 160 trials of the testing session ($t(6) = -2.83$, $p = 0.0299$, paired t-test) (**Figure 2.8B**). However, when splitting the testing session into the pre-reversal and post-reversal trials, this significance disappeared, although RTs remained relatively consistent in that RTs after likely states were slightly faster than those after unlikely states across participants (pre-reversal: $t(6) = -2.44$, $p = 0.0502$; post-reversal: $t(6) = -2.13$, $p = 0.0769$; paired t-tests) (**Figure 2.8C-D**). This slight reduction in the differences between RTs after likely and unlikely states may be due to participants readjusting to the task following the one-hour break, and then moving more tentatively following the reversal.

Figure 2.8

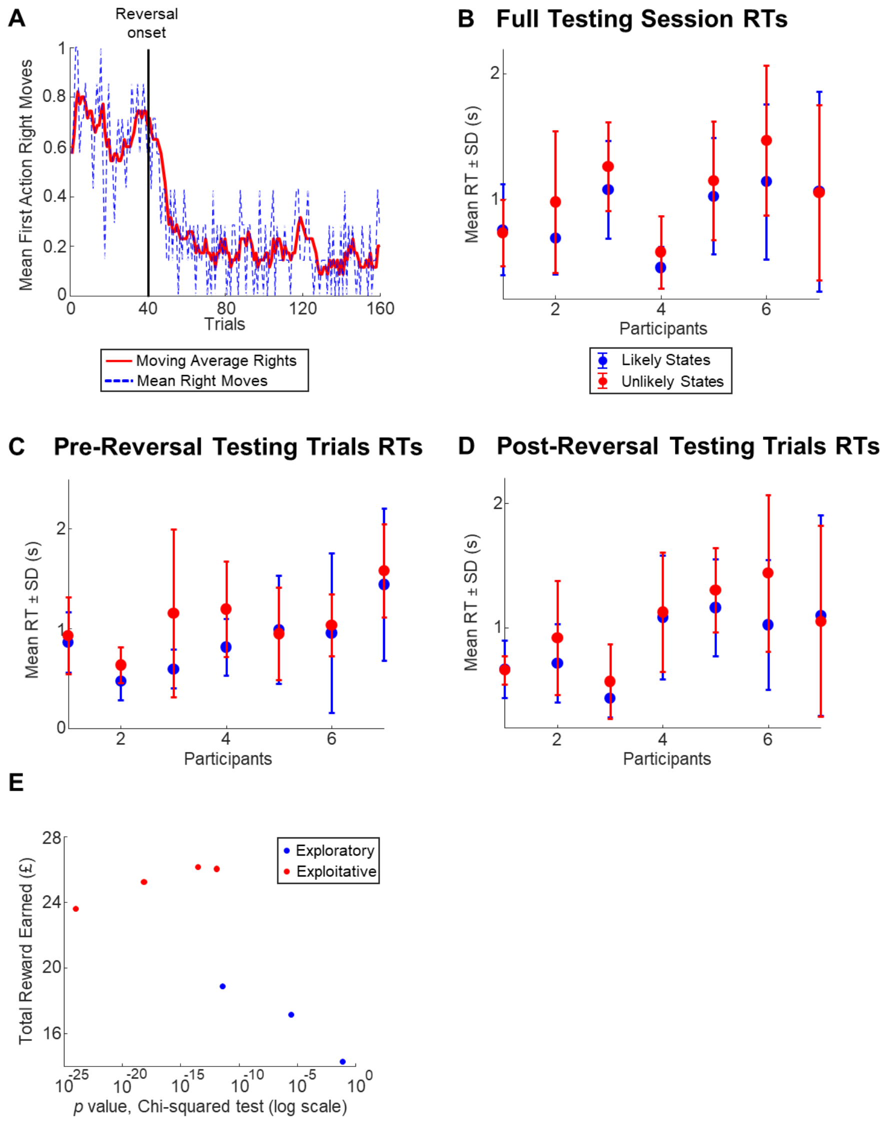


Figure 2.8 Actions in Pilot Three Testing Session. (**A**) Moving average across the cohort of right moves taken in the first action choice of each trial (red solid line), and the mean of right moves taken at $t = 1$ per trial (blue dashed line) over all trials in the training session of pilot three. Moving average used a window of five trials. The moving average of right actions significantly decreased immediately following the reversal onset at trial 40, and gradually reached a plateau (*rho* = -0.838, $p < 0.001$). (**B**) RT across all 160 trials for actions in likely (blue) and unlikely (red) intermediary states in the testing session of pilot three, mean ± SD. Likely

states had significantly faster RTs than unlikely states over all 160 trials across the cohort (*t*(6) = -2.83, *p* = 0.0299; paired *t*-test). (**C**) RT across the first 40 trials for actions in likely and unlikely intermediary states, mean ± SD. There was no significant difference between RTs over the first 40 trials in likely versus unlikely states across the cohort (*t*(6) = -2.44, *p* = 0.0502; paired *t*-test). (**D**) RT across the last 120 trials for actions in likely and unlikely intermediary states, mean ± SD. There was no significant difference between RTs over the last 120 trials in likely versus unlikely states across the cohort (*t*(6) = -2.13, *p* = 0.0769; paired *t*-test). Participants are ordered by performance (*p* values of policy selection chi-squared tests) in all RT plots. (**E**) Relationship between policy choice (chi-squared *p* value) and total reward earned across pilot three testing session. Participants that were categorised as displaying exploratory (blue) and exploitative (red) behaviours in the training session showed a reduced difference in chi-squared *p* value between the groups compared with the training session, but the strong negative relationship between chi-squared *p* value (i.e. policy selection) and total reward earned across the testing session was evident, similarly to the training session. SD = standard deviation.

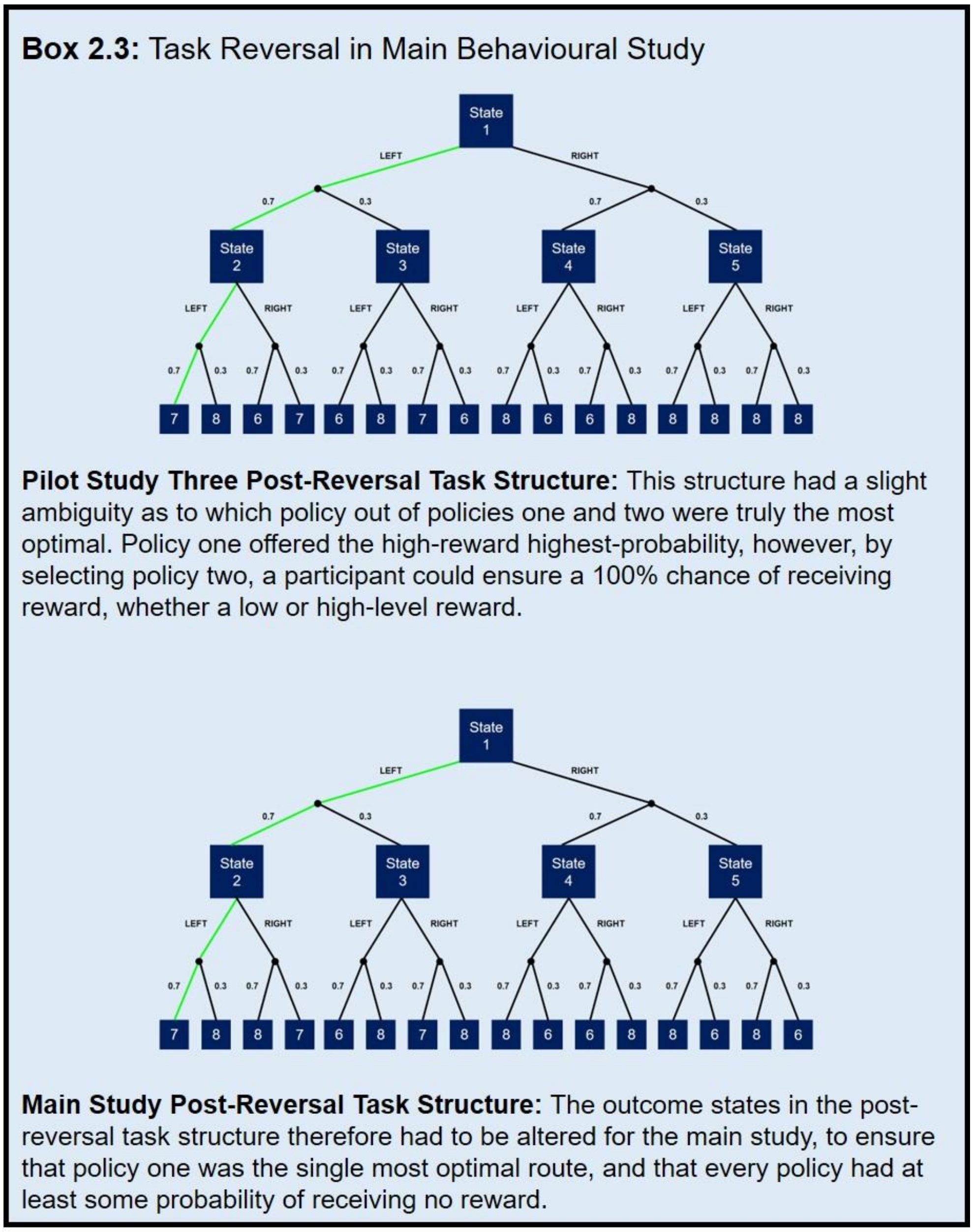
Relationship Between Policy Choice and Total Reward Earned

I then examined the *p* values of the chi-squared tests on policy selection per participant, and the total reward earned across all 160 trials of the testing session. Unlike the training session, participants could not be separated as distinctly into an *exploratory* group and an *exploitative* group compared with the training session, however a strong negative relationship between the chi-squared *p* values and total reward remained evident, in that total reward earned was higher when a participant had a stronger preference for one policy, i.e. had a smaller chi-squared *p* value (**Figure 2.8E**).

2.5 Main Behavioural Study

Based on the successful findings of pilot three, in that most participants, whether they displayed either exploitative or exploratory behaviours in the training session, were able to notice the reversal as it occurred and learned to adapt to the new optimal policy. Therefore, further investigation with a larger participant cohort was required (as pilot three consisted of only seven participants), with the addition of a minor change to the outcome states after the reversal, to ensure that there was no ambiguity as to which policy was the most optimal, i.e. which policy was associated with the highest-level highest-probability reward (**Box 2.3**).

Box 2.3



Box 2.3: Task Reversal in Main Behavioural Study

Pilot Study Three Post-Reversal Task Structure: This structure had a slight ambiguity as to which policy out of policies one and two were truly the most optimal. Policy one offered the high-reward highest-probability, however, by selecting policy two, a participant could ensure a 100% chance of receiving reward, whether a low or high-level reward.

Main Study Post-Reversal Task Structure: The outcome states in the post-reversal task structure therefore had to be altered for the main study, to ensure that policy one was the single most optimal route, and that every policy had at least some probability of receiving no reward.

Box 2.3 Outline of the change in outcome state structure between pilot three and the main behavioural study.

2.5.1 Methods

Participants

Twenty-five participants (14 females) were recruited and tested in the main behavioural study. All participants were recruited using the same exclusion criteria and tested on the behavioural paradigm in the exact same way as in the three pilot studies, and were aged 18 or over (mean = 23.6 ± 3.05 SD).

Behavioural Paradigm

In the main behavioural study, the task structure was very similar to pilot three, with a small change in the outcome states after the reversal in the testing session (**Figure 2.2D, Box 2.3**). This change was to ensure that there was one single unambiguous optimal policy after the reversal: policy one. I therefore altered the outcome states so that there remained a greater-than-zero chance of no reward in every possible route in the task structure. By choosing policy one, participants had both the highest probability of winning 25p and could also earn the highest average reward of 14.35p per trial by continuously selecting this policy.

Exclusion Criteria

For the main study, the upper 95th percentile of the Monte Carlo simulation was £28.20 with a mean of £25.32 (expected reward of £25.32) (**Appendix Figure A1C**). Two participants did not score greater than the mean, and no additional participants scored below the 95th percentile; 23 participants scored above the 95th percentile.

Statistical Analyses

All statistical analyses performed for pilot three were replicated exactly for the main behavioural study.

2.5.2 Results

Training Session

Policy Frequency Showed Significant Optimal Policy Selection Across Trials

In the training session, policy three was again the optimal policy throughout. The mean frequency across all trials and participants of policy three choice was significantly above chance (p = 9.33 x 10^{-12}; sign test) (**Figure 2.9A**). On an individual basis, 20 out of 25 participants chose policy three significantly more than chance across trials ($p < 0.05$, sign test). By conducting individual chi-squared tests, I found that 24 out of 25 participants chose one of the policies significantly more than chance. Of the three participants that had a preferred policy that was not policy three, the optimal policy, two preferred policy four and one preferred policy one.

Figure 2.9

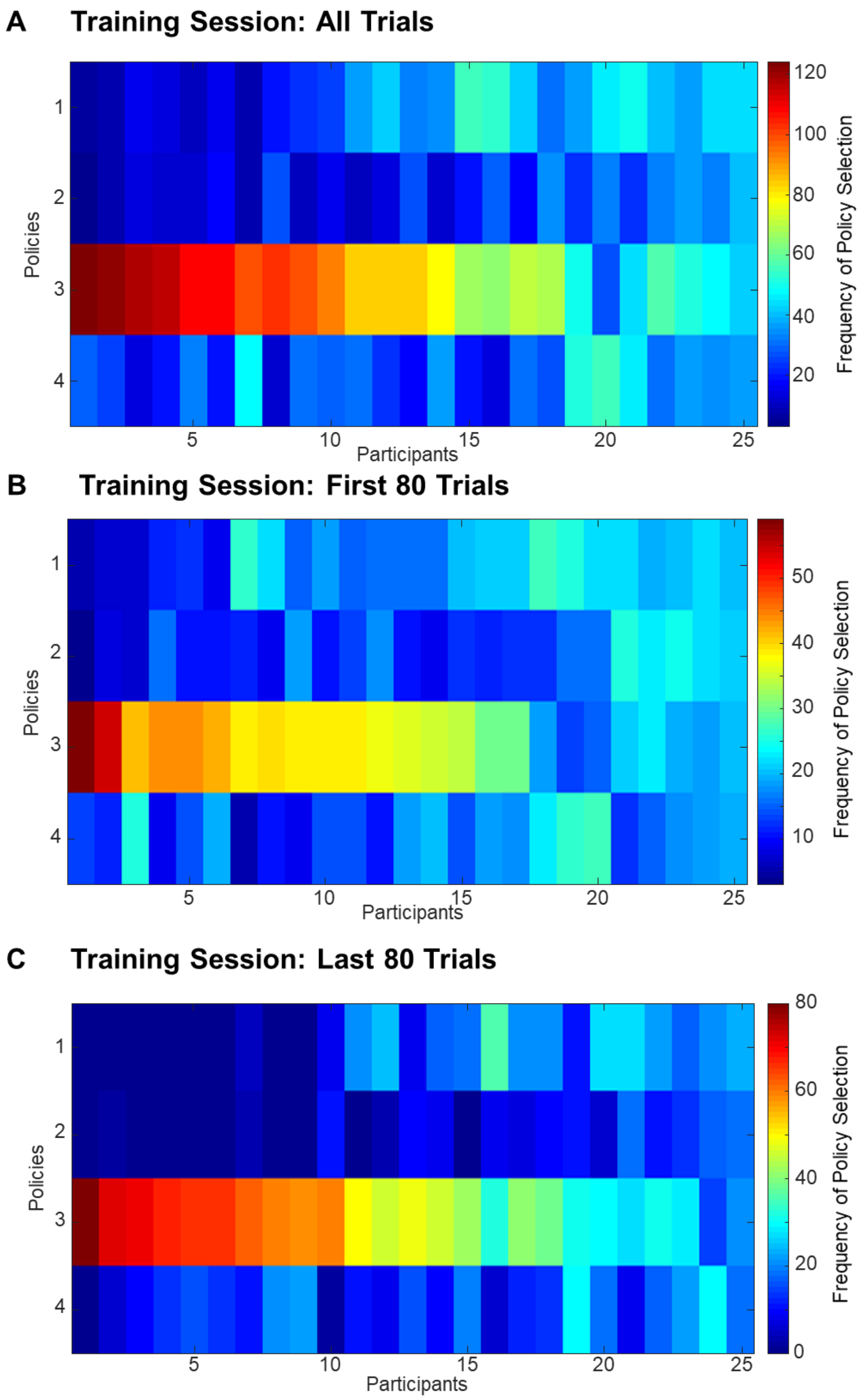
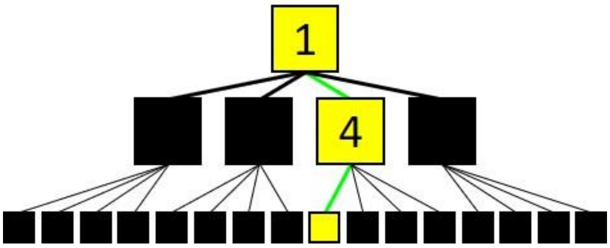


Figure 2.9 Behaviour in Main Study Training Session. (**A**) Heatmap of policy frequency over all 160 trials across the cohort in the training session of the main behavioural study, ordered

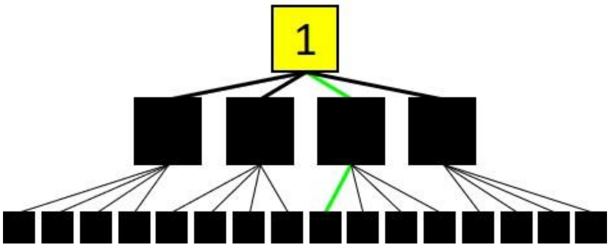
according to their chi-squared test *p* value. Colour bar indicates frequency of policy selection across all trials in the testing session. Mean of policy three choice frequency across the cohort was significantly different to chance level: (p = 9.33 x 10^{-12}, sign test). (**B**) Heatmap of policy frequency over the first 80 trials across the cohort in the main study training session, ordered according to their chi-squared test *p* value. Colour bar indicates frequency of policy selection. Mean of policy three choice frequency across the cohort was significantly different to chance level: (p = 0.00290, sign test). (**C**) Heatmap of policy frequency over the last 80 trials across the cohort in the main study training session, ordered according to individual chi-squared test *p* values. Colour bar indicates frequency of policy selection. Mean of policy three choice frequency across the cohort was significantly different to chance level: (p = 1.57 x 10^{-10}, sign test). For actions at t = 2 from unlikely level two states, four participants moved right significantly more than chance from state three, and 11 participants moved significantly different to chance from state five ($p < 0.05$, individual chi-squared tests). Participants did not show a strong overall preference for moving in any direction over the other from state five (seven preferred left, four preferred right). This, and most participants being at chance level for actions from these states, is likely due to the infrequency of visiting these low-probability states and the low probability of reward from moving in either direction from both states.

I then looked at policy selection in the first and last 80 trials of the training session to investigate participants' learning across the training session. I expected to observe more random action choices at the start of the session while participants were exploring and learning the structure of the environment, and towards the end I expected participants to choose the optimal policy most frequently, as with pilot three. The frequency of policy choice between the first 80 and last 80 trials increased across the cohort (first 80: p = 0.00290, last 80: p = 1.57 x 10^{-10}; sign tests), and policy three remained the most frequent policy choice when examining the first and last 80 trials separately (**Figure 2.9B-C**). In the first 80 trials, 17 participants chose policy three significantly more than chance, and in the last 80 trials this increased to 21. Individual chi-squared tests also revealed that in the first 80 trials, 21 participants selected a policy significantly more than chance, and this increased to 24 during the last 80 trials. Therefore, most participants were able to successfully learn the task structure and navigate to the highest available reward (***Insert:*** *initial state at level one numbered as state one. To take the optimal policy, policy three, the participant must move right to reach state four, then left to earn 25p (70% chance); optimal states in yellow. Green path lines indicate the optimal policy*).



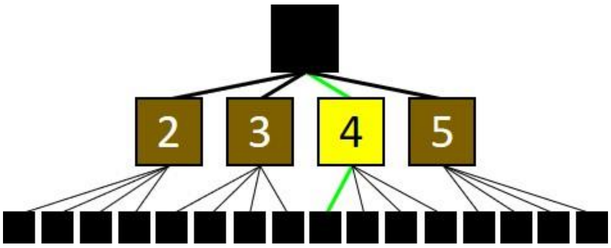
Participants Made Optimal Action Choices Following Initial State and Likely Intermediary States

I then examined individual action choices at each timepoint in the trial, in addition to policy choices. In the training session, 17 out of 25 participants moved right significantly more than moving left for the first action at $t = 1$ across all 160 trials ($p < 0.05$, individual chi-squared tests), which aligns with policy three being chosen most frequently across the cohort. The remaining eight participants were at chance; no participants moved left at $t = 1$ significantly greater than chance. Looking at the first and last 80 trials, 11 participants moved right significantly more than chance in the first 80 trials, and this increased to 19 in the last 80 trials ($p < 0.05$, individual chi-squared



tests), showing that participants learned over the course of the training session that moving right first was optimal. Consistent with previous pilot studies, the moving average of right moves at t = 1 steadily increased across the 160-trial session (*rho* = 0.859, p = 1.10 x 10^{-47}; Pearson's correlation), also indicating that participants gradually learned the optimal policy and increased their frequency of right moves at t = 1 across the training session (**Figure 2.10A**) (***Insert:*** *initial state at level one numbered as state one, in yellow. To take the optimal route from state one, the participant must move right, pre-reversal. Green path lines indicate the pre-reversal optimal policy*).

For likely intermediary states (*see **Figure 2.9** legend for unlikely states*), 10 out of 25 participants moved left from state two significantly more than chance, and no participants moved right from state two more than chance (p < 0.05, individual chi-squared tests). From state four, 20 participants moved left significantly more than chance (p < 0.05, individual chi-squared tests), therefore following the optimal policy. Of the five remaining participants, one moved right significantly more than chance and four were at chance level (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state pre-reversal is state four, in yellow. Green path lines indicate the pre-reversal optimal policy*).



In summary, most participants were able to learn the optimal route in the training session of the task, and there was clear evidence of learning increasing over the course of the training session. Only one participant did not show a significant preference for any policy, therefore this participant may have required more experience of the environment, i.e. more trials, to successfully learn the task structure.

Figure 2.10

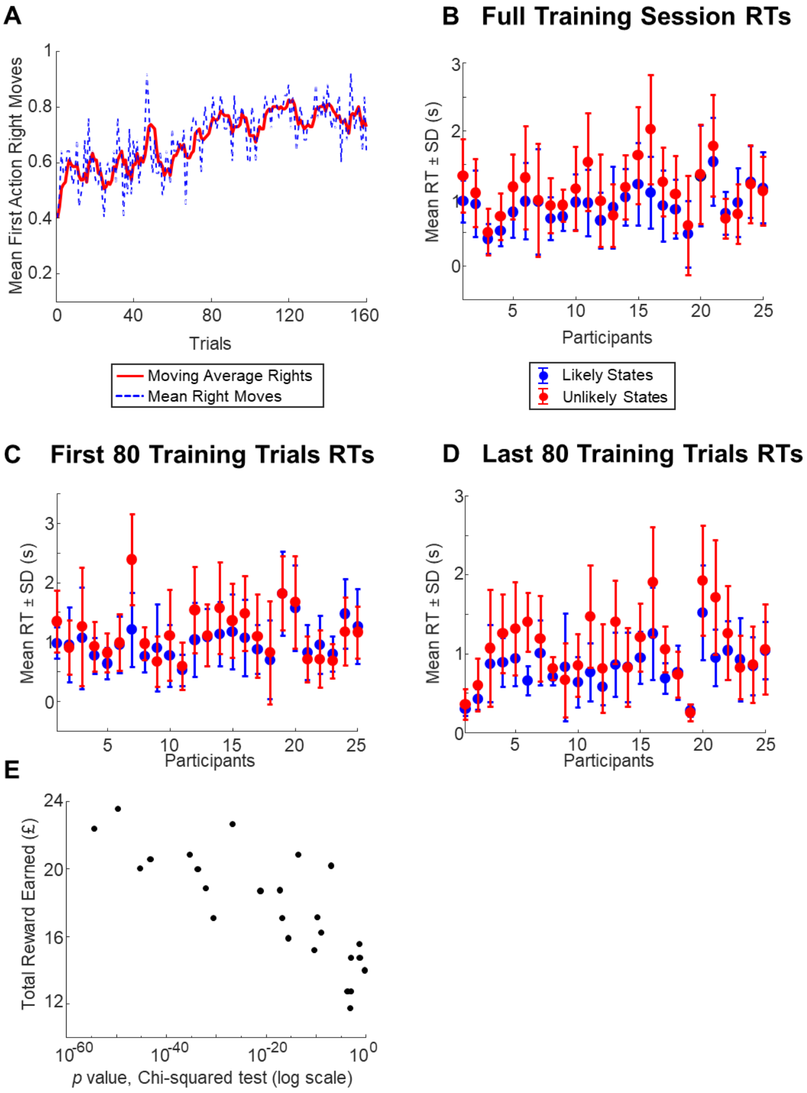


Figure 2.10 Actions in Main Study Training Session. (**A**) Moving average across participants of right moves taken in the first action choice of each trial (red solid line), and the mean of right moves taken at t = 1 per trial (blue dashed line) over all trials in the training session of the main study. Moving average used a window of five trials. The moving average of right actions significantly increased as trials progressed through time (*rho* = 0.859, p = 1.10 x 10^{-47}). (**B**) RT across all 160 trials for actions in likely (blue) and unlikely (red) intermediary states in the training session of the main study, mean ± SD. Likely states had significantly faster RTs than unlikely states over all 160 trials across participants ($t(24)$ = -4.22, p = 3.03 x 10^{-4}; paired

t-test). (**C**) RT across the first 80 trials for actions in likely and unlikely intermediary states, mean ± SD. Likely states had significantly faster RTs than unlikely states in the first 80 trials across the cohort ($t(24) = -2.36$, $p = 0.0266$; paired *t*-test). (**D**) RT across the last 80 trials for actions in likely and unlikely intermediary states, mean ± SD. Likely states had significantly faster RTs than unlikely states in the last 80 trials across the cohort ($t(24) = -4.69$, $p = 9.12 \times 10^{-5}$; paired *t*-test). Participants are ordered by performance (*p* values of policy selection chi-squared tests) in all RT plots. (**E**) Relationship between policy choice (chi-squared *p* value) and total reward earned across the main study training session. There was a strong negative relationship between the chi-squared *p* value (i.e. policy selection) and total reward earned across the training session. SD = standard deviation.

Significantly Faster Reaction Times Following Likely Versus Unlikely Level Two States

In the training session, across all 160 trials RTs for actions at t = 2 following states two and four (level two likely states) were significantly faster than that following states three and five (level two unlikely states) across the cohort ($t(24) = -4.22$, $p = 3.03 \times 10^{-4}$; paired t-test) (**Figure 2.10B**). This was consistent when examining the first 80 and last 80 trials separately, with an increase in the difference between RTs after likely and unlikely states (first 80: $t(24) = -2.36$, $p = 0.0266$; last 80: $t(24) = -4.69$, $p = 9.12 \times 10^{-5}$; paired t-tests) (**Figure 2.10C-D**). Therefore, participants seemed to learn the task structure within the first 80 trials, and were able to navigate the environment having learned which intermediary level-two states they were more likely or less likely to enter for the remaining 80 trials of the training session.

Relationship Between Policy Choice and Total Reward Earned

As with pilot three, I also examined the relationship between total reward earned by participants in the training session and their chi-squared p values for policy selection (**Figure 2.10E**). Unlike the pilot three training session, participants could not be separated into distinct groups but the strong negative relationship between p value and total reward remained consistent with that of the pilot three testing session, i.e. participants who had stronger preferences for one policy (had smaller chi-squared p values) earned higher rewards overall.

Testing Session

Policy Frequency Across Trials Showed Adaptation To Task Reversal

For all testing session analyses, trials were split into the first 40 trials (pre-reversal) and the last 120 trials (post-reversal), in the same way as for pilot three. In the first 40 trials, 22 out of 25 participants chose policy three significantly more than chance, indicating that they successfully recalled the optimal policy from the training session and were able to apply this to the testing session ($p < 0.05$, individual sign tests) (**Figure 2.11A**). This was also significant across the cohort ($p = 1.23 \times 10^{-5}$; sign test, across the cohort). By conducting individual chi-squared tests, I found that 23 participants had a significant preference for one policy over others ($p < 0.05$, individual chi-squared tests).

Figure 2.11

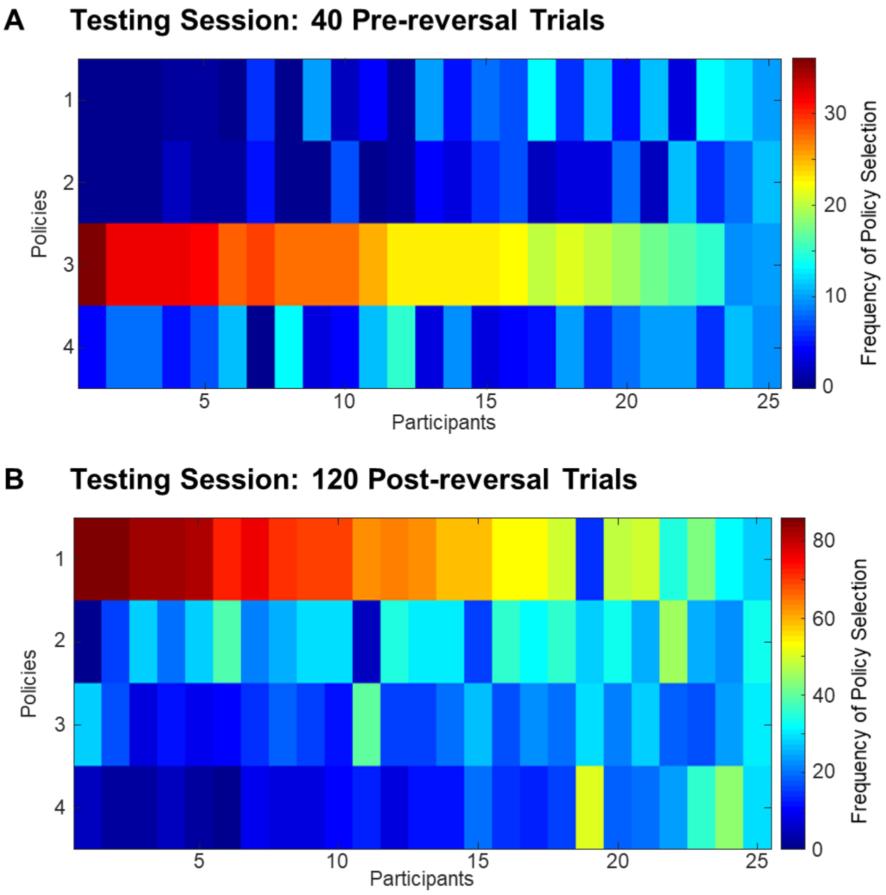
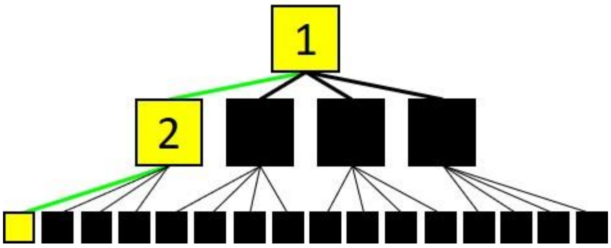


Figure 2.11 Behaviour in Main Study Testing Session. (**A**) Heatmap of policy frequency over the first 40 trials (pre-reversal) across the cohort in the main study testing session, ordered according to their chi-squared test *p* value. Colour bar indicates frequency of policy selection. Mean of policy three choice frequency across participants was significantly different to chance level: (p = 1.23 x 10^{-5}, sign test). (**B**) Heatmap of policy frequency over the last 120 trials (post-reversal) across the cohort in the main study testing session, ordered according to their chi-squared test *p* value. Colour bar indicates frequency of policy selection. Mean of policy one choice frequency across the cohort was significantly different to chance level: (p = 1.03 x 10^{-8}, sign test). For actions at t = 2 from unlikely level-two states: from state three, only one participant moved significantly different to chance before the reversal, and eight participants did not visit state three at all before the reversal. After the reversal, this increased to 15 participants who moved right significantly more than chance from state three. Four

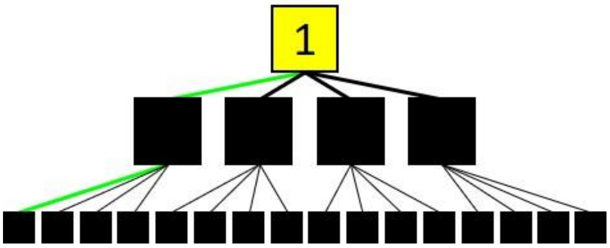
participants moved left significantly more than chance, and six were at chance ($p < 0.05$, individual chi-squared tests). From state five, six participants moved from state five differently to chance before the reversal, and after the reversal two participants moved differently to chance, with participants either preferring to move left or right; there was no consensus between those with a significant preference. This is likely because the most likely outcome states both pre- and post-reversal from state five did not offer any reward, and participants were not expected to have any action preference due to the low reward probabilities.

After the reversal in the last 120 trials, 22 participants chose policy one, the new optimal policy, significantly more than chance ($p < 0.05$, individual sign tests) (**Figure 2.11B**). One participant, however, selected policy one significantly less than chance and appeared to persist in exploring the right arm of the task even after the reversal, showing a significant preference for policy four. The preference for policy one in the last 120 trials was significant across the cohort ($p = 1.03 \times 10^{-8}$; sign test, across the cohort). The individual chi-squared tests revealed that 24 participants had a significant policy preference, and only one participant continued to explore each policy equally throughout the testing session, showing that most participants successfully learned the new task structure after the reversal and were able to infer the new optimal policy (***Insert:*** *To take the post-reversal optimal route from state one, the participant must move left to reach state two, in yellow, then left to earn 25p (70% chance). Green path lines indicate the current optimal policy*).



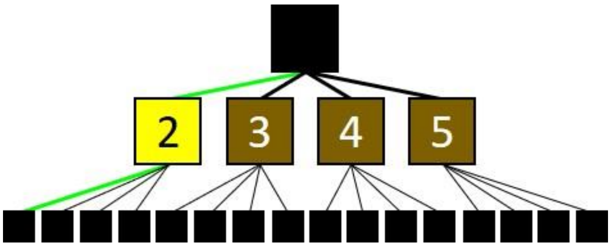
Participants Made Optimal Action Choices Following Initial State and Likely Intermediary States, Before and After the Reversal

Consistent with my previous analyses, I then examined action choices at each timepoint in the trial. In the first 40 trials, 16 out of 25 participants moved right significantly more than chance at $t = 1$ thus following the optimal route, and no participants moved left more than chance; nine were at chance level ($p < 0.05$, individual chi-squared tests). After the reversal, 20 participants moved left at $t = 1$ significantly more than chance ($p < 0.05$, individual chi-squared tests), indicating that most participants learned the reversal and were able to switch their strategies, and also that four participants who had not followed the optimal policy pre-reversal were able to learn and employ the new optimal policy. Four participants did not move reliably different to chance and one participant moved right significantly more than chance at $t = 1$. When looking at the moving average of right moves at $t = 1$, from trial one to 40 the moving average steadily increased, but after trial 40 once the



reversal has occurred, this dropped steeply immediately after the reversal, showing that participants noticed the change in task structure very quickly and changed their actions accordingly. Overall, there was a strong negative correlation with the moving average of right moves at t = 1 as the trials progressed through the testing session (*rho* = -0.882, p = 1.32 x 10^{-53}; Pearson's correlation) (**Figure 2.12A**) (***Insert:*** *initial state at level one numbered as state one, in yellow. To take the optimal route from state one after the reversal, the participant must move left. Green path lines indicate the post-reversal optimal policy*).

For second action choices at t = 2, participants did not explore the left arm of the task during the first 40 trials and therefore showed no preference for optimal routes from states two or three, which suggests that participants remembered that the right arm of the task was most optimal during the training session. From state two, five participants moved left significantly more than chance prior to the reversal with 20 participants who did not move differently to chance. However, 21 participants moved left from state two significantly more than chance post-reversal ($p < 0.05$, individual chi-squared tests), following the new optimal policy (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state post-reversal is state two, in yellow. Green path lines indicate the post-reversal optimal policy*).



In the right arm of the task, more participants appeared to move at chance level at t = 2 following the reversal, as they switched from moving right at t = 1 before the reversal to moving left after the reversal, and therefore were not exploring the right arm of the task to the same extent. From state four, 19 participants moved left at t = 2 significantly more than chance prior to the reversal, following the optimal policy, with six who were not reliably different to chance level ($p < 0.05$, individual chi-squared tests). After the reversal, this reduced to 10 participants who moved left significantly more than chance, and two participants who moved right significantly more than chance. It became more optimal to move right following state four after the reversal, but as the overall frequency of visits to the right arm of the task decreased after the reversal it is unlikely that this was learned by the cohort (*for action choices at t = 2 from unlikely states, see* ***Figure 2.11*** *legend*).

Figure 2.12

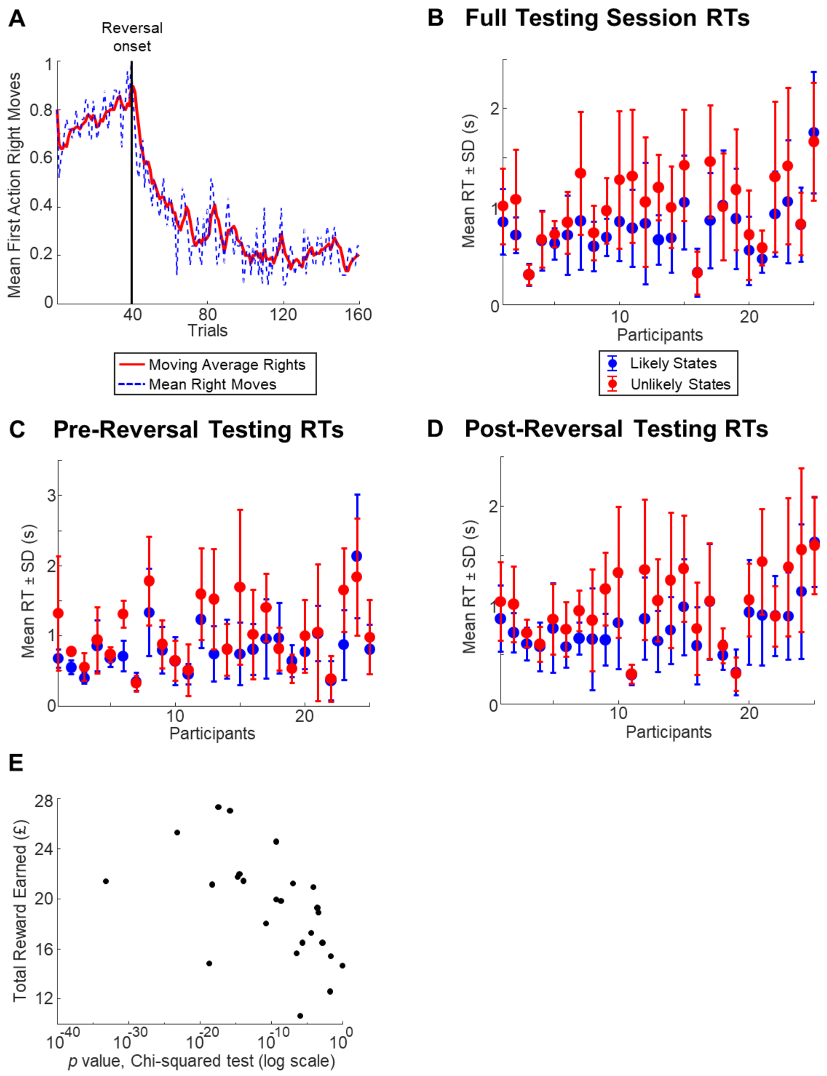


Figure 2.12 Actions in Main Study Testing Session. (**A**) Moving average across the cohort of right moves taken in the first action choice of each trial (red solid line), and the mean of right moves taken at t = 1 per trial (blue dashed line) over all trials in the training session. Moving average used a window of five trials. The moving average of right actions significantly decreased immediately following the reversal onset at trial 40, and gradually reached a plateau (*rho* = -0.882, p = 1.32 x 10^{-53}). (**B**) RT across all 160 trials for actions in likely (blue) and unlikely (red) intermediary states in the testing session of the main study, mean ± SD. Likely states had significantly faster RTs than unlikely states over all 160 trials across

participants ($t(24) = -5.91$, $p = 4.21 \times 10^{-6}$; paired *t*-test). (**C**) RT across the first 40 trials for actions in likely and unlikely intermediary states, mean ± SD. Likely states had significantly faster RTs than unlikely states over the first 40 trials across the cohort ($t(24) = -3.61$, $p = 0.00140$; paired *t*-test). (**D**) RT across the last 120 trials for actions in likely and unlikely intermediary states, mean ± SD. Likely states had significantly faster RTs than unlikely states over the last 120 trials across the cohort ($t(24) = -5.94$, $p = 3.96 \times 10^{-6}$; paired *t*-test). Participants are ordered by performance (*p* values of policy selection chi-squared tests) in all RT plots. (**E**) Relationship between policy choice (chi-squared *p* value) and total reward earned across the main study testing session. The strong negative relationship between chi-squared *p* value (i.e. policy selection) and total reward earned across the testing session was consistent with the training session, although more skewed towards higher *p* values. SD = standard deviation.

Overall, the behaviour of participants in this study was very similar to that in pilot three, as most participants successfully recalled the optimal policy from the training session and employed this strategy in the pre-reversal trials of the testing session. After the reversal, once again most participants were quick to recognise the reversal and seek out the new optimal policy, and proceed to exploit it.

Reaction Times were Significantly Faster Following Likely versus Unlikely Intermediary States Throughout Testing Session

Across all 160 trials of the testing session, RTs after likely level-two states, i.e. at t = 2, were significantly faster than after unlikely level-two states (t(24) = -5.91, p = 4.21 x 10^{-6}; paired t-test) (**Figure 2.12B**), consistent with the training session. In the first 40 trials, RTs following likely states at t = 2 were also faster than that following unlikely states (t(24) = -3.61, p = 0.00140; paired t-test) (**Figure 2.12C**), and this significance increased in the last 120 trials of the testing session (t(24) = -5.94, p = 3.96 x 10^{-6}; paired t-test), suggesting that even after the reversal, participants were able to adapt to the changes and relearn the task environment (**Figure 2.12D**).

Relationship Between Policy Choice and Total Reward Earned

When looking at the relationship between policy choice and total reward earned across the testing session of the main study, I once again saw a strong negative association similar to that in the testing session of pilot three and the training session of the main study, although here this was slightly more skewed towards higher chi-squared p values (**Figure 2.12E**). This further reinforces the suggestion that participants who took a more exploitative strategy to this task, i.e. those who showed a strong preference for one policy over others, earned the highest overall levels of reward.

2.6 Discussion

From conducting these behavioural studies, I found that participants were able to learn the structure of the virtual environment well and were able to form and apply preferred strategies to the task in order to obtain reward, although these strategies may not be the overall optimal strategy. In pilots one and two, all participants showed a significant preference for one policy, regardless of whether this was the optimal policy. Fourteen participants in pilot one (70% of participants), seven in pilot two (77.8%), four in pilot three training session (57.1%) and 21 in the main behavioural study training session (84%) preferred the optimal policy, suggesting successful learning of the location of high-level high-probability reward. Also, in all studies apart from pilot two, participants' RTs from likely states significantly decreased after training and responded faster to familiar states compared to those that are less familiar.

The number of participants that scored above the threshold for minimal learning as used by Gläscher *et al*. (the 95th percentile of the Monte-Carlo simulation), greatly increased between pilot one and the main study; 12 out of 20 participants in pilot one passed this threshold, but 23 out of 25 participants in the main study scored above the threshold. This indicates that the improved task structure in the main study was effective in increasing participants' learning of the task and their ability to infer the optimal policy and earn rewards, and was effective in the investigation of decision-making and uncertainty.

The study by Gläscher *et al*. focused primarily on using fMRI to identify neural signatures of state prediction errors, and distinguishing these from the already well-established neural signatures of reward prediction errors (McClure *et al.*, 2003; O'Doherty *et al.*, 2003), in the context of RL. They also used computational models of model-based and model-free learning in addition to a hybrid model, which comprised a combination of model-free and model-based learning. They found that, through examining behavioural data, that the actions of their participants could not be completely explained by model-free learning theory.

In line with this, their modelling and fMRI results point towards both learning approaches acting together for optimal action selection, as their HYBRID learner (which combined aspects of model-free and model-based learning) better explained participants' behavioural data (Gläscher *et al.*, 2010). The findings in my behavioural studies also point towards participants using aspects of both model-free and model-based learning approaches, as some participants successfully learned optimal routes from fixed training, but other participants chose to explore the environment freely before settling on selecting the optimal policy. Examining the aspects of model-free and model-based learning was the main motivation behind the analysis of individual actions at each timepoint, as optimal actions at t = 1 suggests use of model-based learning. In pilot two, I observed an increase in optimal action selection in the testing session, suggesting that model-free learning could not solely explain the behaviour of participants in this study.

These findings are in line with Daw *et al*. who investigated model-free versus model-based learning models in combination with human fMRI and behavioural data (Daw *et al.*, 2011). In this study, hallmarks of both model-free and model-based learning were found in human behaviour, alongside the combined recruitment of brain networks associated with both learning approaches.

In the Gläscher *et al*. study, there was a reduced emphasis on analysis of behavioural data alone, as they only examined the first action taken by participants at t = 1 in the very first trial of session two, using a one-tailed sign test. Here, I build on this and provide a more extensive behavioural analysis of participants' actions and policy selection profiles, which informed the various adaptations made to the task structure between studies as described above.

A second key difference to this study, is that in Gläscher *et al*., the images used in their paradigm were abstract fractal images, therefore their task was non-spatial. In these studies, I chose to instead use images of natural environments in an aim to replicate navigation through a realistic real-world environment, thus adding a spatial navigation component to the decision-making paradigm.

Throughout the analyses here, I frequently describe particular policies as 'optimal'. Here, I use the word 'optimal' to describe the policies which have the highest probability of the highest level of reward, i.e. the route that follows the path with 70% probability after each action and ends with a 25p reward. Gläscher *et al*. (2010), to examine the relationships between behaviour and the neural state prediction error signals, defined the 'correct choice' in their task as the choice of action associated with the highest expected reward value (reward magnitude x true transition probability) (Gläscher *et al.*, 2010). However, it may be the case that some participants found alternative routes that were optimal for them in a different capacity.

While most participants across all four experimental cohorts exploit the 'optimal' policy significantly more than other policies, a small number of participants show exploitative preferences for policies other than the policy I have defined as optimal. For example, a participant may have initially explored the environment and, due to the probabilistic task structure, may have entered the low-reward (state six) from policy one more frequently than the high-reward state (state seven) following from state three. This participant may have therefore built stronger prior beliefs about state six, and having less uncertainty about the rewards presenting in state six compared with state seven, chose to exploit policy one. Alternatively, less physical effort may have been required from participants to press the same key for both actions within a trial, as opposed to switching the key presses between states. Thus they may have felt that either policy one or policy four were more optimal for them as less motor control/effort was expended in the task.

One possibility is that some participants may have experienced a primacy effect. A recent study (Rey *et al.*, 2020) investigated this primacy effect in a complex decision-making task, by presenting individuals with different objects (in this case, cars) which were described by either positive or negative attributes. Participants were asked immediately following the task which car they would hypothetically buy/which car they thought was the 'best' based on these attributes. Rey *et al*. (2020) found that the car that was presented alongside all of its positive attributes first, followed by its negative attributes, was significantly preferred to another car with identical attributes, but the negative attributes were systematically presented first, with the positive attributes presented after, showing that the attributes presented earlier in the task had greater influence over the decisions made by participants.

Another study by Park and Melamed (2016) examined the effects of presentation order on justice evaluations in a financial investment-based paradigm and showed a strong primacy effect, in that negative reward instability appeared to have a stronger effect on participants' justice evaluations when presented earlier in the task compared to stable rewards (Park and Melamed, 2016). Further, under-rewarded trials were perceived as more strongly negative compared with the positive effect of over-rewarded trials based on participants' self-report, indicating an asymmetry in participants' perceptions of loss versus gain, in line with prospect theory (Levy, 1992).

This could be applied to my studies presented here, as a participant whose initial experience of the 'optimal' policy may not have entered the most rewarding state (30% chance of earning 0p by taking policy three), and their first experience of a less rewarding policy may have offered them the low-level reward. Therefore they may have believed early in the task that receiving the lower reward via policy one was more likely than receiving the higher reward via policy three, despite the transition probabilities of the two states being identical.

In pilot one, I recruited 20 participants to complete the behavioural task. For the initial study, I decided to recruit a large cohort, both to examine varying behavioural profiles across the group and to confirm that participants were able to successfully learn the task, and assess if any modifications to the task structure were necessary. I subsequently recruited much smaller cohorts for the following two confirmatory pilot studies, pilots two and three, since minor modifications were made to the task structure and the purpose of these pilot studies was to confirm if the modified task structures were effective in improving the learning of optimal policies in participants. Once the task structure had been finalised, I then ran a larger-scale behavioural study recruiting 25 participants, here named the 'main study', to identify a range of behavioural profiles similar to pilot one, and how these changed in response to the contextual reversal. This task structure was then used in all subsequent Active Inference modelling and in the fMRI study, with a small reduction in trial number for the testing session of the fMRI study due to scanning time restrictions.

A key limitation of pilot one regarding the presentation of the task to participants was that the images were not congruent. Therefore, as described above, this was changed in pilot two for more congruent images. These different types of environments were easier to distinguish due to their vastly different colour scheme and components of the scenes.

Another limitation of the task design in pilot one was that the three different rewards were represented by different objects; the high-level and low-level rewards were both represented by images of jewels (colour of the jewel indicated reward level), however the no-reward state was represented by an image of an empty treasure chest. This may have resulted in inconsistency between participants in remembering certain routes over others in the absence of knowledge of reward-state mappings during the training session. In pilot two and all subsequent studies, the treasure chest was replaced by a silver jewel to standardise all images relating to reward or lack thereof.

In summary, these behavioural studies were successful in optimising a probabilistic Markov decision task to be used in subsequent studies, using fMRI and pharmacological manipulation. I also observed a wide range of behavioural profiles ranging from highly exploitative to exploratory participants, and showed that participants were successfully able to notice and adapt to a task reversal. Furthermore, by examining actions at each timepoint in the trials, I showed that participants' learning of the task could not be fully explained by model-free learning theory alone, as participants were able to select optimal actions at each timepoint, both before and after task reversal.

Chapter Three

Using Active Inference To Estimate Subject-Specific Parameters of Internal Model Volatility and Reward

3.1 Introduction

A key aim of computational psychiatry is to characterise individual human participants or patients based on their prior beliefs or preferences, for example, identifying and ameliorating negative bias in depression. Through the use of a computational model, one may also, ideally, infer the underlying neuronal mechanisms that result in or contribute to each person's individual choice behaviour. Such subject-specific characterisation may offer insights into brain functioning in both healthy individuals and patients with psychiatric disorders and may also prove a vital contribution to the development of individualised therapeutics and treatments for various psychiatric conditions. One method of achieving this, is through building a computational model under an Active Inference scheme.

Active Inference is defined by Schwartenbeck and Friston (2016) as "Learning and inference via minimisation of variational free energy, under the prior belief that sequential policies minimise expected free energy" (Schwartenbeck and Friston, 2016). It is based on the hypothesis that an agent, or any self-organising system, must act to minimise uncertainty and thus variational free energy to optimise its internal model, rather than acting solely to maximise reward. The free energy principle states that any self-organising system must always act to minimise its free energy in order to maintain equilibrium with its (dynamic) environment, thus minimising surprise, where surprise is defined as the negative log-probability of any outcome (Friston, 2010).

In Active Inference, an agent constructs a generative model of the world, and continually updates this model in order to bring this internal model closer to the 'true' state of the world. The difference between the agent's generative model and the true or 'perfect' model of the world can be described mathematically using free energy. Therefore, under

Active Inference, all actions taken by the agent aim to minimise variational free energy, as this minimises the uncertainty the agent has about the world (Friston *et al.*, 2006).

In order to survive, the agent is constantly in a state of updating its generative model and its estimate of free energy. To minimise free energy, the agent can either choose actions which are believed to have the lowest free energy, i.e. actions which lead to states believed to have the least uncertainty, or the agent may adjust the parameters of its internal model. Thus, the agent has the ability to both change its plans of action and optimise its generative model. Sets of multiple actions chosen by the agent to traverse across states to reach a desired state or outcome are known as policies. The agent's beliefs about actions and states in the world, as described by the agent's generative model, are known as posterior probability distributions.

A number of studies have utilised this method to simulate behaviour in navigational and visual foraging tasks, including epistemic exploration through a maze task and simulated electrophysiological responses of an artificial agent during the task (Kaplan and Friston, 2018). Studies have also employed various modifications of OpenAI gym environments in the context of Active Inference, such as Doom or FrozenLake paradigms (Cullen *et al.*, 2018; Sajid *et al.*, 2021). Mirza *et al*. (2016) applied an Active Inference scheme to simulate saccadic eye movements, to model epistemic foraging in a visual scene and also simulate electrophysiological responses, with the main incentive of characterising and phenotyping human visual foraging activity in terms of eye movements (Mirza *et al.*, 2016).

More recently, a study by Sales *et al*. (2019) employed an Active Inference model to examine noradrenergic firing patterns of the LC in an artificial agent during simulated three-arm explore/exploit and classic go/no-go reward learning behavioural tasks (Sales *et al.*, 2019). This study probed both the phasic and tonic firing modes of the LC and generated models of LC responses under an Active Inference scheme. In the go/no-go task, a well-trained agent experienced SAPEs, in that the agent would expect to receive the more likely 'no-go' cue but instead received the rare 'go' cue and updated its predictions for which state it would occupy at the end of the task. When the agent entered this unexpected state/made an unexpected observation, this resulted in a sudden burst of phasic activity in the LC where a large update was required. Sales *et al*. used an Active Inference model to link trial-by-trial SAPEs to simulated LC activity through the introduction of a parameter (α), encoding internal model volatility.

In the explore/exploit task, the agent had the option to either exploit a known source of reward or explore the environment for new rewarding sources. The task structure comprised a three-arm maze, where in one arm there was a high reward probability and low reward probability in the remaining two arms. The agent first explored the arms until it found a reward in one particular arm, then built a strong prior probability on the availability of the reward in this highly rewarding arm once it had been rewarded in that location multiple times.

These tasks also included contextual reversals, in which the location of the high-reward arm in the case of the explore/exploit task, or the tone-to-reward mapping in the case of the go/no-go task (i.e. a reversal of cue meanings), were switched so that the agent updated its beliefs about the environment and therefore adjusted its priors and behaviour. In the explore/exploit task, increased tonic firing in the LC was seen under this model as the agent learned new priors and altered its behaviour to adapt to the changing reward probabilities. Sales *et al*. therefore postulated that LC activity, and thus subsequent release of NA, may be driven by SAPEs when unexpected observations are made.

Schwartenbeck and Friston (2016) also describe an Active Inference Bayesian model inversion scheme, in which the preferences or hyperpriors of a particular agent or human participant can be estimated based on their responses to a task. This model inversion aims to formally fit the choice behaviour of a participant/agent in the task to a computational model, in order to estimate their preferences and beliefs on an individual level.

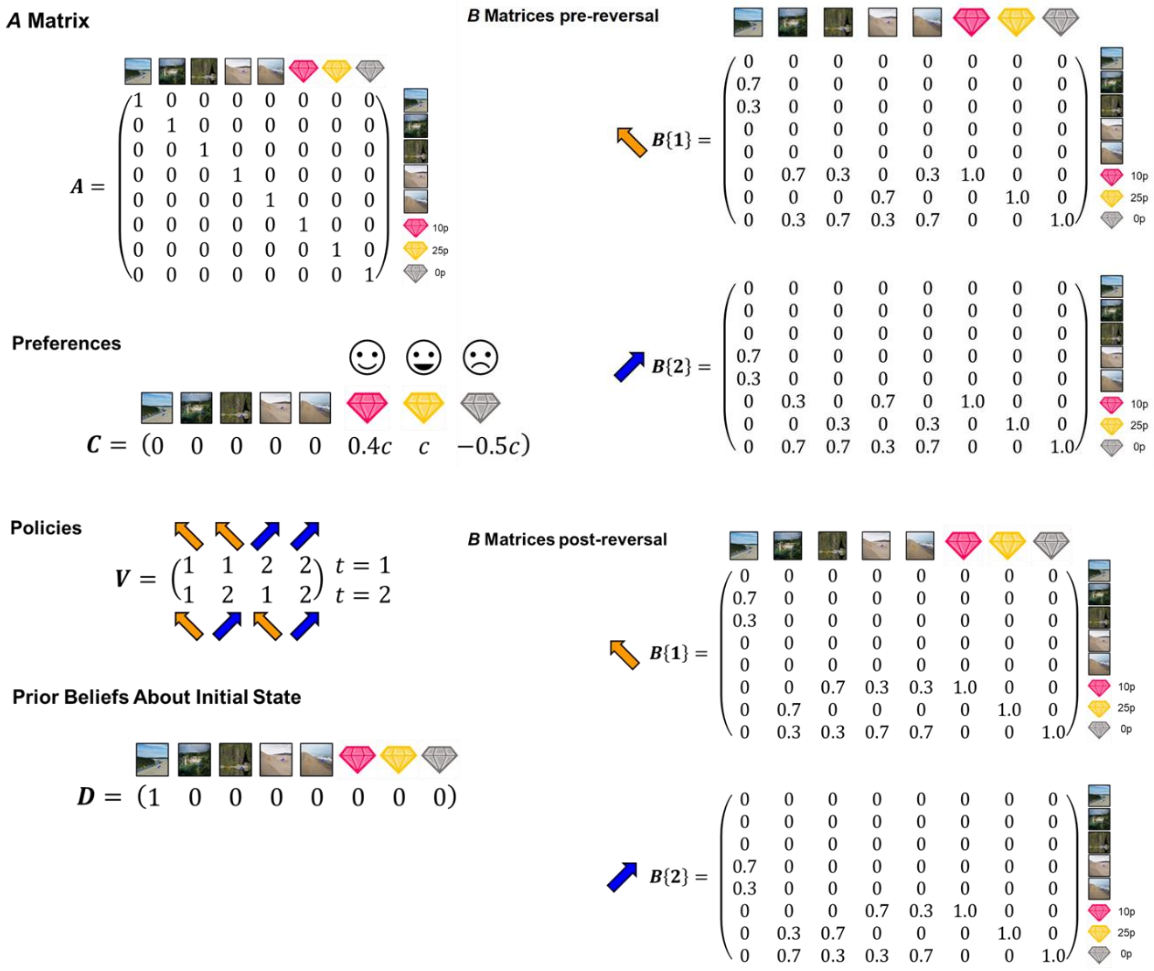
Here, I present an Active Inference model of a probabilistic Markov decision task, in which subject-specific parameters describing internal model volatility (i.e. how flexible an agent's generative model is; how much an agent relies on its prior beliefs) and precision over reward (i.e. how sensitive an agent is to rewarding states) are estimated for human participants through model inversion. In this study, my main aim and rationale for using an Active Inference framework was to optimise a generative model which could be inverted to estimate subject-specific parameters describing internal model volatility (k) and precision over rewarding states (c). My hypotheses were that: 1) true parameter values could be accurately retrieved from simulated data with known parameters, through Active Inference model inversion; 2) this model inversion framework would be able to estimate these parameters of internal model volatility and precision over rewards on a subject-specific level in human participants; and 3) through conducting a Parametric Empirical Bayes (PEB) analysis in addition to classical analyses, these parameters could be used to predict broad measures of behavioural performance in the task. I used an adapted version of the model described by Sales *et al*. (2019) to generate synthetic data in the probabilistic decision-making task designed and investigated in *Chapter Two*, and built upon this by introducing a Bayesian model inversion scheme, to estimate internal model volatility and model decay on an individual level, both in simulated and human behavioural data.

3.2 Methods

3.2.1 Participants and Task

The behavioural data used here was taken from the 25 healthy adult participants as studied in the main behavioural study; participant recruitment and eligibility described in *Chapter Two*. The task structure modelled here, regarding trial number and task reversal, is also identical to that studied in the main behavioural study, but combined the 160 training session trials and 160 testing session trials into a single 320-trial session (*Chapter Two*). In summary; the task in the model simulation consisted of a total of 320 trials: 200 trials used the pre-reversal task structure, with the reversal occurring after trial 200 and the remaining 120 trials used the post-reversal task structure (see **Figure 3.1** for state-transition matrices both pre- and post-reversal).

Figure 3.1



$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\boldsymbol{C} = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0.4c \quad c \quad -0.5c)$$

$$\boldsymbol{V} = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 2 \end{pmatrix} \begin{matrix} t = 1 \\ t = 2 \end{matrix}$$

$$\boldsymbol{D} = (1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)$$

$$\boldsymbol{B}\{1\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.7 & 0.3 & 0 & 0.3 & 1.0 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0 & 0 & 1.0 & 0 \\ 0 & 0.3 & 0.7 & 0.3 & 0.7 & 0 & 0 & 1.0 \end{pmatrix}$$

$$\boldsymbol{B}\{2\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0.7 & 0 & 1.0 & 0 & 0 \\ 0 & 0 & 0.3 & 0 & 0.3 & 0 & 1.0 & 0 \\ 0 & 0.7 & 0.7 & 0.3 & 0.7 & 0 & 0 & 1.0 \end{pmatrix}$$

$$\boldsymbol{B}\{1\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 & 0.3 & 1.0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0.3 & 0.3 & 0.7 & 0.7 & 0 & 0 & 1.0 \end{pmatrix}$$

$$\boldsymbol{B}\{2\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0.3 & 1.0 & 0 & 0 \\ 0 & 0.3 & 0.7 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0.7 & 0.3 & 0.3 & 0.7 & 0 & 0 & 1.0 \end{pmatrix}$$

Figure 3.1 Matrices used to define the generative model. Left: The *A* matrix defines mappings between states and observations, the *C* vector defines prior preferences over states, and the *V* matrix defines available policies, in which a left action is denoted by 1, and a right action is denoted by 2. In this task, there are two timepoints at which an action choice must be made and two available actions (left and right), therefore there are four available policies. The *D* vector defines the agent's beliefs about which state is the initial state in the task. Right: *B* matrices define state transitions, where rows represent states at time *t*, and columns represent states at time *t* + 1. The model contains one *B* matrix for every available action, and in this model, the state transitions change post-reversal.

3.2.2 Model Specification

The model used here is constructed as a discrete state-space model, specifically a (partially observable) Markov Decision Process (MDP), in which stimuli or observations (in this task, visual) and actions/choices are categorized as a set of discrete states and outcomes (Schwartenbeck and Friston, 2016). In an MDP, the agent must make inferences about its current state, predict the outcomes of any actions it may take, and make postdictions (postdiction refers to explanation after the event; in the case of Active Inference, a postdiction occurs when the agent updates its beliefs about the state it occupied in the past) about states the agent has previously entered. The agent then uses this information to minimise its variational free energy, thus optimising its generative model. To do this, the agent may adjust both its behaviour to reach states with the lowest free energy and/or key parameters of its internal model. A Partially Observable MDP (POMDP), is an MDP in which some states are not observable by the agent, i.e. they are *hidden states*. There are eight available states in this task, which describe particular combinations of (visual) features relevant to the agent (described in **Figure 3.1**, **Appendix 1**). To model this task in the context of Active Inference, one must first specify the generative model. Here, the generative model consists of three key matrices: the *A*, *B*, and *C* matrices. The *A* matrix describes the mapping from hidden states to observations. In this task, all eight states are fully observable and there is no uncertainty as to which state results in a particular observation, therefore the *A* matrix takes the form of an 8 x 8 identity matrix.

The *B* matrices represent transition probabilities from the current state at time *t*, to future states at time *t* + 1, dependent on the action taken; there is one *B* matrix for every available action. In this task, the participant or agent can move either left or right, therefore I specified two 8 x 8 *B* matrices corresponding to left and right actions respectively. These transition probabilities depend solely on the current state and action, and do not depend on the history of any previous states. This is a key feature of MDPs, in that they possess the Markov property, or 'memoryless' property. In this task, the agent's generative model is supplied with the environmental *A* and *C* matrices, but must explore the state space to learn the transition probabilities defined in the *B* matrices and update its internal model accordingly. I made the agent naïve to these state transition probabilities by providing the agent with naïve state-transition (*b*) matrices of uniformly distributed random numbers between 0 and 0.01, excluding the values representing state transitions from rewarded states (states six-to-eight) which remained as one-to-one mappings, given that they are fully observable absorbing states.

The *C* vector describes the prior preferences over outcome states, and in this task, this directly reflects the level of monetary reward obtained by the participant or agent in the three absorbing states. The *C* vector for this task is a 1 x 8 vector, in which the preferences reflect the level or lack of reward available to the agent in a given state (**Figure 3.1**).

In addition to these three key matrices, I also provided the generative model with prior beliefs about the agent's initial state (*D* vector), and available policies (*V* matrix) (**Figure 3.1**). In this case, the agent is always certain when they occupy the initial state, therefore in this *D* vector, the likelihood of the agent occupying the initial state in state one equals 1, and this likelihood equals 0 for all remaining states. The agent has four available policies to choose from, and each policy consists of two actions at two timepoints per trial composed of all possible combinations of the two actions, left and right. As such, the *V* matrix is formed of a 2 x 4 matrix (**Figure 3.1**).

By modelling this decision-making paradigm under Active Inference, I can also investigate the agent's internal model volatility/flexibility, or model decay. Sales *et al*. (2019) described an Active Inference model of NA and LC function by examining SAPEs (Sales *et al.*, 2019).

SAPEs are defined as large differences in the BMA of policy dependent states, at successive time steps, indicating unexpected changes in the environment or that the agent has made unexpected observations. The BMA essentially consolidates all the beliefs that the agent has about its place in the environment and its beliefs about its past, present and future states, which the agent then uses to inform its action or policy selection and subsequently enter a new state, aiming to minimise uncertainty.

In this model, SAPEs were calculated as the KL divergence between the BMA at the previous time step (*t* – 1), and the BMA at the current time step (*t*). Therefore, if the agent made a surprising observation, i.e. it did not believe that it would make a particular observation or enter a particular state at the previous time step, the change in BMAs from one time step to the next would increase and the SAPE for the current time step would be larger. The BMA (S_τ) is calculated as follows:

$$S_\tau = \sum_p \pi_p \cdot S_\tau^p \qquad \textit{Eq. 3.1}$$

S_τ^p represents a vector of probabilities for states at time τ under policy *p*, with a probability π. Pi (π) represents the probability of policies and is given by a softmax function (σ) over expected free energies in the past and future, to ensure that policies with lower expected free energies have higher probabilities. Pi is calculated based on the *A*, *B*, and *C* matrices, and is given by the following:

$$\pi = \sigma(-F - \gamma \cdot G) \qquad \textit{Eq. 3.2}$$

F represents past free energy, G represents future free energy, and γ represents the precision parameter, which describes the 'confidence' the agent has in its predictions. The BMAs for the current time step (*t*) and previous time step (*t* – 1) are then used to calculate the SAPE at time *t*:

$$SAPE(t) = \sum_\tau D_{KL}[(S_\tau^t)||(S_\tau^{t-1})] \qquad \textit{Eq. 3.3}$$

The model used by Sales *et al*. (2019) also included an activation function which used the maximum value of SAPEs for a given trial to calculate a decay (or volatility) factor, α, as shown below:

$$\alpha = \alpha_{min} + \frac{\alpha_{max}}{1 + e^{k(SAPE-m)}} \qquad \textit{Eq. 3.4}$$

By endowing the model with this decay factor, the agent was given the ability to forget past observations that may have become irrelevant if the environment became more volatile and/or unexpected changes occurred within the environment. Here, k is a gradient or scalar of the activation function, which weights the effects of SAPEs, and m is the midpoint of the function; m is fixed according to the task being modelled (Sales *et al.*, 2019).

This decay was implemented in this task, in which the agent must learn the *B* matrices, by introducing a modification to the trial-by-trial update equations for the *b* matrices:

$$b(u) = b(u) + \sum_{\tau, p(\tau)=u} \pi_p S_\tau^{\pi_p} \otimes S_{\tau-1}^{\pi_p} - \frac{\big(b\big(\pi_p(\tau) = u\big) - 1\big)}{\alpha} \qquad \textit{Eq. 3.5}$$

Therefore, when the decay factor (α) is low as a result of increased volatility in the environment (thus higher SAPEs), the resulting *b* matrices will predominantly rely on the agent's most recent observations and forget past experiences. Conversely, when α is high as a result of low environmental volatility and low SAPEs, this reduces the decay of values in the *b* matrices, and therefore the *b* matrices utilise the agent's past experiences more and the model becomes more stable.

3.2.3 Data Simulations

To simulate behaviour in this task applying the generative model as described above, I used a modified version of spm_MDP_VB_X.m which can be found in the SPM12 toolbox in MATLAB; code for the application of Active Inference is available at https://www.fil.ion.ucl.ac.uk/spm/software/spm12/. Functions were modified according to the modifications made by Sales *et al.* to an earlier version of this code, spm_MDP_VB.m, to implement the addition of the decay factor and inclusion of SAPEs in the trial-by-trial updating of *b* matrices; code used by Sales *et al.* can be found here: https://github.com/AnnaCSales/ActiveInference (Sales *et al.*, 2019).

I initially simulated task data to examine the behavioural profiles generated by varying parameter values such as the gradient/gain of the activation function, *k*, and the precision over prior preferences of outcome states, *c*. The parameter *k* modulates the effect that changes in SAPEs exert on the model decay factor on a trial-by-trial basis, however the value of *k* remains constant across trials for any particular agent/participant. The effect of SAPE changes, as the input for the activation function, on the output (α), increases as *k* increases and drives a more binary output. However, a low value of *k* reduces the effect of changing values of SAPEs on α, and therefore on model volatility and *b* matrix updating.

The parameter *c* is used in the *C* vector as a precision over the preferences of agents for the three absorbing states in which the agent either receives a low-level reward, a high-level reward, or no reward (**Figure 3.1**). The prior preferences for the five non-absorbing states, states one to five, are assigned a value of 0, as there is no reward associated with each of these states, but the agent may still obtain a reward in future states. The high-level reward state is assigned the value of *c*, the low-level reward state is given the value 0.4*c*, and the no-reward absorbing state is given the value -0.5*c*, because the agent does not receive any reward in this state and has no future possibility of receiving any reward during the current trial. Therefore, if the agent has a high value of *c*, rewarding states are very strongly preferred and the non-rewarded absorbing state is highly undesirable, since these preferences over states are defined in log space. The agent would also have a much greater preference for the high-reward state compared to the low-reward state, whereas with a lower *c* value there would be a reduced distinction in preference between the two rewarding states, whether that level of reward is high or low.

In my simulations, I varied the values of parameters *k*, *c*, *α min*, and *α max* to examine how the synthetic behavioural profiles changed based on these parameters, and which combination of parameters produced the most optimal behaviour in this task. Here, the optimal policy is defined in the same way as for the analyses in *Chapter Two*: the policy which has the highest probability of high-level rewarding outcome. Values for each parameter used in the parameter search are shown in **Table 3.1**. I fixed the values of parameter *m* = 1.7 (midpoint of activation function, see **Equation 3.4** above), and hyperparameters *alpha* = 1 (scale of precision parameter distribution) and *beta* = 1 (rate of precision parameter distribution) in these simulations. MDPs for each parameter combination were replicated five times, i.e. five simulated agents, across 400 trials with the reversal occurring after 200 trials. To examine 'optimal' behaviour, in the parameter search I examined which parameter combination generated behaviour with the highest frequency of optimal policy selection pre-reversal, the highest frequency of optimal policy selection post-reversal, the highest level of reward earned across the task, the highest transition probabilities in the *b* matrices for the optimal policy state transitions pre-reversal (state four-to-seven transition probability) and post-reversal (state two-to-seven), and the KL divergences between the environmental *B* matrices and the agents' estimated *b* matrices (**Table 3.2**). For the step-by-step data simulation, inversion, and analysis pipeline, see **Figure 3.2**.

Table 3.1

Parameter	**Definition**	**Value(s) tested in simulations**
k	gradient	[1 2 4 8 16]
c	precision over preferences	[2 4 8 16 32]
α min	minimum value of decay factor	[2 4 8 16 32 64 128]
α max	maximum value of decay factor	[256 512 1024 2048 4096]
m	midpoint of activation function	1.7
Alpha	scale hyperprior on precision	1
Beta	rate hyperprior on precision	1

Table 3.1 Parameters used in mass data simulations for parameter search.

Table 3.2

Property of Optimal Behaviour	***k***	***c***	***α min***	***α max***
Lowest KL Divergence between *B* (true) and *b* (agent) matrices	1	2	2	256
Highest Total Reward	4	32	2	1024
Highest optimal policy selection pre-reversal	2	32	128	2048
Highest optimal policy selection post-reversal	4	32	8	2048
Highest transition probabilities for state four-to-seven pre-reversal	1	4	64	2048
Highest transition probabilities for state two-to-seven post-reversal	4	2	2	4096
Greatest change in transition probabilities for state four-to-seven pre- to post-reversal	8	32	32	1024

Table 3.2 Parameter combinations to produce optimal behaviour.

Figure 3.2

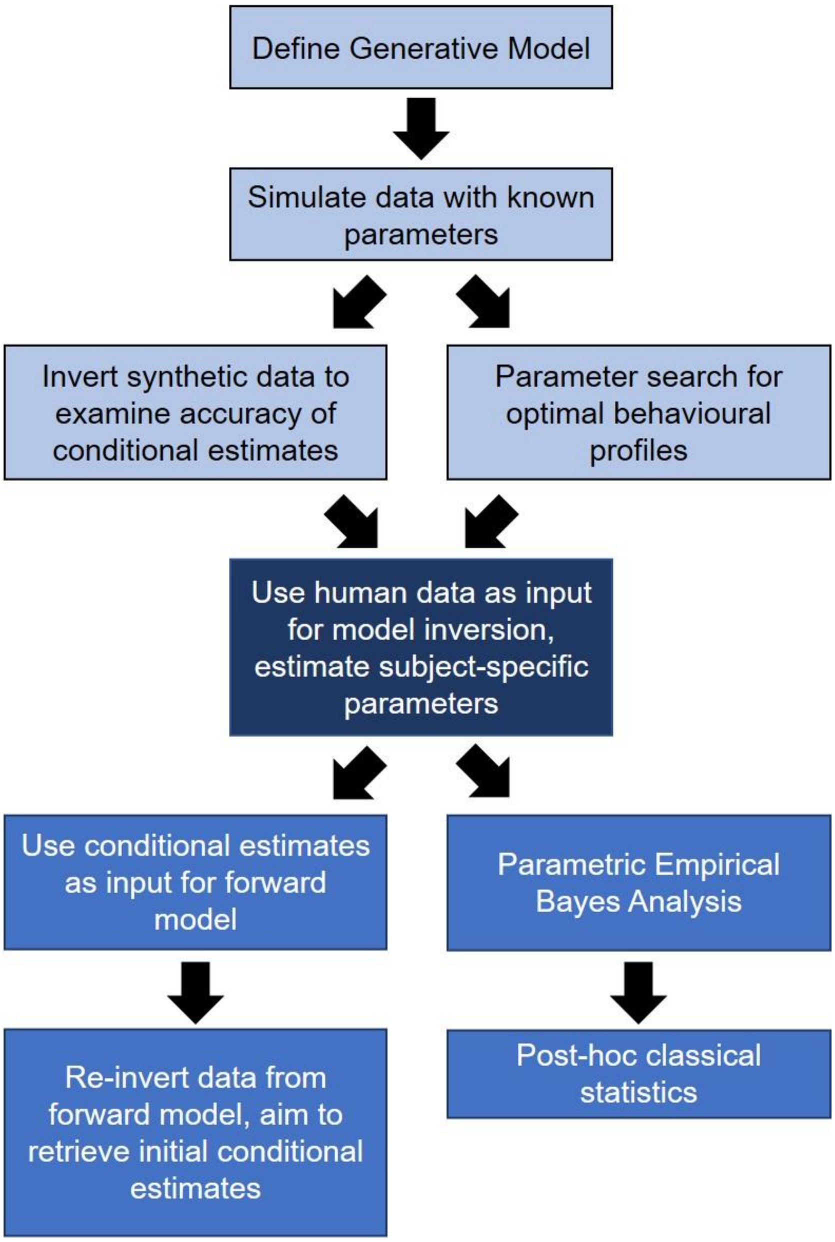

Define Generative Model
Simulate data with known parameters
Invert synthetic data to examine accuracy of conditional estimates
Parameter search for optimal behavioural profiles
Use human data as input for model inversion, estimate subject-specific parameters
Use conditional estimates as input for forward model
Parametric Empirical Bayes Analysis
Re-invert data from forward model, aim to retrieve initial conditional estimates
Post-hoc classical statistics

Figure 3.2 Flowchart detailing the data simulation, model inversion and analysis pipeline.

3.2.4 Model Inversions

In the data simulations described above, I used the generative model and sets of hidden parameters to generate synthetic behavioural data. To investigate and explain the behavioural profiles of human participants in the task, I used a model inversion which estimated hidden parameters on a subject-specific level, given the model and observed data. I used the spm_nlsi_Newton.m pipeline in SPM12, with the addition of the decay factor activation function introduced by Sales *et al*. as described above. In this model inversion, I specified shrinkage priors over the free (unknown) parameters in order to estimate participants' generative models. For the free parameters in each of the model inversions, I used priors with a mean of 0 and variance of 0.5, thus inducing shrinkage towards zero. This shrinkage towards the prior mean acts to prohibit model overfitting.

I initially conducted model inversions for simulated data with known fixed values for hidden parameters described above, to assess if the values of these free parameters could be accurately retrieved through model inversion. I used two models, one in which the parameters *k* and *c* were free parameters, and a second model with the free parameters *k*, *c*, and *m* (midpoint of the activation function, see **Equation 3.4**). For both model inversions, I fixed the values of *α min* = 2, *alpha* = 1, *beta* = 1, and compared inversions using different values of *α max*: 1024 and 2048. Values for *α min* and *α max* were selected based on the mass parameter search described above (see **Table 3.2**). For the two-parameter inversion, *m* was fixed at 1.7. This value was calculated from the mean SAPE over 100 simulated trials of this task when *α* = 16, following the methods of Sales *et al*. (2019). True parameter values versus conditional MAP estimates from the model inversions are presented in log space in **Figure 3.3**.

Figure 3.3

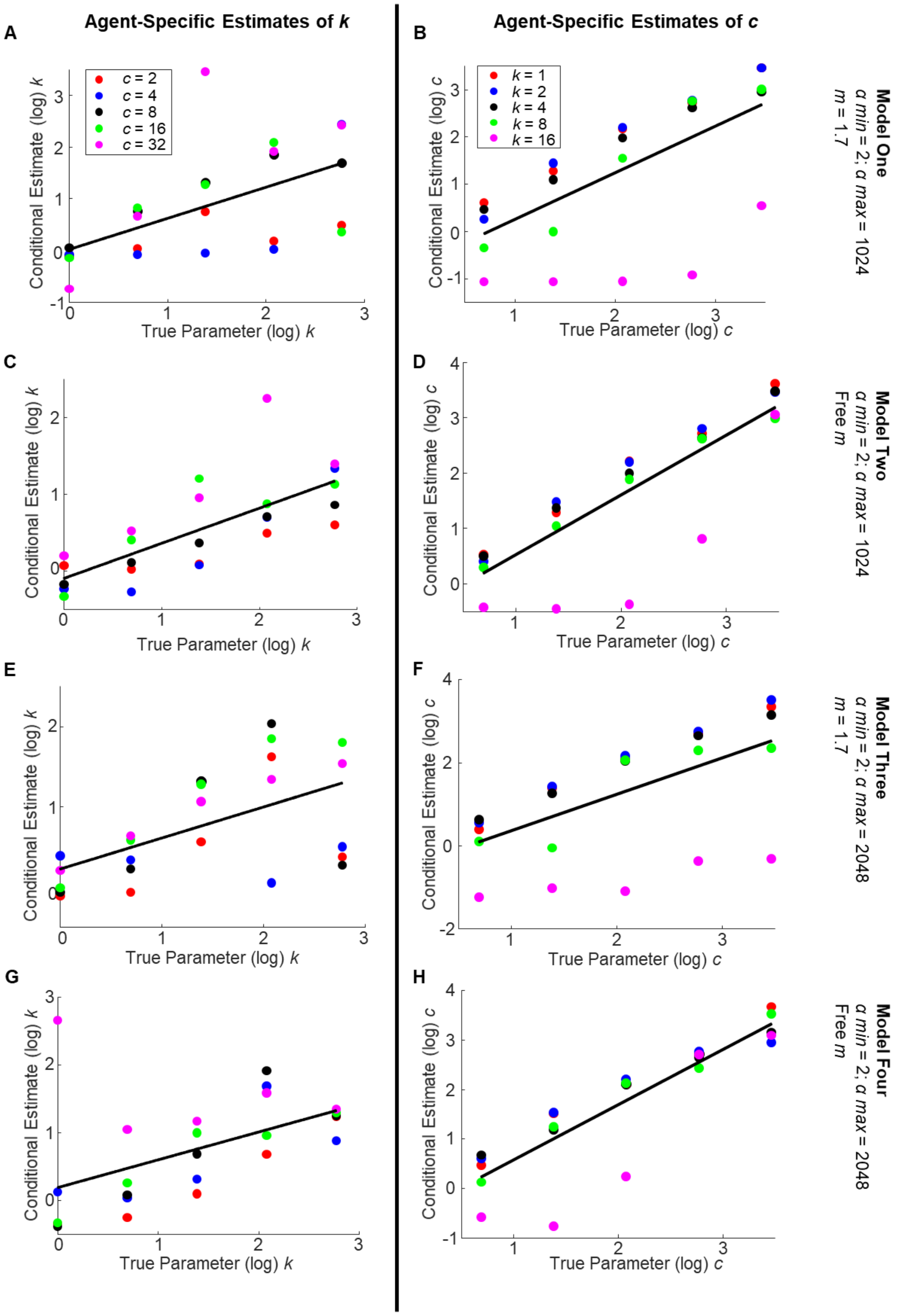


Figure 3.3 True parameter values versus conditional estimates generated through model inversion, for parameters *k* (left) and *c* (right), for each of the four models.

Using the conditional parameter estimates generated through model inversion, I ran four different forward models to produce synthetic data (MDPs). If the model inversions above were successful, the behaviour simulated using the parameter estimates from the inversion should be similar to the behaviour of the initial data simulations used for model inversion. I produced MDPs for parameter estimates from the two-parameter and three-parameter models, using fixed *α max* values of 1024 and 2048.

I then applied human behavioural data, specifically actions taken and states experienced in the task, to the model inversion scheme to estimate these parameters on a subject-specific level. Based on my findings from inverting simulated data, I used *α min* = 2, *α max* = 1024, *alpha* = 1, and *beta* = 1. I tested both the two-parameter inversion with *k* and *c* as free parameters, and the three-parameter inversion, with the addition of *m* as a free parameter. For two-parameter inversions, *m* was fixed at 1.7. For the free parameters, I again used prior means of 0 and variance of 0.5. To ensure that the conditional MAP estimates were accurate and retrievable, I then inputted the conditional estimates in the (forward) generative model to generate synthetic behavioural data and re-inverted these data to examine the re-inverted conditional estimates. If the initial model inversion was successful, the conditional estimates generated from the re-inversion should be very similar to those estimated in the initial inversion.

3.2.5 Parametric Empirical Bayes Analysis

To examine how the conditional estimates of these model parameters may predict coarse behavioural metrics on a group level, I used a PEB for a random-effects analysis, based on participants' behavioural profiles, specifically optimal policy selection frequency and total reward earned across the task. I conducted this analysis in order to identify the underlying mechanisms relevant to these behavioural differences, i.e. which model parameter(s) were associated with these behavioural metrics. To construct the Bayesian General Linear Model (GLM), I included an average mean effect and two second-level covariates to quantify task performance.

This random-effects design matrix comprised three columns, with one column for each covariate: the first column was a column of ones representing the mean across participants, the second column contained the total reward earned across the task per participant, and the third column contained the total percentage of optimal policy choices summed across pre-reversal and post-reversal trials, which revealed participants' behavioural profiles in terms of exploitative or exploratory behaviour. Total reward across the task is indicative of the *c* parameter, in that higher total reward suggests higher precision over state preferences.

High percentages of optimal policy selection not only indicated exploitative behaviour across the task, but also that participants were able to learn the new optimal policy following the reversal. This covariate may be influenced by a participant's *k* parameter, as this parameter plays an important role in internal model flexibility. If a participant's model was highly flexible, i.e. if their *k* parameter was low, they would be expected to exhibit exploratory behaviour and their percentage of optimal policy selection would be low. Similarly, if a participant had a hyper-rigid model, if *k* was high, I would expect exploitative behaviour pre-reversal followed by a reduced ability to adapt to the new task structure post-reversal. Therefore, their percentage of optimal policy selection would be at an intermediate level. A participant that was able to identify and exploit optimal policies both pre-reversal and post-reversal and successfully adapt to the task reversal would therefore have a high percentage of optimal policy selection, as a result of an optimal intermediary-value *k* parameter.

3.2.6 Statistical Analyses

For model inversion of data simulations, I used Pearson's correlations to examine the relationship between true parameter values and the conditional estimates (MAP values) generated through model inversion. These parameters are probability density functions: the parameter's true value and the conditional MAP estimate. I also used Pearson's correlations to compare the conditional estimates from the initial inversion and re-inversions of human data, and also in post-hoc classical analyses of conditional parameter estimates and task performance metrics following the PEB. Goodness-of-fit analyses were also conducted to examine associations between conditional MAP estimates and task performance, with summed square error (SSE) and adjusted R-square values reported.

3.3. Results

3.3.1 Conditional Parameter Estimates Display Strong Correlation with True Parameter Values in Model Inversions of Simulated Data

By conducting a mass parameter search generating MDPs for 875 unique parameter combinations, each with five replications, I revealed a set of parameters that produced optimal behaviour in the task. Optimal behaviour was assessed through seven properties of optimal behaviour, as described above in **Table 3.2**, and the parameter combination that most frequently produced optimal behaviour in the mass simulation consisted of: $k = 4$, $c = 32$, α *min* = 2, and α *max* = 2048. Parameters $k = 1$ and α *max* = 1024 also produced similarly high-performance behaviour. Based on these findings, I ran model inversions on these simulated data and aimed to accurately retrieve the parameters.

I inverted four different models: model one consisted of two free parameters (k, c) and a fixed α *max* = 1024, model two had three free parameters (k, c, m) and fixed α *max* = 1024, model three had two free parameters (k, c) and α *max* = 2048, and finally model four had three free parameters (k, c, m) and α *max* = 2048. For the models with α *max* = 1024, model two consistently performed better than model one, in that it showed stronger correlations between true parameter values and conditional estimates for both k (model one: *rho* = 0.571, $p = 0.0029$; model two: *rho* = 0.733, $p = 3.05 \times 10^{-5}$) and c (model one: *rho* = 0.658, $p = 3.54 \times 10^{-4}$; model two: *rho* = 0.843, $p = 1.24 \times 10^{-7}$) (**Figure 3.3A-D**).

For the models using α *max* = 2048, model three was able to estimate k parameters more closely to the true parameters compared to model four, as there was a stronger correlation between the true parameters and conditional estimates of k when m was a fixed parameter (model three: *rho* = 0.564, $p = 0.0033$; model four: *rho* = 0.520, $p = 0.0077$) (**Figure 3.3E,G**). However, the reverse was true for estimates of c; there was a stronger correlation between true values of c and conditional estimates in model four where α *max* = 2048 (model three: *rho* = 0.600, $p = 0.0016$; model four: *rho* = 0.886, $p = 3.83 \times 10^{-9}$) (**Figure 3.3F,H**).

Overall, the models in which α *max* = 1024 (models one and two) showed the strongest correlations between true parameter values and conditional estimates across both model inversions, compared with the α *max* = 2048 models (models three and four).

Therefore, in my model inversions of human behavioural data, α *max* = 1024 was used as a fixed parameter, in addition to a fixed α *min* = 2.

I also used the conditional estimates for *k* and *c* from each of the model inversions to run forward models and generate synthetic data, and compared these data to the initial simulated data used for the inversions. Accurately retrieved *k* and *c* parameters (and *m* parameters for the three-parameter models) were expected to produce synthetic data close to that of the original simulations. To investigate differences in synthetic behaviour between the initial simulations and post-inversion simulations, I tested the differences between total reward earned across the task per agent, total frequency of optimal policy selection per agent, and the sum of all agents' KL divergences between *b* matrices in initial versus post-inversion simulations, at both pre-reversal (trial 200) and post-reversal (trial 400).

All four models showed very strong significant correlation in total reward between the initial versus post-inversion simulations, with the strongest correlation in model three: the two-parameter model using α *max* = 2048, with model two showing the second-strongest correlation (model one: *rho* = 0.664, *p* = 2.95 x 10^{-4}; model two: *rho* = 0.8440, *p* = 1.15 x 10^{-7}; model three: *rho* = 0.887, *p* = 3.64 x 10^{-9}; model four: *rho* = 0.819, *p* = 5.53 x 10^{-7}) (**Figure 3.4A, C, E, G**). There was also strong correlation in optimal policy selection frequency between the initial and post-inversion simulations, with the strongest correlation in model two; the three-parameter model using α *max* = 1024 (model one: *rho* = 0.469, *p* = 0.0179; model two: *rho* = 0.780, *p* = 4.35 x 10^{-6}; model three: *rho* = 0.714, *p* = 6.16 x 10^{-5}; model four: *rho* = 0.770, *p* = 6.65 x 10^{-6}) (**Figure 3.4B, D, F, H**).

Figure 3.4

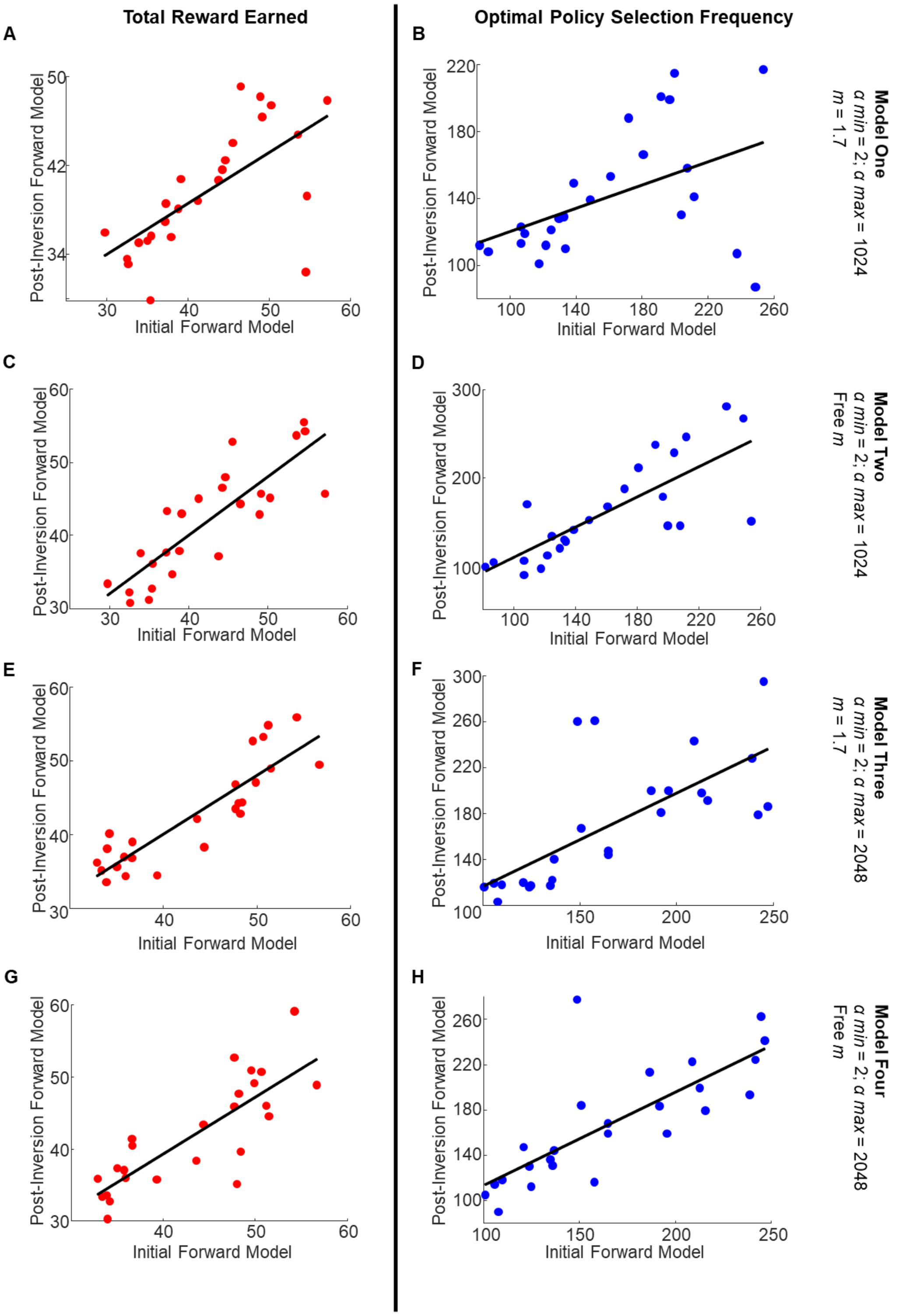


Figure 3.4 Behavioural profiles of initial data simulations compared with forward models produced using conditional estimates of model inversion. Total reward (left) and frequency of optimal policy selection (right) were examined, for each of the four models.

Model two also showed the lowest summed KL divergence in agents' *b* matrices between the initial and post-inversion simulations (model one: 142.43, model two: 96.40, model three: 197.17, model four: 126.39). From the four models, model two produced policy selection profiles closest to that of the original data compared to the other three models, supporting the correlation analysis findings that this model produced conditional estimates closest to that of the initial simulated data. Models two and four also showed stronger correlation for optimal policy selection between initial and post-inversion forward models compared with models one and three, indicating that adding *m* as a free parameter in the model improves the accuracy of policy selection retrieval.

3.3.2 Model Inversion Reliably Estimates Subject-Specific Parameters for Human Participants, and are Retrievable Through Re-inversion of Forward Models

Using fixed parameters *α min* = 2 and *α max* = 1024, I then fed the human behavioural data, i.e. the states encountered by the 25 human participants in the task and their actions taken, into my model inversion to estimate model parameters on a subject-specific level. I inverted the data using both the two-parameter and three-parameter models, and inputted the conditional estimates into the forward model to simulate behavioural data which, given that the inversion was successful, should closely emulate the real human data. I then re-inverted the human data, by inverting the forward models produced from the initial data inversion, in order to accurately retrieve the initial conditional estimates. If both inversions were successful, the conditional estimates for each parameter should be very similar across both inversions. For an exemplary human participant, see **Figure 3.5**. I also tested a four-parameter model, with *k*, *c*, *α min* and *α max* as free parameters. However, this proved to be relatively unsuccessful, as when inverted, parameters *c* and *α min* appeared to bear most of the inter-participant differences, whereas conditional MAP estimates of *k* and *α max* did not deviate from their prior means across the cohort, and all four parameters had very large 90% confidence intervals. Therefore, investigation into this model inversion did not progress beyond the initial inversion.

Figure 3.5

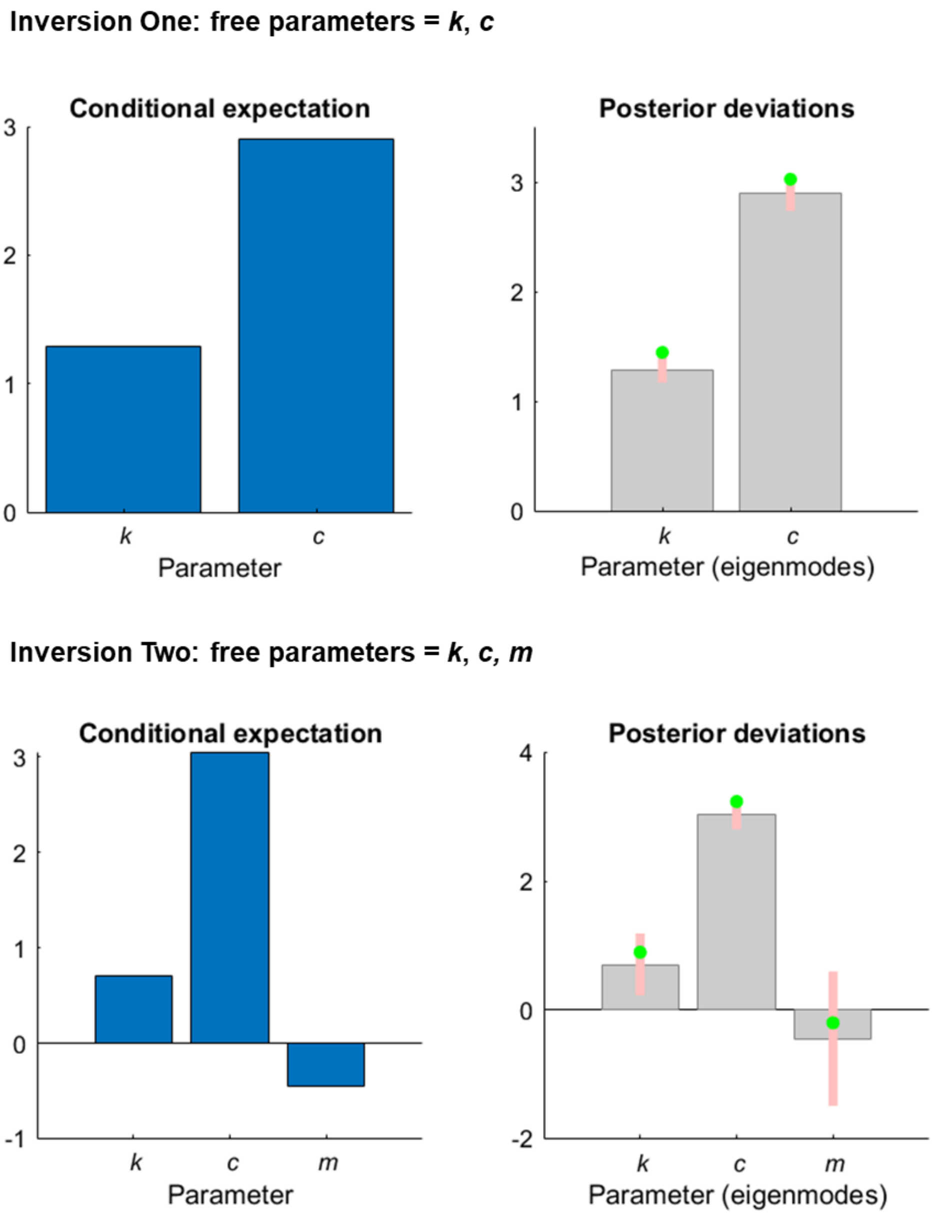


Figure 3.5 Conditional estimates and posterior deviations for an exemplary human participant. Second two-parameter (top) and three-parameter (bottom) model inversions. Green dot = conditional estimates from the initial inversions. Here, the conditional MAP estimates from the first and second inversions are extremely similar in both inversions, displaying highly accurate retrieval of parameter estimates. Pink error bars represent 90% confidence intervals.

For the two-parameter inversion, there was a strong correlation between the conditional estimates of *k* for both inversions (*rho* = 0.710, *p* = 7.11 x 10^{-5}), and also a very strong correlation between the conditional estimates of *c* (*rho* = 0.935, *p* = 7.38 x 10^{-12}), suggesting that the model inversions were able to accurately retrieve parameter estimates across the cohort (**Figure 3.6A-B**).

Figure 3.6

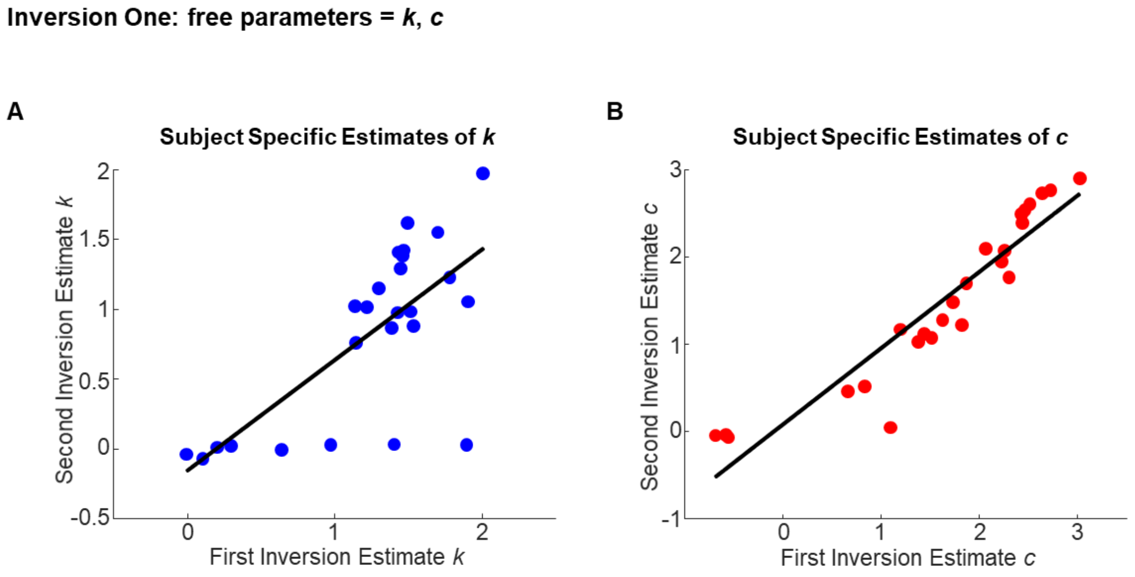


Figure 3.6 Conditional estimates of parameters *k* (A) and *c* (B) comparing the first and second model inversions, for the two-parameter inversion.

In the three-parameter inversion, the conditional estimates for *k* showed a strong correlation between both inversions (*rho* = 0.635, *p* = 6.48 x 10^{-4}), and the conditional estimates for *c* showed a very strong correlation (*rho* = 0.970, *p* = 1.67 x 10^{-15}), demonstrating that estimates for *k* and *c* were successfully retrieved. However, the conditional estimates of *m* were moderately correlated across inversions (*rho* = 0.484, *p* = 0.0142) (**Figure 3.7A-C**).

When comparing the mean free energy between the different models, i.e. two-parameter versus three-parameter inversion, there was a small difference between the two models (log Bayes factor = 2.5), with the three-parameter model having a slightly higher log model evidence across the initial inversions for both models (**Figure 3.7E**). Therefore, the three-parameter model was able to retrieve the parameter estimates with slightly greater success compared to the two-parameter model.

I then examined the behavioural data produced by the forward models, and compared this to the real behaviour of participants (**Figure 3.8**). The frequencies of optimal policy selection across the whole task (frequency of policy three selection pre-reversal + frequency of policy one selection post-reversal) showed strong correlations between observed human behaviour and synthetic data in both the two-parameter model (*rho* = 0.693, *p* = 1.23 x 10^{-4}) and the three-parameter model (*rho* = 0.733, *p* = 3.07 x 10^{-5}) (**Figure 3.9A-B**). Similarly, the total reward was strongly correlated between reward earned by the human participants and in the synthetic data, in both models (two-parameter model: *rho* = 0.764, *p* = 9.00 x 10^{-6}; three-parameter model: *rho* = 0.699, *p* = 1.01 x 10^{-4}) (**Figure 3.9C-D**). Therefore, these parameter estimates generated through model inversion of participant data can be used to produce accurate simulations of human behaviour, where both models investigated above show similar accuracies.

Figure 3.7

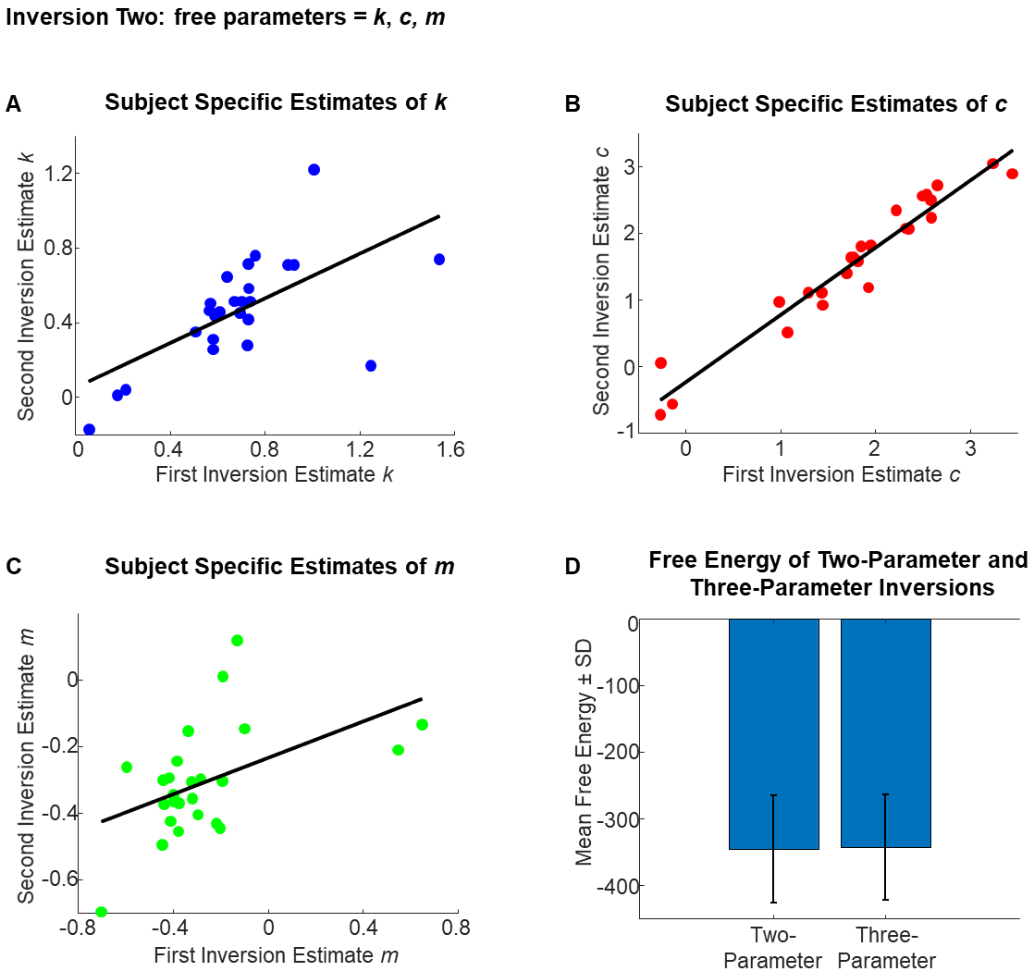


Figure 3.7 Conditional estimates of parameters *k* (A), *c* (B), and *m* (C) comparing the first and second model inversions, for the three-parameter inversion. (**D**) Free energy of the two-parameter (see **Figure 3.6**) versus the three-parameter inversion. The three-parameter inversion has slightly higher log model evidence compared to the two-parameter inversion (log Bayes factor = 2.5), therefore provides a better model fit. SD = standard deviation.

Figure 3.8

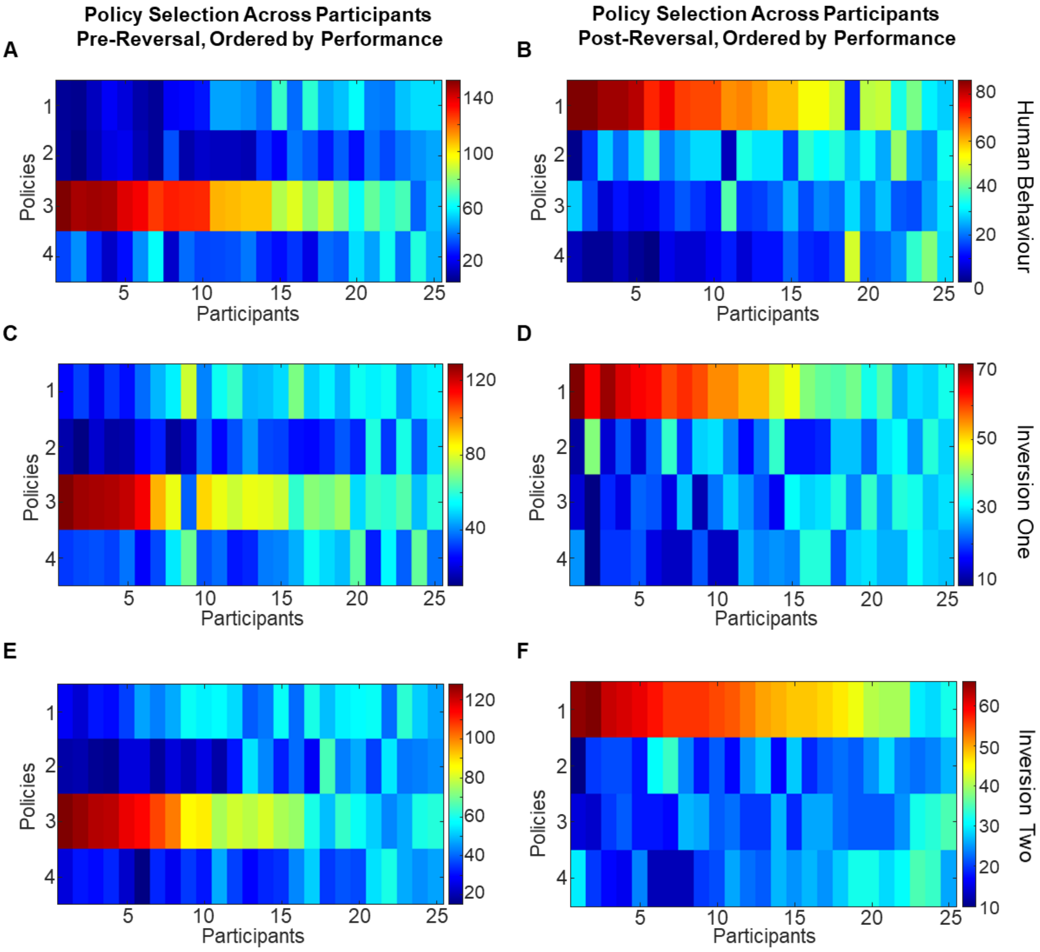


Figure 3.8 Comparison of behaviour between humans and simulations. Policy selection heatmaps describing the behavioural profiles of human participants (**A**-**B**), simulations produced from the conditional estimates generated in the two-parameter model inversion (**C**-**D**), and the three-parameter inversion (**E**-**F**), showing both pre-reversal (left) and post-reversal policy selection (right). Participants in the heatmaps are ordered based on their chi-squared test *p* values, with policy selection most significantly different to chance on the far left, and policy selection closest to chance level on the far right.

Figure 3.9

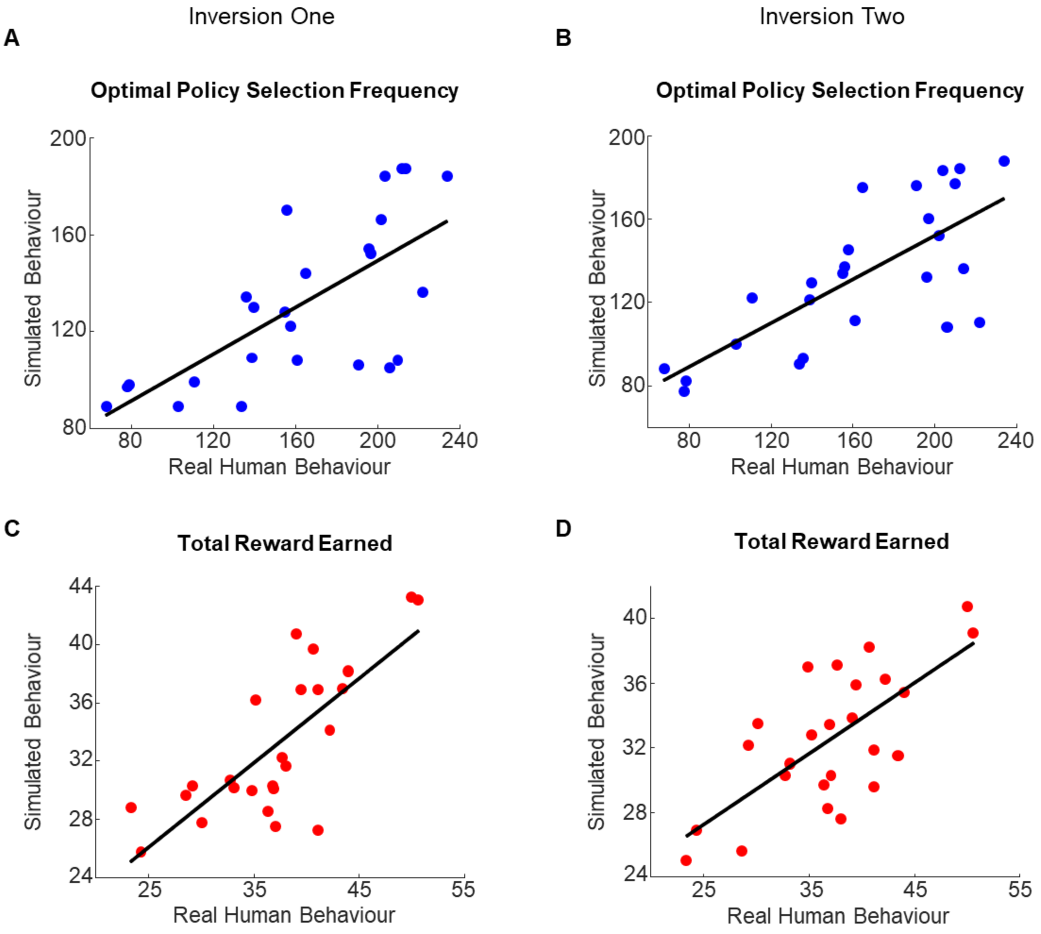


Figure 3.9 Associations in optimal policy selection frequency (A-B) and total reward earned across the task (C-D) between human participant behaviour and simulated behaviour. Generated by running the forward model using the conditional MAP estimates produced through model inversion, for the two-parameter (**A**,**C**) and three-parameter inversions (**B**,**D**).

3.3.3 Parametric Empirical Bayes and Post-Hoc Classical Analyses Reveal Strong Positive Association Between *c* and Task Performance, With an Inverted-U Shaped Effect of *k*

In my PEB analyses, I first examined the effects of total reward on parameters *k* and *c* (and *m* in the three-parameter model), then examined the effects of optimal policy selection frequency across the task. In the two-parameter model, there was a significant association between *c* and the total reward earned (Effect size (*Ep*) = 0.0310, Posterior probability (*Pp*) = 0.948), and also between *k* and total reward (*Ep* = 0.0247, *Pp* = 0.907). However, there was reduced association between the frequency of optimal policy selection, and both *k* and *c* (*k*: *Pp* = 0.500; *c*: *Pp* = 0.500).

In the three-parameter model, there was a significant correlation between *c* and the total reward, in that the total reward increased as *c* increased (*Ep* = 0.0399, *Pp* = 0.9996). Also, there was a reduced association between parameter *m* and the total reward, in that *m* decreased as the total reward increased (*Ep* = -0.0083, *Pp* = 0.584). There were no associations between *k* and either the total reward (*Pp* = 0) or the optimal policy selection frequency (*Pp* = 0), and there were also no associations between either *c* or *m* and the optimal policy selection frequency (*c*: *Pp* = 0; *m*: *Pp* = 0).

In the two-parameter model there was a significant linear correlation between *k* parameter estimates and optimal policy selection frequency (*rho* = 0.413, *p* = 0.0401), however, an inverted-U shaped relationship is evident (**Figure 3.10A**). By fitting polynomials of varying degrees, it is evident that a second-degree polynomial fit the spread of *k* values better compared to a linear fit (linear fit: SSE = 4.66 x 10^{-4}, adjusted R-square = 0.13, d.f. = 23; second-degree polynomial: SSE = 3.19 x 10^{-4}, adjusted R-square = 0.38, d.f. = 22). When fitting a third-degree polynomial, the SSE did not improve further, and the adjusted R-square decreased compared with the second-degree fit (SSE = 3.19 x 10^{-4}, adjusted R-square = 0.35, d.f. = 21). Therefore, a second-degree polynomial curve best described the association between conditional estimates of *k* and optimal policy selection in the two-parameter model, forming an inverted-U shaped relationship.

Figure 3.10

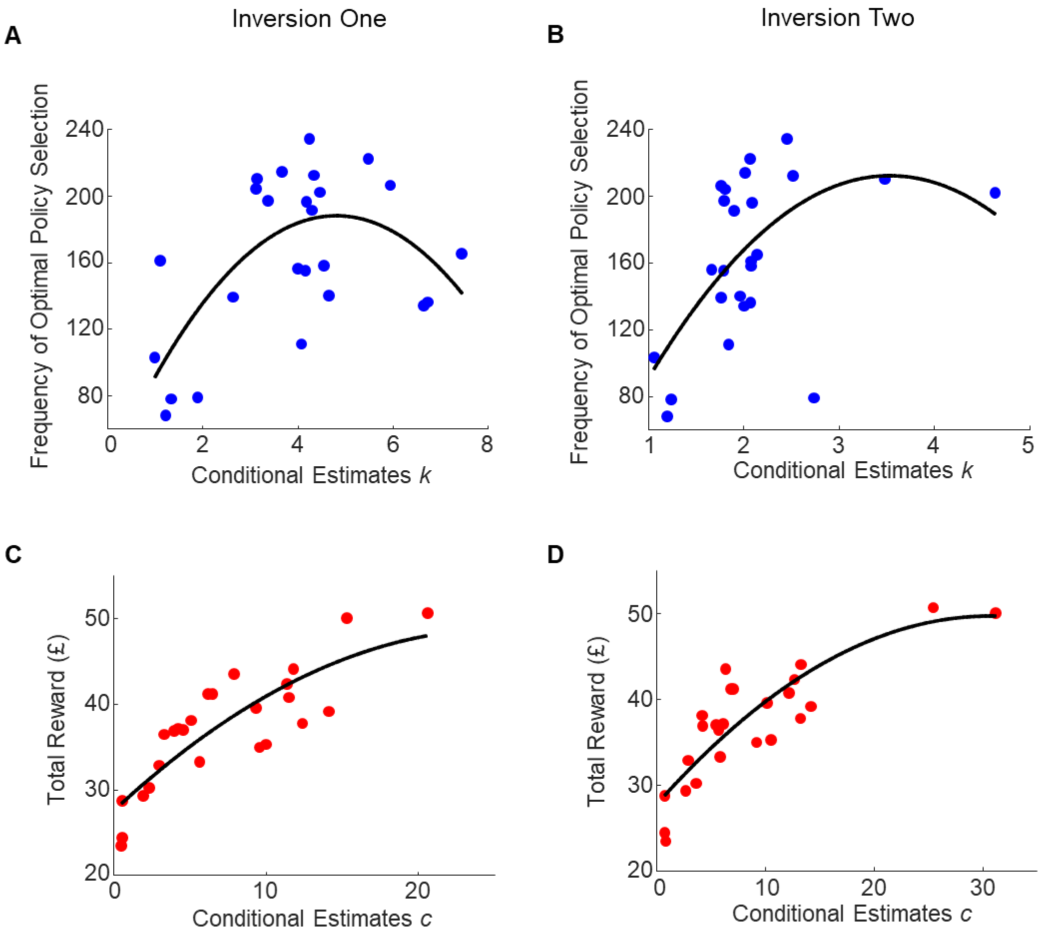


Figure 3.10 Associations between conditional estimates of *k* and *c* as estimated through model inversion, and behavioural metrics describing task performance: optimal policy selection frequency (**A**-**B**) and total reward (**C**-**D**), in the two-parameter (**A**,**C**) and three-parameter inversions (**B**,**D**). There is a clear inverted-U shaped relationship between conditional estimates of *k* and optimal policy selection frequency in the two-parameter inversion, but this effect is reduced in the three-parameter inversion, as the conditional estimates of *k* appear to be shifted to the left.

Similarly in the three-parameter inversion, a second-degree polynomial also provided a better fit for the conditional estimates of *k* compared with a linear fit (linear fit: SSE = 4.58 x 10^{-4}, adjusted R-square = 0.151, d.f. = 23; second-degree polynomial: SSE = 3.90 x 10^{-4}, adjusted R-square = 0.307, d.f. 22) (**Figure 3.10B**). However, a third-degree polynomial provided a slightly better fit compared with the second-degree fit, contrasting the reduced model (SSE = 3.58 x 10^{-4}, adjusted R-square = 0.363, d.f. = 21).

In the two-parameter inversion, there was a very strong significant correlation between the total reward earned across the task and conditional estimates of *c* (*rho* = 0.829, *p* = 3.03 x 10^{-7}). However, when fitting first- and second-degree polynomials, surprisingly the second-degree curve proved a better fit for the data (linear fit: SSE = 3.42 x 10^{-6}, adjusted R-square = 0.674, d.f. = 23; second-degree polynomial: SSE = 3.17 x 10^{-6}, adjusted R-square = 0.684, d.f. = 22). This more accurately describes the saturation effect observed when *c* values exceed 20, which occurs as the maximum possible reward value in the task is reached (**Figure 3.10C**).

There was also very strong significant linear correlation between conditional estimates of *c* in the three-parameter model and total reward (*rho* = 0.816, *p* = 6.63 x 10^{-7}), and as with the reduced model, a second-degree polynomial proved a better fit for the data compared with the linear fit (linear fit: SSE = 3.66 x 10^{-6}, adjusted R-square = 0.666, d.f. = 23; second-degree polynomial: SSE = 3.00 x 10^{-6}, adjusted R-square = 0.729, d.f. = 22), which more effectively captures the saturation effect of *c* (**Figure 3.10D**).

3.4 Discussion

In this study, I have shown that this Active Inference model of a navigational probabilistic decision-making paradigm can successfully estimate subject-specific parameters that describe the internal model volatility (*k*) and precision over rewarding states (*c*) of human participants with relative accuracy, through model inversion. Also, I was able to unpack how these parameters, which described model volatility and precision over rewards, may influence coarse behavioural metrics such as optimal policy selection and total reward earned, respectively.

Across the group, I showed that conditional estimates of *c* were strongly associated with total reward, and also that conditional estimates of *k* exhibited an inverted-U shaped relationship with total optimal policy selection. This inverted-U shaped relationship may reflect the effect of *k* on internal model volatility: hyper-flexible models would produce highly exploratory behaviour, and therefore have a low frequency of optimal policy selection in the task. A hyper-rigid model, however, would produce highly exploitative behaviour initially, but would express strong perseveration behaviour in response to the reversal and may not adapt to the new optimal policy, thus reducing its overall optimal policy selection across the whole task. Therefore, this observed association between estimates of *k* and optimal policy selection may demonstrate this flexibility-rigidity trade-off.

Although the human behavioural participants showed high levels of learning of optimal routes across the cohort, the group showed some behavioural variability in terms of exploitative or exploratory policy selection profiles, both pre-reversal and post-reversal. This enabled me to use the model to characterise the participants' behaviours by estimating the gradient of the activation function *k*, used to calculate a trial-by-trial model decay factor α, estimating a single unique parameter value per participant.

A number of studies have demonstrated the ability to simulate behaviour under Active Inference (Mirza *et al.*, 2016; Parr and Friston, 2017; Sales *et al.*, 2019), and have also exhibited model inversion to retrieve parameter estimates from simulated data (Schwartenbeck and Friston, 2016). For example, FitzGerald *et al*. (2015) investigated evidence accumulation in the commonly-used urn task (FitzGerald *et al.*, 2015). By constructing this task as an MDP, they examined the effects of altering threshold criterion for key parameters in their model, and how this was reflected in expressed behaviour by agents. In addition, they simulated dopaminergic responses alongside modelling the agent's expected precision, and showed how manipulating these model parameters generated distinct behavioural profiles. Mirza *et al*. (2019) also modelled a selective attention task in the context of Active Inference and was able to closely replicate saccadic eye movement patterns as seen in a previous study of human exploration (Mirza *et al.*, 2019). However, the above examples were executed on synthetic behaviour of simulated agents. In this chapter I present the application of the same general schemes to real human behavioural data and estimate subject-specific parameters that can be used to characterise individual participants based on their computational and behavioural profiles.

In my analyses of human behaviour compared with synthetic behavioural data produced using this MDP, I also showed that task performance, specifically total reward earned and the frequency of optimal policy selection both pre- and post-reversal, can be accurately replicated through this Active Inference model. I did, however, expect some level of variation between the real human and simulated datasets. As the task used here is constructed as a probabilistic decision tree, the exact sets of actions and states between datasets are very unlikely to be identical. An agent, endowed with highly accurate k and c parameters, may take the same actions as its human counterpart for the first few trials, but due to the probabilistic and uncertain nature of the task, there will inevitably be variations in the states experienced by the agent and will therefore lead to subsequent deviations in action and policy selection from the original data. Different combinations of k and c may also produce similar behaviours; high values of k and low values of c are both characteristic of highly exploratory/empirical agents, therefore it is possible that a range of parameter combinations could be used to describe the same set of actions. Despite this, I observed strong correlations in behavioural metrics between the datasets and was able to closely re-estimate the initial conditional MAP estimates through a second model inversion.

I also found that, in one of my model inversions, the conditional estimates of k, the parameter which described the gradient of the activation function used to calculate the model decay factor α, formed an inverted-U shape with respect to the frequency of optimal policy selection. This suggests that a mid-value of k produces optimal behaviour, in that participants with a k parameter in this range are able to exploit the optimal policy both pre-

and post-reversal. The inverted-U shaped function was originally coined by Yerkes and Dodson (1908) to describe the effects of too little or too much arousal on performance in difficult cognitive tasks. They postulated that when arousal is too low, an individual will show reduced concentration and motivation, whereas when arousal is too high, the individual will suffer from impaired decision-making and working memory due to excessive stimulation and divided attention, specifically in more difficult cognitive tasks (Yerkes and Dodson, 1908). Therefore, an optimal participant must have a mid-level of arousal to perform optimally in difficult tasks.

This inverted-U shape has since been observed in a number of studies, in both genetic studies of single nucleotide polymorphisms (SNPs) in the NA transporter NET1 (norepinephrine transporter 1), and in studies examining DA modulation of task performance (Cools and D'Esposito, 2011). Cools and Robbins (2004) suggested that different tasks may be described by different inverted U-shaped functions, as in the context of DA, different tasks may be either enhanced or compromised by different levels of DA (Cools and Robbins, 2004). Therefore, tasks may also respond differently to drug manipulation of DA, depending on the optimal levels of neurotransmitter for each task. This may also apply to NA; as reported by Aston-Jones and Cohen (Aston-Jones and Cohen, 2005). Aston-Jones and Cohen described an inverted U-shaped relationship between tonic LC activity and cognitive performance, stating that optimal performance in a task occurs with a moderate tonic LC firing profile, with strong phasic LC firing in response to task-relevant stimuli. Similarly to the original Yerkes-Dodson theory, low levels of tonic LC activity occur in an inattentive and non-alert participant, whereas at the opposite end of the curve, high levels of tonic LC activity are evident in highly distractable participants, and the effects of any phasic bursts of activity are dulled by consistently high tonic discharges.

Furthermore, a genetics study by Sigurdardottir *et al*. (2016) used positron emission tomography (PET) imaging to investigate the effects of SNPs in the NA transporter gene in patients with ADHD and healthy controls (Sigurdardottir *et al.*, 2016). They found an inverted-U shaped effect of NET1 receptor binding, in that an inverse effect of NET1 availability was reported for the major and minor alleles: higher ADHD symptom scores were associated with greater NET1 availability in major allele-carrier patients, whereas for minor allele carriers, higher ADHD symptoms scores were associated with reduced NET1 availability. More recently, Nemoda *et al*. (2018) proposed an inverted-U shaped mechanism for the effect of SNPs in the NET-encoding gene *SLC6A2* on cognitive functions in children with or without ADHD (Nemoda *et al.*, 2018). Future work may combine such genetics studies with computational modelling, by examining subject-specific *k* parameter estimates in individuals expressing different NET1 gene variants, and how this may relate to psychiatric diagnoses and cognitive functioning. Alternatively, pharmacology studies may be conducted in which the effect of NET1 receptor blockade on task performance and model flexibility could be interrogated.

A limitation of this study may be that the two of the free parameters estimated in the model, *k* and *c*, may not be completely independent of one another, which may result in differences between true parameters and the conditional MAP estimates produced through model inversion. For example, I saw in the inversion of simulated data that values of *c* were not accurately retrieved through model inversions when *k* had a fixed value of 16. It could be the case that a high *k* value such as this produced hyper-flexible internal models (as a high *k* would produce lower values of α), therefore the agents would display hyper-exploratory behaviours. This would lead to a reduced capacity to exploit rewarding routes. Such behaviour can also be expected from agents with low precision over state preferences, defined by low values of *c*.

A key aspect of the task design is the distinction between high-level and low-level rewards, and the ability of participants to actively seek the routines containing high-level rewards. Reduced motivation, diminished interest and anhedonia are some of the core depressive symptoms in Major Depressive Disorder (MDD) (Conradi *et al.*, 2011). A number of studies have shown that individuals suffering from MDD or those experiencing anhedonia have a reduced sensitivity to rewarding outcomes. For example, a study by Clery-Melin *et al*. (2011) used a behavioural paradigm which involved handgrip force to investigate incentive modulation, and found that depressed patients displayed a deficit in incentive modulation compared to healthy participants, as they showed a reduced ability to increase handgrip force in response to higher potential rewards (Clery-Melin *et al.*, 2011). Another more recent study found that, in patients suffering from MDD treated with vortioxetine, there was an association between a measure of cognitive function and increased physical effort for higher magnitude and probability rewards, using the Effort Expenditure for Rewards Task (EEfRT) (Subramaniapillai *et al.*, 2019). Such studies indicate that individuals with MDD may experience reduced reward sensitivity or increased emotional blunting towards higher levels of reward. In my model, such individuals may therefore have lower conditional estimates of the *c* parameter, in line with their reduced motivation to actively seek the higher-level lower-probability rewards in the task.

Low values of *c* produce more empirical behavioural profiles, and such agents show reduced preferences/ability to distinguish between the high and low reward levels. Future studies may involve participants with MDD and could examine how subject-specific precision over state preferences varies in depression, and how the neural signatures of SAPEs may be diminished in individuals with MDD.

In summary, using this Active Inference modelling framework I was able to retrieve with accuracy conditional MAP estimates of model parameters signalling internal model volatility (*k*), and precision over rewarding states (*c*) for real human participants performing a decision-making task, through a model inversion scheme. Furthermore, I revealed associations between these conditional estimates and participants' behavioural data, in that total reward was significantly associated with precision over rewarding states, and an inverted-U shaped relationship emerged between optimal policy selection and internal model volatility.

Chapter Four

Using fMRI to Investigate Associations Between Model Volatility and Reward, and Activations in Locus Coeruleus and Anterior Cingulate Cortex

4.1 Introduction

In this chapter, I combine the behavioural paradigm designed in *Chapter Two* and the Active Inference model inversion pipeline constructed in *Chapter Three*, with fMRI and a pharmacological manipulation to investigate the neural signatures of SAPEs and how these neural signals may relate to drug-induced selective NA reuptake inhibition. Focusing on the model parameters *k* and *c* as introduced in *Chapter Three*, I also probe how such neural activity and behaviour in the reversal paradigm may be associated with conditional parameter estimates of internal model volatility (*k*) and precision over rewarding states (*c*), and how these may be influenced by drug manipulation.

NA is a catecholamine neurotransmitter with vital roles in high-level cognitive functioning such as arousal, executive control, and the formation and consolidation of memories. The importance of NA in such functions manifests in the wide distribution of NA-containing axons across the central nervous system (Ramos and Arnsten, 2007; Sara, 2009).

NA exerts its effects by binding to membrane-bound, G protein-coupled adrenergic receptors known as adrenoceptors. NA binds to three different adrenoceptor types: alpha-1 adrenoceptors (α_1), alpha-2 adrenoceptors (α_2) and beta adrenoceptors (β). Of these, there are further subdivisions of receptor types: α_1 is subdivided into α_{1A}, α_{1B}, and α_{1D}; and α_2 is subdivided into α_{2A}, α_{2B}, α_{2C}, and α_{2D}. β adrenoceptors are also subdivided into β_1, β_2 and β_3 receptor subtypes. α_1 and β adrenoceptors are thought only to exist at postsynaptic sites, whereas α_2 adrenoceptors (mainly α_{2A} and less frequently, α_{2C}) are said to be found at both pre- and post-synaptic locations (Maletic *et al.*, 2017).

NA displays the highest affinity for α_2 adrenoceptors, which have an inhibitory effect on cell signalling pathways as these receptors are known to reduce intracellular cAMP (cyclic adenosine monophosphate) (ZhuGe *et al.*, 1997; Maletic *et al.*, 2017). This contrasts the effects of α_1 and β adrenoceptors, which have been shown to increase cellular levels of cAMP, and therefore exert a stimulatory effect on cell signalling (Cotecchia *et al.*, 1990). However, NA binding at these β and α_1 adrenoceptors only occurs during periods of high NA concentrations as NA has lower levels of affinity for α_1 and β adrenoceptors compared to α_2. Therefore, low levels of NA release may have an overall inhibitory effect on neuronal activity due to α_2 adrenoceptor binding, and as NA increases, neuronal activity may increase as NA begins binding to α_1 and β adrenoceptors. This describes a modulatory role for adrenoceptors in the relationship between NA release and neuronal transmission.

The major source of NA in the brain is the LC. The LC is a cluster of neurons positioned in the pontine brainstem, adjacent to the fourth ventricle. This nucleus has a relatively small size, in adult humans measuring ~14.5 mm in length and 2.5 mm thick, with approx. 10,000-15,000 neurons (Berridge and Waterhouse, 2003). Despite this, the LC possesses expansive forebrain connections and is responsible for the majority of NA projections to the spinal cord and the neocortex, including regions such as the hippocampus, thalamus, hypothalamus, amygdala and cerebellum (Chamberlain and Robbins, 2013; Maletic *et al.*, 2017), supplying NA across the central nervous system. As a result of the expansive connections of the LC-NA network, any dysregulation of this system may culminate in the dysfunction of many cognitive processes and therefore may lead to various cognitive and affective disorders including stress-related disorders such as post-traumatic stress disorder (PTSD), and ADHD (Berridge and Waterhouse, 2003).

For example, Chandley *et al*. examined astrocyte pathology in the noradrenergic LC in both individuals with MDD and healthy controls, and found reduced expression of glutamate transporter genes in the astrocytes of patients with MDD (Chandley *et al.*, 2013). A pharmacological study also showed impaired memory consolidation across both MDD patients and healthy controls following administration of clonidine, an α_2 receptor agonist, leading to suppression of noradrenergic activity (Kuffel *et al.*, 2014). More recently, Nemoda *et al*. (2018) conducted a genetic association study to investigate SNPs in the *SLC6A2* gene, which encodes the NA transporter NET1, in children expressing symptoms of ADHD (Nemoda *et al.*, 2018). They suggested an inverted-U shaped modulatory effect of these SNPs over the relationship between NA levels in prefrontal cortex and cognitive functioning, in that the SNPs showed associations with attention deficits, but in contradictory directions between the community sample versus the child psychiatry sample.

Two key cortical structures from which the most dominant projections to the LC descend are the OFC and the ACC. These structures have been implicated in the processing of task-related utility, e.g. in the processing of rewards and punishment, respectively, in the context of behavioural tasks (Aston-Jones and Cohen, 2005; Apps and Ramnani, 2014; Wang *et al.*, 2017). The ACC receives a wide range of converging inputs from various brain structures, such as the ventral striatum and amygdala, and has also specifically been associated with evaluation of costs, pain and aversive stimuli (Fuchs *et al.*, 2014). In particular, increased ACC activation has been observed in response to monetary loss and performance errors in decision-making tasks (Shenhav *et al.*, 2013; Foti *et al.*, 2015; Kolling *et al.*, 2016).

Reboxetine is a selective NA reuptake inhibitor, as it binds specifically to the NA transporter protein, NET1, and prevents the reuptake of extracellular NA into synaptic terminals through the NET1 transporter. A single oral dose of reboxetine has been shown in a number of studies to increase levels of salivary cortisol, which is an indicator of increased central noradrenergic activity (Hennig *et al.*, 2000; Hill *et al.*, 2003; Miskowiak *et al.*, 2007). Reboxetine has been widely studied in the context of working memory and temporal processing (Rammsayer *et al.*, 2001), and also as an antidepressant (Harmer *et al.*, 2003).

In this chapter, I conducted an fMRI study in combination with manipulating NA by blocking NA reuptake with a single 4 mg dose of reboxetine, and investigated how this manipulation can influence belief updating and SAPEs in the decision-making task as described in *Chapter Two*, under an Active Inference scheme. I also replicated the three-parameter Active Inference model inversion on the participants in this fMRI study, as previously conducted on the 25 healthy volunteers from the main behavioural study, detailed in *Chapter Three*. Finally, using task-based fMRI, I identified neural signatures of SAPEs in this spatial memory and decision-making task, and how these neural signatures relate to parameter estimates from the model inversion, which describe the mechanisms underlying behavioural differences in each participant.

I hypothesized that: 1) the participants would successfully learn the structure of the behavioural task but would show a range of behavioural profiles across the group, reflecting my findings in the main behavioural study in *Chapter Two*; 2) following estimation of subject-specific parameters of internal model volatility (*k*) and precision over reward (*c*), I would observe a group-wide effect of the SNRI, in that parameter MAP estimates would change between the pre-drug training session and the post-drug testing session; and 3) I would observe task-related significant activations in the LC and ACC associated with conditional MAP estimates of behaviourally-relevant parameters, as strongly task-relevant brain regions.

Figure 4.1

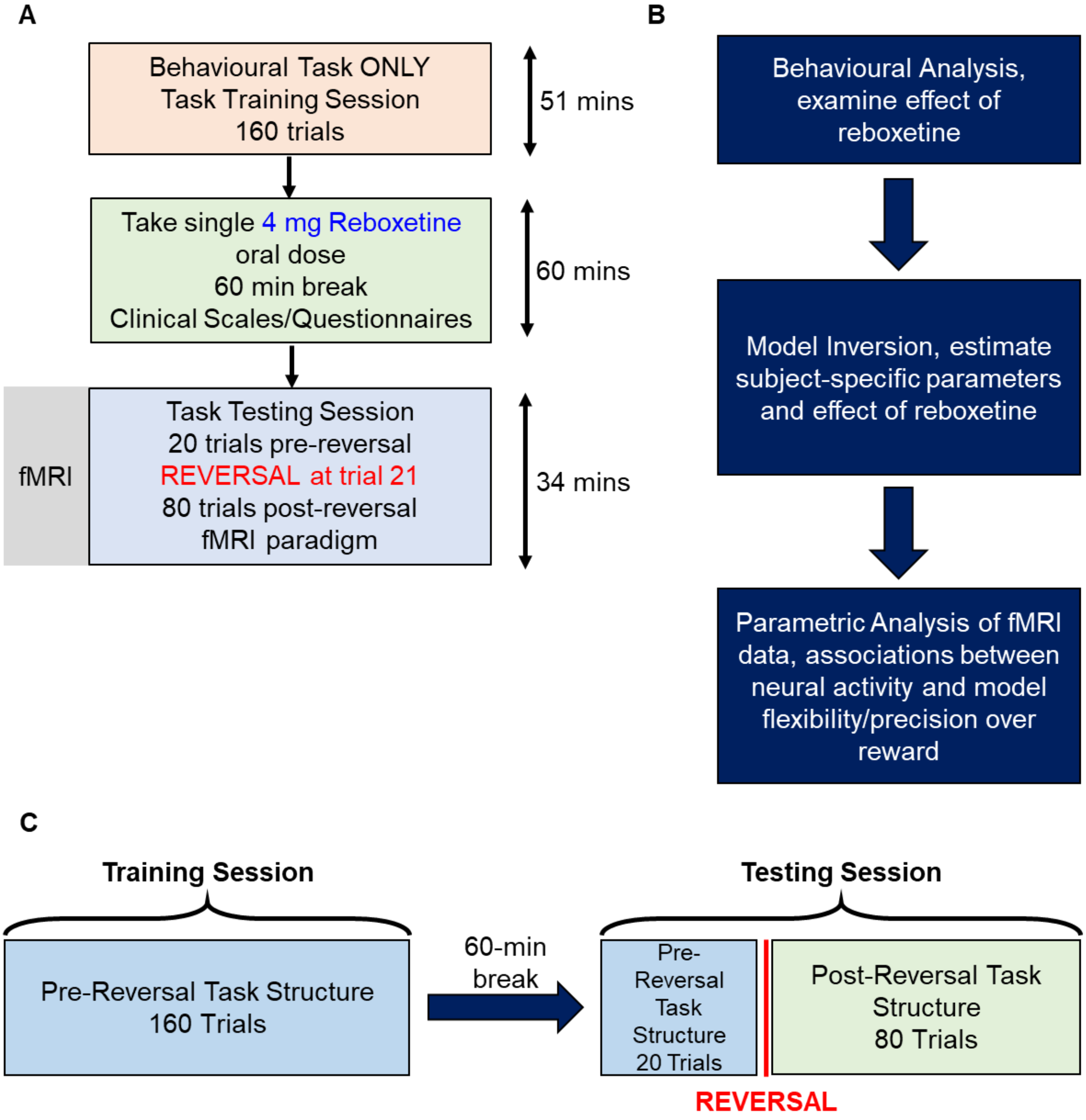

A
Behavioural Task ONLY
Task Training Session
160 trials
51 mins
Take single 4 mg Reboxetine
oral dose
60 min break
Clinical Scales/Questionnaires
60 mins
fMRI
Task Testing Session
20 trials pre-reversal
REVERSAL at trial 21
80 trials post-reversal
fMRI paradigm
34 mins
B
Behavioural Analysis, examine effect of reboxetine
Model Inversion, estimate subject-specific parameters and effect of reboxetine
Parametric Analysis of fMRI data, associations between neural activity and model flexibility/precision over reward
C
Training Session
Pre-Reversal Task Structure
160 Trials
60-min break
Testing Session
Pre-Reversal Task Structure
20 Trials
Post-Reversal Task Structure
80 Trials
REVERSAL

Figure 4.1 Study and Analysis Pipeline. (**A**) Structure of fMRI experiment. Each participant attended for one experimental appointment, which consisted of the task training session, oral administration of reboxetine, then completion of the task testing session during fMRI scanning. (**B**) Analysis pipeline for data collected from 16 healthy adult volunteers. Behavioural data, demographic and psychological battery data, and fMRI data were collected for each participant. (**C**) Task structure. Reversal occurred after trial 20 in the testing session only, during MRI scanning.

4.2 Methods

4.2.1 Participants

Sixteen healthy adult volunteers (mean age = 29.6 ± 11.9 SD, range 18-58, nine females) were recruited from the staff and student population of King’s College London, and from the general population of south-west London, through online and email adverts. Sample size was set to reflect that which is typical in the published literature (O'Doherty *et al.*, 2003; Chadwick *et al.*, 2015). Written consent was obtained at the start of the study appointment following application of exclusion criteria and verbal COVID-19 symptom screening, which included having their temperature taken and recorded. This study was approved by University of Bristol School of Psychological Sciences Research Ethics Committee (reference: 95123).

Any participants with contraindications to reboxetine, contraindications to MRI, any current or history of neurological or psychiatric illness including depression, regular or recent use of psychoactive drugs, use of any medication that may interact with reboxetine including anti-depressants, and any pregnant or breastfeeding participants were excluded from the study. Urine samples were collected from all participants to test for drugs of abuse (cocaine, amphetamine, methamphetamine, cannabis, opiates, benzodiazapine) and, if female, pregnancy. All participants were right-handed. One participant was unable to complete the full fMRI session of the task, and only completed 64 trials out of 100, due to a scanner failure. This participant was not excluded from my analyses, but the shorter task length was accounted for in all analyses.

4.2.2 Behavioural Paradigm

The task structure was identical to that used in the main behavioural study (see *Chapter Two*), with minor changes in timings, and task length of the testing session (**Figure 4.1A,C**). The training session lasted for 160 trials with a constant pre-reversal task structure, identical to that of the main behavioural study, with altered timings. The testing session started with 20 trials using this same task structure, then a reversal occurred after trial 20, switching to the post-reversal task structure as used in the main behavioural study. The testing session was shortened due to scanning constraints on the total task length, and to ensure the comfort of the participant. For each trial, participants had 3 s to respond by pressing one of the two buttons on the keyboard or button box. If participants failed to make a key/button press within this time, the restart screen was presented for 2 s, and the trial restarted. The rewarding scene at the end of each trial was presented for 2 s. The inter-state interval was randomly sampled from a uniform distribution from 3-5 s, and a fixation cross was displayed between each trial and between each state transition. In this study, each trial had a fixed length of 19 s, therefore the inter-trial interval was determined on a trial-by-trial basis depending on the participants' RTs and the inter-state interval, to make the total trial length up to 19 s. The training session consisted of 160 trials, and lasted approx. 51 minutes, not including missed trials.

The testing session took place during MRI data acquisition, and was shortened to 100 trials, with a two-minute break after the 20th trial. During the short break, a black screen with white text 'Short Break' was presented and no responses were recorded from the participant; the break was introduced for the comfort of the participant in the scanner. The reversal in task structure occurred after this short break for the remaining 80 trials. The reversal followed the same task structure change as in the main behavioural study. The inter-state and inter-trial intervals were calculated as above, with each trial having a fixed duration of 19 s. The testing session lasted approx. 34 minutes, including the two-minute break, not including missed trials. For behavioural analyses, intermediary states (states two-to-five) are referred to as level two states. The initial state (state one) is referred to as the level one state. Outcome states are referred to as level 3 states (*see* ***Inserts*** *in* ***Results***).

4.2.3 Active Inference Modelling

Model Inversion

To estimate subject-specific parameters across the MRI cohort, I applied the same Active Inference model as described in *Chapter Three* with the actions and states experienced by the MRI study participants. Once again, I focused on the three free model parameters encoding: a scalar on internal model volatility (k), a precision parameter over rewarding outcome states (c), and the midpoint of the activation function that defines internal model volatility (m). Based on the model comparison analyses conducted in *Chapter Three*, I ran the three-parameter model inversion scheme, with free parameters *k*, *c* and *m*. To examine the effect of drug manipulation on the conditional MAP parameter estimates, I ran the model inversions for the pre-drug trials (trials 1-160) and the post-drug trials (161-260) separately.

Parametric Empirical Bayes Analysis

I then aimed to identify how the parameter estimates generated through model inversion predicted behavioural changes on a group-level, and how they varied with respect to the drug manipulation. Following the same pipeline as in *Chapter Three*, I ran PEB analyses for the pre-drug training session and the post-drug testing session separately, based on the total reward earned and total percentage of optimal policy selection per participant. The structure of the Bayesian GLM was identical to that constructed in *Chapter Three* but with training and testing sessions separated: a three-column design matrix consisting of a column of ones to denote the average mean effect across participants, a second column containing the total reward earned in pounds per session, and a third column containing the percentage of optimal policy selection for each session in the task.

4.2.4 MRI Data Analysis

MRI Data Acquisition

All imaging data were collected at the Centre for Neuroimaging Sciences, King's College London, using a Discovery™ MR750 3.0T MR scanner (General Electric, Boston, MA, USA) with a 32-channel head coil. Functional imaging data were collected with TE/TR = 26/2000, FOV = 20 cm, flip angle = 80°, and resolution = 2.1 x 2.1 x 3.3 mm, acquiring 40 axial oblique slices parallel to the AC/PC line to cover the whole head. Functional imaging was preceded by T2 FLAIR (Fluid Attenuated Inversion Recovery) structural scans and followed by acquisition of FieldMap data (TE1/TE2/TR = 4.90/7.30/500, FOV = 240 cm, flip angle = 60°, resolution = 1.9 x 1.9 x 3.3 mm). Resting state data were also acquired, with TE/TR = 26/2000 ms. Participants also underwent physiological monitoring during scanning; respiratory movement and cardiac pulse waveform were monitored using respiratory bellows and a pulse oximeter.

Image Pre-Processing

Functional images were realigned to correct for motion and unwarped using voxel displacement maps produced from fieldmap data, then coregistered to each participant's anatomical scan and brought into alignment with tissue probability maps (TPMs) to transform the data into MNI space. The raw fieldmap images of two participants were corrupted, therefore the functional images of these participants simply underwent realignment for motion correction prior to coregistration without unwarping. TPMs used were released with the SPM12 toolbox (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) with tissues for grey matter, white matter, cerebrospinal fluid, bone, soft tissue, and air/background. For normalisation of images in MNI space, a voxel size of 2 x 2 x 2 mm was used for functional images, and a smaller voxel size of 1 x 1 x 1 mm was used for structural images; spatial normalisation and resampling were performed together. Functional images were then spatially smoothed with a kernel size of 3 x 3 x 3 mm (FWHM) and 4th degree B-spline interpolation.

First-level Analysis

I first aimed to examine the differences in neural activation between likely and unlikely states including both intermediary states (level two) and outcome states (level three); and between all pre-reversal and post-reversal states. Onset timings for six state conditions (initial state pre-reversal, initial state post-reversal, likely pre-reversal, unlikely pre-reversal, likely post-reversal, unlikely post-reversal) were defined for all states experienced for each participant in the task (*see* ***Box 4.1***), and modelled using a canonical haemodynamic response function and additional temporal derivative using SPM12. Motion parameters estimated from image realignment were added to the GLM as regressors, in addition to physiological noise correction regressors for cardiac pulse and respiratory movement. Individual GLMs with five contrasts were generated on a within-subject first level: one with all state conditions averaged, the second and third contrasting all pre-reversal states to all post-reversal states (pre > post reversal, then post > pre-reversal), and the fourth and fifth contrasting all likely states (plus the initial state, state one) to all unlikely states (likely + state one > unlikely states, then unlikely > likely + state one).

Second-level Analysis

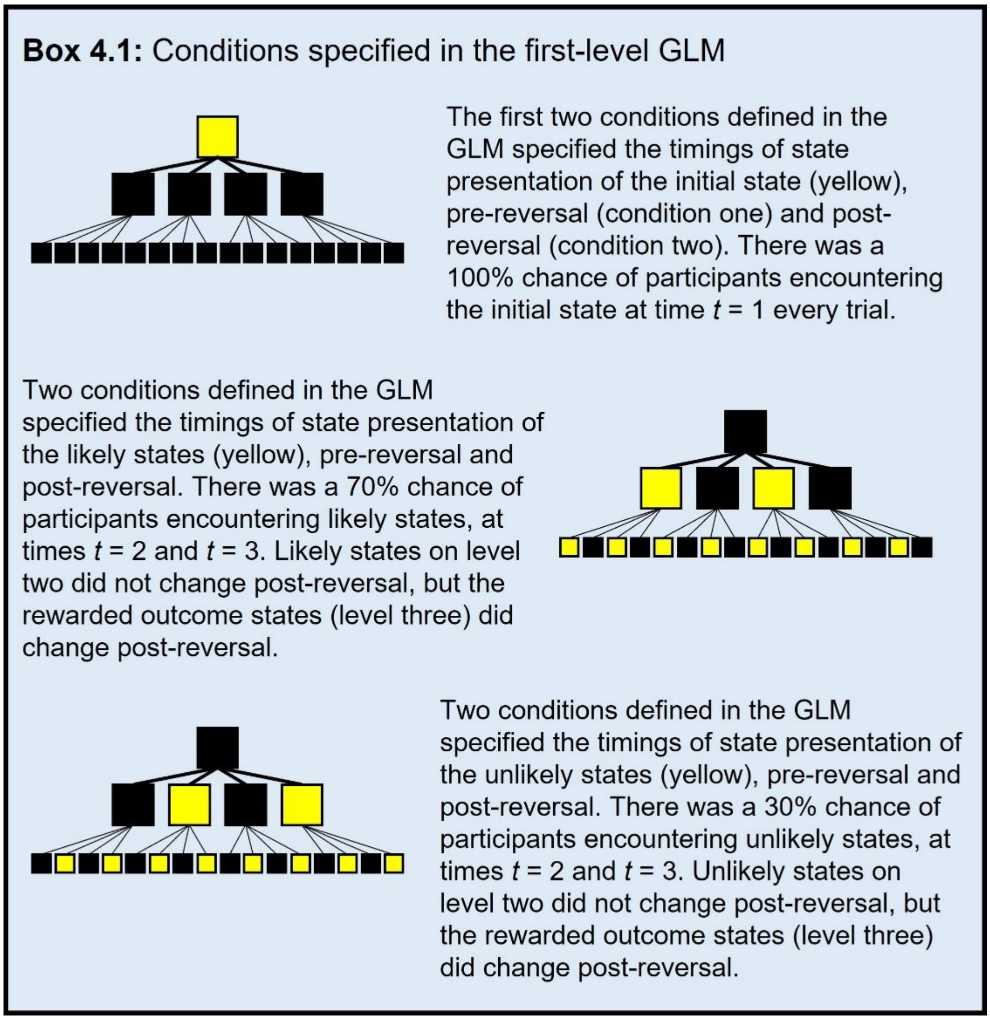
I then aimed to investigate the group-level differences in neural activation associated with state probability and effect of reversal, specifically in task-relevant brain regions. To achieve this, I used each contrast from the within-subject analysis to conduct one-sample *t*-tests for each of the GLMs, both with and without an explicit mask. A mask covering the ACC, midbrain and pons was generated using the WFU-Pick Atlas toolbox (Maldjian *et al.*, 2003; Maldjian *et al.*, 2004) alongside SPM12, selected regions as defined in the AAL atlas (Tzourio-Mazoyer *et al.*, 2002), and applied to the data. These regions were selected for the mask due to strong task relevance (Aston-Jones and Cohen, 2005). To examine group-level effects, I used a one-sample *t*-test. I initially used a simple group average to investigate average effects of state condition (effect of reversal or state probability) on a group level. I then introduced subject-specific parameter estimates of *k*, *c*, and *m*, generated in log space through the three-parameter model inversion, as covariates to examine how conditional MAP estimates of model parameters may predict BOLD responses, at both the whole-brain cluster level, using an extent threshold of 25 voxels, and peak-level using the explicit ACC/pons/midbrain mask. For peak-level covariate analyses, significance was determined at $p < 0.05$ small volume corrected (sphere radii = 8 mm); *p* values are reported for peak-level, unless specified as cluster-level.

4.2.5 Statistical Analyses

To investigate the behavioural profiles in the task, individual chi-squared tests were conducted on participants' policy selection and action selection frequencies within-subject. I also conducted binomial (sign) tests on policy selection to look specifically at the selection frequencies of optimal policies in the task. Paired *t*-tests were used to highlight significant differences in participants' RTs between likely (states two and four) and unlikely (states three and five) intermediary states at time $t = 2$. I also used a two-way ANOVA to examine the effect of reboxetine on RTs in likely versus unlikely intermediary (level-two t = 2) states, i.e. comparing training session and testing session RTs, using MATLAB R2020b.

For classical analyses of conditional MAP parameter estimates generated through model inversion, Pearson's correlations were used to examine relationships between conditional MAP estimates and task performance metrics, i.e. total reward and optimal policy selection. Optimal policy selection was calculated as the percentage of trials in which the current optimal policy was selected by the participant. Paired *t*-tests were used to examine the change in conditional MAP estimates from the training session (pre-drug) versus the testing session (post-drug). Goodness-of-fit analyses were also conducted to investigate relationships between conditional MAP estimates and behavioural metrics of task performance, with SSE and adjusted R-square values reported. To examine significant activations in neuroimaging data, one-sample *t*-tests were conducted using the SPM12 toolbox in MATLAB, with Family-Wise Error (FWE) corrected and uncorrected *p* values reported. For full data analysis pipeline, see **Figure 4.1B**.

Box 4.1



Box 4.1: Conditions specified in the first-level GLM

The first two conditions defined in the GLM specified the timings of state presentation of the initial state (yellow), pre-reversal (condition one) and post-reversal (condition two). There was a 100% chance of participants encountering the initial state at time *t* = 1 every trial.

Two conditions defined in the GLM specified the timings of state presentation of the likely states (yellow), pre-reversal and post-reversal. There was a 70% chance of participants encountering likely states, at times *t* = 2 and *t* = 3. Likely states on level two did not change post-reversal, but the rewarded outcome states (level three) did change post-reversal.

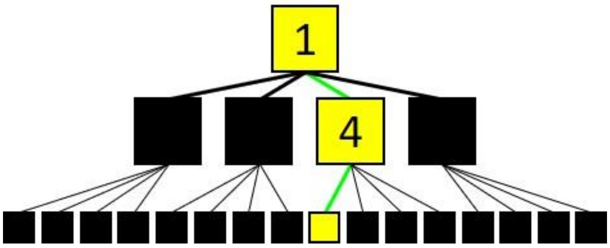
Two conditions defined in the GLM specified the timings of state presentation of the unlikely states (yellow), pre-reversal and post-reversal. There was a 30% chance of participants encountering unlikely states, at times *t* = 2 and *t* = 3. Unlikely states on level two did not change post-reversal, but the rewarded outcome states (level three) did change post-reversal.

Box 4.1 Outline of the conditions specified in the first-level GLM in fMRI data analysis.

4.3 Results

4.3.1 Behaviour: Training Session, Prior to Drug Administration in Pre-Reversal Task Structure

Participants Successfully Learned Optimal Policy Selection



In the training session, which took place prior to drug administration, all 16 participants selected any policy significantly more than chance level ($p < 0.05$, individual chi-squared tests). By conducting binomial sign tests per participant, I found that 14 participants selected policy three, the pre-reversal optimal policy, significantly more than chance, and one participant selected this policy significantly less than chance, showing significant preference for policy four instead. When dividing the total trials into the first and last 80 trials, there is evidence that most participants learned the optimal policy in the first half of the training session, and remaining participants continued learning over the course of the task and learned the optimal policy in the last 80 trials. Ten participants selected policy three significantly more or less than chance in the first 80 trials ($p < 0.05$, binomial sign tests), with 12 participants showing significant preference for any policy, i.e. they did not select policies at random ($p < 0.05$, chi-squared tests) (**Figure 4.2**). However, in the last 80 trials, 15 participants selected policy three significantly differently to chance, and all 16 participants showed significant preference for any particular policy, demonstrating the continued learning of participants across the whole task. (***Insert:** initial state at level one numbered as state one. To take the optimal policy, policy three, the participant must move right to reach state four, then left to earn 25p (70% chance); optimal states in yellow. Green path lines indicate the optimal policy*).

Figure 4.2

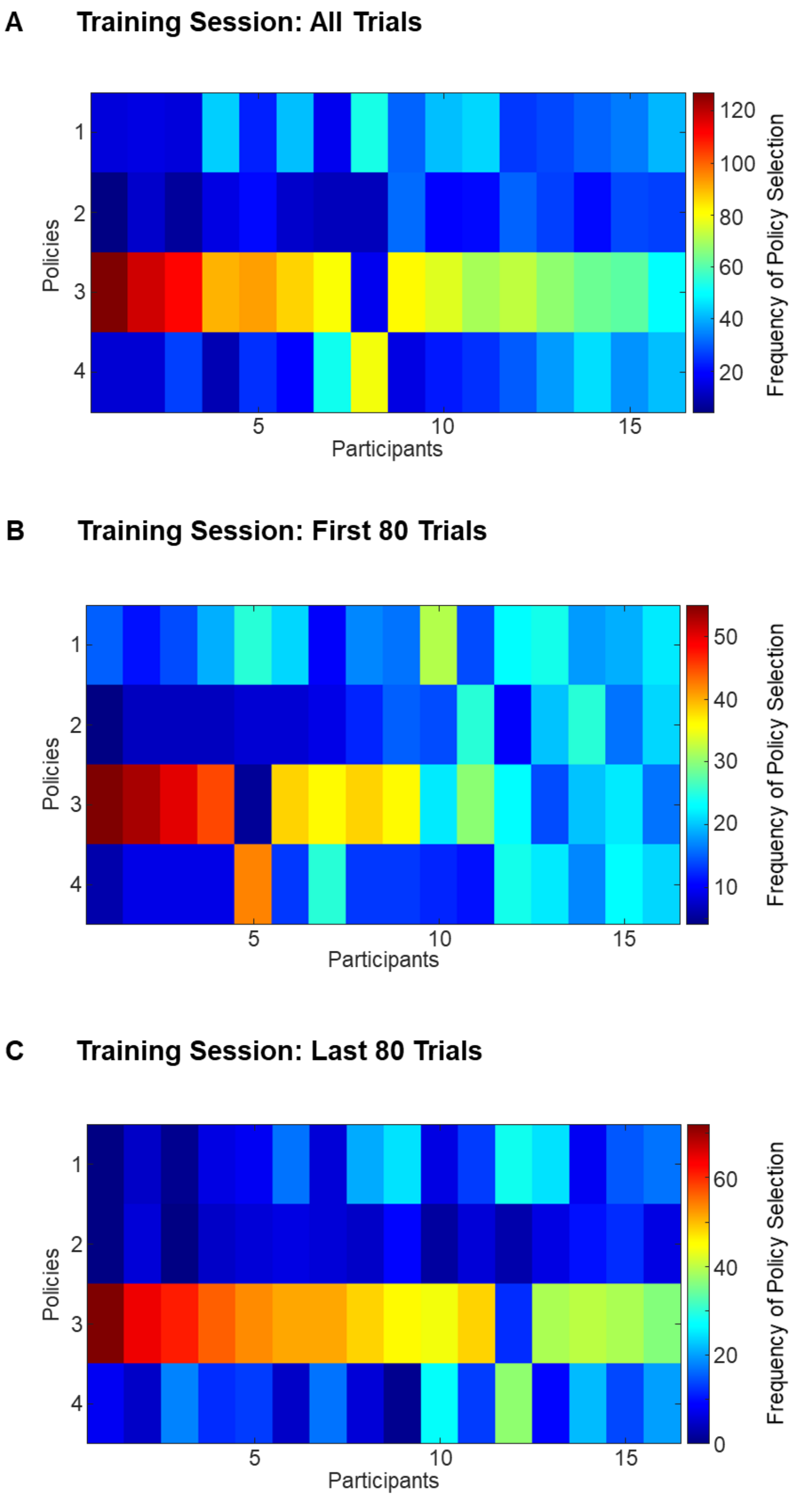


Figure 4.2 Behaviour in Training Session. Policy selection frequencies across all trials (**A**), the first 80 trials (**B**), and the last 80 trials (**C**) of the training session ordered by performance (*p*

values of policy selection chi-squared tests). Colour bars indicate frequency of policy selection. For individual action choices from unlikely level two (intermediary) states, only one participant moved left significantly more than chance from state three at *t* = 2, the unlikely intermediary state in the left, less rewarding arm of the task. From state five, the unlikely level two state in the right arm of the task, eight participants chose a particular action significantly different to chance, with four moving left and four moving right more than chance. Similar to state three, this state is less likely to be reached, and the probability of receiving any reward following this state is also very unlikely, therefore participants are not expected to learn which action leads to the most rewarding outcome from states three and five.

This continued learning is further reflected in the moving average of right moves at *t* = 1. Across all 160 trials, the moving average of right moves at *t* = 1 significantly increases over time (*rho* = 0.813, *p* < 0.001), indicating that participants move right first more consistently as the task progresses (**Figure 4.3A**).

Figure 4.3

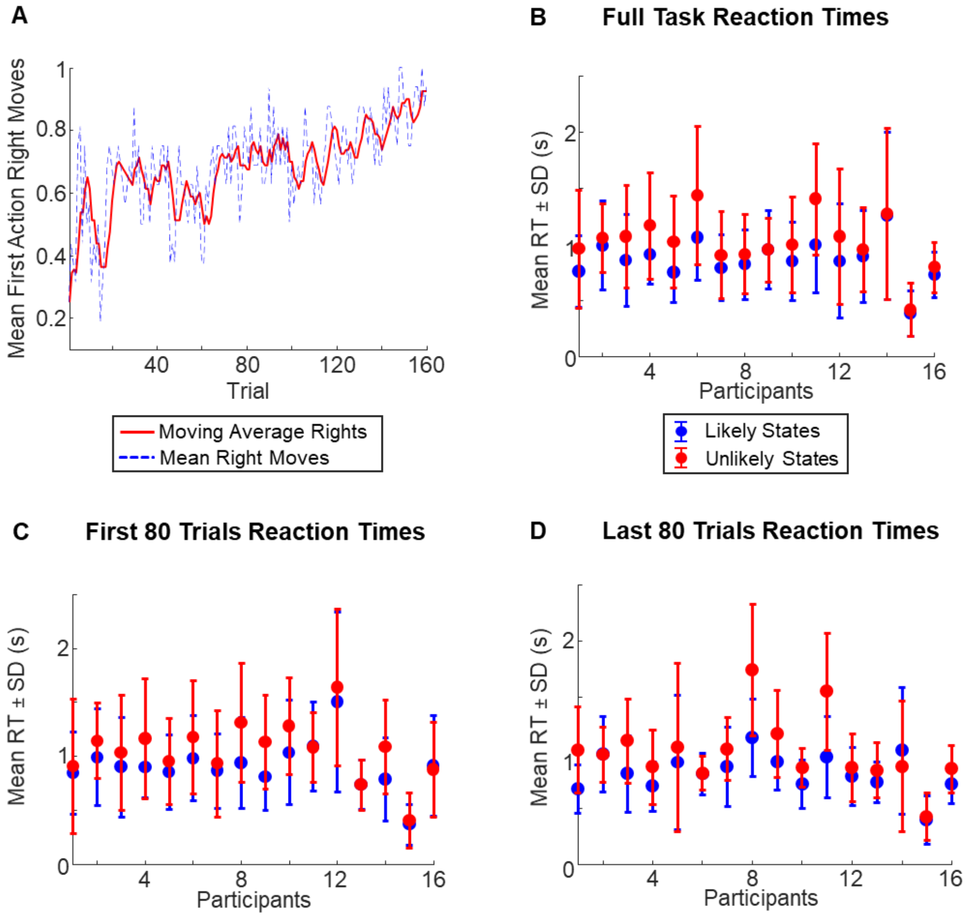
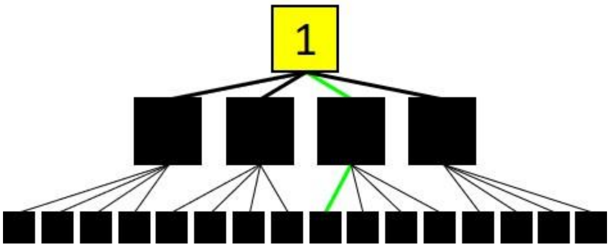


Figure 4.3 Actions in Training Session. (**A**) Moving average of right moves at *t* = 1, from state one, in the training session. (**B**-**D**) RT differences between likely and unlikely level-two (intermediary) states, across the whole training session (**B**), during the first 80 trials (**C**), and during the last 80 trials (**D**). Participants are ordered by performance (*p* values of policy selection chi-squared tests). RTs following likely states are consistently faster than RTs following unlikely states. SD = standard deviation.

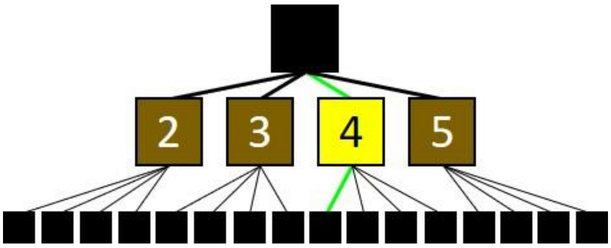
Participants Consistently Made Optimal Action Choices At Levels One And Two Of The Task Structure, Further Supporting Model-Based Learning

I also examined the individual action choices at each time step in the trial by conducting chi-squared tests on a within-participant level, in addition to policy selection. For the initial action choice from state one at *t* = 1, 15 participants chose to move right significantly more than chance, with one participant at chance level. When examining the first and last 80 trials separately, the frequency of initial right moves increases as the task session progresses, in line with the moving average described above. In the first 80 trials, eight participants moved right at *t* = 1 significantly more than chance, and eight participants were not significantly different to chance. In the last 80 trials, this increased to 14 participants who moved right at *t* = 1 significantly more than chance. Across the training session, no participants moved left at *t* = 1 significantly more than chance level. (***Insert:*** *initial state at level one numbered as state one, in yellow. To take the optimal route from state one, the participant must move right, pre-reversal. Green path lines indicate the optimal policy*).



For the intermediary (level two) states, only six participants chose to move left from state two at *t* = 2. This is likely due to the fact that most participants learned the optimal policy which required them to move right at *t* = 1 rather than left, therefore were less likely to find themselves in state two.

To follow the optimal policy, participants would expect to find themselves in state four at *t* = 2, the most likely state reached after moving right at *t* = 1. In the training session, 14 participants moved left from state four at *t* = 2 significantly more than chance, with the remaining two participants at chance level, thus following the optimal policy. (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state pre-reversal is state four, in yellow. Green path lines indicate the optimal policy*).



Reaction Times Following Likely States Were Significantly Faster Than That Following Unlikely States, Consistently Across The Training Session

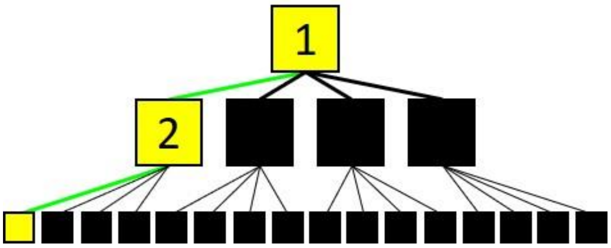
RTs across the training session, and in the first and last 80 trials separately, were also examined in likely versus unlikely intermediary (level two) states. Over all 160 trials, RTs following unlikely intermediary states were significantly slower than RTs following likely intermediary states ($t(15) = -5.04$, $p = 1.46 \times 10^{-4}$) (**Figure 4.3B**). When dividing the first and last 80 trials, RTs between likely and unlikely states followed the same pattern. In the first 80 trials, RTs following unlikely states were significantly slower than RTs following likely states ($t(15) = -4.62$, $p = 3.36 \times 10^{-4}$), and this was also the case in the last 80 trials, although to a slightly lesser extent ($t(15) = -3.62$, $p = 0.0025$) (**Figure 4.3C-D**). These differences suggest that participants were able to work out which states were the likely and unlikely intermediary states, but learning of the unlikely states may have improved during the second half of the training session.

4.3.2 Behaviour: Testing Session (Post-drug) with Task Reversal and fMRI

Participants Successfully Learned The Post-Reversal Optimal Policy And Changed Strategy Rapidly Following Reversal Onset

In the testing session, the task consisted of 20 trials pre-reversal, and a further 80 trials post-reversal. Prior to the reversal, six participants selected policy three significantly more than chance, retaining the information they had acquired during the training session (*p* < 0.05, binomial sign tests). Ten participants, however, chose any policy significantly different to chance, suggesting that some participants had explored the environment and selected alternative policies more than chance level (*p* < 0.05, chi-squared tests) (**Figure 4.4A**).

Following the reversal, 14 participants selected the new optimal policy, policy one (*see **Insert***), significantly different to chance, with 13 of these participants selecting policy one more than chance, and one participant selecting policy one significantly less than chance, opting for policies two and three instead. This indicates that this participant may have persevered with the pre-reversal optimal policy after the reversal (policy three), then explored the task searching for the new optimal policy, but failed to discover that the new optimal policy was policy one (***Insert***). By conducting individual chi-squared tests, I found that 15 participants selected any policy significantly different to chance; only one participant selected policies at chance level. However, across the cohort, most participants were able to successfully find the new optimal policy post-reversal and take that route significantly more than alternative routes (**Figure 4.4B**). (***Insert:** To take the post-reversal optimal route from state one, the participant must move left to reach state two, in yellow, then left to earn 25p (70% chance). Green path lines indicate the post-reversal optimal policy*).



When examining the moving average of right moves from state one at *t* = 1, the switch in preferred policy is evident, as the number of right moves from state one continues to increase up to trial 20, the onset of the reversal, which is immediately followed by a significant sharp drop in right moves from state one (*rho* = -0.875, *p* < 0.001) (**Figure 4.5A**).

Figure 4.4

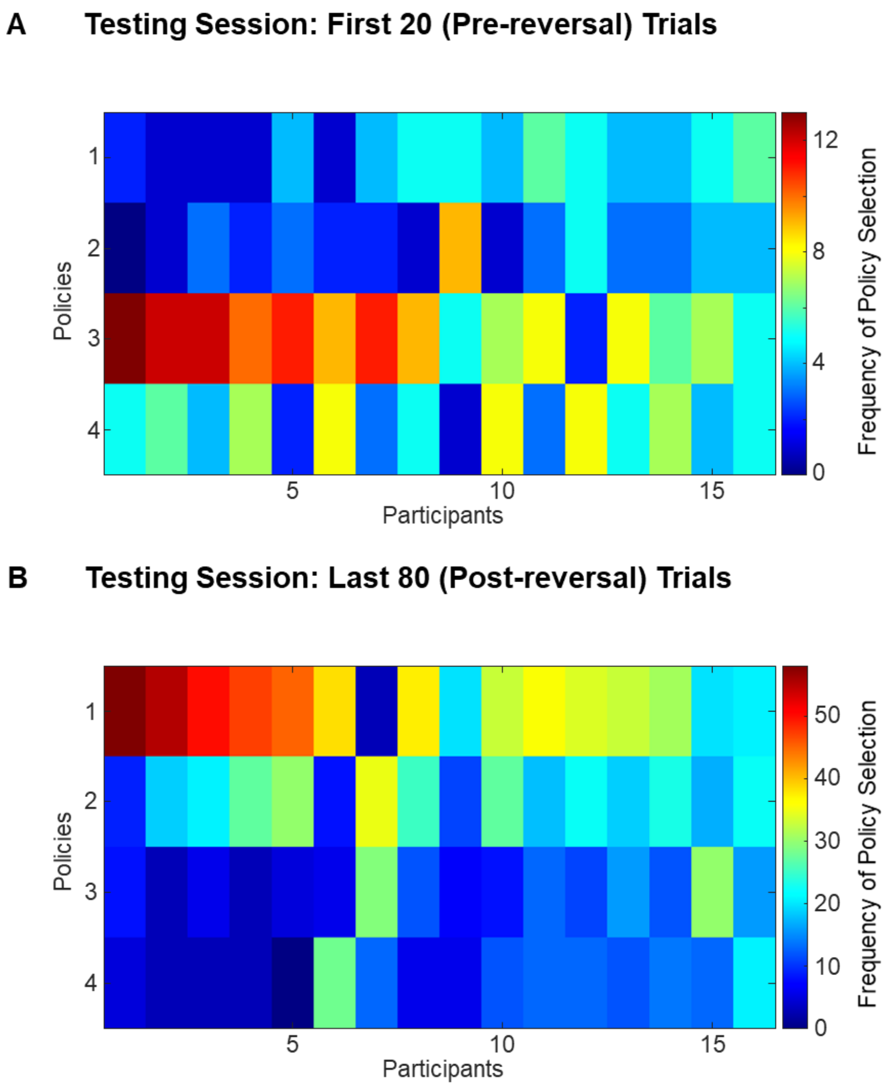
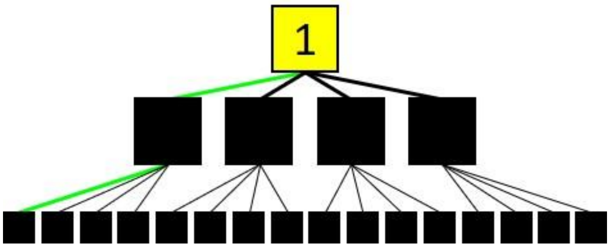


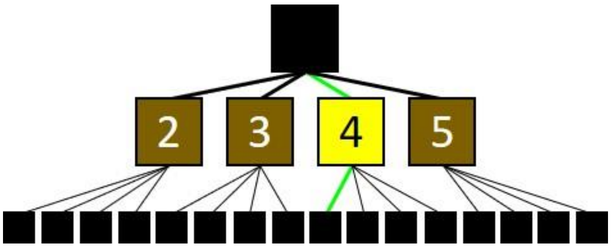
Figure 4.4 Behaviour in Testing Session. Policy selection frequencies across the first 20 trials which took place before the reversal (**A**), and the last 80 (post-reversal) trials (**B**) of the testing session ordered by performance (*p* values of policy selection chi-squared tests). Colour bars indicate frequency of policy selection. For individual actions taken from unlikely level two states pre-reversal, no participants moved differently to chance from state three (left arm), and only one participant moved differently to chance from state five (right arm). Following the reversal, again only one participant moved differently to chance from state five, but 12 participants moved right significantly more than chance, thus following the most optimal route from this unlikely state, and one participant moved left significantly more than chance.

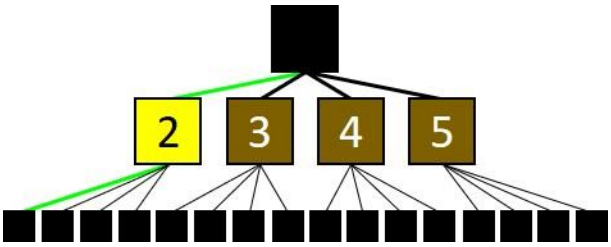
Individual Action Analysis Showed Adaptation to Reversal And Continued Model-Based Navigation Of The Task, In Line With Previous Analyses

For the initial actions from state one at *t* = 1 pre-reversal, six participants moved right significantly more than chance level and no participants moved left significantly more than chance (*p* < 0.05, individual chi-squared tests), thus continuing to select the optimal arm of the task. Most participants, however, appeared to act in a more exploratory fashion. Following the reversal, this changed to 12 participants who moved left significantly more than chance, with no participants who moved right significantly more than chance, indicating that by the end of the task, most participants had noticed the task reversal and were either continuing to explore the environment to locate the new location of high reward, or were exploiting the new optimal policy, policy one. (***Insert:*** *initial state at level one numbered as state one, in yellow. To take the optimal route from state one after the reversal, the participant must move left. Green path lines indicate the optimal policy*).



For actions from the most optimal level two state (state four, see ***Insert***) at *t* = 2 prior to the reversal, nine participants moved left significantly more than chance, and the remaining seven participants were at chance level; no participants moved right significantly more than chance, in line with the suggestion above that participants either explored during the first 20 trials of the task, or continued to exploit the current optimal route. Following the reversal, however, this dropped to four participants; three of which continued to take the pre-reversal optimal policy, and one who moved right from this state significantly more than chance. This is likely due to most participants preferring to explore the left arm of the task, therefore exploiting the new optimal policy rather than persevering with the former optimal policy. (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state pre-reversal was state four, in yellow. Green path lines indicate the pre-reversal optimal policy*).





For actions taken from the post-reversal optimal level two state (state two) at t = 2, 13 participants moved left significantly more than chance, thus following the optimal policy, but one participant moved right significantly more than chance from this state, instead selecting policy two (the second most optimal policy). This may be a result of a delayed switching from the formerly-optimal right arm, to the newly-optimal left arm of the task, which may have delayed learning of optimal actions from level two states. Prior to the reversal, only one participant moved differently to chance from this state, as expected, since this arm of the task was not optimal pre-reversal. (***Insert:*** *intermediary states at level two numbered two-to-five. Optimal level-two state post-reversal was state two, in yellow. Green path lines indicate the post-reversal optimal policy*).

Overall, this examination of individual actions demonstrates that participants were able to rapidly notice and respond to the task reversal, particularly with significant action switching at the first task level at t = 1, further supporting the conclusion that action selection cannot be explained completely by model-free learning, in line with my findings described in *Chapter Two*.

Reaction Times Remained Significantly Faster Following Likely Versus Unlikely States

Across the full testing session, similarly to the training session, RTs following likely level-two states were significantly faster than those following unlikely level-two states ($t(15) = -5.15$, $p = 1.19 \times 10^{-4}$) (**Figure 4.5B**). Before the reversal occurred, this effect that was seen in the training session was seen again, as RTs following unlikely intermediary states were significantly slower than RTs following likely states ($t(15) = -3.60$, $p = 0.0026$) (**Figure 4.5C**). Post-reversal, this effect is still seen and becomes even greater still ($t(15) = -4.04$, $p = 0.0011$) (**Figure 4.5D**). This suggests that, even after the reversal and participants had to readjust their beliefs about the structure of the task, they retained the knowledge of the likely versus unlikely states, and continued to learn the transition probabilities of the intermediary states.

Figure 4.5

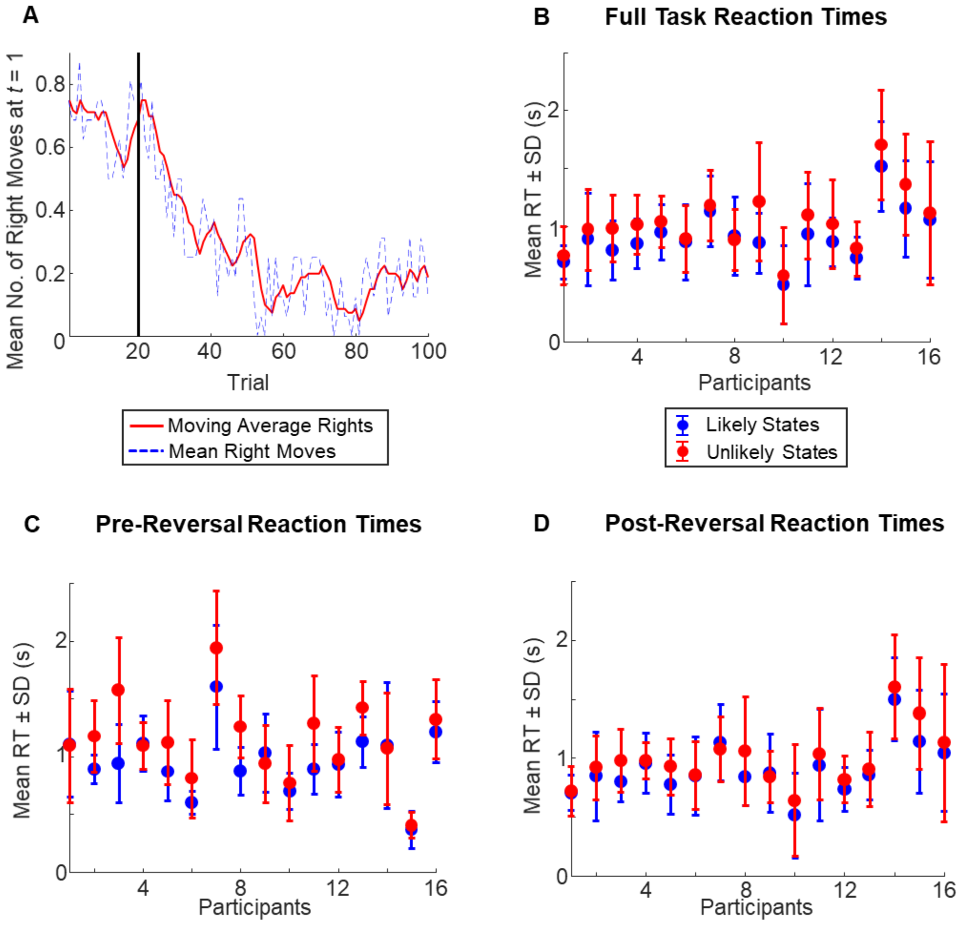


Figure 4.5 Actions in Testing Session. (**A**) Moving average of right moves at *t* = 1, from state one, in the testing session. (**B**-**D**) RT differences between likely and unlikely level-two (intermediary) states, across the whole testing session (**B**), during the first 20 pre-reversal trials (**C**), and during the last 80 post-reversal trials (**D**). Participants are ordered by performance (*p* values of policy selection chi-squared tests). RTs following likely states are consistently faster than RTs following unlikely states, both before and after the reversal. SD = standard deviation.

No Significant Effect Of Reboxetine On Behaviour At The Group Level

To investigate the effect of reboxetine on participants' behaviour in the task, I looked at optimal policy selection (%) and reward rate between the training and testing sessions. I also looked at the RT differences between the training and testing sessions, to examine whether differences between participants' RTs for likely versus unlikely states were affected by reboxetine, and to ensure that reboxetine did not have a significant influence on participants' overall reaction speed. Reward rate was calculated as the average amount of money earned per trial in pence (total reward/trial number), to account for the shorter length of the testing session. A paired *t*-test showed no significant difference in participants' optimal policy selection between training and testing sessions ($t(15) = 0.838$, $p = 0.415$). There was also no significant difference in reward rate between the training and testing sessions ($t(15) = -1.20$, $p = 0.250$, paired *t*-test). This may be a result of subject-specific mixed effects of the drug manipulation, in that participants may have been affected differently by NA reuptake inhibition, depending on their performance level in the pre-drug training session (**Figure 4.6A-B**).

When looking at RTs, a two-way ANOVA showed that there was no significant effect of drug on overall RT ($F(1,63) = 0.31$, $p = 0.578$), and no significant interaction between the effect of drug and state probability ($F(1,63) = 0.11$, $p = 0.737$). This suggests that reboxetine did not have a significant overall effect on RTs across participants, and did not influence RTs following likely versus unlikely level-two states (**Figure 4.6C**).

Figure 4.6

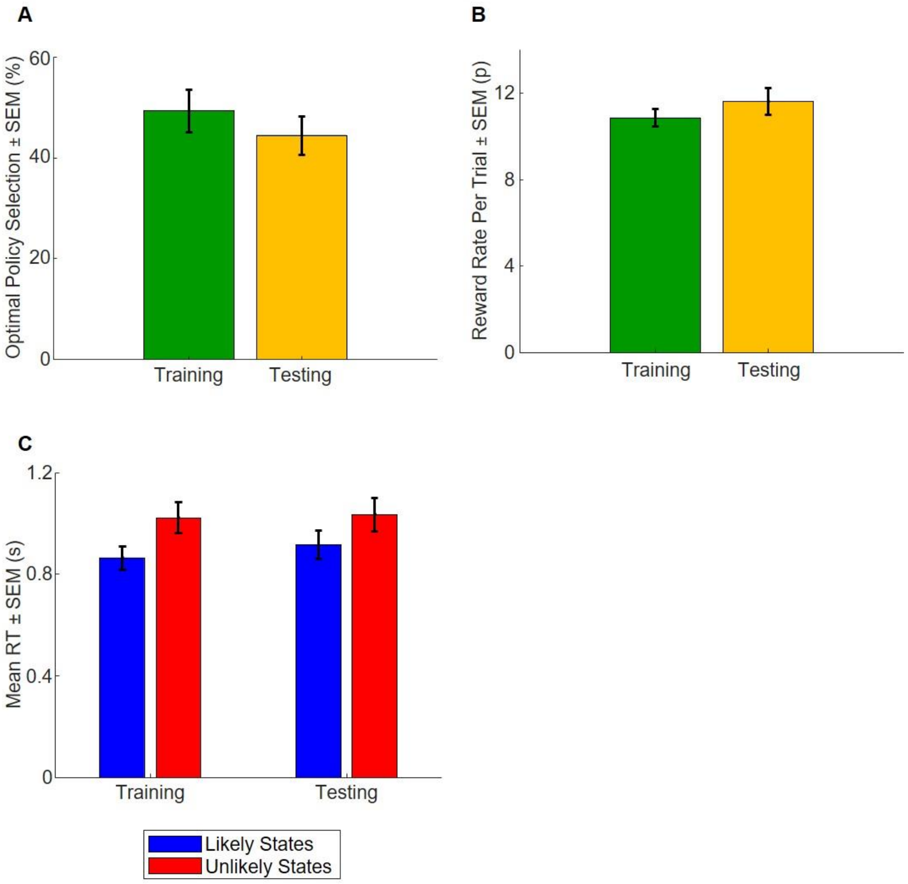


Figure 4.6 Comparisons of Training and Testing Sessions. (**A**) Optimal policy selection (%) for participants in the pre-drug training session versus post-drug testing session, between which was no significant difference. (**B**) Reward rate (pence) for participants; there was no significant difference in participants' reward rate between the training and testing sessions. (**C**) RTs following likely and unlikely intermediary states. There was no significant effect of reboxetine on RTs, and no significant interaction between state probability (likely versus unlikely states) and drug. SEM = standard error of mean.

4.3.3 Classical Analyses of Model Inversion Parameters Revealed Effect of Reboxetine on Internal Model Volatility, but not Precision Over Reward

Using the same model inversion pipeline as that used in *Chapter Three*, I inputted the sets of states and actions from the 16 human participants in the MRI study into the model inversion, to estimate subject-specific parameters of model volatility and precision over reward. Based on my results as described in *Chapter Three* which found that the three-parameter model best represented the data, I used the three-parameter inversion with fixed parameters *α min* = 2 and *α max* = 1024 for these participants, including *m* as a free parameter. I ran model inversions separating sets of states and actions from the training session (prior to drug administration) and testing session (post-drug administration), to examine the effect of reboxetine on these parameter estimates.

I first examined differences in conditional MAP estimates of *k*, *c*, and *m* in the pre-drug training session versus the post-drug testing session, to look for any effect of reboxetine across participants. A paired *t*-test demonstrated no significant difference between estimates of *c* pre-drug versus post-drug (*t*(15) = -0.236, *p* = 0.817). This indicates that reboxetine did not have an effect on the participants' precision over rewarding states. Similarly, there was no significant difference between estimates of *m* pre-drug versus post-drug (*t*(15) = -0.214, *p* = 0.833). However, a slight difference in *k* parameter estimates emerged between task sessions, though non-statistically significant (*t*(15) = 1.92, *p* = 0.0735), in that *k* decreased across the cohort following reboxetine administration. This non-significant result may be due to mixed effects of reboxetine on optimal policy selection, as on an individual basis, participants who were more exploratory in the training session (low optimal policy selection) appear to show increases in optimal policy selection, and vice versa (**Figure 4.7A**).

Figure 4.7

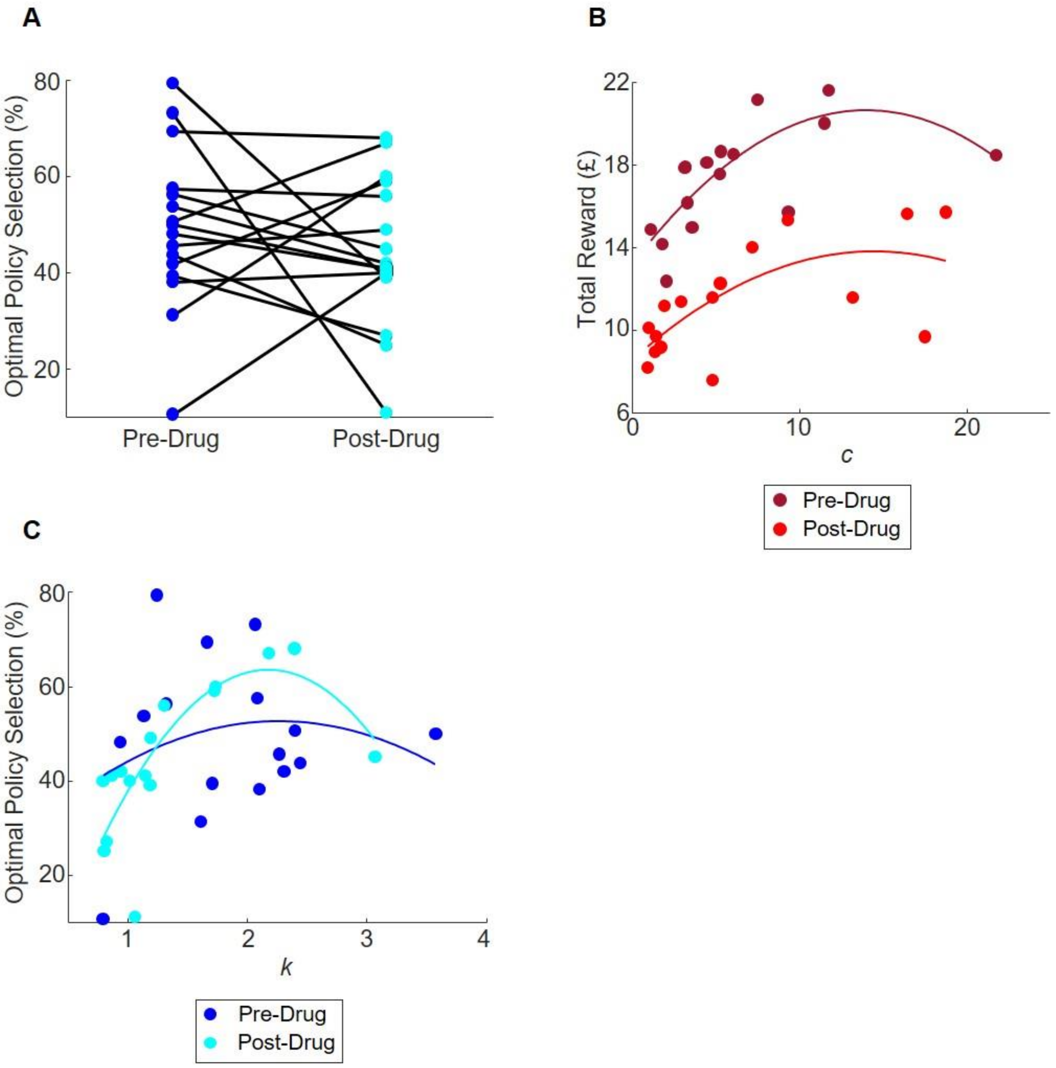


Figure 4.7 Conditional MAP Estimates Associated with Behaviour. (**A**) Change in optimal policy selection (%) from the pre-drug training session to the post-drug testing session. On an individual level, some participants show large increases or decreases in optimal policy selection, whereas others show smaller changes, displaying mixed effects. (**B**) Relationship between conditional MAP estimates of *c* (precision over reward) and total reward earned by participants in the pre-drug training session and post-drug testing session. (**C**) Relationship between conditional MAP estimates of *k* (internal model volatility) and optimal policy selection (as percentage of trials in which the optimal policy was selected), in the pre-drug training session and the post-drug testing session.

I then examined the parameter estimates in relation to coarse behavioural metrics using classical statistics. Across both the training and testing sessions, there were significant moderate correlations between estimates of *c* and total reward per session (training: *rho* = 0.529, *p* = 0.0350; testing: *rho* = 0.616, *p* = 0.0110) (**Figure 4.7B**). Conversely, there were no significant correlations between estimates of *m* and optimal policy selection (percentage of optimal policy selection of out of total number of trials) in either task session (training: *rho* = 0.375, *p* = 0.152; testing: *rho* = 0.0467, *p* = 0.864). Interestingly, there was no significant correlation between *k* estimates and optimal policy selection for the training session (*rho* = 0.0957, *p* = 0.724). However, this became significant during the testing session (*rho* = 0.603, *p* = 0.0134) (**Figure 4.7C**).

In my PEB analyses, I initially examined the effects of total reward on the *k*, *c*, and *m* parameters, then examined the effects of optimal policy selection (%) after accounting for reward, across the training and testing sessions separately. In the pre-drug training session, there was a moderate association between *c* and total reward earned (*Ep* = 0.0465, *Pp* = 0.781), but no associations between *k* or *m*, and either behavioural metric (*Pp* = 0). In the testing session, there was moderate association between *m* and total reward earned (*Ep* = 0.0511, *Pp* = 0.751), with a reduced association between *c* and total reward earned (*Ep* = 0.0215, *Pp* = 0.539), again with no association between *k* and either total reward or optimal policy selection (*Pp* = 0).

When examining curve fitting for *k*, a change in the inverted-U shape, as seen in the main behavioural study modelling described in *Chapter Three*, is evident pre- versus post-drug. Prior to reboxetine administration, there was a flattening of the curve, and greatly reduced data fit when fitting polynomials of degrees one to three (linear fit: SSE = 4.12 x 10^3, adjusted R-square = -0.0616, d.f. = 14; second-degree polynomial: SSE = 3.92 x 10^3, adjusted R-square = -0.0882, d.f. = 13; third-degree polynomial: SSE = 2.96 x 10^3, adjusted R-square = 0.109, d.f. = 12). However, post-drug, curve fitting greatly improved with strong fits for second and third degree polynomials, and the inverted-U shaped relationship between *k* and optimal policy selection steepened, with *k* estimates shifting to the left across the group (linear fit: SSE = 2.27 x 10^3, adjusted R-square = 0.318, d.f. = 14; second-degree polynomial: SSE = 1.31 x 10^3, adjusted R-square = 0.576, d.f. = 13; third-degree polynomial: SSE = 1.16 x 10^3, adjusted R-square = 0.594, d.f. = 12) (**Figure 4.7C**).

By also looking at curve fitting for *c*, it is clear that this steepening and shifting of the inverted-U occurs exclusively for *k*. Similarly to *c* estimates from the main behavioural study in *Chapter Three*, a saturation in *c* estimates is evident: reward increases as *c* estimates increase, then as *c* increases beyond 15, a plateau in reward is observed (**Figure 4.7B**). For both the training session (pre-drug) and testing session (post-drug), second-degree polynomials best fit the relationship between *c* and total reward per session, in line with the observation of the saturation effect (*Pre-drug*: linear fit: SSE = 6.93 x 10^5, adjusted R-square = 0.229, d.f. = 14; second-degree polynomial: SSE = 4.21 x 10^5, adjusted R-square = 0.496, d.f. = 13; third-degree polynomial: SSE = 4.06 x 10^5, adjusted R-square = 0.473, d.f. = 12. *Post-drug*: linear fit: SSE = 6.40 x 10^5, adjusted R-square = 0.335, d.f. = 14; second-degree polynomial: SSE = 5.71 x 10^5, adjusted R-square = 0.362, d.f. = 13; third-degree polynomial: SSE = 5.59 x 10^5, adjusted R-square = 0.324, d.f. = 12.). This provides evidence that the SNRI reboxetine specifically affects internal model volatility in human participants, without affecting precision over reward.

4.3.4 Task-Based Magnetic Resonance Imaging Reveals Activation in Locus Coeruleus And Anterior Cingulate Cortex Associated With Parameters of Precision Over Reward and Internal Model Volatility, Respectively

An ACC/Pons/Midbrain mask was applied to the functional imaging data, and effects of reversal and state probability were investigated using a one-sample *t*-test. There was a significant increase in activation in the LC region in pre-reversal trials compared with post-reversal trials across the group (*p* = 0.014, FWE corrected) (**Figure 4.8A**). Activation in the region of the LC was also observed in an average across all states, and for the positive effect of likely states, but these activations were non-significant.

I then introduced the conditional MAP estimates of *k*, *c*, and *m* generated in log space through model inversion (described above) into the one-sample *t*-test as covariates. Examining the data on a whole-brain level (in the absence of the mask) using a cluster threshold of 25 voxels for a positive effect of likely states, a significant cluster was found in the left supramarginal gyrus, correlated with these subject-specific conditional MAP estimates of *k* (cluster-level: $p < 0.01$, FWE corrected) (**Figure 4.8B**).

Figure 4.8

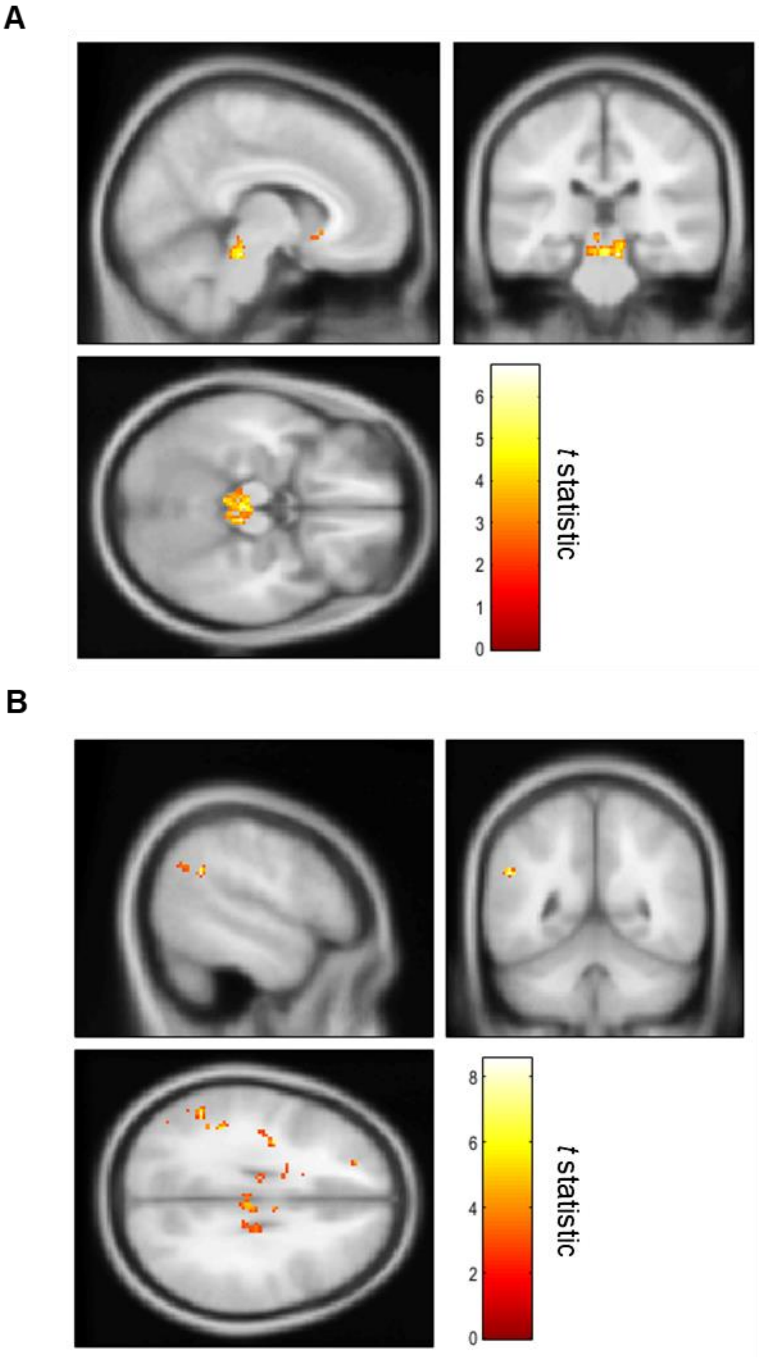


Figure 4.8 Task-associated Activations. (**A**) Significant activation in the LC for the positive pre-reversal condition, across all participants. (**B**) Significant cluster-level activation in the left supramarginal gyrus for the positive effect of likely states, associated with subject-specific *k* parameter estimates. Colour bars indicate *t* statistic.

Using small volume *p* value correction with the application of the ACC/pons/midbrain mask, significant activation was also found in the ACC associated with MAP estimates of *k* for a positive effect of likely states across the task (*p* = 0.006, FWE corrected) (**Figure 4.9A**). Associated with MAP estimates of *c*, also using small volume *p* value correction, significant activation was found in the LC region, again for a positive effect of likely states (*p* = 0.003, uncorrected). On a subject-specific level, this LC activation appears to fluctuate per-participant based on performance in the task (**Figure 4.9B**). There was also additional activation in the region of the VTA associated with estimates of *c* (*p* = 0.001, uncorrected), a region previously noted for its relevance to reward processing (Schultz *et al.*, 1997).

Figure 4.9

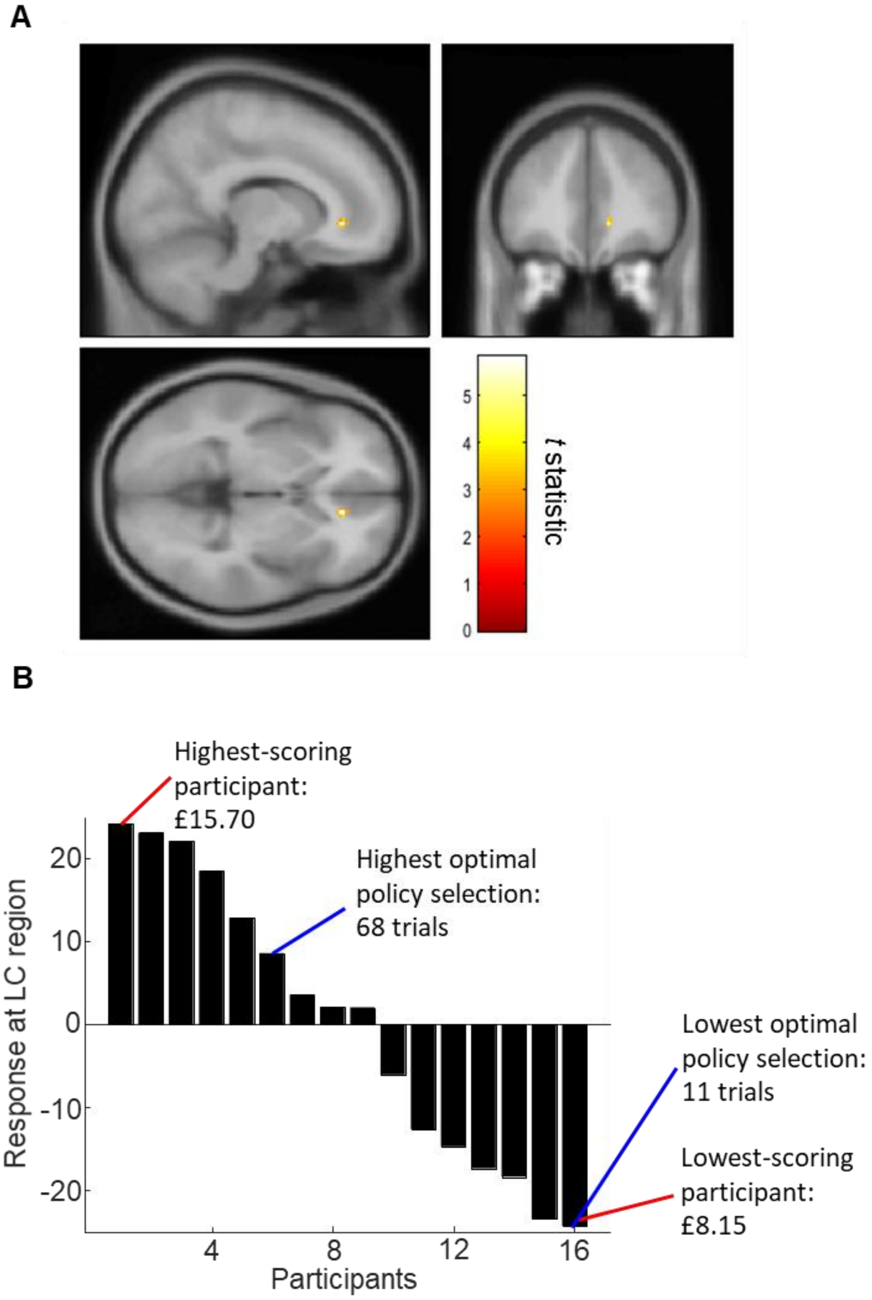


Figure 4.9 Task-associated Activations. (**A**) Significant activation in the ACC for the positive effect of likely states, associated with *k* parameters. (**B**) Fitted predicted response at LC region associated with the covariate *c* (precision over reward conditional MAP estimates) for the positive likely states condition. Participants are ordered by conditional MAP estimates of *c*, from highest to lowest. Highest-scoring participants showed highest LC activation, and lowest-scoring participants showed lowest LC activation in this condition, associated with parameter estimates of precision over reward (*c*). Colour bar indicates *t* statistic.

4.4 Discussion

In this study, I found significant activity in the LC associated with precision over rewards, under the condition of likely task states. I also found significant activation in the ACC associated with internal model flexibility, again under the condition of likely task states. This suggests a role for NA in motivation and potentially behavioural energising, particularly in a cognitively difficult task such as the one I used here.

Similarly to my previous main behavioural study as detailed in *Chapter Two*, most participants were able to successfully learn the structure of the probabilistic decision-making task, and identified the optimal policy both pre- and post-reversal. However, once again I also observed a wide spectrum of behavioural phenotypes, with the highest performing participants displaying highly exploitative policy preference but were flexible enough to adapt to the new optimal policy post-reversal, and some lower-performing participants either displaying highly exploratory behaviour, or exhibiting perseverative behaviour (persisting with the pre-reversal optimal policy for many trials after the reversal).

Crucially, although the difference in *k* parameter estimates between the pre-drug training session and the post-drug testing session was not statistically significant on the group level, by examining optimal policy selection on an individual basis a mixed effect of reboxetine was revealed. Participants who had low optimal policy selection in the training session, i.e. more *exploratory* participants, displayed sharp *increases* in optimal policy selection following selective NA reuptake inhibition. Conversely, participants who showed high levels of optimal policy selection in the training session, i.e. more *exploitative* participants, displayed sharp *decreases* in their optimal policy selection during the testing session. Therefore, in line with my results describing an inverted-U shaped relationship between task performance and model volatility, this mixed effect of reboxetine suggests that subject-specific changes in participants' levels of NA can either improve or diminish task performance, depending on initial pre-drug behaviour.

I was also able to estimate subject-specific parameter estimates of internal model volatility and precision over rewarding states, and I examined these in relation to coarse behavioural metrics, i.e. optimal policy selection and total reward, in an effort to identify the underlying mechanisms which explain the behavioural differences observed in the task. I found that administration of reboxetine exerted a slight yet specific effect on participants' model volatility/flexibility, in that participants' estimates of model volatility decreased across the cohort, although not statistically significant. Following drug manipulation, the distribution of model volatility estimates also changed, as the inverted-U shaped distribution (as observed and analysed in *Chapter Three*) steepened in the task testing session after the SNRI had taken effect. Neither of these effects were seen in participants' estimates of precision over reward, indicating that the *k* parameter in this model, representing model volatility, is specifically affected by NA reuptake inhibition. This finding is in line with previous work, which suggest that NA modulates internal model volatility (Sales *et al.*, 2019).

Recent work has suggested that NA plays multiple roles, encoding both behavioural flexibility and motivation in cognitively challenging tasks. Jahn *et al*. described a cost/benefit decision-making task combined with α_2 NA receptor agonism using clonidine in non-human primates (Jahn *et al.*, 2018). They revealed a dose-dependent effect of the NA receptor agonist, which acts to reduce central levels of NA, in that both choice variability and physical force exertion were reduced, without the cost/benefit trade-off suffering as a result. Another non-human primate study examined a reward/effort trade-off paradigm, in the context of NA and DA (Varazzani *et al.*, 2015). Their findings suggest roles of DA not only in reward, as previously established, but also in the anticipation of action cost. NA, conversely, was linked again to behavioural energising and action motivation. This work implies a specific role of NA in effort processing, which ultimately influences the perception of upcoming reward value, dependent on the effort or difficultly required to obtain the reward.

Manipulations of NA have also been previously investigated during a probabilistic serial RT task. A study by Marshall *et al*. (2016) used the NA receptor antagonist prazosin to examine contextual uncertainty and how manipulation of noradrenergic pathways may modulate uncertainty and belief updating, studying in particular volatility uncertainty. Volatility uncertainty describes our beliefs about how stable our environment is, and therefore the rate of change of probabilistic relationships between different contexts (Marshall *et al.*, 2016). This study found that prazosin increased the rate that individuals updated their volatility estimates, suggesting that NA antagonism influences the rate that individuals update their beliefs about the volatility of the environment. However, prazosin did not have a significant effect on the tonic learning rate about probabilistic contexts. From this, they concluded that NA is a key modulator in the learning of uncertainty emerging from unexpected changes in the environment. This concurs with Yu and Dayan, who theorised that NA encodes unexpected uncertainty (Yu and Dayan, 2005).

Another study examined how NA may be involved in modulating the flexibility of neural models using fMRI and pupillometry with human volunteers while they performed a four-arm bandit task (Muller *et al.*, 2019). They identified distinct neural signatures specific to exploitative and exploratory behaviours: the dorsomedial prefrontal cortex and a fronto-parietal network were found to be activated during exploratory phases, and the hippocampus and medial orbitofrontal cortex (mOFC) were found to be associated with exploitation phases. Using baseline pupil size as an indicator of neuromodulatory state, they also found, importantly, that changes in mOFC representation strength were predicted by pupil dilation, and pupil dilation was itself predicted by ACC activity. This was a key finding, as the ACC has been reported previously to have strong projections descending to the LC (Aston-Jones and Cohen, 2005), and via these interactions with the LC may modulate belief uncertainty. This links with my findings of association between subject-specific internal model volatility and activation in the ACC, as the ACC was found to show increased activity in response to increased model entropy (Muller *et al.*, 2019).

Changes in pupil diameter have been associated with LC activity in both animal and human studies (Reimer *et al.*, 2016; de Gee *et al.*, 2017). Future work may incorporate pupillometry into a similar fMRI and drug manipulation study to examine noradrenergic projections between the LC and the ACC in response to task reversal in a probabilistic decision-making paradigm.

In summary, I observed LC activity in association with participants' subject-specific precision over reward, indicating a potential role of NA in motivation within the task. Activations in the ACC were also observed in association with subject-specific internal model flexibility, which supports previous findings that the ACC plays vital roles in belief updating and changes in environmental entropy (Muller *et al.*, 2019).

Chapter Five

Dynamic Causal Modelling for EEG Identifies Lateralised Memory Circuit Dropout in Alzheimer's Disease Patients

Some of the work detailed in this chapter has been published in *Brain Communications* (Tyrer *et al.*, 2020). Alzheimer's patients and healthy aged controls were recruited by Sarah Adams. Task design, collection of EEG data and behavioural measures, e.g. RTs, and calculation of recognition accuracy scores were conducted by Jessica R. Gilbert. ACEs were conducted by Sarah Adams, and demographic data were collected by Sarah Adams and Jessica Gilbert. All writing in this chapter and all subsequent analyses (behavioural data analyses, demographic data analyses, source localisation/identification, DCMs, PEB, source extraction, time-frequency analyses, statistical analyses) were conducted by Ashley Tyrer; see Tyrer *et al*. (Tyrer *et al.*, 2020).

5.1 Introduction

In this thesis thus far, I have demonstrated computational modelling of cognition: I have been able to generate models of *behaviour* to effectively characterise human participants based on their neural activity (recorded through BOLD signals) and broad measures of behavioural performance, in the context of a decision-making task. However, modelling both behaviour and biological mechanisms play important roles in the understanding of pathology. While previous chapters focused on the modelling of cognition, specifically decision-making in the young healthy brain, this chapter outlines how to model the *biological circuits* of cognition, in the context of aging and neurodegeneration.

Alzheimer's disease is the most prevalent cause of dementia in older adults, accounting for approximately two-thirds of dementia cases (Zhang *et al.*, 2016). Key histopathological hallmarks of Alzheimer's disease, including extracellular Aβ aggregates and intracellular hyperphosphorylated tau neurofibrillary tangles (Buckner *et al.*, 2005), have distinct deposition patterns that may relate to aberrant patterns of network connectivity in the brains of Alzheimer’s patients. Tau pathology is most prominent in the entorhinal cortex of the medial temporal lobes (MTL) in early Alzheimer’s disease stages, then progresses outwards, with hippocampal hyper- and hypo-connections both reported features of disease progression (Marks *et al.*, 2017; Pasquini *et al.*, 2019). Aβ, distributed more broadly, may relate to effects in the default mode network (Sperling *et al.*, 2009; Palmqvist *et al.*, 2017), where both enhanced and reduced functional connections have been reported in resting-state imaging studies (Hedden *et al.*, 2009; Chang *et al.*, 2018). However, despite clear evidence for widespread disruption of neural connectivity, there are limited consistent reports of compensatory connections (Gould *et al.*, 2006). By identifying functional connections that support cognition, the development of interventions that target and bolster these regional interactions could potentially delay or ameliorate disease progression.

Recent studies have reported increased right-lateralized activity as a putative compensatory mechanism in at-risk allele carriers who have not yet developed symptoms of dementia (Han *et al.*, 2007). The putative role of right-lateralized activations as a compensatory network is supported by findings showing early asymmetric alterations in Alzheimer’s disease pathology, where cortical atrophy and deposition of Aβ have been shown to be more pronounced in left medial temporal regions (Derflinger *et al.*, 2011; Frings *et al.*, 2015). Similarly, in patients with MCI, left-lateralized abnormalities may predominate. For example, functional imaging markers of novelty responses in the left hippocampal formation showed a positive predictive association with subsequent cognitive decline in MCI patients (Miller *et al.*, 2008). Also, a recent study by Weise *et al*. (2018) examined cerebral glucose metabolism in Aβ-positive subjects with MCI, and showed asymmetric declines in the left MTL compared to Aβ-negative controls, with evidence of reduced asymmetry once the disease progressed to dementia (Weise *et al.*, 2018). A recent study by Penny *et al*. (2018) used DCM to investigate effective connectivity during a semantic naming task in carriers of the *PSEN1* mutation, which results in early-onset familial Alzheimer’s disease, with carriers scanned pre-symptomatically and followed for over a decade. It was found that increased effective connectivity from left medial temporal to right inferotemporal sources predicted subsequent decline in Mini-Mental State Examination (MMSE) score (Penny *et al.*, 2018). Thus while laterality might be ‘an old idea’ in cognitive neuroscience, it may have a particular importance in dementia and is worth exploration in Late-Onset Alzheimer’s Disease (LOAD).

DCM is a computational method well-suited for studying putative compensatory mechanisms, as it estimates effective connectivity both within and between sources of activity, meaning that connections are examined in the context of regional activity changes, *while participants perform a task.* Moreover, with DCM one can derive the way in which experimental conditions or manipulations, such as cognitive tasks, recruit specific connections. Compensatory connections have been observed using DCM for EEG in healthy older adults (Gilbert and Moran, 2016). In a study of implicit (repetition priming) memory, older adults were found to recruit prefrontal-sourced top-down connections, contrasting with younger subjects who recruited a more traditional bottom-up connectivity hierarchy with feedforward input from early visual cortex only. During this task, bilateral visual cortex, temporal and parietal regions, and inferior frontal cortex were included as sources of activation in the DCMs. Here, I use both an implicit memory task as well as an explicit memory task to examine changes in connectivity within this network in patients with Alzheimer's disease. MTL-dependent explicit (recognition) memory has been shown to be impaired in early Alzheimer's disease (Wang *et al.*, 2014), whereas implicit memory processing has been shown to be preserved, allowing for a range of performance metrics in patients (Golby *et al.*, 2005).

In this chapter, I used DCM and group-level PEB analyses to investigate how inter-regional connectivity and within-region dynamics during implicit and explicit memory tasks are affected in Alzheimer's disease. I hypothesized that hierarchical left-hemisphere specific connections may be weakened in the Alzheimer's patient cohort compared to healthy aged controls. I also aimed to measure whether connections in the right hemisphere provided compensation during these memory tasks. High-density EEG and behavioural data were collected from Alzheimer's disease patients and healthy controls. Based on my findings in the PEB analysis, I focused on left and right hemisphere connectivity, examining putative left-hemisphere circuit dropout and right-hemisphere compensation in Alzheimer's disease.

5.2 Materials and Methods

5.2.1 Participants

Twenty-three Alzheimer's disease patients and 21 healthy controls (patients: mean age = 80 years, range = 68-89 years, 13 females; controls: mean age = 74 years, range = 66-91 years, 12 females) were asked to complete two mnemonic tasks while 64-channel EEG recordings were collected, preceded by a behavioural encoding phase completed prior to recording. Two patients (both females) were excluded from all analyses described below as the patients were not able to key-press independently during data collection. All control participants were free from neurological or psychiatric disorders. Patients were recruited from outpatient clinics at the Carilion Centre for Healthy Aging, Roanoke, VA, USA. Patients had a presumed diagnosis that met Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for clinical Alzheimer's disease. Study protocols were approved by the Carilion Clinic Institutional Review Board and the Virginia Polytechnic and State University.

5.2.2. Experimental Design

Two separate tasks were collected during the test phase, preceded by a single encoding phase (**Figure 5.1B-C**). During the encoding phase, EEG recordings were not taken. A total of 200 full-colour images were used, comprising nameable objects from well-known categories including a mix of both living and non-living stimuli (84 animals, 74 foods, 32 plants, and 10 body parts). Images were presented centrally on a 1024 x 768 pixel viewing screen, were 17.8 x 19.1 cm (7 x 7.5 in) in size, subtending a visual angle of five degrees, with the longest dimension covering 300 pixels. Participants were seated approximately 101.6 cm (40 in) from the screen. In the encoding phase, participants were shown 100 images. Participants were asked to covertly name each item as quickly as possible and press the spacebar on a computer keyboard as they named each item to record RT. Each image was presented for 2 s with a variable 1.5-2.5 s interstimulus interval in which a fixation cross was presented. After a delay period (following EEG system set-up), participants performed the priming and recognition tasks. Task order (priming versus recognition tasks) was randomized across participants. During both the priming and recognition tasks, task timing was identical to that in the encoding phase (**Figure 5.1C**).

Figure 5.1

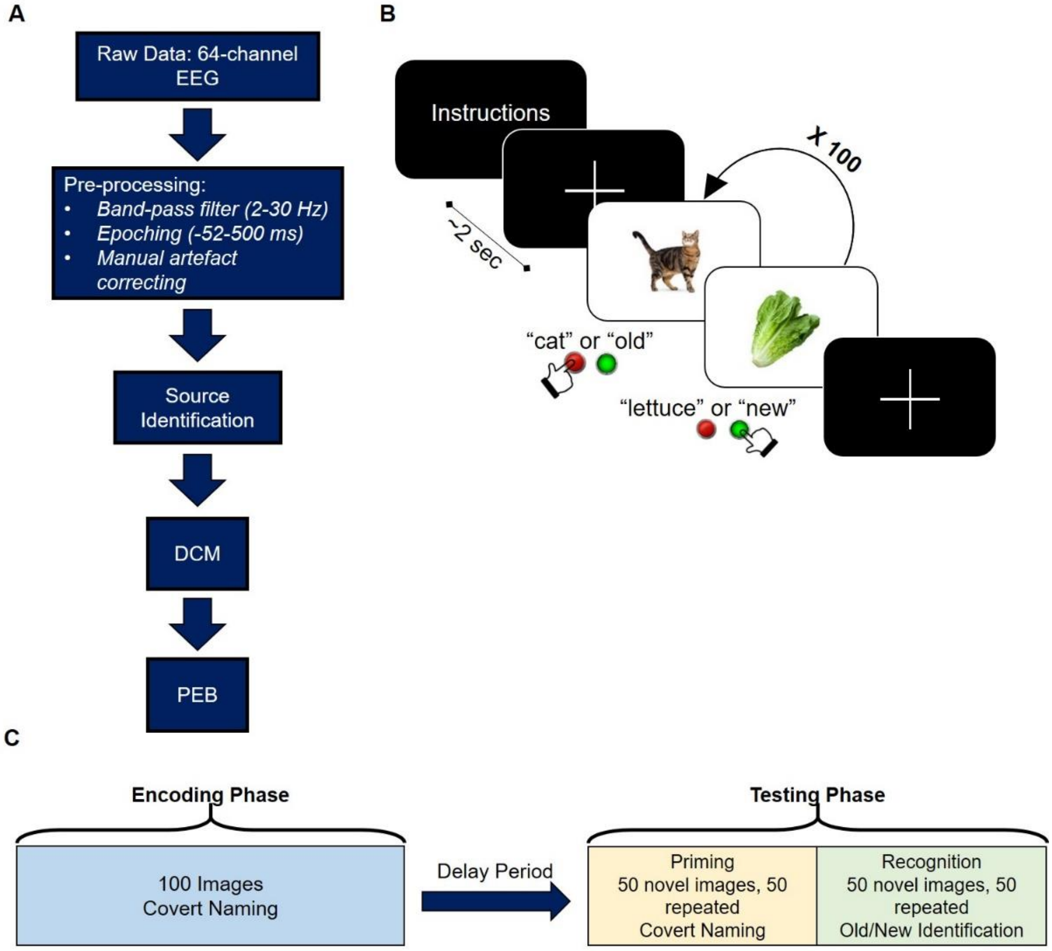


Figure 5.1 Analysis pipeline and task structure. (**A**) Schematic of the EEG data analysis pipeline, from collection and pre-processing of raw EEG data, through source identification, to constructing DCMs and analysing the DCMs using PEB. (**B**) Visual mnemonic priming and recognition task structure. Participants were presented with an image of an object and were instructed to covertly name the object (priming task) or indicate whether the object was old or new (recognition task), for 100 trials per task. (**C**) Task structure, which was preceded by a single encoding phase in the absence of EEG recordings. Order of priming and recognition tasks were counterbalanced across participants in the testing phase.

During the priming task, participants covertly named the 100 objects presented as quickly as possible while concurrently key-pressing to measure RT. Fifty images had not been seen before (novel) and 50 were repeated from the encoding phase (repeated), with image order randomized across participants. In line with task designs from previous picture-naming studies, covert naming was used to reduce EEG artefacts (Kan and Thompson-Schill, 2004; Gilbert *et al.*, 2010). During the recognition task, participants were again shown 100 images, with 50 repeated from the encoding phase (but not the same repeated images used in the priming task) and 50 novel images. Participants were instructed to indicate which objects were not seen previously (novel items) and which were presented earlier in the encoding phase (repeated items) by pressing one of two keyboard keys as quickly as possible, which were randomized across participants (**Figure 5.1B**).

5.2.3 Behavioural and Demographic Data Analyses

During both tasks, RTs were recorded, and accuracy scores were calculated for the recognition task. Accuracy was calculated as the percentage of correct key presses (i.e. correctly identifying if the image shown was novel or repeated and pressing the correct corresponding key) of the total number of key-presses in the task; missed trials were not counted towards the accuracy score.

A selection of demographic data was also collected from both patients and controls (**Table 5.1**), as well as the Addenbrooke's Cognitive Examination (ACE): a written neuropsychological test which examines attention, fluency, language, memory, and visuospatial ability (Addenbrooke's Cognitive Examination Revised Version, 2005) (Mioshi *et al.*, 2006). The ACE, which was initially designed as an extension of the MMSE, aims to pinpoint cognitive impairment in dementia and other neuropsychiatric conditions, including Alzheimer's disease. For the wide selection of demographic data collected from all participants, see **Table 5.1**.

	Controls	**AD Patients**
Participants	21	21
Female	12 (57.14%)	11 (52.38%)
Left-Handed	3 (14.29%)	0
Age (years)	73.71 ± 6.37 (66-91)	80.05 ± 6.18 (68-89)
ACE Score	91.90 ± 4.17 (80-99)	60.86 ± 10.91 (37-75)
MMSE Score	29.76 ± 0.436 (29-30)	22.62 ± 4.46 (15-30)
Education (years)	16.10 ± 2.61 (12-22)	13.19 ± 1.94 (11-18)
Social Network Score	7.90 ± 2.61 (4-12)	6.81 ± 3.60 (2-12)
Travel Score	4.10 ± 1.34 (1-6)	2.29 ± 1.19 (0-5)
Exercise Score	2.43 ± 0.507 (2-3)	1.43 ± 0.676 (1-3)
Diagnosis - Scan (days)		376.6 ± 720.2 (14-3192)
Depressive Symptoms		5 (23.8%)
Diabetes Mellitus		3 (14.3%)
Hypertension		8 (38.1%)

Table 5.1 Descriptive Statistics for Demographic Data. Data includes number of participants (*N*), percentage of total participant number in each group (%), mean ± standard deviation and range (*N*-*N*) of demographic variables. AD = Alzheimer's disease. To calculate the 'exercise' score, participants were asked whether they would describe their current level of exercise as 'sedentary', 'moderate' or 'vigorous'. Participants were given the score of one, two, or three, respectively. The 'social network' score was calculated as the sum of the number of close friends and close relatives the participant claimed to currently have. The 'travel' score was calculated as the sum of the number of times the participant claimed to travel out of the state per month, travel out of the country per year, whether they had ever lived out of the country (and how long for), and how frequently they currently travel per month.

5.2.4 EEG Data Acquisition and Pre-processing

EEG recordings were collected using a DC amplifier (BrainAmp MR Plus, Brain Products GmbH Gilching, Germany) and a 64-channel electrode system (actiCAP, Brain Products GmbH), referenced to the average of 64 channels, as described in Gilbert and Moran, 2016 (Gilbert and Moran, 2016). Impedances of <5 kΩ for all electrodes were confirmed prior to data collection. Data were sampled at 1000 Hz and online filtered at DC-250 Hz during data acquisition.

EEG data were analysed using the academic freeware SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). Pre-processing involved band-pass filtering to retain signals from 2-30 Hz, segmenting the continuous EEG signal into 552 ms epochs (-52-500 ms peristimulus time), and manually artefact-correcting to remove bad trials and channels, for example, trials containing remnant artefacts or eyeblinks. Data were then averaged based on the stimulus condition, i.e. novel images and repeated images, following baseline correction. The final pre-processed data features thus comprised event-related potentials (ERPs) over each of the 64 sensor electrodes for each condition and for each participant (see **Figure 5.2C** for uncoregistered EEG sensor positions; **Figure 5.3A-B** for ERP grand means). A schematic of the data analysis pipeline is shown in **Figure 5.1A**.

5.2.5 Source Localisation and Identification

Three-dimensional spatiotemporal source reconstruction was performed using SPM's multiple spare priors routines, to infer the network of active sources generating the ERPs to inform my network model. This source reconstruction optimises sources using a parameterised lead field, and constrained sparse 'minimum norm'-type regression model (though constraints embody multiple (512) patches *a priori* precluding source smearing). Sources were estimated for broadband power (2-30 Hz) over the ERP time window from 0-450 ms. For each participant and condition, a 3D volumetric image of sources was obtained. From these, second-level (i.e. group) analyses were performed using one-sample *t*-tests. These *t*-tests were conducted separately for the priming task and recognition task, and included both patients and controls, and both task conditions (**Figure 5.2A-B**). I analysed the groups together in order to obtain the most general solution for subsequent DCM.

Figure 5.2

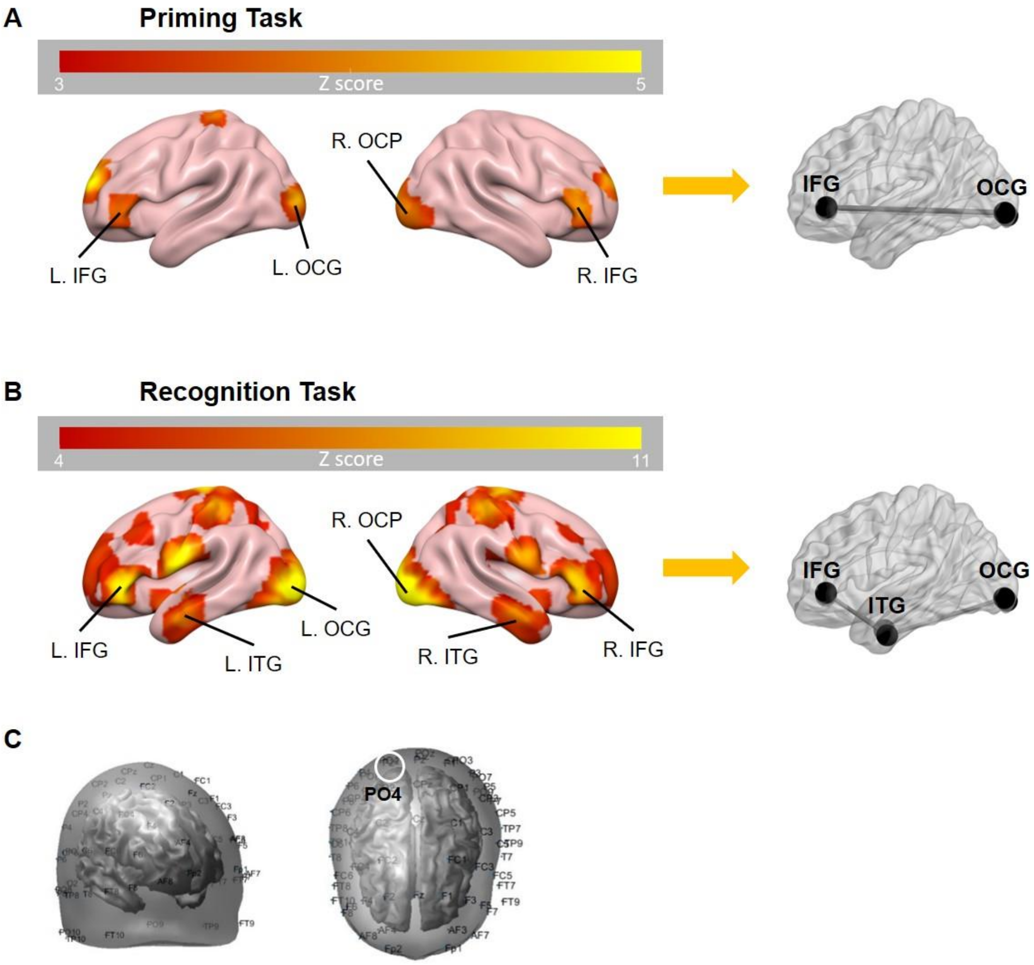


Figure 5.2 Source identification in priming and recognition tasks. (**A**) Bilateral four-source model identified using 3D source reconstruction for the priming task. (**B**) Bilateral six-source model identified using 3D source reconstruction for the recognition task. Colour bars indicate Z scores. (**C**) Uncoregistered EEG sensor positions; front-right side view (left) and top view (right). Approximate location of channel PO4 circled in white. L = left; R = right.

Figure 5.3

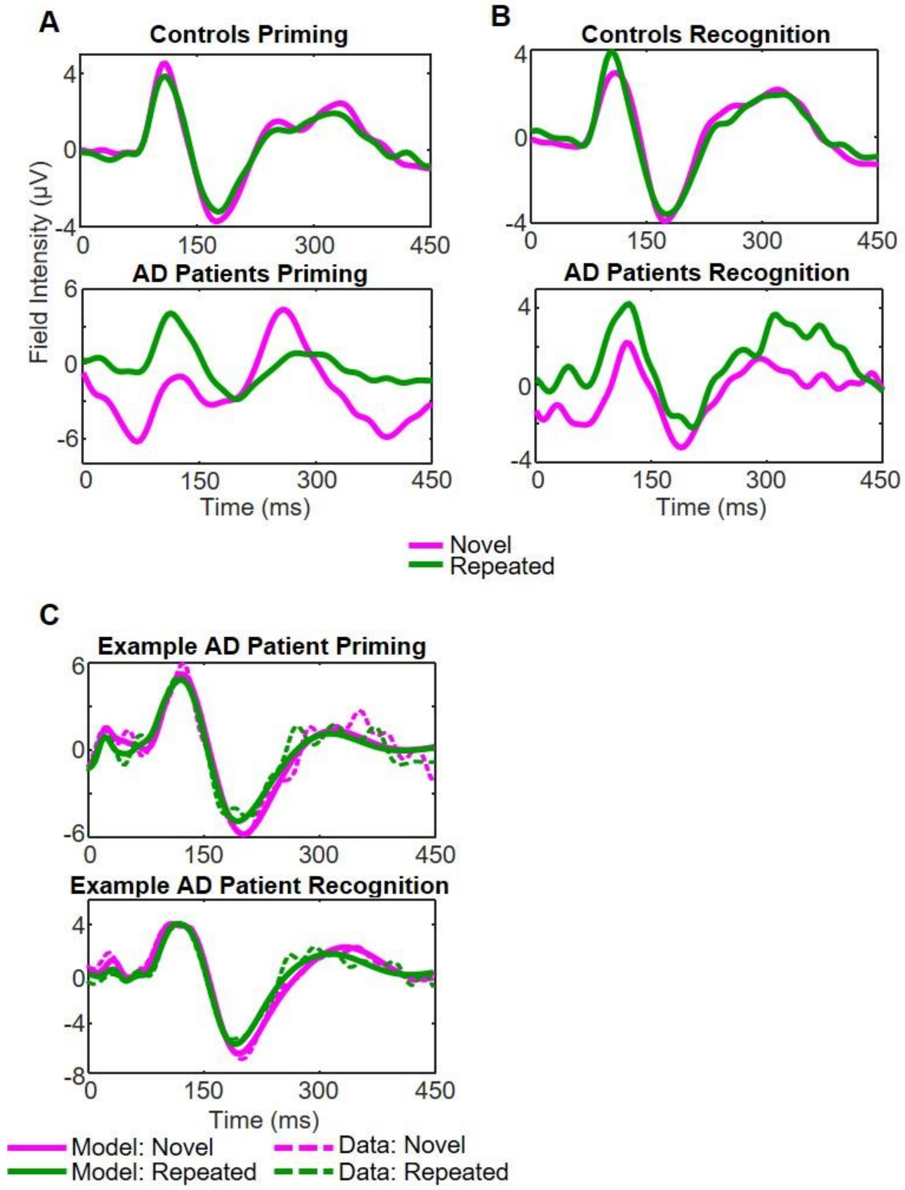


Figure 5.3 ERP grand means for patients and controls, and exemplary patient DCM fits. (**A**) Grand mean of controls (top) and patients (bottom), showing ERPs for averaged novel (magenta) and repeated trials (green) in the priming task for channel PO4 (right occipital pole). (**B**) Grand mean of controls and patients, showing ERPs for averaged novel and repeated trials in the recognition task for channel PO4. (**C**) DCM fits (solid line) and real data (dashed line) from the first mode of an exemplary Alzheimer's disease patient in the priming (top) and recognition (bottom) tasks. AD = Alzheimer's disease.

5.2.6 Dynamic Causal Modelling

DCM served as my framework for a model-based assay of source connectivity. Originally developed for analysing connectivity in fMRI data, and later for M/EEG data, with DCM, one makes inferences about parameters that may not be observed with fMRI or M/EEG data directly, known as the latent parameters. DCM is hypothesis-driven and can be used to test specific hypotheses about the activity between sources in a network, rather than being limited to asking questions about the strength of sources. DCM estimates effective connectivity, i.e. the influence that one source or neuronal system has over another, in that connections are examined in a context-dependent manner. Therefore, one may ask questions such as “How does the forward connection from region A to region B change between a novel and repeated condition?”, rather than deriving connectivity itself. With DCM one can investigate how experimental conditions or manipulations modulate connectivity. The inferences made using DCM with EEG data also describe more neurobiologically plausible parameters, as EEG data is highly resolved in time and can therefore relate to the causes of underlying neuronal and synaptic dynamics more directly.

This study utilises conductance-based neuronal mass modelling. The conductance-based model utilises dynamic Morris-Lecar-type equations, i.e. reduced two-dimensional form of the original four-dimensional Hodgkin-Huxley model, which describe the flow of ions at the synapse:

$$C\dot{V} = g(V_{rev} - V) \qquad \textit{Eq. 5.1}$$

where $C\dot{V}$ = current (capacitance x change in membrane potential), g = conductance, V_{rev} = reversal potential, and:

$$\dot{g} = \kappa\big(\gamma_{aff}\sigma\big(\mu_{aff} - V_{threshold}, \Sigma_{aff}\big) - g\big) \qquad \textit{Eq. 5.2}$$

where $\dot{g}$ = conductance, κ = time constant, σ represents a sigmoid function which describes the cumulative distribution function of the normal distribution $\mathcal{N}(\mu_{aff}, \Sigma_{aff})$, in which the proportion of afferent cell-firing is determined by the threshold potential $V_{threshold}$. Σ_{aff} = firing variance, g = number of open channels, and γ parameterises the connection strengths between cellular layers (Moran *et al.*, 2011; Moran *et al.*, 2013).

Equations 5.1 and **5.2** can be expanded to describe the single-cell currents in each of the cell subpopulations included in the model: inhibitory interneurons (extra-granular layers), excitatory spiny stellate cells (granular layers), and excitatory pyramidal cells (extra-granular layers), as follows:

Inhibitory Interneurons

$$C\dot{V}^{(2)} = g_L\left(V_L - V^{(2)}\right) + g_E^{(2)}\left(V_E - V^{(2)}\right) + g_I^{(2)}\left(V_I - V^{(2)}\right)$$

$$\dot{g}_E^{(2)} = \kappa_E\left(\gamma_{23}^E\sigma\left(\mu_V^{(3)} - V_R, \Sigma^{(3)}\right) - g_E^{(2)}\right)$$

$$\dot{g}_I^{(2)} = \kappa_I\left(\gamma_{22}^I\sigma\left(\mu_V^{(2)} - V_R, \Sigma^{(2)}\right) - g_I^{(2)}\right)$$

Excitatory Spiny Stellate Cells

$$C\dot{V}^{(1)} = g_L\left(V_L - V^{(1)}\right) + g_E^{(1)}\left(V_E - V^{(1)}\right) + I$$

$$\dot{g}_E^{(1)} = \kappa_E\left(\gamma_{13}^E\sigma\left(\mu_V^{(3)} - V_R, \Sigma^{(3)}\right) - g_E^{(1)}\right)$$

Excitatory Pyramidal Cells

$$C\dot{V}^{(3)} = g_L\left(V_L - V^{(3)}\right) + g_E^{(3)}\left(V_E - V^{(3)}\right) + g_I^{(3)}\left(V_I - V^{(3)}\right)$$

$$\dot{g}_E^{(3)} = \kappa_E\left(\gamma_{31}^E\sigma\left(\mu_V^{(1)} - V_R, \Sigma^{(1)}\right) - g_E^{(3)}\right)$$

$$\dot{g}_I^{(3)} = \kappa_I\left(\gamma_{32}^I\sigma\left(\mu_V^{(2)} - V_R, \Sigma^{(2)}\right) - g_I^{(3)}\right)$$

Eq. 5.3

where g comprises matrices of size 4 x 3 (number of sources x number of cell subpopulations), for each of the leak (g_L), excitatory (g_E) and inhibitory (g_I) conductances. Connections between cellular layers are described by γ, for example, γ_{13}^E parameterizes excitatory connections from excitatory pyramidal cells (cell subpopulation 3) to excitatory spiny stellate cells (cell subpopulation 1). Here, spiny stellate cells receive excitatory input from pyramidal cells, and inhibitory interneurons receive input from pyramidal cells and also via self-connections. Pyramidal cells receive input from both spiny stellate cells and inhibitory interneurons. These non-linear differential equations describe how the current of a single cell ($C\dot{V}$) and its conductance g evolve over time (Marreiros *et al.*, 2008). All cells in this model express AMPA and GABA$_A$ receptors, with ion channel time constants ($^1/_{\kappa_{e/i}}$) (Moran *et al.*, 2013).

Extrinsic connections, i.e. region-to-region connections, also enter at specific cortical layers: forward connections project from pyramidal cells of one region onto spiny stellates of another region, backward connections project from pyramidal cells to both inhibitory interneurons and pyramidal cells, and lateral connections project from pyramidal cells onto all cell subpopulations of the connecting region. Though connections between within-region cell subpopulations have been denoted γ as above, between-region connectivity, i.e. forward/backward/lateral connections, have the same mathematical form and role as the within-region connection strengths. These between-region connections comprised the so-called *A* matrix (Friston *et al.*, 2003).

I used a neural mass model to describe the activity at each source. Specifically, I employed the NMDA (*N*-methyl-ᴅ-aspartate) model (Moran *et al.*, 2011). To specify the network, I allowed for connections between these neural masses.

For this study, DCMs were specified for each individual participant to examine the modulation of extrinsic connectivity between patients and controls, and between the novel and repeated conditions. The DCMs were fit to the scalp-related ERPs from 0-450 ms peristimulus time. Based on my group-level source activity maps generated in the 3D source reconstruction analyses described above, I identified two network structures: one for each task.

Priming Network

A four-source model was used to describe the network dynamics during the implicit priming task. The sources included left inferior occipital gyrus (OCG) (MNI coordinates: -32 -94 -6), right occipital pole (OCP) (MNI coordinates: 28 -96 -8), and bilateral sources in the inferior frontal gyrus pars triangularis (IFG) (MNI coordinates left: -40 40 -2 and right: 40 40 -4) (**Figure 5.2A**), as previously reported (Gilbert and Moran, 2016).

Recognition Network

Given that the source localisation results found temporal regions of activation in addition to frontal and occipital sources, I used an extended six-region network comprising occipital, temporal and frontal sources for the recognition task data. These consisted of left inferior OCG (MNI coordinates: -30 -96 -4), right OCP (MNI coordinates: 28 -96 -8), bilateral sources in the inferior temporal gyrus (ITG) (MNI coordinates left: -46 -6 -34 and right: 46 -4 -30), and bilateral sources in the IFG pars triangularis (MNI coordinates left: -46 40 2 and right: 42 38 0) (**Figure 5.2B**). As expected, the explicit recognition phase in the task recruited additional brain regions. Bilateral anterior temporal sources were selected for the extended explicit memory network due to their strong task relevance, in line with previous analyses (Gilbert and Moran, 2016).

I optimised DCMs for evoked responses (DCM for ERPs) for each participant individually for both models. For the priming task DCM, I specified both bottom-up and top-down hierarchical connections between the occipital sources and IFG, bilaterally for the *A* matrix, without lateral connections (**Figure 5.2A**). For the recognition task DCM, I defined both forward and backward connections from occipital sources to ITG, and ITG to IFG bilaterally for the *A* matrix, without lateral connections (**Figure 5.2B**). For both tasks, I assumed that no crosstalk between hemispheres would occur via lateral connections.

We then defined the *B* matrix: a connectivity matrix similar to the *A* matrix which defines task-dependent modulatory connections, i.e. the difference in novel versus repeated image trials on specified connections. For my models, I defined the *B* matrices with the same connections as in the *A* matrices, but with added self-connections for all sources. The input vector *C* defines the activity sources receiving subcortical sensory input, which here were the left and right occipital sources in the models for both tasks. The models also included parameters describing local glutamate connectivity (*G*), the time constant of post-synaptic responses (*T*), and delays between sources (*D*) (**Tables 5.2** and **5.3**). These parameters constitute a multivariate set, θ.

These generative models of interacting sources were inverted according to a Variational Bayesian scheme to examine the likelihood of parameters, given the model and data for each participant individually, using an off-the-shelf Variational Bayes algorithm

(Friston, 2002). Inversion of the models was performed for each task, and each subject, individually. This approximates the posterior probability of model parameters $p(\theta|y,m)$, i.e. the probability of the model parameters given the data and the model. The ERP scalp response is represented by y, and m represents the model, i.e. which regions in the brain are connected and how these connections are modulated by the tasks; θ represents the model parameters (see **Figure 5.3C** for example patient DCM fits). For the priming task, the model had 29 parameters (**Table 5.2**), and for the recognition task, the model had 43 parameters due to the increased number of sources in the network (**Table 5.3**). Given these inferred parameter sets, I next sought to determine those parameters associated with task performance, and those associated with disease *per se*. To study these group effects, these posterior parameters were then passed into the PEB analysis outlined below.

Parameter	Parameter Description
S(1)	Overall excitability
T(1)	AMPA time constant left iOCG
T(2)	AMPA time constant right OCP
T(3)	AMPA time constant left IFG
T(4)	AMPA time constant right IFG
G(1)	Intrinsic glutamate connectivity within left iOCG
G(2)	Intrinsic glutamate connectivity within right OCP
G(3)	Intrinsic glutamate connectivity within left IFG
G(4)	Intrinsic glutamate connectivity within right IFG
A{1}(3,1)	Forward connections from left iOCG to left IFG
A{1}(4,2)	Forward connections from right OCP to right IFG
A{2}(1,3)	Backward connections from left IFG to left iOCG
A{2}(2,4)	Backward connections from right IFG to right OCP
B{1}(1,1)	Trial-dependent self-connections of left iOCG
B{1}(3,1)	Trial-dependent connections from left iOCG to left IFG
B{1}(2,2)	Trial-dependent self-connections of right OCP
B{1}(4,2)	Trial-dependent connections from right OCP to right IFG
B{1}(1,3)	Trial-dependent connections from left IFG to left iOCG
B{1}(3,3)	Trial-dependent self-connections of left IFG
B{1}(2,4)	Trial-dependent connections from right IFG to right OCP
B{1}(4,4)	Trial-dependent self-connections of right IFG
C(1)	Subcortical input into left iOCG
C(2)	Subcortical input into right OCP
R(1)	Stimulus onset parameter
R(2)	Stimulus dispersion parameter
D(1)	Within-region cell-to-cell population signal delay
D(2)	Between-region signal delay
U(1)	Exogenous background activity
CV(1)	Membrane capacitance

Table 5.2 PEB parameters in the priming task.

Parameter	**Parameter Description**
S(1)	Overall excitability
T(1)	AMPA time constant left iOCG
T(2)	AMPA time constant right OCP
T(3)	AMPA time constant left ITG
T(4)	AMPA time constant right ITG
T(5)	AMPA time constant left IFG
T(6)	AMPA time constant right IFG
G(1)	Intrinsic glutamate connectivity within left iOCG
G(2)	Intrinsic glutamate connectivity within right OCP
G(3)	Intrinsic glutamate connectivity within left ITG
G(4)	Intrinsic glutamate connectivity within right ITG
G(5)	Intrinsic glutamate connectivity within left IFG
G(6)	Intrinsic glutamate connectivity within right IFG
A{1}(3,1)	Forward connections from left iOCG to left ITG
A{1}(4,2)	Forward connections from right OCP to right ITG
A{1}(5,3)	Forward connections from left ITG to left IFG
A{1}(6,4)	Forward connections from right ITG to right IFG
A{2}(1,3)	Backward connections from left ITG to left iOCG
A{2}(2,4)	Backward connections from right ITG to right OCP
A{2}(3,5)	Backward connections from left IFG to left ITG
A{2}(4,6)	Backward connections from right IFG to right ITG
B{1}(1,1)	Trial-dependent self-connections of left iOCG
B{1}(3,1)	Trial-dependent connections from left iOCG to left ITG
B{1}(2,2)	Trial-dependent self-connections of right OCP
B{1}(4,2)	Trial-dependent connections from right OCP to right ITG
B{1}1(1,3)	Trial-dependent connections from left ITG to left iOCG
B{1}(3,3)	Trial-dependent self-connections of left ITG
B{1}(5,3)	Trial-dependent connections from left ITG to left IFG
B{1}(2,4)	Trial-dependent connections from right ITG to right OCP
B{1}(4,4)	Trial-dependent self-connections of right ITG
B{1}(6,4)	Trial-dependent connections from right ITG to right IFG
B{1}(3,5)	Trial-dependent connections from left IFG to left ITG
B{1}(5,5)	Trial-dependent self-connections of left IFG
B{1}(4,6)	Trial-dependent connections from right IFG to right ITG
B{1}(6,6)	Trial-dependent self-connections of right IFG
C(1)	Subcortical input into left iOCG
C(2)	Subcortical input into right OCP
R(1)	Stimulus onset parameter
R(2)	Stimulus dispersion parameter
D(1)	Within-region cell-to-cell population signal delay
D(2)	Between-region signal delay
U(1)	Exogenous background activity
CV(1)	Membrane capacitance

Table 5.3 PEB parameters in the recognition task.

5.2.7 Parametric Empirical Bayes and Classical Analyses

PEB was used for a random-effects analysis over model parameters, based on the presence or absence of Alzheimer's disease and task performance for both tasks separately. The PEB comprises a Bayesian GLM at the second level. Here I constructed the Bayesian GLM, using two second-level covariates as well as including an average mean effect. A random effects design matrix was generated containing three separate columns, one for each covariate: the first column was the average over all subjects (a column of ones), the second column defined disease, i.e. patient or control (one or zero respectively), and the third column defined the parametric task performance (either mean RT difference for the priming task, calculated as the mean novel RT minus mean repeated RT for each participant, or accuracy score for the recognition task). From this one can compute which parameters show group-level differences based on disease state (patients versus controls), and which are also affected by task performance, as well as their probabilities. Thus, the GLM allows us to examine the network correlates of task performance while accounting for disease state. I describe which connections in the models were strengthened/weakened as a result of disease, the directionality of such connections, and whether this was specific to a particular hemisphere. Also, these analyses inform us about whether such connections are modulated by task performance, and whether these connections may be performing compensatory roles in patients based on task performance. The PEB analysis essentially re-estimates model parameters at the level of individual DCMs by conducting a search over all possible parameter combinations that emulate the design matrix. The final analyses report the effect size, direction, and probability. This approach aims to reduce the second-level effects using Occam's razor until only meaningful parameters that contribute to group differences remain (**Figure 5.5-5.6**).

5.2.8 Source Extraction and Time-Frequency Analyses

We also examined spectrograms within regions in an exploratory analysis. Once sources had been identified, source data were extracted as single trials from the output of inverse source reconstruction for each task separately, with VOI radii of 8 mm, across the frequency band 2-58 Hz. In priming task data, the left and right IFG were extracted (MNI coordinates left: -40 40 -2 and right: 40 40 -4). In recognition task data, the left and right IFG (MNI coordinates left: -46 40 2 and right: 42 38 0) and the left and right ITG (MNI coordinates left: -46 -6 -34 and right: 46 -4 -30) were extracted as bilateral pairs.

Time-frequency analysis of EEG activity was based on Morlet Wave transform using the SPM12 toolbox in MATLAB, which allows the transformation of EEG data into the time-frequency domain. This was used for the decomposition of single-trial time-frequency values from 2-58 Hz for the extracted sources. The time-frequency data were then rescaled using the log ratio method, in which all power values are divided by the mean power from -52 ms to 0 ms, i.e. pre-stimulus time, and the log taken. To produce time-frequency spectrograms, the mean differences between novel and repeated trials were calculated, then averaged across participants.

5.2.9 Statistical Analyses

A two-way mixed effects ANOVA and Wilcoxon signed rank tests were conducted to examine RT differences across patients and controls in the priming task, and a one-way ANOVA was conducted to examine differences in accuracy scores between patients and controls in the recognition task. A MANOVA was also conducted to investigate group differences in demographic variables, comparing patients and controls. Three-way ANOVAs were then conducted to investigate effects of depression, hypertension and diabetes mellitus on recognition accuracy scores and ACE scores within the patient group, in addition to a post-hoc one-tailed two-sample *t*-test to specifically examine differences in accuracy score for patients with and without depressive symptoms. These statistical tests were conducted using IBM SPSS Statistics 24 software. Spearman’s rank correlations were conducted using MATLAB software to examine associations between ACE scores and mean RT difference/accuracy scores.

Following the PEB, I also conducted post-hoc classical statistical tests using MATLAB software on group differences in parameter values between patients and controls using two-tailed two-sample *t*-tests, and correlations between specific parameter values and the behavioural measures calculated previously (i.e. mean RT difference for the priming task and accuracy score for the recognition task) using Pearson’s correlation.

Two-sample *t*-tests were conducted on the time-frequency data to confirm differences between patients and controls, uncorrected for multiple comparisons. Repeated measures ANOVAs were also conducted between novel and repeated trials within control subjects, also uncorrected. All analyses of time-frequency data were conducted using the SPM12 toolbox in MATLAB.

5.3 Results

5.3.1 Controls Consistently Outperformed Patients in Both Implicit and Explicit Memory Tasks, With High Variability in Patients' Task Performance

I conducted a two-way mixed effects ANOVA on the RTs in novel and repeated trials in patients and controls. This revealed a significant between-subjects main effect of disease ($F(1,40) = 44.4$, $p < 0.001$, $\eta_p^2 = 0.526$) indicating that RTs were significantly higher in patients compared to controls. There was also a significant within-subject main effect of trial type ($F(1,40) = 16.7$, $p < 0.001$, $\eta_p^2 = 0.294$), suggesting that RTs in novel trials were overall significantly higher than RTs in repeated trials (**Figure 5.4A**). However, the disease x trial type interaction was not significant ($F(1,40) = 1.02$, $p = 0.318$), indicating that the RT differences in novel versus repeated trials did not differ significantly between patients and controls (**Figure 5.4A**). This suggests that implicit memory may be preserved in patients in the priming task, as well as in controls.

For RTs in both novel and repeated trials, however, the variances were unequal for patients compared with controls, which may result in inflated *p* values (novel: $F(1,40) = 28.0$, $p < 0.001$; repeated: $F(1,40) = 30.7$, $p < 0.001$; Levene's test of equality of error variances). Therefore, to supplement the ANOVA, I conducted Wilcoxon signed rank tests to examine RT differences between novel and repeated trials in patients and controls separately. In controls, the median of novel RTs was significantly higher than that of repeated RTs ($Z = -4.02$, $p < 0.001$). In patients, the medians of novel and repeated RTs were not significantly different ($Z = -1.48$, $p = 0.140$).

To assess how task performance relates more broadly to cognitive decline, I examined the relationship between performance during this task and the ACE exclusively in patients.

There was no correlation between ACE scores and mean RT difference (difference between mean novel and mean repeated RT) for patients only (*rho* = 0.0741, *p* = 0.750) (**Figure 5.4B**).

Figure 5.4

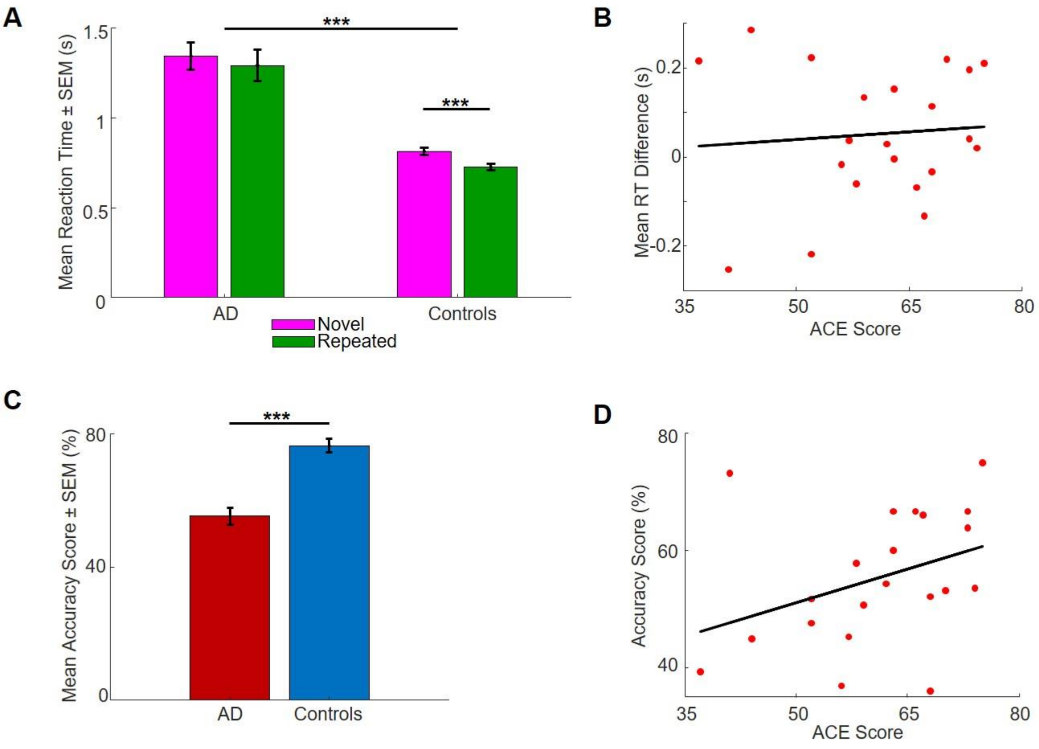


Figure 5.4 Task performance across patients and controls. (**A**) Mean RTs ± SEM for novel and repeated trials in the priming task, in patients and controls. Controls had significantly faster RTs across trial types than patients, and controls had significantly faster RTs in repeated trials compared to novel trials. (**B**) No correlation between ACE scores and mean RT differences for patients only in the priming task. (**C**) Mean accuracy scores ± SEM in the recognition task, in patients and controls. Controls had significantly higher accuracy scores compared to patients. (**D**) Strong correlation between ACE scores and accuracy score for patients only in the recognition task. AD = Alzheimer's disease; SEM = standard error of mean. ******* = *p* < 0.001.

Task performance was measured in the recognition task by calculating accuracy scores, i.e. the number of successful responses out of total responses in the task. Behaviourally, accuracy scores for the recognition task were significantly higher for controls (mean = 0.765, SEM = 0.0210) than for patients (mean = 0.553, SEM = 0.0249) (*F*(1,40) = 42.1, *p* < 0.001) (**Figure 5.2C**). Once again, I investigated the relationship between ACE scores and task performance, in this case, recognition accuracy scores. There was a moderate correlation between accuracy scores and ACE scores for patients (*rho* = 0.434, *p* = 0.0492), with high variability in the spread of accuracy scores and ACE scores (**Figure 5.4D**).

Overall, these group-level results showed the typical decline in mnemonic processing seen in patients with Alzheimer’s disease. Moreover, I also observed high variability in accuracy scores and priming performance (i.e. RT), where variability in the explicit task was related to established clinical scales. I therefore aimed to understand how this variability is related to network connectivity using my DCMs.

5.3.2 Years of Education, Travel, and Exercise Scores were Significantly Lower in Patients, but No Effect of Social Network Scores

A MANOVA was conducted to examine differences between patients and controls across the following demographic variables: years of education, travel score, social network score and exercise score (**Table 5.1**). Using Pillai’s trace, I found a significant effect of disease state, in that the demographic variables tested were significantly different between patients and controls (*V* = 0.606, *F*(4,37) = 14.2, *p* < 0.001). Separate univariate ANOVAs on each demographic variable revealed significant effects of years of education (*F*(1,40) = 16.8, *p* < 0.001), travel score (*F*(1,40) = 21.5, *p* < 0.001), and exercise score (*F*(1,40) = 29.4, *p* < 0.001), but did not reveal a significant effect of social network score (*F*(1,40) = 1.28, *p* = 0.265).

The travel score contained a historical element, as participants were asked if they had previously lived abroad during their lives, suggesting that the amount of travelling an individual did during their life may influence their susceptibility to Alzheimer’s disease in later life. In contrast to the travel score, the social network score only considered each participant’s current number of social networks, rather than historical social networks prior to diagnosis, and therefore cannot be used as an accurate indication of the role of social networks in the risk of developing the disease.

Demographic variables such as years of education, travel score, and exercise may have an impact on the likelihood of an individual developing dementia later in life, however, historical data for exercise levels would be required for this to be conclusive. Rather than solely the travel scores and years of education directly affecting the susceptibility of an individual to suffering from dementia, it is more likely that these factors play roles in a complex socioeconomic interaction with additional factors.

The medical histories of patients were then examined. 23.8% of patients were also diagnosed with depression or displayed depressive symptoms (n = 5), however no significant difference was shown in recognition accuracy score between patients that displayed depressive symptoms and patients without depressive symptoms (*F*(1,15) = 3.56, *p* = 0.079). However, interestingly, a post-hoc one-tailed *t*-test highlighted a slight trend toward higher recognition accuracy scores in patients with depressive symptoms compared to non-depressive patients, though non-significant (*t*(19) = 1.72, *p* = 0.0511, one-tailed). This non-statistically significant finding may be due to lack of statistical power, as only five patients in this cohort suffered from depressive symptoms. Patients that displayed depressive symptoms also had no significant difference between ACE scores in patients with and without depressive symptoms (*F*(1,15) = 0.971, *p* = 0.340).

Similar findings are seen when examining patients with hypertension and diabetes mellitus. 38.1% of patients suffered from hypertension (n = 8), and 14.3% of patients suffered from diabetes (n = 3). However, there was no significant difference in accuracy scores between patients with and without hypertension (*F*(1,15) = 0.097, *p* = 0.759), nor those with and without diabetes (*F*(1,15) = 0.589, *p* = 0.455). Similar results were seen when comparing ACE scores in patients with and without hypertension (*F*(1,15) = 0.125, *p* = 0.729) or diabetes (*F*(1,15) = 0.803, *p* = 0.384), but these findings may be due to very low n numbers. Intercepts between depression and hypertension, and between hypertension and diabetes were also non-significant for accuracy scores (*F*(1,15) = 0.309, *p* = 0.587; *F*(1,15) = 0.692, *p* = 0.418) and for ACE scores (*F*(1,15) = 0.393, *p* = 0.540; *F*(1,15) = 0.746, *p* = 0.401); no participants suffered from both depression and diabetes mellitus. This suggests that hypertension and diabetes mellitus may not exert significant effects on cognitive ability or behavioural memory performance in patients suffering from Alzheimer's disease, although in this sample the numbers of patients suffering from hypertension and diabetes mellitus are very low, therefore this cannot be conclusive.

5.3.3 Patients Showed Within-Region Slowing in Implicit Memory Network

In my PEB analysis, I first tested whether there was a group difference in the connectivity strengths and modulated connectivity strengths between patients and controls, and then examined the effects of mnemonic task performance, while accounting for effects of disease state.

Controls had significantly increased strengths of intrinsic glutamate connectivity (*G*) in right OCP compared to the patients in the priming task (*Ep* = -0.482, *Pp* = 1.00) (**Figure 5.5**). Additionally, patients had significantly reduced aggregate excitatory receptor activity (i.e. increased excitatory time constant *T*) in the left IFG (*Ep* = 0.367, *Pp* = 1.00) and a greater delay in signal transmission (*D*), i.e. the time taken for signals to transmit from region to region including axonal delays (*Ep* = 0.194, *Pp* = 1.00), indicating a general slowing in memory processing in patients. The PEB analysis also found significant associations between intrinsic glutamate connectivity in right IFG and mean RT difference, in that this connectivity was increased for larger RT differences (i.e. a better priming effect) (*Ep* = 1.52, *Pp* = 1.00), suggesting that task-related increases in local glutamate connectivity in the right hemisphere may have a modulatory effect on implicit memory task performance in patients only (**Figure 5.5**).

5.3.4 Left Hemisphere Circuit Deficits in Explicit Memory Sub-Network in Patients, With Task-Associated Connectivity Increases in Right Hemisphere

In the recognition task, the PEB showed increased forward connectivity (*A* matrix) from left ITG to left IFG in controls compared to patients (*Ep* = -0.352, *Pp* = 1.00) (**Figure 5.6**). The subcortical input (*C*) into left inferior OCG was also increased in controls compared to patients (*Ep* = -0.177, *Pp* = 1.00), indicating that patients may suffer from reduced visual input into their left-hemisphere memory circuit (**Figure 5.6**). Top-down connections from right ITG to right OCP were found in the PEB analysis to have significant correlation with accuracy score (**Figure 5.6**), as these connections increased with higher accuracy scores (*Ep* = 0.670, *Pp* = 1.00), in addition to the recurrent bottom-up connectivity from right OCP to right ITG showing similarly strong positive associations with accuracy score (*Ep* = 0.553, *Pp* = 1.00). This implies a strong association between increased right hemisphere connectivity and improved task performance in this more taxing memory recall task, while accounting for group differences.

Figure 5.5

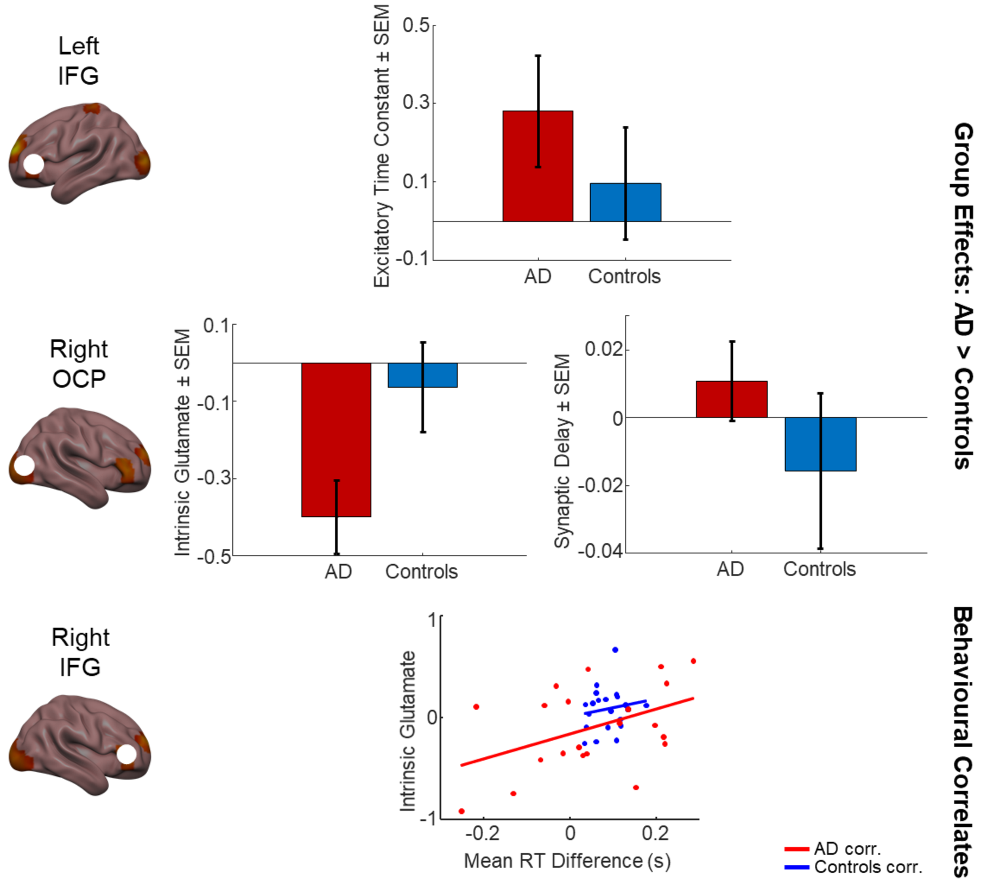


Figure 5.5 Model parameters estimated using PEB, for group-level differences and effects of task performance in the priming task. Top: group differences in excitatory time constant between patients and controls in the left IFG, showing mean ± SEM of parameter estimates across participants. Centre: group differences in intrinsic glutamate (centre left) and synaptic delay (centre right) between patients and controls in the right OCP. Bottom: correlation between intrinsic glutamate and implicit memory task performance in the right IFG. Patients: red; controls: blue. AD = Alzheimer's disease; corr. = correlation; SEM = standard error of mean.

Figure 5.6

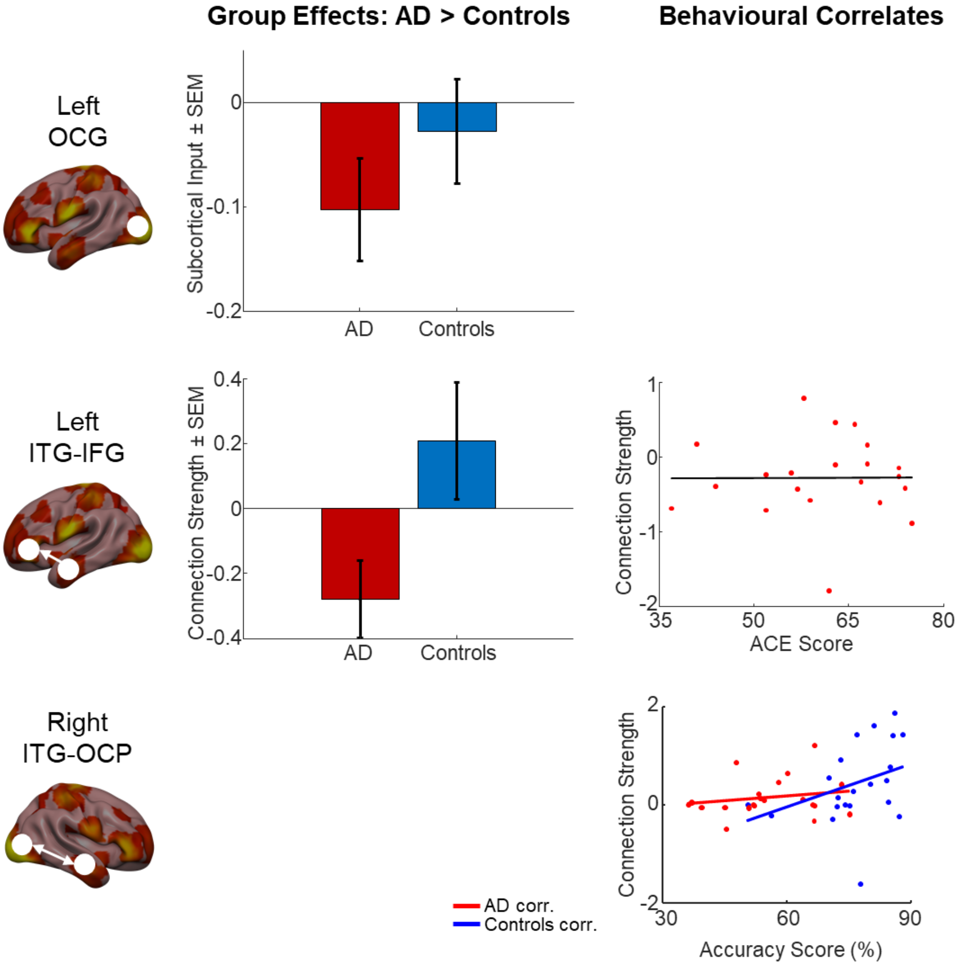


Figure 5.6 Model parameters estimated using PEB, for group-level differences and effects of task performance in the recognition task. Top left: group differences in subcortical input into the left OCG between patients and controls, showing mean ± SEM of parameter estimates across participants. Centre left: group differences in forward connectivity strengths from the left ITG to left IFG between patients and controls, showing mean ± SEM of parameter estimates across participants. Centre right: no significant correlation between forward connectivity strengths from the left ITG to left IFG, and ACE scores in patients only (*rho* = 0.0039, *p* = 0.987, Spearman's rank correlation). Bottom right: correlation between backward connectivity strengths from the right ITG to right OCP, and explicit memory task performance. Patients: red; controls: blue. AD = Alzheimer's disease; corr. = correlation; SEM = standard error of mean.

5.3.5 Post-hoc Classical Analyses Confirmed PEB Findings, and Revealed That Association Between Implicit Memory Performance and Intrinsic Glutamate Activity is Patient-Driven

To confirm my findings from the PEB analysis, I then used classical statistics to further interrogate the effects of disease and task performance on the above parameter estimates.

Classical inference on these parameter estimates confirmed that in the priming task, controls had significantly increased intrinsic glutamate connectivity in right OCP compared to patients (*t*(40) = 2.22, *p* = 0.0319), as seen in the PEB (**Figure 5.5**). Also, intrinsic glutamate connectivity within right IFG showed significant positive correlation with the RT difference (*rho* = 0.421, *p* = 0.00550) (**Figure 5.5**). However, this correlation is primarily driven by patients (patients only: *rho* = 0.439, *p* = 0.0465), rather than the controls (controls only: *rho* = 0.167, *p* = 0.468), implying that this glutamate connectivity in the right hemisphere may be playing a task-related compensatory role specifically in Alzheimer's disease patients.

In the recognition task, classical inference on parameter estimates displayed an increase in forward connectivity from left ITG to left IFG in controls compared to patients (*t*(40) = -2.26, *p* = 0.0295) (**Figure 5.6**). Furthermore, backward connectivity from right ITG to right OCP showed significant positive correlation with recognition accuracy score (*rho* = 0.354, *p* = 0.0216), further confirming my findings in the PEB analysis of left-hemisphere dropout in patients, and task-related increases in right hemisphere connectivity in explicit memory processing (**Figure 5.6**).

5.3.6 Time-Frequency Analyses Uncovered Greater Theta Band Power in Controls Across Both Tasks, With Increased Lateralised High-Gamma Band Power in Patients Exclusively in Recognition Task

Between patients and controls in the priming task, controls showed greater theta activity at 5 Hz in the left IFG throughout the trial compared to patients, peaking after 228 ms (*t*(1,40) = 2.60, *p* = 0.006). I also saw increased low-gamma early in the trial at 28 ms post-stimulus at 33 Hz in patients, also in the left IFG (*t*(1,40) = 1.95, *p* = 0.027) (**Figure 5.7A**).

Figure 5.7

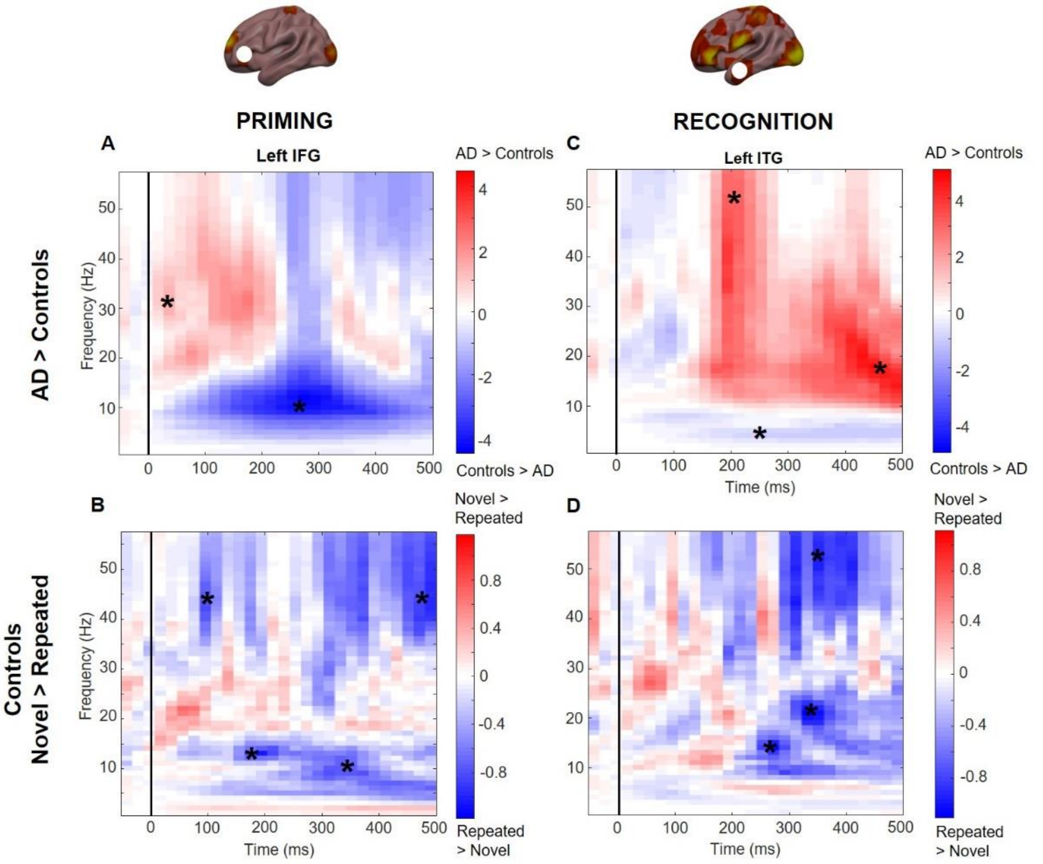


Figure 5.7 Time-frequency analysis of source extracted data in left IFG in the priming task, and left ITG in the recognition task. (**A**) Differences between patients and controls in the left IFG in the priming task, mean across participants and trial type in each group. Red indicates power greater in patients than controls, blue indicates power greater in controls than patients. (**B**) Differences between novel and repeated trials in the priming task, controls only, in the left IFG, mean across all control participants and trials per trial type. Red indicates power greater in novel trials than repeated trials, blue indicates power greater in repeated trials than novel trials. (**C**) Differences between patients and controls in the left ITG in the recognition task, mean across participants and trial type in each group. (**D**) Differences between novel and repeated trials in the recognition task, controls only, in the left ITG, mean across all control participants and trials per trial type. Colour bars indicate spectral power. Vertical black line at $t = 0$ indicates stimulus onset. * = $p < 0.05$, uncorrected for multiple comparisons. AD = Alzheimer's disease.

When examining novel versus repeated trials exclusively in controls, left IFG broadband activity was overall greater in repeated trials compared to novel trials. I observed increases in beta activity mid-to-late trial in response to repeated stimuli at 14 Hz after 168 ms and 13 Hz after 348 ms (*t*(1,60) = 4.11, *p* < 0.001; *t*(1,60) = 1.86, *p* = 0.038), and also increases in early- and late-trial gamma activity at 47 Hz and 41 Hz, after 108 ms and 468 ms, respectively (*t*(1,60) = 2.46, *p* = 0.012; *t*(1,60) = 2.77, *p* = 0.006) (**Figure 5.7B**).

In the recognition task, patients showed significantly greater late-trial beta activity compared to controls at 17 Hz, 488 ms (*t*(1,40) = 3.28, *p* = 0.001), in addition to mid-trial high-gamma activity at 54 Hz, 188 ms (*t*(1,40) = 2.43, *p* = 0.009) in the left ITG. Such high-gamma activity was not observed in the priming task. Controls also displayed the same increase in theta activity at 5 Hz compared to patients as in the priming task, peaking at 248 ms (*t*(1,40) = 2.67, *p* = 0.005) also in left ITG (**Figure 5.7C**).

Repeated measures ANOVAs highlighted overall increased activity in left ITG in repeated compared to novel trials in controls, similarly to the priming task. This showed significantly greater late-trial high frequency gamma activity in repeated trials after 348 ms at 52 Hz compared with novel trials in controls' left ITG (*t*(1,60) = 3.59, *p* = 0.001). Mid-trial low frequency gamma was also significantly greater in repeated than novel trials, at 15 Hz, 268 ms and 24 Hz, 328 ms (*t*(1,60) = 1.96, *p* = 0.025; *t*(1,60) = 1.81, *p* = 0.035) (**Figure 5.7D**). The significant differences described above, however, did not survive correction for multiple comparisons.

5.4 Discussion

While previous studies have shown left hemisphere-specific effects in patients with Alzheimer's disease at rest (Scahill *et al.*, 2002; Miller *et al.*, 2008), here I used DCM of task-based EEG and PEB to demonstrate that in simple priming memory tasks, Alzheimer's disease patients suffer from slowing of implicit memory processes in the left hemisphere but display task-related right hemisphere-specific upregulation of local glutamate connectivity, which may play a compensatory role in implicit memory circuits. In the more taxing explicit memory task, I found source-level memory circuit dropout in the left hemisphere of patients which were preserved in the priming task. Also using PEB, I showed task-associated increases in connectivity strengths and local excitation specifically in the right hemisphere, implying compensatory mechanisms are being performed by the right hemisphere in these patients. Importantly, the PEB analyses enabled me to examine both group effects and the effects of each task while accounting for disease state.

I further showed that the tasks used in this study are effective in testing implicit and explicit memory, as controls consistently showed significantly better performance compared to patients as expected. These behavioural results revealed a high level of variability in task performance of Alzheimer's patients for both tasks; task performance was quantified as mean RT difference in the priming task and accuracy score in the recognition task. Recognition accuracy scores correlated strongly with patients' ACE scores, an established clinical score which can be used to indicate the presence of dementia and disease progression. This correlation suggests a direct relationship between patients' explicit memory and severity of cognitive deficits.

These two behavioural tasks examined distinct types of memory recall: implicit priming memory and explicit recognition memory. Previous studies have demonstrated a slight preservation of implicit memory in patients with mild Alzheimer's disease using similar priming and recognition memory tasks, specifically using picture stimuli. A study by Deason *et al*. (2015) showed preserved implicit conceptual priming memory in subjects suffering from mild Alzheimer's disease comparing the priming effect when pictures and words were presented as visual stimuli. They found intact priming in Alzheimer's disease patients only when pictures were used as stimuli, and also used an explicit recognition memory task to demonstrate a decline in recognition memory in patients with mild Alzheimer's disease compared to healthy aged controls (Deason *et al.*, 2015). Another study by Martins and Lloyd-Jones (2006) showed similar effects using a fragmented picture paradigm, demonstrating preserved perceptual closure in Alzheimer's disease patients (Martins and Lloyd-Jones, 2006). The patient group examined here suffered from mild to moderate Alzheimer's disease and displayed varying severities of cognitive decline, reflected by a wide range of clinical ACE scores. This variation in cognitive ability within the patient group could explain why, although patients were slightly slower in novel trials compared with repeated trials, this preservation of implicit memory is not statistically significant.

Using source localisation, I identified distinct networks in each task: a six-source bilateral network in the recognition task; and a simplified four-source network in the priming task. These networks included left and right occipital sources and left and right frontal sources, with the addition of left and right temporal sources in the recognition task only. I expected to identify a more complex network for the recognition task, in that there is greater recruitment of medial temporal regions in explicit memory recall, as compared to priming. It naturally follows, therefore, to include bilateral ITG sources here as an extension of the network used for the priming task. As in all DCM studies, my findings are dependent on my selection of sources included in the models, however my source selection is justified and well-supported by source localisation analyses and previous work (Gilbert and Moran, 2016). My six-source network represents a sub-network of the full explicit memory network which may employ additional regions. However, to reduce model complexity and prevent over-fitting of the model, I selected a maximum of six sources in the sub-network. The deposition of tau neurofibrillary tangles has been reported to initiate in the MTL and spread outwards as the disease progresses (Marks *et al.*, 2017; Pasquini *et al.*, 2019), and studies have shown that Aβ has increased deposition in the left MTL during early Alzheimer's disease (Frings *et al.*, 2015). I then generated DCMs for the tasks using these two different networks and conducted PEB analyses to examine both group effects and task performance.

The PEB analyses revealed a slowing of signal transmission generally and with further slowing (increased time constants) specifically in the left hemisphere of Alzheimer's patients. This was observed in the priming task, along with strong associations between local within-region glutamate connectivity (*G*) and task performance in the right frontal lobe, specifically the right IFG. This correlation was predominantly driven by patients only, revealed by post-hoc classical analyses – indicating compensatory, right hemisphere recruitment. The spread of parameter values for glutamatergic local connectivity was much greater in patients, and showed strong correlation with the mean RT difference (i.e. task performance) in the priming task, whereas in healthy controls the range of *G* parameter values was much narrower and did not correlate with task performance, exhibiting a low variability relative to patients. This glutamate connectivity, or gain, also showed group-level differences, not related to task performance *per se*: with enhanced gain in the right OCP which was significantly greater in healthy controls compared to patients in the priming task. These findings suggest that intrinsic glutamate connectivity in the right hemisphere may act as a compensatory mechanism in Alzheimer's patients while performing simpler implicit memory tasks, but controls have overall higher levels of this connectivity as a group and are still able to outperform patients.

In the recognition task, however, much larger-scale network differences between patients and controls are evident. There was significant network dropout in the left hemisphere: forward connectivity from the left ITG to left IFG was reduced in patients compared with controls, and subcortical input into the left inferior OCG was significantly reduced in patients. This therefore indicates that left hemisphere memory circuits are compromised in patients, and in the more difficult explicit memory task patients have a reduced capacity to compensate for this loss, as seen in their significantly reduced task performance versus controls. It may be that these networks are preserved in implicit memory processing in patients, at least early in the disease, as this dropout is absent from the PEB analysis of the simpler priming task. In terms of task-based effects, I observed a strong relationship between recurrent right hemisphere connectivity, namely forward right OCP to right ITG and backward right ITG to right OCP, and recognition accuracy score across participants, with a higher PEB effect size for the backward connections. Thus, this right hemisphere region-to-region connectivity may play a role in explicit memory recall.

Many studies have found that, at rest, patients suffering from MCI or early stages of Alzheimer's disease have lateralized atrophy specifically in the left hemisphere (Miller *et al.*, 2008), and a resting-state MRI study by Thompson *et al*. (2001) found increased left hemisphere grey matter atrophy in patients with mild to moderate Alzheimer's disease compared to healthy elderly controls (Thompson *et al.*, 2001). Another resting-state MRI study (Fox *et al.*, 1996) showed significant asymmetry in the left and right hippocampal formation of pre-symptomatic individuals at risk of familial Alzheimer's disease who were followed for three years and later developed symptoms of Alzheimer's disease. The right hippocampal formation showed no significant differences to that of controls, whereas left hippocampal formations were significantly smaller than that of controls during this pre-symptomatic period (Fox *et al.*, 1996). Here, I show specific left-lateralized slowing and depletion of connectivity in patients with Alzheimer's disease in two different memory tasks, with potential task-related compensation for implicit memory circuits in the right hemisphere.

Regarding demographic data collected from participants, it has been widely documented that level or years of education have a significant impact on the risk of individuals suffering from dementia. This link between education and dementia was initially suggested by Mortimer (Mortimer and Graves, 1993), who concluded that education may increase levels of 'intellectual reserve', which may play a protective role against dementia (Mortimer and Graves, 1993; Sharp and Gatz, 2011). Another study by Almeida *et al.* (2015) found that individuals with high levels of cognitive reserve (measured by educational attainment) displayed reduced levels of pathological tau and Aβ protein in cerebrospinal fluid compared to those with low cognitive reserve (Almeida *et al.*, 2015). Also, a more recent study conducted a Mendelian randomisation analysis and found that higher levels of education (i.e. attending further education institutions, and/or increased years of education) was linked to reduced risk of Alzheimer's disease (Larsson *et al.*, 2017). This study therefore supports previous findings that individuals with higher years of education may have a reduced risk of Alzheimer's disease.

Interestingly, in this study patients that displayed depressive symptoms performed slightly better in the recognition task than patients that did not suffer from depressive disorders, though not statistically significant. However, there is a large body of evidence suggesting that depression is a risk factor for developing Alzheimer's disease and may even be an early symptom of Alzheimer's disease and MCI. Alzheimer's disease has been linked to a number of psychiatric disorders, and depression and apathy are common presentations of psychosis in Alzheimer's patients (Lee and Lyketsos, 2003). Depression, alongside anxiety disorders, has been suggested to worsen cognitive decline in Alzheimer's patients (Starkstein *et al.*, 2008). In addition, levels of pro-inflammatory cytokines, which can induce the generation of neurodegenerative factors and reactive oxygen species, have been found to increase in both Alzheimer's disease and depression (Leonard and Myint, 2006; Wuwongse *et al.*, 2010). It may be the case that the dementia patients in this study that also showed depressive symptoms may have been misdiagnosed as having dementia, however the exact mechanisms of how Alzheimer's disease and depression are linked are unknown. The non-statistically significant result for recognition accuracy scores may also be due to a very low n number, as only five Alzheimer's patients suffered from depressive symptoms. Therefore, further investigation into how these neuropsychiatric conditions and their pathological mechanisms are associated is required, though these findings indicate an alternate, intriguing direction and one worth further consideration.

Furthermore, exercise is widely reported to reduce the risk of developing dementia in later life but is also linked to reducing the risk of depressive disorders. In several studies, particularly in the past 15 years, a strong link between physical exercise and cognitive health has been suggested. Exercise and aerobic activity have been found to show positive effects on brain structure in both Alzheimer's patients and healthy aged adults in structural imaging studies (Burns *et al.*, 2008; Bugg and Head, 2011). A study by Liang *et al.* (2010) also demonstrated a novel association between levels of physical exercise and Alzheimer's-associated biomarkers in healthy aged adults (Liang *et al.*, 2010). Furthermore, dementia patients taking part in exercise study programs have shown reductions in depressive symptoms (Williams and Tappen, 2008), and exercise may slow down the cognitive decline that occurs in Alzheimer's patients (Rolland *et al.*, 2007; Paillard *et al.*, 2015). Therefore, exercise levels may directly affect cognitive ability, but may indirectly mitigate cognitive decline through reducing depressive symptoms. However, without retrospective data, self or family report (as relied upon here) may be unreliable or at least noisy.

In the time-frequency analyses of source extracted data from frontal regions in both tasks, and temporal regions for the recognition task, I revealed increased high gamma-band power mid-trial in patients' left ITG compared to controls exclusively in the recognition task, with increased mid-trial low gamma-band power in patients' left IFG in the priming task. Controls also displayed greater theta-band power than patients across the whole trial in both tasks. When examining differences between novel and repeated trials in controls only, I saw an overall increased power in response to repeated stimuli in both tasks, particularly in the high-gamma and high-alpha/beta frequency bands.

Han *et al*. (2017) examined spectral power in combination with functional connectivity during an object-location memory task, and found that theta- and alpha-band power was significantly increased in Alzheimer's patients compared to healthy aged controls during the memory retrieval phase of the task (Han *et al.*, 2017). However, no significant differences in delta-, beta- or gamma-band power were identified during this task session. This contrasts with my findings, as I observed increased theta-band power across the trials in controls compared with patients in both tasks and increased gamma-band power in patients. However, my findings became non-significant when correcting for multiple comparisons.

This ambiguity in the time-frequency analysis literature may reflect distinct task demands. Furthermore, my results may be due to the range of disease severities from mild to moderate within the patient group, evidenced by the large variation in task performance and ACE scores observed across patients, as there may be variation in spectral power pertaining to disease severity which wasn't accounted for in my analyses. Differences in resting-state EEG amplitude modulation have been identified between patients with differing severities of Alzheimer's disease (Fraga *et al.*, 2013). Fraga *et al*. (2013) reported differences in the delta modulation of the beta frequency band between patients suffering from mild versus moderate Alzheimer's disease, in that those with moderate Alzheimer's disease displayed significant decreases in modulation in the beta band and increases in delta and theta modulations of the theta band compared to those with mild Alzheimer's disease. This suggests that EEG amplitude modulation can not only be an indicator of disease presence, but also of disease severity. Regarding EEG spectral power, they also found significant increases in theta-band power in parietal and occipital locations in patients with mild Alzheimer's disease compared with healthy aged controls, again contrasting my findings of increased theta-band power in frontal regions in controls. Therefore, future work may involve categorising Alzheimer's patients based on their disease severity or cognitive performance metrics and examining differences in power across multiple patient groups.

A potential limitation of this paradigm is the use of covert rather than overt naming during the priming task. However, the explicit identification of an object during priming does not affect long-term object priming as shown in a recent study (Gomes and Mayes, 2015), and whether subjects were consistent in their naming of objects is of greater importance than the particular name given to the object in the context of the priming task.

The use of EEG in this paradigm was essential for estimating parameters such as excitatory time constants and signal delays. EEG delivers a high temporal resolution at timescales constant with that of synaptic transmission; this is in contrast with other human neuroimaging techniques such as fMRI. While fMRI may offer much more powerful spatial resolution, in this experiment it was crucial to obtain more temporally resolved, direct measures of neuronal activity in order to scrutinise parameters inferred using my DCMs which span from macroscale region-to-region connectivity, to mesoscale ensembles of cellular and synaptic dynamics.

Future work may examine potential lateralized compensatory mechanisms and cognitive reserve in bilingualism. Bilingualism has been widely reported to delay the onset of many forms of dementia, including Alzheimer’s disease (Bialystok *et al.*, 2007; Craik *et al.*, 2010). Bilingual brains have also been shown to undergo experience-associated neuro-structural alterations, particularly in left-hemispheric regions such as the left IFG (Stein *et al.*, 2012) and left inferior parietal lobule (Della Rosa *et al.*, 2013). These changes may be neuroprotective in age-related cognitive decline as compared with brains of monolinguals (Abutalebi *et al.*, 2014). Bilingual brains may therefore be able to better compensate the loss of connectivity in the left hemisphere that is observed in dementias such as Alzheimer’s disease. Such work could offer powerful insights into compensatory and neuroprotective mechanisms against Alzheimer’s disease. Overall, these results speak to a relative specificity of functional pathology in regional circuit-level signal integration and how compensatory measures may be in play and may be identified.

Chapter Six

General Discussion

In this thesis, I applied the Bayesian computational approaches of Active Inference and DCM to investigate neuronal mechanisms underlying behavioural phenotypes and neural signals involved in decision-making and visual memory, using fMRI and EEG, in healthy adult volunteers and Alzheimer's disease patients, respectively.

I initially conducted four behavioural studies: three exploratory pilot studies and finally a main behavioural study, consisting of 25 healthy young-adult participants (*Chapter Two*). By conducting these studies, I aimed to optimise a probabilistic decision-making task, which would probe different aspects of uncertainty: namely, *expected uncertainty*, which arises as a result of consistent, known unreliability or known lack of precision in an environment; and *unexpected uncertainty*, which results from drastic unsignalled changes in the environment that require large belief updating and model rebuilding (Yu and Dayan, 2005). The task structure was based on that used by Gläscher *et al.*, with modifications to enhance the effect of spatial foraging, and in the latter two behavioural studies, the introduction of a contextual reversal to specifically probe *unexpected uncertainty* (Gläscher *et al.*, 2010). Additional aims of these studies were to examine the policy selection and behavioural phenotypes on a within-subject level, and determine whether the choice behaviour of participants could be explained by model-free or model-based learning methods, by analysing participants' individual action choices at each timepoint in the trials in a free-choice testing session, which followed a fixed training session in which participants could view the environment, but not navigate freely.

Over the course of conducting this set of studies, the paradigm was altered and fine-tuned to continuously improve the task. Following the first pilot study, it was clear that participants were viewing each scene as a discrete state or location, rather than a single point on a continuous route. This is what I was aiming to achieve in order to emulate spatial foraging, therefore in the second pilot study I replaced the images in the task with more congruent images, which appeared to be taken from different points along the same route. This improved my findings, as there was a slight increase in 'correct' first actions in the second pilot study compared with the first. This suggests a slight increase in model-based learning mechanisms in the second pilot study compared with pilot one. However, overall selection of the optimal policy did not increase significantly. This led to the third pilot study, in which I aimed to investigate optimal trial length. This third study also saw the introduction of the task reversal: the optimal policy was switched from the right arm to the left arm of the task during the testing session. Here, I wanted to examine how participants would respond and adapt to unexpected uncertainty in the task, as a result of the reversal.

Based on my findings in pilot studies one and two, the main focuses of the third pilot study and subsequent main behavioural study were model-based SAPEs, therefore the fixed training session was switched to a free-choice training session, which I hypothesized would lead to more rapid learning of the optimal policy, with information about rewarding states revealed at the start of the task. I also hypothesized that I would observe similar ranges in behavioural profiles as in pilots one and two, which may change on a single-subject level following the reversal.

In the third pilot study, I found that the increased task length and free-choice training session revealed an interesting division in behaviour across the cohort. Out of seven participants, four could be labelled as *exploiters*, in that they showed highly significant preference for the optimal policy. Conversely, three participants could be labelled as *explorers*, as their choice behaviour remained at chance for the duration of the free-choice training session. Following the reversal in the testing session, however, most of the previously exploratory participants noticed that a reversal had taken place, and due to their experience of the task structure from previous exploration, chose to exploit the new, post-reversal optimal policy. This was a particularly intriguing observation, as in the previous two pilot studies there were more continuous spectra of behavioural phenotypes across participants rather than two distinct groups of *explorers* and *exploiters,* which may have emerged as a result of the new free-choice training session.

Due to the success of the third pilot study, it was necessary to investigate this longer task structure, with the inclusion of a testing-session reversal, in a larger cohort: the main behavioural study. Following a small additional adaptation to the outcome states in the task structure (removing any ambiguity of which post-reversal policy was truly ‘optimal’), participants’ learning of the optimal policy increased even further: participants in the main behavioural study showed increased learning of optimal routes compared with all three previous studies, therefore demonstrating that this probabilistic decision-making paradigm had been improved and optimised through previous structural adaptations.

Overall, I found that participants were able to learn the task structures successfully, but also displayed a wide spectrum of behavioural characteristics, in that some participants were able to learn and *exploit* the so-called ‘optimal’ policy (the policy which lead to the highest-probability high-level reward) within the first half of the task, whereas other participants seemed to prefer a more *exploratory* strategy, by selecting each policy relatively close to chance level. Similarly to the results reported by Gläscher *et al*., I also found that, consistently across all four studies, participants’ behaviour on a group-level could not be fully explained solely by model-free learning mechanisms, as participants were able to learn optimal actions at each timepoint in the trials, and displayed perseverative behaviour in the optimal ‘arm’ of the task, following unlikely events of reward absence (Gläscher *et al.*, 2010). These findings are also in line with Daw *et al*., who demonstrated that human participants displayed hallmarks of both model-free and model-based learning approaches in a two-step decision-making task (Daw *et al.*, 2011). As the four studies progressed, I observed that the proportion of participants able to learn the optimal policy increased as adaptations were made to the task (apart from pilot study three, in which participants could be split into *explorers* and *exploiters*), with the main behavioural study generating the highest percentage of participants who were able to sufficiently exploit the optimal policy (84%).

Once this paradigm had been fully optimised, I then applied the task structure of 320 trials (in the behavioural tasks, this was divided into a 160-trial training session and a 160-trial testing session, which were combined into a single 320-trial session for the modelling simulation task) and a task reversal after trial 200 (in the behavioural studies, this was following trial 40 in the testing session), to an Active Inference framework, by constructing a generative model of this task (*Chapter Three*). The key aim of this study, and main rationale behind using Active Inference, was to execute an inversion of the generative model, using the state-action data collected in the main behavioural study, in order to accurately estimate subject-specific parameters of internal model flexibility and precision over rewarding states. The behavioural task also particularly lent itself to being structured as a POMDP, as the task was a probabilistic Markov decision tree with hidden states, and participants were expected to develop beliefs about state-transitions, i.e. update their initially naïve state-transition (*b*) matrices.

I then wanted to use these parameter estimates to phenotype individuals using coarse behavioural metrics, and characterise how these parameters influence behaviour on the group-level. I experimented with two different model inversions: one with two free parameters that described a scalar over internal model volatility (*k*) and precision over rewarding states (*c*); and one with three free parameters, adding a further parameter that described the midpoint of the function defining model volatility (*m*) to *k* and *c*. The purpose of also estimating *m* was to examine whether conditional MAP estimates of *k* and *c* would be more accurately retrieved by allowing more flexibility in the updating of the internal model decay parameter α, which was calculated on a trial-by-trial basis based on values of *k* and *m*. This model was originally adapted by Sales *et al.*, who introduced the model decay parameter α to investigate SAPEs and how these affected simulated firing patterns of LC activity (Sales *et al.*, 2019).

The parameters *k*, *c*, and *m* were selected for this inversion because they describe aspects of participants' choice behaviour relevant to my hypotheses, and a single parameter value is calculated per participant (rather than per-participant per-trial dynamic parameters such as α), which greatly simplifies subject-specific phenotyping. I hypothesized that the model inversion would be able to successfully retrieve parameter values of *k* and *c* in simulated data, and that conditional MAP estimates of *k* and *c* would predict the frequency of optimal policy selection and total reward earned in the task, respectively.

Through data simulation using fixed values of *k* and *c*, model inversion of the simulated data, then inputting conditional MAP estimates of *k* and *c* into the forward model and subsequently reinverting the data from the forward model, I was able to accurately re-estimate values of *k* and *c* for simulated data, when the fixed values were known. Then, by inputting the states and actions experienced by my human participants in the main behavioural study, I used the model to generate conditional MAP estimates of *k*, *c*, and *m*. I confirmed the accuracy of these estimates by inputting these estimates into the forward model, simulating behaviour in the task, and subsequently re-inverting the model in much the same way as with the initial data simulations. The data simulated using values of *k*, *c*, and *m* was remarkably similar to that produced by real human participants, and model re-inversion retrieved the initial conditional MAP estimates with relative accuracy.

Further, I ran a PEB analysis, followed by classical statistics, to unpack how these parameters predicted broad behavioural differences on a group level. I found that *c* showed strong associations with total reward earned, and appeared to have a linear relationship up to a threshold, after which reward did not increase as *c* increased, demonstrating a saturation effect. There were no significant linear associations between *k* and optimal policy selection across the cohort, however, an inverted-U shaped relationship between *k* and optimal policy selection was clearly evident, in that intermediate values of *k* were associated with high frequencies of optimal policy selection, whereas *k* values at high and low extremes were associated with low or moderate levels of optimal policy selection. This could reflect the influence of *k* on internal model flexibility: a hyper-flexible model would result in a highly exploratory participant unable to detect the optimal policy and exploit it, and a hyper-rigid model would result in a participant who initially locates and exploits the optimal policy, but would struggle to adapt to the task reversal. This reflects previous work which has identified inverted-U shaped functions in relation to arousal, NA levels and tonic LC activity (Yerkes and Dodson, 1908; Aston-Jones and Cohen, 2005).

Therefore, this Active Inference model inversion framework was highly successful in accurately estimating subject-specific parameter values describing internal model volatility and precision over rewarding states, in addition to revealing interesting links between conditional MAP estimates of model parameters and broad behavioural metrics.

Building upon both my behavioural and computational findings, I then conducted an fMRI-pharmacology study, in which healthy adult volunteers completed the probabilistic decision-making task as optimised in my behavioural studies (*Chapter Two*), in combination with fMRI and the SNRI, reboxetine (*Chapter Four*). In this study, participants completed the training session of the task, took a single oral dose of reboxetine, then completed the testing session of the task during collection of fMRI images. I also replicated the Active Inference model inversion pipeline, as conducted in *Chapter Three*, to estimate subject-specific parameters of precision over rewarding states and internal model volatility, in the same way as with the participants in the main behavioural study. I also used these parameter estimates as covariates in group-level parametric analysis of fMRI data, to investigate how these parameters may predict neural activity during the task.

In this study, I aimed to investigate how selective NA reuptake inhibition influenced SAPEs and belief updating in the previously-studied task, and how this drug manipulation might also affect how participants responded to the reversal. I also aimed to examine how neural signatures of SAPEs observed during fMRI scanning could relate to the conditional parameter estimates generated through model inversion. The use of fMRI in this study was vital for the inspection of LC activity, as an extremely task-relevant structure, and the high spatial resolution of fMRI was of great benefit due to the relatively small size of the LC. I hypothesized that the behaviour displayed by participants during training would closely reflect that observed in my behavioural studies, in that most participants would be able to locate the optimal policy and exploit it accordingly, but participants would nevertheless display a range of behavioural phenotypes. I also hypothesized that the drug manipulation would have a group-level effect on conditional estimates of precision over reward and internal model volatility, and also that neural activity previously reported to be implicated in decision-making, SAPEs, and reversal learning would be observed in relation to different task conditions (Aston-Jones and Cohen, 2005; Apps and Ramnani, 2014; Wang *et al.*, 2017).

Here, I found significant increases in activation in the LC across the group, for the positive effect of pre-reversal trials. With the addition of conditional MAP estimates as covariates, I found significant neural activity in the LC associated with the likely state condition, and precision over rewarding states (*c*), where subject-specific activations appeared to reflect behavioural performance: highest levels of activity occurred in the highest-performing participants and vice versa. Significant activations in the ACC were also observed in relation to internal model volatility (*k*), similarly associated with the likely state condition in the task. Additionally, participants displayed a broad spectra of behavioural phenotypes ranging from highly exploitative of the optimal policy to chance-level exploration of all possible routes, as expected and as previously observed. Although group-level changes in optimal policy selection as a result of pharmacological manipulation were non-significant, participants who were exploitative in the training session did display greater exploration during the testing session, and those who were more exploratory initially became more exploitative following drug administration, indicating a mixed effect of reboxetine on exploration/exploitation patterns, depending on pre-drug behaviour. This finding supports my previous observation of an inverted-U shaped relationship in optimal policy selection across participants.

Interestingly, the effect of reboxetine on conditional MAP estimates of *k* and *c* appeared to be parameter-specific. Where *c* showed consistent significant correlation with total reward both pre- and post-drug, and where *m* showed consistent lack of correlation with optimal policy selection both pre- and post-drug, *k* showed *no* significant correlation with optimal policy selection pre-drug, but *did* show strong significant correlation with optimal policy selection post-drug. This change as a result of reboxetine was also reflected in a slight (non-significant) difference between conditional estimates of *k* pre- versus post-drug, and a steepening (and left-shifting) effect of the inverted-U shaped function. Neither of these effects were observed in *c* or *m*, indicating that *k*, denoting internal model volatility, was specifically affected by NA reuptake inhibition. This, in combination with my imaging findings described above that link LC activation with precision over rewarding states, point towards a role of NA in motivation and behavioural energising, as previously investigated in non-human primates (Varazzani *et al.*, 2015; Jahn *et al.*, 2018).

The above findings detail how computational models can be used effectively to characterise human participants based on their coarse behavioural metrics and neural activity, recorded as BOLD signals through fMRI, in relation to decision-making in the healthy brain. In my final study, I aimed to apply similar computational Bayesian methods in the context of neurodegenerative disease, by investigating aberrant connectivity within neural networks in Alzheimer's disease patients, and how this network dysfunction relates to pro-cognitive compensatory changes (*Chapter Five*). I hypothesized that Alzheimer's disease patients would display a weakening of hierarchical left hemisphere-specific connectivity, which could be worsened as task difficulty increased. I also aimed to examine whether compensatory mechanisms were employed by right hemisphere connections in implicit and explicit memory tasks.

I first pre-processed and analysed high-density EEG data collected from patients suffering from Alzheimer's disease and healthy aged controls, while they completed visual priming (implicit memory) and recognition (explicit memory) tasks. I then examined source localised network activity exhibited by participants, and used DCM and PEB analyses to interrogate changes in effective connectivity and underlying synaptic dynamics between patients and controls. The use of EEG enabled the collection of more direct measures of neural activity compared with data that would be obtained through alternative neuroimaging methods such as fMRI, with high temporal resolution on the timescale of synapses, which enabled the interrogation of parameters inferred through DCM spanning macroscale inter-region connectivity, to mesoscale cellular and synaptic dynamics. This neuroimaging method is also much preferred for patients suffering from neurodegenerative disease, as it is much less stressful for patients compared with MRI scanning.

I found significant reductions in subcortical visual input and in temporo-frontal connectivity in patients in the explicit memory task, specifically in the left hemisphere. Significant slowing in left hemisphere signal transmission was also observed during the implicit priming task, with significantly more distinct dropout in connectivity during the recognition task, suggesting that these network drop-out effects are affected by task difficulty.

Furthermore, during the implicit memory task, increased right frontal activity was correlated with improved task performance in patients only, suggesting that right hemisphere compensatory mechanisms may be employed to mitigate left-lateralized network dropout in Alzheimer’s disease. Taken together, these findings suggest that Alzheimer’s disease is associated with lateralized memory circuit dropout and potential compensation from the right hemisphere, at least for simpler memory tasks.

Recent work by Ian Robertson has suggested a potential role of NA in cognitive reserve (Robertson, 2013). Many studies have reported significant cell atrophy in the LC in Alzheimer’s disease pathology (Szot *et al.*, 2006), and degeneration of the LC has been observed during early stages of Alzheimer’s disease, in addition to during pre-Alzheimer’s conditions such as MCI. NA also plays important roles in neurobiological pathways that have been linked to Alzheimer’s disease pathology, including anti-inflammatory mechanisms such as stimulation of production of Brain-Derived Neurotrophic Factor (BDNF) (Mannari *et al.*, 2008), and reduction of amyloid toxicity (Heneka *et al.*, 2010). NA has also been shown to play vital roles in age-related declines in working memory (Wang *et al.*, 2011); well-known to be detrimentally affected in Alzheimer’s disease.

Robertson also described mechanisms specifically in the right hemisphere associated with cognitive reserve (Robertson, 2014). Cognitive reserve, as termed by Stern, refers to differences in cognitive processing performance, which may increase the likelihood of an individual to preserve cognitive ability or functioning, despite the presence of damage and/or neurological disease (Stern, 2012). He postulated that both structural and functional connectivity in the right hemisphere, particularly pre-frontal cortex, could predict cognitive reserve (**Figure 6.1**). This is in line with my findings in *Chapter Five* that the right hemisphere played compensatory roles in implicit priming memory in patients suffering from Alzheimer's disease.

Something that I found to be particularly compelling in my studies was the phenomenon of the inverted-U shaped function which can be used to describe the relationship between NA levels, or simply arousal, and performance in decision-making paradigms. A review by Arnsten (1998) highlighted the importance of the pre-frontal cortex in working memory, and its high sensitivity to noradrenergic and dopaminergic inputs. Arnsten also outlined that many studies have observed inverted-U shaped functions between levels of catecholamines and behavioural task performance, particularly DA and NA (Arnsten, 1998), with inverted-U shaped relationships between cognitive control and DA levels reported in a number of studies (Cools and Robbins, 2004; Cools and D'Esposito, 2011), in addition to the inverted-U shaped relationship between tonic LC activity and task performance (Aston-Jones and Cohen, 2005). Robertson postulated a potential link between arousal, modulated by NA, and cognitive reserve, in that optimal levels of arousal could potentially alleviate symptoms of cognitive decline, such as those seen in Alzheimer's disease (Robertson, 2014).

Figure 6.1

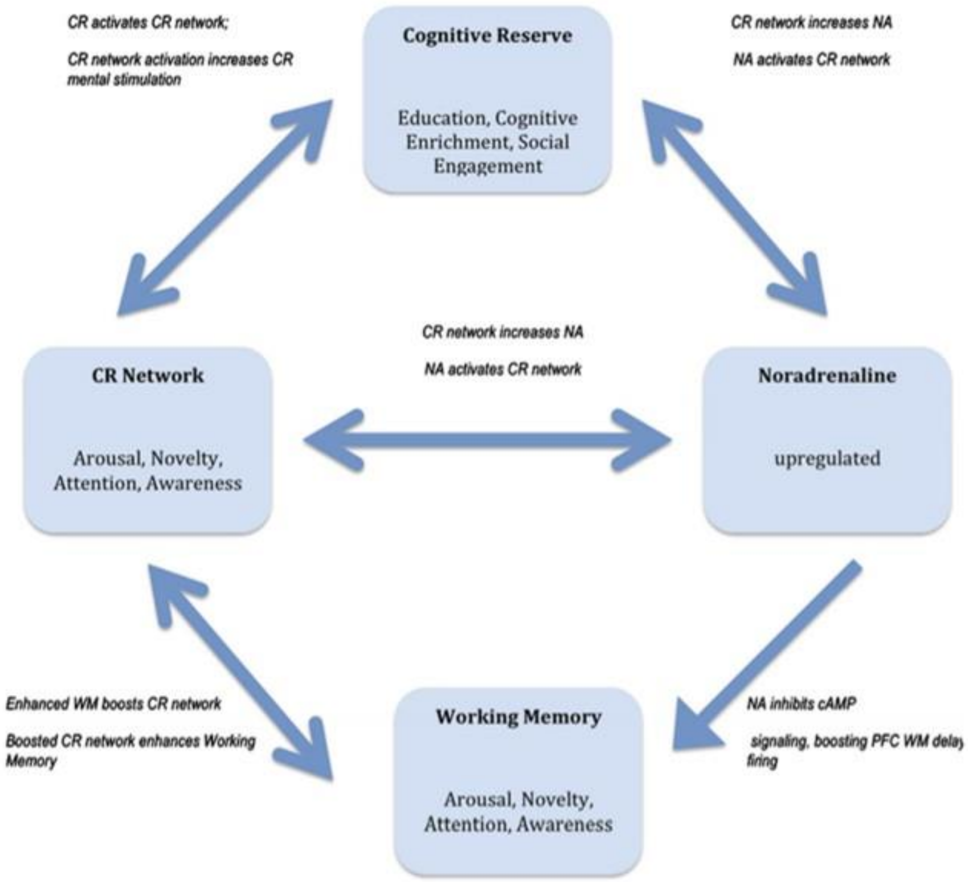

CR activates CR network;
CR network activation increases CR mental stimulation
Cognitive Reserve
Education, Cognitive Enrichment, Social Engagement
CR network increases NA
NA activates CR network
CR network increases NA
NA activates CR network
CR Network
Arousal, Novelty, Attention, Awareness
Noradrenaline
upregulated
Enhanced WM boosts CR network
Boosted CR network enhances Working Memory
Working Memory
Arousal, Novelty, Attention, Awareness
NA inhibits cAMP
signaling, boosting PFC WM dela
firing

Figure 6.1 Cognitive Reserve, Working Memory and NA. A hypothetical network linking cognitive reserve to NA modulation and working memory, taken from (Robertson, 2014). Abbreviations: CR = cognitive reserve; NA = noradrenaline; WM = white matter; cAMP = cyclic adenosine monophosphate.

Future work may involve the investigation of NA in decision-making, in the context of healthy aging compared with Alzheimer's disease, and how activity in the LC and ACC may potentially link to cognitive reserve. Decision-making and visual memory paradigms could be conducted in tandem to examine how disease-related deficits in visual memory may also be reflected in noradrenergic decision-making pathways. This could potentially be combined with pupillometry, to obtain additional insights into LC activity alongside task-based neuroimaging.

In summary, each component of this thesis highlights the importance of applying computational models to problems in neuroscience and psychiatry, in both health and disease. I initially optimized a behavioural paradigm to target decision-making pathways, and demonstrate model-based learning of the environment. I subsequently used the optimized task structure to construct a generative model, which I then inverted to generate subject-specific parameter estimates of precision over rewarding states and internal model volatility. Then, I combined these with fMRI and selective NA reuptake inhibition to investigate belief updating and how neural signatures of SAPEs relate to internal model volatility and precision over reward on a within-subject level. Finally, I applied the computational approaches of DCM and PEB to Alzheimer's patient data, and revealed left-lateralized memory circuit dropout, combined with right-hemisphere specific compensatory mechanisms in implicit memory, in comparison to healthy aged controls. Each of these computational methods have provided valuable insight into the neural dynamics underlying the observable behavioural and neural phenotypes, in the context of vital cognitive functions, decision-making and memory, in both the healthy and diseased brain.

Chapter Seven

References

Abutalebi J, Canini M, Della Rosa PA, Sheung LP, Green DW, Weekes BS. Bilingualism protects anterior temporal lobe integrity in aging. Neurobiol Aging 2014; 35: 2126-33.

Almeida RP, Schultz SA, Austin BP, Boots EA, Dowling NM, Gleason CE*, et al.* Effect of cognitive reserve on age-related changes in cerebrospinal fluid biomarkers of alzheimer disease. JAMA Neurol 2015; 72: 699-706.

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

Amodeo LR, Mcmurray MS, Roitman JD. Orbitofrontal cortex reflects changes in response-outcome contingencies during probabilistic reversal learning. Neuroscience 2017; 345: 27-37.

Apps MA, Ramnani N. The anterior cingulate gyrus signals the net value of others' rewards. J Neurosci 2014; 34: 6190-200.

Arnsten AF. Catecholamine modulation of prefrontal cortical cognitive function. Trends Cogn Sci 1998; 2: 436-47.

Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. Annu Rev Neurosci 2005; 28: 403-50.

Aston-Jones G, Rajkowski J, Kubiak P. Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. Neuroscience 1997; 80: 697-715.

Aston-Jones G, Rajkowski J, Kubiak P, Alexinsky T. Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. J Neurosci 1994; 14: 4467-80.

Bayes T. Lii. An essay towards solving a problem in the doctrine of chances. By the late rev. Mr. Bayes, frs communicated by mr. Price, in a letter to john canton, amfr s. Philosophical transactions of the Royal Society of London 1763: 370-418.

Berridge CW, Waterhouse BD. The locus coeruleus-noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. Brain Res Brain Res Rev 2003; 42: 33-84.

Bialystok E, Craik FI, Freedman M. Bilingualism as a protection against the onset of symptoms of dementia. Neuropsychologia 2007; 45: 459-64.

Box GEP. Science and statistics. Journal of the American Statistical Association 1976; 71: 791-799.

Brodersen KH, Deserno L, Schlagenhauf F, Lin Z, Penny WD, Buhmann JM*, et al.* Dissecting psychiatric spectrum disorders by generative embedding. Neuroimage Clin 2014; 4: 98-111.

Brodersen KH, Schofield TM, Leff AP, Ong CS, Lomakina EI, Buhmann JM*, et al.* Generative embedding for model-based classification of fmri data. PLoS Comput Biol 2011; 7: e1002079.

Buckner RL, Snyder AZ, Shannon BJ, Larossa G, Sachs R, Fotenos AF*, et al.* Molecular, structural, and functional characterization of alzheimer's disease: Evidence for a relationship between default activity, amyloid, and memory. J Neurosci 2005; 25: 7709-17.

Bugg JM, Head D. Exercise moderates age-related atrophy of the medial temporal lobe. Neurobiol Aging 2011; 32: 506-14.

Burns JM, Cronk BB, Anderson HS, Donnelly JE, Thomas GP, Harsha A*, et al.* Cardiorespiratory fitness and brain atrophy in early alzheimer disease. Neurology 2008; 71: 210-6.

Chadwick MJ, Jolly AE, Amos DP, Hassabis D, Spiers HJ. A goal direction signal in the human entorhinal/subicular region. Curr Biol 2015; 25: 87-92.

Chamberlain SR, Robbins TW. Noradrenergic modulation of cognition: Therapeutic implications. J Psychopharmacol 2013; 27: 694-718.

Chandley MJ, Szebeni K, Szebeni A, Crawford J, Stockmeier CA, Turecki G*, et al.* Gene expression deficits in pontine locus coeruleus astrocytes in men with major depressive disorder. J Psychiatry Neurosci 2013; 38: 276-84.

Chang YT, Huang CW, Chang WN, Lee JJ, Chang CC. Altered functional network affects amyloid and structural covariance in alzheimer's disease. Biomed Res Int 2018; 2018: 8565620.

Clery-Melin ML, Schmidt L, Lafargue G, Baup N, Fossati P, Pessiglione M. Why don't you try harder? An investigation of effort production in major depression. PLoS One 2011; 6: e23178.

Conradi HJ, Ormel J, De Jonge P. Presence of individual (residual) symptoms during depressive episodes and periods of remission: A 3-year prospective study. Psychol Med 2011; 41: 1165-74.

Cools R, Clark L, Owen AM, Robbins TW. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. J Neurosci 2002; 22: 4563-7.

Cools R, D'esposito M. Inverted-u-shaped dopamine actions on human working memory and cognitive control. Biol Psychiatry 2011; 69: e113-25.

Cools R, Robbins TW. Chemistry of the adaptive mind. Philos Trans A Math Phys Eng Sci 2004; 362: 2871-88.

Costa VD, Tran VL, Turchi J, Averbeck BB. Reversal learning and dopamine: A bayesian perspective. J Neurosci 2015; 35: 2407-16.

Cotecchia S, Kobilka BK, Daniel KW, Nolan RD, Lapetina EY, Caron MG*, et al.* Multiple second messenger pathways of alpha-adrenergic receptor subtypes expressed in eukaryotic cells. J Biol Chem 1990; 265: 63-9.

Craik FI, Bialystok E, Freedman M. Delaying the onset of alzheimer disease: Bilingualism as a form of cognitive reserve. Neurology 2010; 75: 1726-9.

Cullen M, Davey B, Friston KJ, Moran RJ. Active inference in openai gym: A paradigm for computational investigations into psychiatric illness. Biol Psychiatry Cogn Neurosci Neuroimaging 2018; 3: 809-818.

Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging 2008; 29: 514-23.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. Neuron 2011; 69: 1204-15.

Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 2005; 8: 1704-11.

Dayan P, Yu AJ. Phasic norepinephrine: A neural interrupt signal for unexpected events. Network 2006; 17: 335-50.

De Gee JW, Colizoli O, Kloosterman NA, Knapen T, Nieuwenhuis S, Donner TH. Dynamic modulation of decision biases by brainstem arousal systems. Elife 2017; 6.

Dearden R, Friedman N, Russell S. Bayesian q-learning. Aaai/iaai, 1998. 761-768.

Deason RG, Hussey EP, Flannery S, Ally BA. Preserved conceptual implicit memory for pictures in patients with alzheimer's disease. Brain Cogn 2015; 99: 112-7.

Della Rosa PA, Videsott G, Borsa VM, Canini M, Weekes BS, Franceschini R*, et al.* A neural interactive location for multilingual talent. Cortex 2013; 49: 605-8.

Derflinger S, Sorg C, Gaser C, Myers N, Arsic M, Kurz A*, et al.* Grey-matter atrophy in alzheimer's disease is asymmetric but not lateralized. J Alzheimers Dis 2011; 25: 347-57.

Fitzgerald TH, Schwartenbeck P, Moutoussis M, Dolan RJ, Friston K. Active inference, evidence accumulation, and the urn task. Neural Comput 2015; 27: 306-28.

Foti D, Weinberg A, Bernat EM, Proudfit GH. Anterior cingulate activity to monetary loss and basal ganglia activity to monetary gain uniquely contribute to the feedback negativity. Clin Neurophysiol 2015; 126: 1338-47.

Fox NC, Warrington EK, Freeborough PA, Hartikainen P, Kennedy AM, Stevens JM*, et al.* Presymptomatic hippocampal atrophy in alzheimer's disease. A longitudinal mri study. Brain 1996; 119 (Pt 6): 2001-7.

Fraga FJ, Falk TH, Kanda PA, Anghinah R. Characterizing alzheimer's disease severity via resting-awake eeg amplitude modulation analysis. PLoS One 2013; 8: e72240.

Frings L, Hellwig S, Spehl TS, Bormann T, Buchert R, Vach W*, et al.* Asymmetries of amyloid-beta burden and neuronal dysfunction are positively correlated in alzheimer's disease. Brain 2015; 138: 3089-99.

Friston K. The free-energy principle: A unified brain theory? Nat Rev Neurosci 2010; 11: 127-38.

Friston K, Fitzgerald T, Rigoli F, Schwartenbeck P, J OD, Pezzulo G. Active inference and learning. Neurosci Biobehav Rev 2016; 68: 862-879.

Friston K, Kilner J, Harrison L. A free energy principle for the brain. J Physiol Paris 2006; 100: 70-87.

Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free energy and the laplace approximation. Neuroimage 2007; 34: 220-34.

Friston KJ. Bayesian estimation of dynamical systems: An application to fmri. Neuroimage 2002; 16: 513-30.

Friston KJ, Harrison L, Penny W. Dynamic causal modelling. Neuroimage 2003; 19: 1273-302.

Fuchs PN, Peng YB, Boyette-Davis JA, Uhelski ML. The anterior cingulate cortex and pain processing. Front Integr Neurosci 2014; 8: 35.

Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM. The functional anatomy of the mmn: A dcm study of the roving paradigm. Neuroimage 2008; 42: 936-44.

Gilbert JR, Gotts SJ, Carver FW, Martin A. Object repetition leads to local increases in the temporal coordination of neural responses. Front Hum Neurosci 2010; 4: 30.

Gilbert JR, Moran RJ. Inputs to prefrontal cortex support visual recognition in the aging brain. Sci Rep 2016; 6: 31943.

Gläscher J, Daw N, Dayan P, O'doherty JP. States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 2010; 66: 585-95.

Golby A, Silverberg G, Race E, Gabrieli S, O'shea J, Knierim K*, et al.* Memory encoding in alzheimer's disease: An fmri study of explicit and implicit memory. Brain 2005; 128: 773-87.

Gomes CA, Mayes A. Does long-term object priming depend on the explicit detection of object identity at encoding? Front Psychol 2015; 6: 270.

Gould RL, Arroyo B, Brown RG, Owen AM, Bullmore ET, Howard RJ. Brain mechanisms of successful compensation during learning in alzheimer disease. Neurology 2006; 67: 1011-7.

Greicius MD, Supekar K, Menon V, Dougherty RF. Resting-state functional connectivity reflects structural connectivity in the default mode network. Cereb Cortex 2009; 19: 72-8.

Hampton AN, Bossaerts P, O'doherty JP. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci 2006; 26: 8360-7.

Han SD, Houston WS, Jak AJ, Eyler LT, Nagel BJ, Fleisher AS*, et al.* Verbal paired-associate learning by apoe genotype in non-demented older adults: Fmri evidence of a right hemispheric compensatory response. Neurobiol Aging 2007; 28: 238-47.

Han Y, Wang K, Jia J, Wu W. Changes of eeg spectra and functional connectivity during an object-location memory task in alzheimer's disease. Front Behav Neurosci 2017; 11: 107.

Harmer CJ, Hill SA, Taylor MJ, Cowen PJ, Goodwin GM. Toward a neuropsychological theory of antidepressant drug action: Increase in positive emotional bias after potentiation of norepinephrine activity. Am J Psychiatry 2003; 160: 990-2.

Hedden T, Van Dijk KR, Becker JA, Mehta A, Sperling RA, Johnson KA*, et al.* Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. J Neurosci 2009; 29: 12686-94.

Heneka MT, Nadrigny F, Regen T, Martinez-Hernandez A, Dumitrescu-Ozimek L, Terwel D*, et al.* Locus ceruleus controls alzheimer's disease pathology by modulating microglial functions through norepinephrine. Proceedings of the National Academy of Sciences 2010; 107: 6058-6063.

Hennig J, Lange N, Haag A, Rohrmann S, Netter P. Reboxetine in a neuroendocrine challenge paradigm: Evidence for high cortisol responses in healthy volunteers scoring high on subclinical depression. Int J Neuropsychopharmacol 2000; 3: 193-201.

Hill SA, Taylor MJ, Harmer CJ, Cowen PJ. Acute reboxetine administration increases plasma and salivary cortisol. J Psychopharmacol 2003; 17: 273-5.

Jahn CI, Gilardeau S, Varazzani C, Blain B, Sallet J, Walton ME*, et al.* Dual contributions of noradrenaline to behavioural flexibility and motivation. Psychopharmacology (Berl) 2018; 235: 2687-2702.

Kan IP, Thompson-Schill SL. Effect of name agreement on prefrontal activity during overt and covert picture naming. Cogn Affect Behav Neurosci 2004; 4: 43-57.

Kane GA, Vazey EM, Wilson RC, Shenhav A, Daw ND, Aston-Jones G*, et al.* Increased locus coeruleus tonic activity causes disengagement from a patch-foraging task. Cogn Affect Behav Neurosci 2017; 17: 1073-1083.

Kaplan R, Friston KJ. Planning and navigation as active inference. Biol Cybern 2018; 112: 323-343.

Kiebel SJ, David O, Friston KJ. Dynamic causal modelling of evoked responses in eeg/meg with lead field parameterization. Neuroimage 2006; 30: 1273-84.

Kiebel SJ, Garrido MI, Moran R, Chen CC, Friston KJ. Dynamic causal modeling for eeg and meg. Hum Brain Mapp 2009; 30: 1866-76.

Kolling N, Behrens T, Wittmann MK, Rushworth M. Multiple signals in anterior cingulate cortex. Curr Opin Neurobiol 2016; 37: 36-43.

Kuffel A, Eikelmann S, Terfehr K, Mau G, Kuehl LK, Otte C*, et al.* Noradrenergic blockade and memory in patients with major depression and healthy participants. Psychoneuroendocrinology 2014; 40: 86-90.

Larsson SC, Traylor M, Malik R, Dichgans M, Burgess S, Markus HS*, et al.* Modifiable pathways in alzheimer's disease: Mendelian randomisation analysis. BMJ 2017; 359: j5375.

Lee HB, Lyketsos CG. Depression in alzheimer's disease: Heterogeneity and related issues. Biol Psychiatry 2003; 54: 353-62.

Lehmann M, Barnes J, Ridgway GR, Ryan NS, Warrington EK, Crutch SJ*, et al.* Global gray matter changes in posterior cortical atrophy: A serial imaging study. Alzheimers Dement 2012; 8: 502-12.

Leonard BE, Myint A. Inflammation and depression: Is there a causal connection with dementia? Neurotox Res 2006; 10: 149-60.

Levy JS. An introduction to prospect theory. Political Psychology 1992: 171-186.

Liang KY, Mintun MA, Fagan AM, Goate AM, Bugg JM, Holtzman DM*, et al.* Exercise and alzheimer's disease biomarkers in cognitively normal older adults. Ann Neurol 2010; 68: 311-8.

Maldjian JA, Laurienti PJ, Burdette JH. Precentral gyrus discrepancy in electronic versions of the talairach atlas. Neuroimage 2004; 21: 450-5.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets. Neuroimage 2003; 19: 1233-9.

Maletic V, Eramo A, Gwin K, Offord SJ, Duffy RA. The role of norepinephrine and its alpha-adrenergic receptors in the pathophysiology and treatment of major depressive disorder and schizophrenia: A systematic review. Front Psychiatry 2017; 8: 42.

Mannari C, Origlia N, Scatena A, Del Debbio A, Catena M, Dell'agnello G*, et al.* Bdnf level in the rat prefrontal cortex increases following chronic but not acute treatment with duloxetine, a dual acting inhibitor of noradrenaline and serotonin re-uptake. Cell Mol Neurobiol 2008; 28: 457-68.

Marks SM, Lockhart SN, Baker SL, Jagust WJ. Tau and beta-amyloid are associated with medial temporal lobe structure, function, and memory encoding in normal aging. J Neurosci 2017; 37: 3192-3201.

Marreiros AC, Daunizeau J, Kiebel SJ, Friston KJ. Population dynamics: Variance and the sigmoid activation function. Neuroimage 2008; 42: 147-57.

Marshall L, Mathys C, Ruge D, De Berker AO, Dayan P, Stephan KE*, et al.* Pharmacological fingerprints of contextual uncertainty. PLoS Biol 2016; 14: e1002575.

Martins CA, Lloyd-Jones TJ. Preserved conceptual priming in alzheimer's disease. Cortex 2006; 42: 995-1004.

Mcclure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. Neuron 2003; 38: 339-46.

Miller SL, Fenstermacher E, Bates J, Blacker D, Sperling RA, Dickerson BC. Hippocampal activation in adults with mild cognitive impairment predicts subsequent cognitive decline. J Neurol Neurosurg Psychiatry 2008; 79: 630-5.

Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The addenbrooke's cognitive examination revised (ace-r): A brief cognitive test battery for dementia screening. Int J Geriatr Psychiatry 2006; 21: 1078-85.

Mirza MB, Adams RA, Friston K, Parr T. Introducing a bayesian model of selective attention based on active inference. Sci Rep 2019; 9: 13915.

Mirza MB, Adams RA, Mathys CD, Friston KJ. Scene construction, visual foraging, and active inference. Front Comput Neurosci 2016; 10: 56.

Miskowiak K, Papadatou-Pastou M, Cowen PJ, Goodwin GM, Norbury R, Harmer CJ. Single dose antidepressant administration modulates the neural processing of self-referent personality trait words. Neuroimage 2007; 37: 904-11.

Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D*, et al.* Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602 2013.

Moran R, Pinotsis DA, Friston K. Neural masses and fields in dynamic causal modeling. Front Comput Neurosci 2013; 7: 57.

Moran RJ, Stephan KE, Dolan RJ, Friston KJ. Consistent spectral predictors for dynamic causal models of steady-state responses. Neuroimage 2011; 55: 1694-708.

Moran RJ, Stephan KE, Seidenbecher T, Pape HC, Dolan RJ, Friston KJ. Dynamic causal models of steady-state responses. Neuroimage 2009; 44: 796-811.

Mortimer JA, Graves AB. Education and other socioeconomic determinants of dementia and alzheimers-disease. Neurology 1993; 43: S39-S44.

Muller TH, Mars RB, Behrens TE, O'reilly JX. Control of entropy in neural models of environmental state. Elife 2019; 8.

Nemoda Z, Angyal N, Tarnok Z, Birkas E, Bognar E, Sasvari-Szekely M*, et al.* Differential genetic effect of the norepinephrine transporter promoter polymorphisms on attention problems in clinical and non-clinical samples. Front Neurosci 2018; 12: 1051.

O'doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 2001; 4: 95-102.

O'doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. Neuron 2003; 38: 329-37.

Paillard T, Rolland Y, De Souto Barreto P. Protective effects of physical exercise in alzheimer's disease and parkinson's disease: A narrative review. J Clin Neurol 2015; 11: 212-9.

Palmqvist S, Scholl M, Strandberg O, Mattsson N, Stomrud E, Zetterberg H*, et al.* Earliest accumulation of beta-amyloid occurs within the default-mode network and concurrently affects brain connectivity. Nat Commun 2017; 8: 1214.

Park H, Melamed D. The effects of stability and presentation order of rewards on justice evaluations. PLoS One 2016; 11: e0168956.

Parr T, Friston KJ. Uncertainty, epistemics and active inference. J R Soc Interface 2017; 14.

Pasquini L, Rahmani F, Maleki-Balajoo S, La Joie R, Zarei M, Sorg C*, et al.* Medial temporal lobe disconnection and hyperexcitability across alzheimer's disease stages. J Alzheimers Dis Rep 2019; 3: 103-112.

Penny W, Iglesias-Fuster J, Quiroz YT, Lopera FJ, Bobes MA. Dynamic causal modeling of preclinical autosomal-dominant alzheimer's disease. J Alzheimers Dis 2018; 65: 697-711.

Rajkowski J, Kubiak P, Aston-Jones G. Locus coeruleus activity in monkey: Phasic and tonic changes are associated with altered vigilance. Brain Res Bull 1994; 35: 607-16.

Rammsayer TH, Hennig J, Haag A, Lange N. Effects of noradrenergic activity on temporal information processing in humans. Q J Exp Psychol B 2001; 54: 247-58.

Ramos BP, Arnsten AF. Adrenergic pharmacology and cognition: Focus on the prefrontal cortex. Pharmacol Ther 2007; 113: 523-36.

Redish AD, Gordon JA 2016. *Computational psychiatry: New perspectives on mental illness*, MIT Press.

Reimer J, Mcginley MJ, Liu Y, Rodenkirch C, Wang Q, Mccormick DA*, et al.* Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nat Commun 2016; 7: 13289.

Rey A, Le Goff K, Abadie M, Courrieu P. The primacy order effect in complex decision making. Psychol Res 2020; 84: 1739-1748.

Robertson IH. A noradrenergic theory of cognitive reserve: Implications for alzheimer's disease. Neurobiol Aging 2013; 34: 298-308.

Robertson IH. Right hemisphere role in cognitive reserve. Neurobiol Aging 2014; 35: 1375-85.

Rolland Y, Pillard F, Klapouszczak A, Reynish E, Thomas D, Andrieu S*, et al.* Exercise program for nursing home residents with alzheimer's disease: A 1-year randomized, controlled trial. J Am Geriatr Soc 2007; 55: 158-65.

Rolls ET. The functions of the orbitofrontal cortex. Neurocase 1999; 5: 301-312.

Rygula R, Walker SC, Clarke HF, Robbins TW, Roberts AC. Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning. J Neurosci 2010; 30: 14552-9.

Sajid N, Ball PJ, Parr T, Friston KJ. Active inference: Demystified and compared. Neural Comput 2021; 33: 674-712.

Sales AC, Friston KJ, Jones MW, Pickering AE, Moran RJ. Locus coeruleus tracking of prediction errors optimises cognitive flexibility: An active inference model. PLoS Comput Biol 2019; 15: e1006267.

Sara SJ. The locus coeruleus and noradrenergic modulation of cognition. Nat Rev Neurosci 2009; 10: 211-23.

Scahill RI, Schott JM, Stevens JM, Rossor MN, Fox NC. Mapping the evolution of regional atrophy in alzheimer's disease: Unbiased analysis of fluid-registered serial mri. Proc Natl Acad Sci U S A 2002; 99: 4703-7.

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997; 275: 1593-9.

Schwartenbeck P, Friston K. Computational phenotyping in psychiatry: A worked example. eNeuro 2016; 3.

Sharp ES, Gatz M. Relationship between education and dementia: An updated systematic review. Alzheimer Dis Assoc Disord 2011; 25: 289-304.

Shenhav A, Botvinick MM, Cohen JD. The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron 2013; 79: 217-40.

Sigurdardottir HL, Kranz GS, Rami-Mark C, James GM, Vanicek T, Gryglewski G*, et al.* Effects of norepinephrine transporter gene variants on net binding in adhd and healthy controls investigated by pet. Hum Brain Mapp 2016; 37: 884-95.

Sperling RA, Laviolette PS, O'keefe K, O'brien J, Rentz DM, Pihlajamaki M*, et al.* Amyloid deposition is associated with impaired default network function in older persons without dementia. Neuron 2009; 63: 178-88.

Starkstein SE, Mizrahi R, Power BD. Depression in alzheimer's disease: Phenomenology, clinical correlates and treatment. Int Rev Psychiatry 2008; 20: 382-8.

Stein M, Federspiel A, Koenig T, Wirth M, Strik W, Wiest R*, et al.* Structural plasticity in the language system related to increased second language proficiency. Cortex 2012; 48: 458-65.

Stephan KE. Bayesian inference and bayesian model selection. Methods & Models for fMRI data analysis 2017.

Stephan KE, Penny WD, Moran RJ, Den Ouden HE, Daunizeau J, Friston KJ. Ten simple rules for dynamic causal modeling. Neuroimage 2010; 49: 3099-109.

Stern Y. Cognitive reserve in ageing and alzheimer's disease. Lancet Neurol 2012; 11: 1006-12.

Subramaniapillai M, Mansur RB, Zuckerman H, Park C, Lee Y, Iacobucci M*, et al.* Association between cognitive function and performance on effort based decision making in patients with major depressive disorder treated with vortioxetine. Compr Psychiatry 2019; 94: 152113.

Sutton RS, Barto AG 2018. *Reinforcement learning: An introduction*, MIT press.

Szot P, White SS, Greenup JL, Leverenz JB, Peskind ER, Raskind MA. Compensatory changes in the noradrenergic nervous system in the locus ceruleus and hippocampus of postmortem subjects with alzheimer's disease and dementia with lewy bodies. J Neurosci 2006; 26: 467-78.

Thompson PM, Mega MS, Woods RP, Zoumalan CI, Lindshield CJ, Blanton RE*, et al.* Cortical change in alzheimer's disease detected with a disease-specific population-based brain atlas. Cereb Cortex 2001; 11: 1-16.

Thorndike EL 1912. Animal intelligence. Experimental studies. LWW.

Tyrer A, Gilbert JR, Adams S, Stiles AB, Bankole AO, Gilchrist ID*, et al.* Lateralized memory circuit dropout in alzheimer's disease patients. Brain Commun 2020; 2: fcaa212.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N*, et al.* Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage 2002; 15: 273-89.

Uddin LQ. Complex relationships between structural and functional brain connectivity. Trends Cogn Sci 2013; 17: 600-2.

Valentin VV, Dickinson A, O'doherty JP. Determining the neural substrates of goal-directed learning in the human brain. J Neurosci 2007; 27: 4019-26.

Varazzani C, San-Galli A, Gilardeau S, Bouret S. Noradrenaline and dopamine neurons in the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. J Neurosci 2015; 35: 7866-77.

Vidaurre D, Smith SM, Woolrich MW. Brain network dynamics are hierarchically organized in time. Proc Natl Acad Sci U S A 2017; 114: 12827-12832.

Wang M, Gamo NJ, Yang Y, Jin LE, Wang XJ, Laubach M*, et al.* Neuronal basis of age-related working memory decline. Nature 2011; 476: 210-3.

Wang S, Shi Y, Li BM. Neural representation of cost-benefit selections in rat anterior cingulate cortex in self-paced decision making. Neurobiol Learn Mem 2017; 139: 1-10.

Wang WC, Ranganath C, Yonelinas AP. Activity reductions in perirhinal cortex predict conceptual priming and familiarity-based recognition. Neuropsychologia 2014; 52: 19-26.

Watkins CJ, Dayan P. Q-learning. Machine learning 1992; 8: 279-292.

Watkins CJCH. Learning from delayed rewards. 1989.

Weise CM, Chen K, Chen Y, Kuang X, Savage CR, Reiman EM*, et al.* Left lateralized cerebral glucose metabolism declines in amyloid-beta positive persons with mild cognitive impairment. Neuroimage Clin 2018; 20: 286-296.

Williams CL, Tappen RM. Exercise training for depressed older adults with alzheimer's disease. Aging Ment Health 2008; 12: 72-80.

Winder-Rhodes SE, Chamberlain SR, Idris MI, Robbins TW, Sahakian BJ, Muller U. Effects of modafinil and prazosin on cognitive and physiological functions in healthy volunteers. J Psychopharmacol 2010; 24: 1649-57.

Wuwongse S, Chang RC, Law AC. The putative neurodegenerative links between depression and alzheimer's disease. Prog Neurobiol 2010; 91: 362-75.

Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology 1908; 18: 459-482.

Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. Neuron 2005; 46: 681-92.

Zald DH, Pardo JV. Emotion, olfaction, and the human amygdala: Amygdala activation during aversive olfactory stimulation. Proc Natl Acad Sci U S A 1997; 94: 4119-24.

Zhang Y, Li P, Feng J, Wu M. Dysfunction of nmda receptors in alzheimer's disease. Neurol Sci 2016; 37: 1039-47.

Zhuge R, Li S, Chen TH, Hsu WH. Alpha2-adrenergic receptor-mediated ca2+ influx and release in porcine myometrial cells. Biol Reprod 1997; 56: 1343-50.

Chapter Eight

Appendix

8.1 Active Inference: States, Observations, Actions and Matrices

States

1 = At location 1; forest/desert. Can move left or right

2 = At location 2; forest. Can move left or right

3 = At location 3; lakeside. Can move left or right

4 = At location 4; dunes/mountains. Can move left or right

5 = At location 5; dunes/ocean. Can move left or right

6 = At location 6, 9, 11, 16 or 19; pink gem. 10p reward, no actions

7 = At location 13, 14 or 21; gold gem. 25p reward, no actions

8 = At location 7, 8, 10, 12, 15, 17, 18 or 20; silver gem. 0p reward, no actions

Observations

One-to-one relationship between states and observations:

1 = Forest/desert, first location visited by agent. Requires a decision/action; left or right

2 = Forest, second location visited by agent. Requires an action; left or right

3 = Lakeside, second location visited by agent. Requires an action; left or right

4 = Dunes/mountains, second location visited by agent. Requires an action; left or right

5 = Dunes/ocean, second location visited by agent. Requires an action; left or right

6 = Third location visited by agent. 10p reward, pink gem. No actions required.

7 = Third location visited by agent. 25p reward, gold gem. No actions required.

8 = Third location visited by agent. 0p reward, silver gem. No actions required.

Actions

1 = left

2 = right

A **Matrix**

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

B **Matrices**

$$\boldsymbol{B}\{\boldsymbol{1}\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.7 & 0.3 & 0 & 0.3 & 1.0 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0 & 0 & 1.0 & 0 \\ 0 & 0.3 & 0.7 & 0.3 & 0.7 & 0 & 0 & 1.0 \end{pmatrix}$$

$$\boldsymbol{B}\{\boldsymbol{2}\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0.7 & 0 & 1.0 & 0 & 0 \\ 0 & 0 & 0.3 & 0 & 0.3 & 0 & 1.0 & 0 \\ 0 & 0.7 & 0.7 & 0.3 & 0.7 & 0 & 0 & 1.0 \end{pmatrix}$$

Preferences

$$\boldsymbol{C} = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0.4c \quad c \quad -0.5c)$$

Prior Beliefs About Initial State

$$\boldsymbol{D} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Policies

$$\boldsymbol{V} = \begin{array}{cc} \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 2 \end{pmatrix} & \begin{array}{l} t = 1 \\ t = 2 \end{array} \end{array}$$

B matrices post-reversal: Pilot Three

$$\boldsymbol{B}\{\boldsymbol{1}\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 & 0 & 1.0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0.3 & 0.3 & 0.7 & 1.0 & 0 & 0 & 1.0 \end{pmatrix}$$

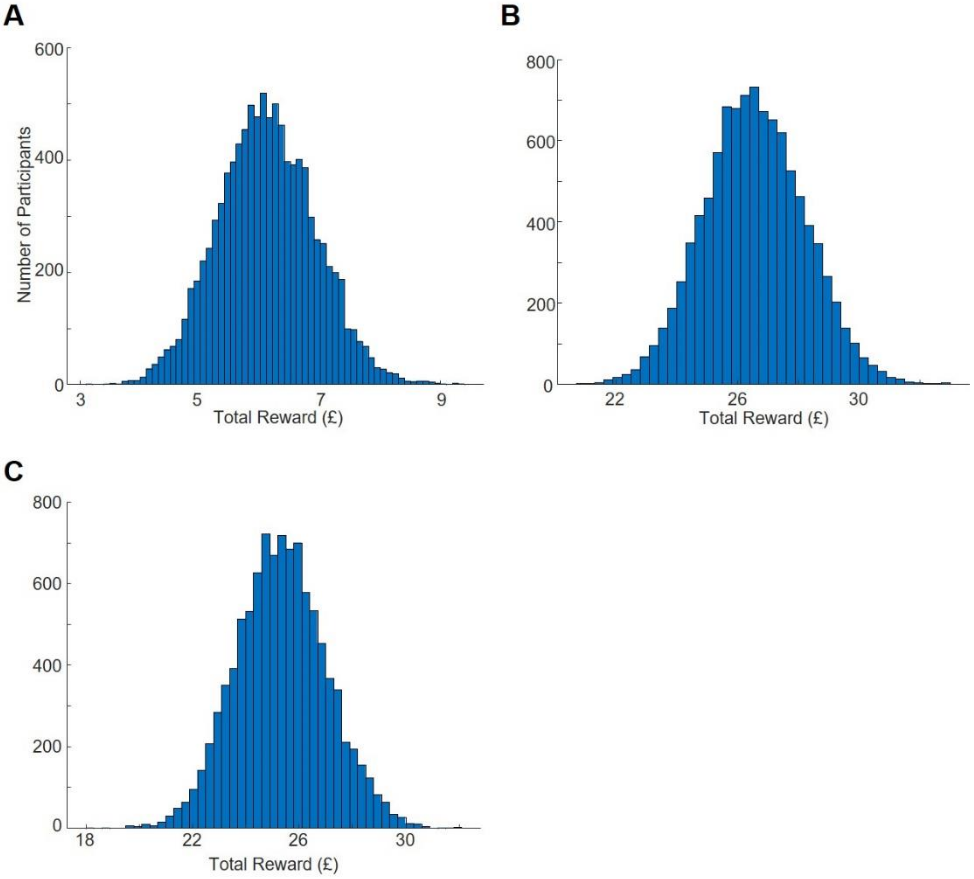
$$\boldsymbol{B}\{\boldsymbol{2}\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.7 & 0.3 & 0.7 & 0 & 1.0 & 0 & 0 \\ 0 & 0.3 & 0.7 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 0.3 & 1.0 & 0 & 0 & 1.0 \end{pmatrix}$$

B matrices post-reversal: Main Behavioural Study

$$\boldsymbol{B}\{\boldsymbol{1}\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 & 0.3 & 1.0 & 0 & 0 \\ 0 & 0.7 & 0 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0.3 & 0.3 & 0.7 & 0.7 & 0 & 0 & 1.0 \end{pmatrix}$$

$$\boldsymbol{B}\{\mathbf{2}\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0.3 & 1.0 & 0 & 0 \\ 0 & 0.3 & 0.7 & 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0.7 & 0.3 & 0.3 & 0.7 & 0 & 0 & 1.0 \end{pmatrix}$$

Appendix Figure A.1



Appendix Figure A.1 (**A**) Histogram of total reward distribution obtained from a Monte-Carlo simulation of 10,000 randomly behaving agents using the testing session task structure of pilots one and two, with 80 trials. Minimal learning threshold (95th percentile): £7.40; mean: £6.08; median: £6.05. (**B**) Histogram of total reward distribution obtained from a Monte-Carlo simulation of 10,000 randomly behaving agents using the task structure of pilot three, with 200 pre-reversal trials and 120 post-reversal trials. Minimal learning threshold: £29.25; mean: £26.53; median: £26.50. (**C**) Histogram of total reward distribution obtained from a Monte-Carlo simulation of 10,000 randomly behaving agents using the task structure of the main study, with 200 pre-reversal trials and 120 post-reversal trials. Minimal learning threshold: £28.20; mean: £25.32; median: £25.30.