Motivation and Reward-Seeking in Dorsal Raphe Nucleus



Luke Priestley

Brasenose College

University of Oxford

A dissertation in partial fulfilment of the requirements for the degree

Doctor of Philosophy

Trinity 2024

Motivation and reward seeking in dorsal raphe nucleus

Abstract

In this thesis, I investigate the role of the dorsal raphe nucleus (DRN) in reward-guided behaviour. DRN is an ancient part of the brainstem and an important source of serotonin in mammals, but fundamental aspects of its function remain elusive. Here, I present a combination of human and non-human primate experiments to argue that DRN controls changes in reward-seeking. In Ch.III, I use recently developed analysis techniques to show that monkeys exhibit motivation-states during a foraging-style behavioural task. These states shaped their baseline predisposition to pursue rewards, and were linked to the richness of the environment: a high-motivation state was occupied when the environment was rich with many high-value opportunities, and a low-motivation state was occupied when the environment was poor and low-value opportunities abounded. In Ch.IV, I show that distinctive patterns of DRN activity occurred when animals transitioned between these states. I then show that disrupting DRN with transcranial ultrasound stimulation reduced the frequency of motivation-state transitions and prevented animals from reconciling their motivation-state with the richness of the environment. Finally, in Ch.V I demonstrate that DRN has a similar role in the human brain. I show that human behaviour during a foraging-style task is characterised by temporally persistent behavioural policies, and that these policies were set according to the richness of the environment. DRN activity signalled changes in behavioural policy, but only for changes that were adaptive given the richness of the environment. Taken together, these experiments suggest that DRN has a conserved and causal role in reconciling general patterns of reward-seeking with general features of the environment.

Luke Priestley

Dissertation submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Experimental Psychology at the University of Oxford

Word count: 45.000

Brasenose College

Trinity Term 2024

Acknowledgements

First and foremost, I would like to thank my supervisors, Nima Khalighinejad and Matthew Rushworth. I will be forever grateful for their kind, patient and percipient guidance on all manner of things scientific and personal over the last four years. The fact that they coaxed an empirical thesis from someone so temperamentally philosophical is a sure testament to their pedagogical abilities!

I would also like to thank all the other people who have been part of my scientific life at Oxford. In particular, I would like to thank Ali Mahmoodi for insightful discussions about the experiments in Ch.III–V and beyond – there was never a dull moment in his firebrand presence. I would also like to thank William Reith for his meticulous contributions to the experiment in Ch.V. In relation to Ch.III–IV, I would like thank Mark Chiew and Mo Shahdloo for their assistance with the fMRI recording and reconstruction protocols, and Xinghao Cheng and Robin Cleveland for their help with planning and validating the ultrasound brain stimulation experiment. I would additionally like to thank all the other members of the Rushworth and Klein-Flugge groups who inhabited the Tinsley building throughout my tenure – I am fortunate to have spent my time with such genial and passionate people.

I would like to acknowledge the four macaque monkeys who participated in the experiments reported in Ch.III–IV: Ultra, Ulrich, Vampire and Winky. They are complex and sensitive creatures, and I did my best to treat them with the respect that they deserve. I will not forget them. I would also like to thank all the staff at the Biomedical Services Building involved in the care, welfare and training of the monkeys – in particular, Urs Schüffelgen and Andy Emberton, for their help with data collection.

I would like to thank the Clarendon Fund, Brasenose College, and the Oxford-Australia Fund, who generously supported my studies.

Beyond the scientific, I would like to thank the people who have been part of my life in the UK; Ella and Ashna for their love and support; Clemence, Ellie, Fionn, Michał and Freddie for their friendship.

Finally, I would above all like to thank my family: Bella, Anna & Mark; Nancy & Harry; Iris & John; Fred. I have enjoyed the rare privilege of pursuing ends of my own design, not just in the last four years but throughout my life in general. None of it would have been possible without them.

This work is dedicated to my family:

Bella, Anna & Mark; Nancy & Harry; Iris & John.

Contents

List of figures **x**

Chapter I: Stage-setting **1**

Chapter II: Action, reward and the brain **3**

2.1 What makes a reward worth pursuing? 4

2.2 Cortico-subcortical circuits for reward-guided behaviour 8

2.2.1 Reward evaluation in dorsal anterior cingulate cortex and anterior insular cortex *9*

2.2.2 Action initiation in the basal ganglia *11*

2.3 Reward-related function of the ascending neuromodulatory systems 12

2.3.1 What is neuromodulation, and why is it important? *12*

2.3.2 Anatomy of the ascending neuromodulatory pathways *14*

2.3.3 Substantia nigra and ventral tegmental area *17*

2.4.3.1: The serotonin-dopamine opponency hypothesis *19*

2.4.3.2 How is dorsal raphe nucleus neural activity related to rewards? *20*

2.4.3.3 Dorsal raphe nucleus and behavioural change *22*

2.3.5: Habenula *24*

2.4 The present work 25

Chapter III: A Hidden Markov Model General Linear Model identifies motivation-states in macaque monkeys **29**

3.1: Introduction 29

3.2: Results 31

3.2.1 Animals are more motivated to pursue rewards in rich environments *31*

3.2.2: A Hidden Markov Model identifies motivation-states in behaviour *37*

3.3: Discussion 49

3.4: Materials and Methods 53

3.4.1 Experimental model and subject details *53*

3.4.2 Behavioural training *53*

3.4.3 Behavioural task *54*

3.4.4 Behavioural analysis *57*

3.4.5 General Linear Model Hidden Markov Model *60*

Chapter IV: Dorsal raphe nucleus controls motivation-state transitions in monkeys 65

4.1 Introduction 65

4.2 Results 66

4.2.1 Brain activity in DRN signals the richness of the environment and changes in motivation state *66*

4.2.2 Non-invasive disruption of DRN perturbs motivation-state transitions *78*

4.2.3 A cortico-subcortical circuit for reconciling behaviour with the environment *86*

4.3 Discussion 92

4.4 Methods 95

4.4.1 Experimental model and subject details *95*

4.4.2 Behavioural training *95*

4.4.3 Behavioural task and analysis of behavioural data *95*

4.4.4 Acquisition, reconstruction and pre-processing of MRI data *96*

4.4.5 fMRI data analysis *97*

4.4.6 Transcranial ultrasound stimulation (TUS) *102*

4.4.7 Acoustic modelling *103*

4.4.8 Resting-state imaging data acquisition, pre-processing, and analysis *104*

4.4.9 TUS data analysis *105*

Chapter V: Dorsal raphe nucleus signals behavioural policy changes in humans 107

5.1 Introduction 107

5.2 Results 109

5.2.1 A behavioural task for manipulating the richness of the environment *109*

5.2.2 Reward-guided decisions are modulated by the richness of the environment *113*

5.2.3 Behaviour is organised in option-specific policies that are reconciled with the richness of the environment *117*

5.2.4 Brain activity in dorsal raphe nucleus represents the richness of the environment, and environment-driven changes in behaviour *121*

5.2.5 Multivariate representations of value in frontal cortex are modulated by the richness of the environment *128*

5.3 Discussion 136

5.4 Methods 140

5.4.1 Participants *140*

5.4.2 Behavioural task *140*

5.4.3 Behavioural analysis *143*

5.4.4 Imaging data acquisition *148*

5.4.5 fMRI data preprocessing *149*

5.4.6 Whole-brain fMRI data analysis *150*

5.4.7 ROI timecourse fMRI analysis *152*

5.4.8 Representational similarity analysis *155*

5.4.9 Analysis of midbrain dopaminergic function *159*

Chapter VI: General discussion **161**

Appendices **167**

Table 1: fMRI cluster locations from GLM5.9 167

Table 2: fMRI cluster locations from GLM5.10 168

References **169**

List of figures

3.1	A behavioural task for controlling the richness and stochasticity of the environment	31
3.2	Animals are more likely to pursue reward opportunities in rich environments.	32
3.3	A General Linear Model Hidden Markov Model for explaining temporal patterns in behaviour.	41
3.4	Validating and implementing a GLM–HMM fitting procedure using 5-fold cross validation.	42
3.5	A 2-state GLM–HMM captures key quantitative and qualitative patterns of animal behaviour.	44
3.6	GLM–HMM states and state-transitions are not reducible to fatigue, satiety or time-on-task.	46
3.7	GLM–HMM shows distinct, persistent, and biologically meaningful motivation-states during behaviour.	47
4.1	Diagram of fMRI scanning apparatus.	70
4.2	Cortical and subcortical regions of interest.	71
4.3	Brain activity in dorsal raphe nucleus represents the richness of the environment and changes in motivation-states.	73
4.4	No evidence of environment-stochasticity representations in ROIs.	76
4.5	Further examination of motivation-state transitions in dorsal raphe nucleus	77
4.6	Non-invasive disruption of dorsal raphe nucleus shows causal role motivation-state transitions	82
4.7	Simulation of the propagation of acoustic waves produced by the TUS protocol targeting DRN and VTA.	84
4.8	Anterior insula represents the richness of the environment with the same pattern as subcortical ROIs.	88
4.9	Habenula and anterior insula encode pursue-vs-reject decisions.	89
4.10	A circuit for reconciling motivation with the environment.	90
4.11	Dorsal raphe nucleus controls motivation-state transitions as a function of the environment.	91
5.1	A behavioural task for controlling the richness of the environment.	112
5.2	Pursue-vs-reject decisions are modulated by the richness of the environment.	116
5.3	Participants exhibit consistent behavioural policies that are shaped by the richness of the environment.	120
5.4	Brain activity in dorsal raphe nucleus represents the richness of the environment and environment-driven changes in behavioural policy.	126
5.5	Activity in the midbrain dopaminergic nuclei represents pursue-vs-reject decisions, and the value-difference between current and recent reward opportunities	128
5.6	Whole-brain fMRI analysis identifies regions of interest for representational similarity analysis.	132
5.7	Multivariate representations of value are modulated by the environment in dorsal anterior cingulate cortex and anterior insular cortex.	133
5.8	Univariate activity in dorsal anterior cingulate cortex and anterior insular cortex does not represent policy-switches.	135
5.9	A cortico-subcortical circuit for reconciling behavioural policies with the environment	136

Chapter I

Stage-setting

What makes something worth doing? According to the philosopher Elizabeth Anscombe, the answer is simple: we do things to get what we want. “The primitive sign of wanting,” she writes, “is trying to get”, and so the teleological purpose of action is to help us achieve our goals and ends (*1*).

In this thesis, I examine the mechanisms of reward-guided action in the brain. Anscombe’s truths are simple, but they are useful starting points, even for a neuroscientist. Sensorimotor control is a basic feature of the nervous system: it allows organisms to identify and act upon opportunities for reward in the world around them, and in complex animals takes the form of decision-making circuits that provide sophisticated voluntary control of behaviour (*2*, *3*). Cognitive and systems neuroscientists have characterised aspects of these circuits in impressive detail, including how movements are organised into programmatic patterns of action, and how different courses of action are evaluated for prospective rewards. However, despite their phylogenetic link, voluntary action and reward-guided decision-making have been studied largely independently (§2.1). This leaves open the most pressing questions raised by Anscombe’s account: how the brain decides which rewards are worth wanting, and which rewards it should try to get.

In this thesis, I attempt to bridge this gap. I report the results of two experiments examining how the brain decides *whether* to engage in reward-guided behaviour using whole-brain imaging techniques capable of characterising many regions at once. I use these recordings to highlight the special role of a specific nucleus in the brainstem: dorsal raphe nucleus. The raphe nuclei are distinguished as conserved sources of serotonin in the mammalian, avian,

reptilian, and piscine brain, suggesting that they originated in an ancestral Euteleostomi animal approximately 400 million years ago (*4–8*). Here, I show that dorsal raphe nucleus in human and non-human primates exhibits two interrelated patterns of activity: (i) a pattern for the general value or 'richness' of the environment as a whole, as distinct from specific sources of reward, and (ii) a pattern for behavioural changes that follow from the richness of the environment, which are shaped by its ebbs and flows. In this way, I argue that dorsal raphe nucleus has a fundamental role in reconciling behaviour with the surrounding world, in keeping with its place in the shared history of vertebrate life.

Chapter II

Action, reward and the brain

In this chapter, I provide the conceptual and empirical background to the experiments reported in Ch.III–V. In §2.1, I briefly introduce historical perspectives on voluntary action and reward-guided decision-making in cognitive neuroscience, and argue that their shortcomings are partly resolved in foraging theory – a branch of behavioural ecology related to food-seeking in animals. I then briefly review patterns of human and non-human primate decision-making that are consistent with foraging theory's precepts, and introduce the concept of behavioural states as an organising principle for natural behaviour.

In §2.2, I review the implementation of reward-guided action and motivation in the brain. To begin, I briefly describe the cortico-basal-ganglia circuit that has been the focus of many experiments. I then move to the ascending neuromodulatory systems. I describe the basic points and purposes of neuromodulation (§2.3.1) and the anatomy of the neuromodulatory pathways in mammals (§2.3.2), before undertaking a detailed review of the midbrain dopaminergic nuclei and their longstanding links with action and reward (§2.3.3). I subsequently turn to the dorsal raphe nucleus (DRN) – the focus of this thesis (§2.3.4). I argue that although DRN has proven recalcitrant to general theories of function, empirical studies of its activity during reward-guided behaviour cohere into two major themes: (1) DRN tracks the statistics of reward outcomes over extended time-horizons, akin to a recency-weighted distribution, and; (2) DRN activity is involved in changes between opposing modes of reward-guided behaviour, like patience and impulsivity, and exploitation and exploration. I end by sketching a hypothesis for DRN function that integrates these themes, and which is tested in this thesis. Finally, I briefly review the function of the habenula (Hb) in the epithalamus. I pay

particular attention to Hb's interconnections with the forebrain, midbrain and brainstem, and argue that it is an important source of reward information to DRN and other subcortical nuclei.

In §2.4, I introduce the experimental work. I briefly explain the key methodological choices and techniques: recording of brain activity with functional magnetic resonance imaging, disruption of brain activity with transcranial ultrasound stimulation, and the use of a non-human primate model alongside human volunteers. I then prefigure the experiments reported in Ch.III–V and link them to the matters raised in the introduction.

2.1 What makes a reward worth pursuing?

This thesis examines how the brain determines *whether* an opportunity for reward is worth pursuing. It is related to a longstanding distinction in cognitive neuroscience between two types of action (*9*, *10*): *exogenous* actions driven by stimuli external to an organism, and *endogenous* or *voluntary* actions generated by processes internal to an organism. Although few real-world behaviours cleave into these categories without proviso, the distinction highlights that complex organisms have control of their own movements and are not simply subject to input from surrounding environment (*3*). These abilities – to decide between different courses of action, and to decide whether action is worthwhile – underpin a variety of goal-directed behaviours that are critical for survival.

Given the connection between voluntary action and reward-guided behaviour, it is perhaps surprising that they are rarely linked in experiments. The scientific study of voluntary action evolved out of human volunteers executing "spontaneous" movements, which explicitly served no purpose at all (*11*). At the other end of the spectrum, reward-guided decision-making studies have historically drawn inspiration from theories of rational choice developed in neoclassical economics (*12*). These experiments have made important contributions to

uncovering how human and non-human animals trade-off parameters like the identity, magnitude, probability, and timing of rewards in their behaviour (*13*). However, in the paradigm case, they involve forced choices between a range of simultaneously available options, akin to the way that consumers decide between economic goods (*14*, *15*). This obviates basic questions about reward-guided behaviour, like whether the participant wants either option, or whether they want to engage in reward-seeking at all.

These lacunae are elegantly resolved in a scenario called *prey selection* set forth in foraging theory – a branch of behavioural ecology concerned with food-seeking choices in animals (*16*, *17*). During prey selection, a predator undergoes sequential encounters with prey (i.e. reward) opportunities and must decide, upon each encounter, whether to seize the presently available opportunity or search the environment for better alternatives. In contrast to economic decision-making paradigms, animals in prey selection do not have to act upon every opportunity they receive. The question at stake, thus, is *whether* an opportunity is worth pursuing. The behavioural experiments in this thesis follow a prey-selection-inspired structure to probe the neural mechanisms of this process.

What makes an opportunity worth engaging with? It is trivial that animals are more likely to pursue opportunities as a function of their anticipated reward-value. However, theories of optimal foraging emphasise that a range of broader contextual factors should also be taken into account (*16*, *17*). Many of these factors relate to the biological constraints that encumber animals, which mean that behaviour must carefully balance the reward value of food with the costs of obtaining it (*18*). Although planning and executing reward pursuit in the wild has many species-specific costs, the cost that features most prominently in theories is common to all creatures – *time*. Normative models of prey selection therefore conceive prey selection as a comparison between (i) the reward value of an opportunity in the foreground, given the time needed to obtain it, with (ii) the time-averaged reward-value of the environment, which

represents what an animal can expect to encounter if equivalent time was spent searching for alternatives. According to this view, the relative richness or sparseness of an animal's broader environment should therefore fundamentally change its propensity for reward-pursuit: in rich environments with many high-value opportunities, animals should be selective in their choices, whereas in poor environments with sparsely distributed rewards they should settle for meagre opportunities that might ordinarily be ignored.

The hypothesis that animals track the richness of the environment is consistent with natural behaviour in many species across both vertebrate and invertebrate phyla (*16*, *18–23*). It has also been useful for explaining behaviour in laboratory-based tasks (*24–27*). For example, Constantino and Daw examined human behaviour in foraging-inspired "patch-leaving" task, where participants collected stochastic rewards in patches that depleted in value over time (*28*). The critical decision was *when* to leave a patch and begin foraging at a new one, and the normative strategy involved pinpointing when the reward-rate in the present patch ebbed below the average patch in the environment. A quantitative model based on this strategy closely approximated human behaviour and – importantly – outperformed well-established algorithms for reward-guided learning and decision-making like temporal difference reinforcement learning. Macaque monkeys performing comparable tasks show the same pattern of sensitivity to reward-rates (*24*, *29*). Indeed, reframing classic reward-guided tasks in foraging-style scenarios ameliorates some of the puzzling biases that animals typically show in these situations (*30*). Taken together, these results suggest that the richness of the environment is a fundamental and conserved influence on reward-guided behaviour across species. In this thesis, I examine how this influence is implemented in the primate brain using behavioural paradigms that feature carefully controlled changes in the richness of the environment.

The final behavioural concept that is important for this thesis concerns *behavioural-states*: temporally extended changes in the way an animal attends to, processes, and acts upon

external inputs from the surrounding world (*31–33*). Behavioural states are elusive phenomena precisely because they involve changes within an animal that affect how it responds to a given external stimulus. Notwithstanding basic survival states like fear-induced avoidance (*32*, *34*), it is therefore difficult to experimentally evoke or manipulate behavioural states, and they instead need to be inferred retrospectively based on patterns of behaviour over time. New techniques for behavioural analysis are beginning to make this feasible, however (*35*, *36*). For example, Ashwood and colleagues report a General Linear Model Hidden Markov Model (GLM–HMM) framework which shows that mice performing perceptual orientation discrimination tasks occupy discrete states of task engagement: (i) an engaged state, in which orientation decisions are strongly weighted to sensory evidence from visual stimuli, and; (ii) a series of disengaged states, in which decisions are loosely related to sensory input and instead arise from biases that cause habitual patterns of responding (*35*). The basic features of engagement states are visible in many perceptual decision-making studies, where errors or 'lapses' tend to cluster together in time (*37–39*). The advantage of the GLM-HMM framework, however, is the ability to quantitatively delineate these periods from one another, and precisely identify their transition points.

Behavioural states like engagement and disengagement likely arise from the same biological constraints emphasised in foraging and behavioural ecology: perceptual decision-making tasks are difficult – especially for mice – and the fact that mice perform the task negligently for long periods of time is partly because they need to recuperate the resources depleted during accurate, engaged performance. In this way, although behavioural states are not formally included in foraging theory, foraging theory nevertheless makes implicit predictions about state-like fluctuations in behaviour. In the case of reward seeking, insofar as organisms aim to maximise their reward intake net of biological costs, an animal's baseline propensity to pursue reward opportunities should vary according to the richness of the

surrounding environment (see above). In Ch.III, we adapt the GLM–HMM framework to study time-varying motivation-states in macaque monkeys and demonstrate that they are indeed shaped by the richness of the environment.

In the next section, I review the neural mechanisms of reward-guided action. Although the experiments in this thesis focus on dorsal raphe nucleus, I begin with a broad overview of the cortico-subcortical circuits implicated in this process. I then describe the basic principles of neuromodulation, and the anatomy of the major ascending neuromodulatory pathways. Finally, I provide a detailed review of the substantia nigra and ventral tegmental area, and dorsal raphe nucleus, which are the neuromodulatory nuclei with the strongest links to reward-guided behaviour.

2.2 Cortico-subcortical circuits for reward-guided behaviour

Historically, research on the neural mechanisms of reward-guided behaviour has focussed on an intricate series of cortico-subcortical loops involving interactions between the cortex and the basal ganglia (*40–43*). The function of the cortex, in this circuit, is to generate and evaluate possible behaviours given an animal's goals and environment. The basal ganglia then translate these plans into organised patterns of movement (*42*). Below, I elaborate on salient aspects of this circuit, including two cortical regions implicated in evaluating *whether* rewards are worth pursuing: dorsal anterior cingulate cortex, and anterior insular cortex.

2.2.1 Reward evaluation in dorsal anterior cingulate cortex and anterior insular cortex

Many areas of the cortex work in concert during reward-guided behaviour to evaluate courses of action and their reward consequences. Here, I review the contributions of dorsal anterior cingulate cortex and anterior insular cortex because they are specifically involved in the aspects of behaviour raised in §2.1.

Dorsal anterior cingulate cortex (dACC) is the rostral and agranular portion of the cingulate cortex comprising Brodmann areas 24 and 32 in the human brain, and area 24 in the macaque monkey brain (*44–48*). Its anatomical connections are marked by prominent links to: (i) cortical and subcortical regions linked to reward, including the ventromedial prefrontal cortex, insular cortex, amygdala, hypothalamus and striatum, and; (ii) regions involved in motor control, including the motor cortex, premotor cortex and even direct projections to the spinal cord (*49–52*). dACC is thus ideally placed to synthesise information about action and reward (*50*, *53*, *54*).

Recording and lesion experiments in many species indicate that this is an important aspect of dACC function (*55*, *56*): for example, human and non-human animals with dACC lesions are unable to use action-outcome relationships to guide their choices (*54*, *57–61*), and are seemingly unable to grasp the reward value of action when physical effort is required (*62*, *63*). In addition to this, a variety of experimental paradigms show that dACC activity encodes higher-order aspects of reward-guided tasks, like counterfactual outcomes in multi-option choices, and estimates of the global reward-rate in foraging-style behaviours (*26*, *29*, *64–66*). This suggests that dACC is fundamental not just for linking actions and rewards, but also for

appraising the value of specific reward opportunities *relative* to other sources of reward in the environment.

In contrast to the medial location of dACC, insular cortex is located deep in the lateral sulcus of the cortex between the frontal and temporal lobes (*44*, *67*). In the human brain it is divided by the central insular sulcus into an agranular/dysgranular anterior lobule, and a granular posterior lobule, which have different anatomical connections and functional properties (*67*). Macaque insular cortex is comparatively lissencephalic which makes demarcating the anterior and posterior aspects difficult, however there are emergent gyri that approximate agranular-to-dysgranular, and dysgranular-to-granular tissue transitions along the anterior-posterior axis (*67–69*). Anterior insular (AI) is marked by prominent connections with (i) reward-related cortical and subcortical regions including the orbitofrontal cortex, cingulate cortex, amygdala and ventral striatum; (ii) motor regions in the medial frontal cortex and subcortical visceromotor regions including the lateral hypothalamus and periaqueductal grey (PAG), and; (iii) strong intra-regional connections from the posterior insular lobule, which convey internal physiological signals from the body (*70–72*, *72–79*).

AI's function is poorly understood relative to many other cortical regions because it is difficult to reach with invasive recording and lesion techniques. Early experiments involving electrical stimulation of insular neurons indicated that it controls visceromotor responses in the face and upper body, in addition to sympathetic and parasympathetic functions like cardiac activity and respiration (*79–81*). In combination with AI's anatomical connections, this led to the hypothesis that AI facilitates *interoception*, or awareness and regulation of internal physiological states (*82*, *83*). Subsequent study of AI with task-based fMRI has reinforced its links with interoception and implicated it in a so-called salience network for identifying and responding to homeostatically relevant stimuli like rewards (*84–87*). In this way, AI and dACC might evaluate reward opportunities with complementary frames of reference: dACC with an

external frame that contrasts reward opportunities to their external context, and AI with an *internal* frame that evaluates opportunities given the state of the body.

2.2.2 Action initiation in the basal ganglia

In contrast to the cortex, the fundamental role of the basal ganglia is movement (*41*, *42*). A rich vein of inputs conveying an animal's external context, internal state and plans for action are sent from the cortex to the striatum – the basal ganglia's main input nucleus (*41*). The striatum serves to organise these inputs, initially by inhibiting the motor and brainstem areas responsible for motor production so that competing possibilities for behaviour can be evaluated without generating disorganised patterns of movement (*88*). Action initiation requires suprathreshold activity in the striatum to overcome this inhibition. The ascending neuromodulatory systems have a critical role in this process, both by controlling neural dynamics in the striatum itself, and by modulating the cortical inputs that drive striatal activity to begin with (*43*).

I review the ascending neuromodulatory systems in the next section. I begin by describing the basic principles of neuromodulation, and the anatomy of the major neuromodulatory pathways. I then provide a detailed review of the substantia nigra and ventral tegmental area, the dorsal raphe nucleus, and the habenula, with an emphasis on the reward-related signals they are purported to transmit.

2.3 Reward-related function of the ascending neuromodulatory systems

The ascending neuromodulatory systems (ANS) are an assemblage of phylogenetically ancient nuclei which are widely conserved throughout all vertebrate subphyla (*89*, *90*). The ANS are so-named because they control the release of fundamental neurotransmitters: the indolamine serotonin [5-hydroxytryptamine; 5-HT], the catecholamines dopamine [4-(2-Aminoethyl)benzene-1,2-diol] and norepinephrine [4-(2-Amino-1-hydroxyethyl)benzene-1,2-diol], and the ester acetylcholine (2-Acetoxy-N,N,N-trimethylethanaminium). Although there are important differences between these chemicals, they have a common ability to engage metabotropic G protein-coupled receptors (GPCRs), which confers them with functional capabilities that differ from neurotransmitters linked to ionotropic receptors (*91*). These are briefly reviewed below.

2.3.1 What is neuromodulation, and why is it important?

The importance of neuromodulation is best illustrated by way of contrast with basic neurotransmission. Neurotransmission the fundamental biochemical process for information communication between neurons (*91*, *92*). It is mediated by signalling molecules called neurotransmitters of which there are two canonical types: (i) excitatory compounds, like glutamic acid (glutamate), which potentiate neuronal activity, and; (ii) inhibitory compounds, like Gamma-aminobutyric acid (GABA), which reduce neuronal excitability. Neurotransmitters are released at the axon terminals of neurons after action-potentials before rapidly binding to ionotropic receptors at the dendrites of postsynaptic neurons. Binding influences the polarisation of the postsynaptic cell, and therefore the likelihood of a subsequent action potential and onward propagation of the signal. The overarching theme of

neurotransmission, thus, is specificity: it occurs on the order of milliseconds, is localised to synaptic contact sites, and involves one-to-one, neuron-to-neuron communication.

In comparison, neuromodulation is diffuse and amorphous (*92*). Many of the details remain poorly understood, and the distinction with basic neurotransmission is blurred by the fact that neuromodulatory chemicals also participate in neurotransmission at times: for example, acetylcholine interacts with the ionotropic receptors described above, while GABA acts upon metabotropic receptors linked to neuromodulation in some circumstances. However, neuromodulatory chemicals have two broad functional capabilities arising from their interactions with metabotropic GCPRs that form bases for a distinction (*91*). Firstly, GCPRs can initiate so-called second messenger intra-cellular signalling cascades which cause temporally extended changes in the receiving cell (*91–93*). In contrast to neurotransmission, neuromodulatory signals can therefore alter neural activity over long periods of time. Secondly, GCPRs occur extra-synaptically at the neuronal membrane. This facilitates an inter-cellular signalling process called volume transmission, in which neuromodulatory molecules diffuse through extracellular fluids and bind to extra-synaptic receptors that are remote from the release location (*91*, *94*). The release of neuromodulatory molecules from a specific axon can thus have a spatially diffuse, wide-ranging impact on many downstream neurons, in contrast to the spatially specific communication accomplished by neurotransmission.

These properties allow the ANS to control fundamental aspects of neural dynamics through direct excitation and/or inhibition of neurons, changes to tonic firing rates, modification of neuronal excitability, and regulation of synaptic plasticity, all at multiple timescales (*95*). This confers neural circuits with a degree of functional and structural plasticity that would be difficult to otherwise attain given that neuroanatomy is predominantly fixed in the developed brain (*96*, *97*). According to a widely held theory, the ANS pathways might therefore be specialised for broadcasting critical, low-dimensional signals of brain-wide

relevance: for example, information about salient internal (e.g. hunger) and/or external (e.g. presence of food) states, which require co-ordinated sequences of behaviour to act upon, and therefore co-ordinated interactions between neural circuits (*98*, *99*). These signals are important influences on a variety of reward-guided behaviours.

2.3.2 Anatomy of the ascending neuromodulatory pathways

Neuromodulatory axons originate from neurons with cell-type-specific genes that enable the synthesis and release of neuromodulatory chemicals (*91*). For the most part, these neurons cluster together in a series of nuclei across the brainstem, midbrain and basal forebrain. I describe the anatomy of each nucleus below.

The main noradrenergic nucleus is the *locus coeruleus* (LC) located in the dorsomedial part of the pons in the brainstem, ventral to the cerebellum and adjacent to the edges of the fourth ventricle (*44*). LC's name, meaning 'blue spot' in Latin, derives from its distinctive neuromelanin pigmentation which arises from byproducts of the noradrenaline synthesis process (*100*). It is the smallest nucleus in the ANS, comprising just 25–50,000 neurons in the developed human brain, and yet contains more than 50% of all noradrenergic neurons in mammals (*101*). Noradrenergic neurons outside LC connect mainly to the spinal cord and lower brainstem, and are therefore unlikely to be involved in cognition (*102*). Conversely, noradrenergic axons from LC reach nearly the entire cortex, the thalamic nuclei, and key aspects of the limbic system like the hippocampus and amygdala (*103–108*).

The cholinergic system is distributed across a basal forebrain complex comprising the medial septum, diagonal band nuclei, ventral pallidum, substantia innominata and nucleus basalis of Mynert (referred to collectively as BF, hence) (*109*). There are further cholinergic

nuclei in the brainstem, including the pedunculopontine nucleus (PPN) in the rostral pons adjacent to the superior cerebellar peduncle, and the laterodorsal tegmental (LDT) nucleus in the tegmentum of the pons, just caudal to PPN (*44*, *110*, *111*). PPN and LDT neurons project mainly to other subcortical regions, including the thalamus, basal ganglia, brainstem and spinal cord, and are linked primarily to movement (*112*, *113*). The basal forebrain, in contrast, is characterised by ascending pathways to frontal and sensory cortex, in addition to the hippocampus, perirhinal cortex and entorhinal cortex, and the basolateral amygdala. (*109*, *114–128*, *128–132*).

The major dopaminergic (DA) nuclei are the substantia nigra (SN) and ventral tegmental area (VTA) in the tegmentum of the mesencephalon. VTA lies below the red nucleus, between the midline and medial forebrain bundle, and between the rostral pons and interpeduncular nucleus on the anterior-posterior axis (*44*). SN is anterior and lateral to VTA, and is divided into a medial pars compacta subdivision (SNc) containing dopamine neurons, and a lateral pars reticulata (SNr) subdivision that is predominated by GABAergic neurons and linked to the direct cortico-basal-ganglia pathway (*44*, *133*, *134*). Dopaminergic SNc and VTA axons form three ascending pathways: (i) a mesocortical pathway linking VTA to the frontal cortex, with particularly strong innervation of cingulate and orbitofrontal cortex (*135–137*); (ii) a mesolimbic pathway linking VTA to the ventral striatum (*138–141*), and; (iii) a nigrostriatal pathway linking SNc to the dorsal striatum (*133*, *142*, *143*). There are additional dopaminergic axons to the hippocampus, amygdala, thalamus, and brainstem, however they are sparser than the canonical pathways described above (*139*, *144*).

Finally, serotonergic neurons occur mainly in the raphe nuclei in the brainstem. The dorsal raphe nucleus (DRN) is situated between the oculomotor nucleus and mid-pons and extends laterally from the midline to form a distinctive wing-like shape (*44*). The median raphe nucleus (MRN) lies ventral to DRN and spans the caudal edge of the superior cerebellar

peduncles to the motor nucleus of V (*44*). Of the two, DRN is more strongly serotonergic: it contains 30–40% of the serotonin neurons in rodent and human central nervous systems compared to just 5-10% in MRN, and approximately 30–50% of DRN's total volume is serotonergic (*145–149*). DRN projects prominently to reward-related areas of the cortex including insular, cingulate, and orbitofrontal cortex (*99*, *145*, *150*). It has additional connections to sensory, somatosensory and parietal cortex, but these are limited compared to their frontal counterparts (*145*). Beyond the cortex, DRN has notable connections to the amygdala and basal ganglia, and to other the ANS nuclei including BF and SN/VTA (*145*, *150–154*). MRN's connections are confined to brainstem structures at the midline, in addition to the thalamus, hypothalamus and hippocampus (*99*, *145*, *146*, *155*, *156*). DRN is therefore the major source of serotonergic innervation for the cortical mantle.

The ANS nuclei, thus, are characterised by broad and arborising patterns of axonal projection, and there is significant overlap in their targets. However, there are general patterns in these connections that hint at the likely function of each pathway. Projections from LC and cholinergic BF cover nearly the entire cortex including occipital, parietal and temporal lobes, in addition to the hippocampal complex. Accordingly, LC and BF alike have been implicated in attention and arousal states which help organisms adapt to unpredictable environments, and to the encoding, retrieval and manipulation of representations in working and long-term memory (*100*, *109*, *157*). In contrast, SN/VTA and DRN are more prominently connected to so-called limbic areas of the cortex and basal ganglia, and therefore have strong links to reward-guided learning and decision-making (*158*, *159*).

In the remainder of this section, I review SN and VTA, and DRN in more detail. I first discuss SN and VTA because of their long-standing association with reward-guided behaviour, and because hypotheses about DRN function have frequently been formulated in contrast to established ideas about dopamine.

2.3.3 Substantia nigra and ventral tegmental area

SN and VTA have a well-characterised relationship with reward-guided behaviour. In perhaps the most influential demonstration of this link, Schultz, Dayan and Montague showed that dopamine neurons in macaque monkey VTA display phasic activity after unanticipated rewards and reward-predictive cues, and phasic depression after reward omissions in a classical conditioning task (*160*). The patterns resembled reward prediction errors – i.e., the discrepancy between the expected and actual reward received by an animal from moment to moment. Importantly, computation of reward prediction errors (RPEs) is a cardinal feature of reinforcement learning – a family of algorithms developed primarily in computer science to facilitate optimal reward-guided learning and decision-making (*161*). The experiment therefore linked three levels of analysis: it took an objective that animals have plausibly evolved to accomplish (reward-maximisation), connected the objective to a formal algorithmic theory (reinforcement learning), and showed that the algorithm was biologically implemented in the activity of dopaminergic neurons (*162*).

The RPE hypothesis has since been affirmed in many different species, and with a variety of recording and causal intervention techniques (*163–169*). However, there are aspects of dopaminergic dynamics with which it is difficult to reconcile. One of the more perplexing is the fact that dopamine concentration in the striatum ‘ramps up’ as animals approach an anticipated reward (*170*). The effect is particularly robust in operant maze tasks, where animals freely navigate structured environments populated by hidden or distal rewards (*170–172*). It is difficult to see how ramping could represent an RPE in these contexts – if dopamine conveys the discrepancy between predicted and actual values of an animal’s external state, why are dopamine signals strongest in close proximity to rewards, where reward-predictions peak? And

why is ramping so prominent for operant behaviour, when animals presumably select actions with the highest predicted reward value?

Despite attempts to explain ramping as a consequence of learning and prediction (*173–175*), the phenomenon evokes an alternative perspective that predates the RPE hypothesis: dopamine as an invigorator of movement. The first selective ablations of dopaminergic pathways in animals showed that the primary consequence was "difficulty in initiating activity" (*176*). This paralleled developments in neurology at the time: although cell death in the SN of Parkinson's patients was first noted in the late 19th century (*177*), it was the 1960s when the relationship between nigrostriatal degeneration and akinesia was formalised and treated with the dopamine precursor L-DOPA (*178–181*).

State-of-the-art techniques for recording and manipulating brain activity in freely moving animals have since confirmed that SN and VTA are causally involved in voluntary actions (*43*, *170*, *182–184*). How might this reconcile with their learning-related functions? One hypothesis holds that movement and learning are subserved by discrete circuits involving different dopaminergic subpopulations (*185–187*), but the anatomical evidence for this dissociation is equivocal at best (*43*, *182*, *188*). An alternative proposal highlights the commonality between functions: in both cases, SN and VTA appear to signal something about *value* – i.e., about what an animal stands to gain either by initiating behaviour in the present (movement) or modifying behaviour in the future (learning) (*189*). This perspective emphasises that movement-related activity in SN and VTA neurons does not code movement *per se*, but rather invigorates movements in proximity to rewards, in the manner of the ramping signal observed in reward guided navigation (*190–192*). Thus construed, SN and VTA are likely to be critical for motivating the pursuit of specific reward opportunities, and this role might unfold on two different timescales – in the present, as during action, and in the future, as in learning.

2.3.4: Dorsal raphe nucleus

In contrast to SN and VTA, DRN has proven recalcitrant to general theories of function. Serotonin's purported physiological roles include sleep (*193*), nociception (*194*), appetite (*195*), cardiovascular rhythms (*196*), and respiration (*197*), and in relation to behaviour specifically, it has been connected to anxiety (*198*, *199*), affect (*98*, *200*), reward (*201*, *202*), controllability (*203*), impulsivity and patience (*204*, *205*), social behaviour and aggression (*206*, *207*), and behavioural change and flexibility (*208*, *209*). It is, to paraphrase Jacobs and Fornal, seemingly implicated in everything and yet essential for nothing (*210*).

Here, I review the relationship between DRN and reward-guided behaviour. Because the experiments in this thesis involve recording and manipulating DRN activity, I predominantly discuss studies with the same focus, although I note where examination of DRN-derived serotonin has provided a complementary perspective.

2.4.3.1: The serotonin-dopamine opponency hypothesis

According to an influential hypothesis, DRN is an opponent motivational system to SN and VTA for reward-guided learning and decision-making (*98*, *158*). This view holds that DRN activity represents negative valanced events like punishments and threats, and mediates aversive forms of learning that enable organisms to avoid undesirable outcomes. It is closely linked to the idea that DRN controls behavioural inhibition, which is often the adaptive strategy for avoiding punishments in the natural world (*98*, *211*).

The opponency hypothesis is appealing on three fronts. Firstly, having separate systems for appetitive and aversive events resolves lingering tensions in the dopaminergic theory of reinforcement learning and motivation: for example, because neurons cannot have negative activity, there are fundamental limits on the precision with which a single class of neurons –

i.e., dopaminergic neurons – can represent both appetitive and aversive outcomes, and control both approach and avoidance behaviour (*158*). This problem is obviated by the existence of an aversion-specific system. Secondly, an opponent interrelation between dopaminergic and serotonergic pathways is consistent with DRN's dense innervation of VTA and SNc (*146*, *150*) and evidence of pharmacological antagonism between the two: infusing dopamine into ventral striatum, for example, increases the likelihood of spontaneous locomotion, but this effect reverses when serotonin is administered to the same region (*212–214*). Finally, the hypothesis explains a longstanding historical association between DRN, aversion and inhibition dating to early recording and ablation experiments (*200*, *215*, *216*). For example, the first selective lesions of DRN's serotonergic forebrain pathways prevented rats from acquiring punishment-induced motor inhibition in an operant lever pressing task (*217*). This prefigured early extracellular recording experiments, where DRN neurons in anesthetised animals displayed phasic activations in response to noxious stimuli (*218*).

2.4.3.2 How is dorsal raphe nucleus neural activity related to rewards?

The problem for the opponency hypothesis is the weight of evidence against it: as the feasibility of recording and manipulating DRN activity has improved, it has become apparent that it is not specific to aversive events and not opposed to SN/VTA activity when both regions are compared (*201*, *202*, *219–225*). Indeed, although DRN's reward-related responses are heterogenous (*222*, *223*), many studies report that the majority of DRN neurons are positively modulated by rewards and reward-predictive cues (*124*, *201*, *202*, *219*, *220*, *222*), and/or show subtle tonic activity changes that are not evoked by specific outcomes (*202*, *219–221*).

There are two general patterns in these experiments. Firstly, DRN activity often signals *general* reward-related features of an animal's context (reward-context, henceforth), as

opposed to properties of specific reward outcomes (*158*, *159*, *226*). The most direct evidence for this is two macaque fMRI studies in which there is a clear relationship between reward-context and behaviour. In the first, Wittman and colleagues recorded whole-brain activity with fMRI while monkeys performed a reversal-learning style task (*227*). In addition to standard reinforcement-learning processes, animals exhibited a 'spread of effect', whereby choices were not linked precisely to the rewards they produced but instead diffusely to an animal's overarching context, like the rewards from several preceding trials. fMRI analysis indicated that compared to the rest of the brain, BOLD signal in DRN prominently covaried with the average value of recent rewards.

Khalighinejad and colleagues demonstrated a qualitatively similar phenomenon during a reward-guided action-timing task (*228*). Here, animals needed to decide *when* to act upon reward opportunities that increased in reward-probability over time. A key influence on waiting was the duration of inter-trial-intervals (ITIs), which modulated the time-averaged rate at which animals could obtain rewards. Whole-brain fMRI analysis indicated that BOLD signal in DRN tracked block-wise changes in ITI duration across the course of the experiment and – by association – changes in an animal's reward-context.

Several other experiments have connected DRN activity with reward context, albeit in tasks where reward context has no clear link with behaviour. For example, in blocked classical conditioning tasks, photometric recordings in mice and electrophysiological recordings in monkeys indicate that tonic DRN activity discriminates appetitive from aversive conditioning blocks – in other words, DRN represents the value of the context the animal finds itself in (*220*, *221*). Similarly, in macaque monkeys performing cue-guided saccade tasks, tonic DRN activity codes an animal's reward expectation given the stage of the task it occupies (*202*, *219*, *229*). This phenomenon spans relatively simple discriminations like distinguishing ITI periods from intra-trial periods where rewards are imminent, to higher-order information like an animal's

stage of progress in multi-step tasks requiring many decisions before reward delivery. This is broadly consistent with DRN neurons coding 'state-value' – the future reward that can be expected given the current context (*230*).

The second recurring observation is DRN activity that represents reward uncertainty (*222*, *231*, *232*). Unlike the reward-context view, this hypothesis holds that DRN codes the *deviation* of specific reward outcomes from reward expectations. Such prediction-error-like responses are present in both photometric and electrophysiological recordings of DRN neurons (*202*, *223*), and include so-called 'unsigned' prediction-errors that approximate statistical measures of uncertainty (*222*). In the most thorough exploration of this phenomenon, Grossman and colleagues recorded genetically identified serotonergic DRN neurons in mice performing a multi-armed bandit task (*231*). Here, tonic DRN activity in inter-trial intervals covaried with 'expected-uncertainty' – a recency weighted average of unsigned prediction-errors capturing the average dispersion of reward-outcomes from the mean. In tandem, phasic responses at reward-delivery covaried with 'unexpected-uncertainty' – the degree to which a particular reward outcome departed from the expected uncertainty distribution.

Clearly, the notion that DRN activity represents the expected uncertainty of reward sources is closely related to the reward-context hypothesis: in the uncertainty experiments, DRN was proposed to track a recency-weighted average over *prediction errors*, whereas in the reward-context experiments it was linked to reward-expectancy which, in its simplest form, is an average over *outcomes*. The common theme of both views is that DRN activity reflects general, slow-fluctuating features of the environment that encompass many individual reward events. This is an important point of contrast with the phasic patterns of SN and VTA activity linked to reward-guided action initiation and learning.

2.4.3.3 Dorsal raphe nucleus and behavioural change

A separate group of experiments has examined how causal manipulations of DRN activity change behaviour in self-paced tasks. Many of these studies involve intertemporal choice: the ability to forego small rewards in the present in return for large rewards in the future, or even the ability to wait for future rewards altogether. In the standard inter-temporal choice paradigm, animals self-initiate trials in which they must wait for delayed rewards that vary in their timing and reward probability (i.e. the probability that reward will be delivered at all). Tonic patterns of DRN activity during these tasks distinguish between patient and impulsive periods of behaviour (*233*), and optogenetic stimulation of serotonergic neurons increases waiting times – i.e., it makes animals more patient (*204*, *234*, *235*). Subsequent experiments have demonstrated that this effect is specifically reward-guided. For example, DRN photostimulation does not make animals less likely to produce motor activity in general, meaning that the link is not reducible to behavioural inhibition (*236*). Similarly, optogenetic activation of DRN neurons does not appear to have primary rewarding qualities, meaning that animals do not wait because they find the interim stimulation reinforcing (*237*).

Does DRN activity control patience, or is patience an expression of broader function? Lottem and colleagues show that optogenetic manipulation of DRN produces a strikingly similar pattern of behaviour in a foraging-inspired reversal learning task, where animals needed to actively nose-poke for rewards at a series of negatively-correlated rewards ports (*209*). Under DRN stimulation, animals performed more continuous nose pokes at a particular reward port – in other words, they tended to *exploit* a reward source for more time before exploring other alternatives. This is remarkably similar to earlier experiments in marmosets, in which removal of serotonergic inputs to prefrontal cortex caused perseveration errors during reversal learning (*208*, *238*).

Taken together, these results suggest that DRN does not control patience for rewards *per se*, but rather how long animals will persist with a form of behaviour: in intertemporal choice, persistence is expressed as waiting for a particular reward, whereas in reversal learning it is expressed as exploitation of a specific reward source.

Importantly, foraging theory emphasises that deciding whether to engage with an opportunity like a delayed reward depends partly on knowing the distribution of alternative rewards elsewhere (see §2.1). It is therefore plausible that DRN's role in behavioural persistence is fundamentally related to reward-evoked patterns of activity signalling general features of recent outcomes, like their average value and/or uncertainty. DRN, thus, might track the general distribution of rewards in the environment to drive changes between different modes or strategies of behaviour. The experiments in this thesis aimed to test this hypothesis.

2.3.5: Habenula

Finally, in reviewing reward-related function of SN, VTA and DRN it is important to highlight the contribution of the habenula (Hb) – a pair of nuclei in the epithalamus, located at the posterior-dorsal-medial aspect of the thalamus proper (*44*). Although Hb is not part of the ANS *per se*, it is a similarly ancient structure with strong and reciprocal connections to several ANS nuclei (*239–242*). It has a particularly well-characterised inhibitory relationship with SN and VTA: Hb activity during reward-guided behaviour is distinguished by a negative value signal which mirrors activity in SN and VTA and, indeed, mediates inhibition of the midbrain dopaminergic nuclei in response to negatively-valanced events (*243–245*). This connection is mediated by a cluster of GABAergic neurons in the rostro-medial tegmentum (RMTg): in brief, negative outcomes produce excitation of habenula neurons, which in turn drive excitation of GABAergic RMTg neurons that subsequently inhibit SN and VTA (*246*, *247*). It has therefore

been argued that Hb is a key precursor to the reward and movement signals imparted by SN and VTA (*239*, *240*, *248*).

Hb activity has a similar form of inhibitory control over DRN, which it connects to via the interpeduncular nucleus (*249–251*). However, it is not reward-induced inhibition *per se* that makes Hb important. The lateral subdivision of Hb (LHb) implicated in reward has two broad classes of afferent connection. Medial LHb receives input from forebrain limbic structures that are directly innervated by the cortex, including the lateral hypothalamic and preoptic areas, the ventral pallidum, substantia innominata, and diagonal band, and the amygdala(*240*, *252*). Lateral LHb, meanwhile, receives input from the globus pallidus internal segment which connects to frontal cortex via the striatum as part of the basal ganglia (*240*, *252*). Hb, thus, receives integrated information from a range of higher-order brain regions which are important for sensing, predicting and evaluating reward opportunities. It is therefore likely that Hb is a critical source of reward-related information to SN, VTA and DRN which is both parallel to, and independent of, major descending pathways in the medial forebrain bundle.

2.4 The present work

The experiments reported in this thesis examine the implementation of reward-seeking in the macaque monkey and human brain. We accomplished this using a series of foraging-inspired behavioural tasks. These tasks had a common basic structure, in which participants encountered a single reward opportunity on each trial that could be pursued or rejected, and in which the distribution of opportunities over time was controlled to engender changes in the richness of the environment. Incorporating these elements of foraging scenarios allowed us to investigate how the brain decides *whether* to engage in reward-seeking in the first instance, and how this process is reconciled with statistical features of the external world.

We recorded brain activity during behaviour using functional magnetic resonance imaging (fMRI). fMRI measures the relative concentration of diamagnetic oxyhaemoglobin and paramagnetic deoxyhaemoglobin – the so-called blood oxygenation level dependent (BOLD) signal – which changes as a function of the metabolic processes that support neural activity (*253*). Although there are limits on its spatial and temporal resolution, a distinct advantage of BOLD signal is that it can be recorded and precisely localised from the whole brain simultaneously. In the present experiments, this allowed us to directly compare activity in DRN with activity in other regions for two purposes: (i) to specify the DRN's role compared to functionally related brain regions, and (ii); to identify patterns of DRN coactivation with other regions, indicating how they might cohere together as neural circuits (*84*, *254*, *255*).

Like all recording techniques, fMRI is a correlational tool that cannot show whether brain activity is causally linked to behaviour. It is possible to make such inferences, however, when fMRI-identified patterns of activity are disrupted. In Ch.IV, we report an experiment that accomplished this with transcranial ultrasound stimulation (TUS): a reversible and minimally invasive technique for perturbing neural activity using focussed ultrasonic waves. The basic mechanism of TUS is kinetic interactions between ultrasonic waves and mechanosensitive ion channels at the neuron and astrocyte membrane (*256*, *257*). These interactions can produce short term changes in neural activity by inducing N-methyl-d-aspartate (NMDA)-dependent plasticity (*258–262*). Previous experiments have shown that short TUS trains of the type used here can transiently change neural activity in deep and spatially specific targets without causing long-term perturbation or structural damage (*263–265*). On account of this, TUS is ideally suited to probing several different brain regions within the same animal.

Finally, we used a combination of non-human primate and human participants across different experiments. The main scientific advantage of model organisms is that a wider range of causal intervention techniques are available for use in animals: in the present experiments,

for example, it would not be possible to perform TUS of the human midbrain and brainstem given the incipient stage of TUS techniques. In addition, it is perhaps more feasible to directly engage the mechanisms of reward-seeking in animal subjects compared to human volunteers. Animals receive primary rewards for their performance on behavioural tasks, which are likely more salient than the secondary reinforcers offered to their human counterparts. Similarly, animals are not subject to demand characteristics – a phenomenon whereby human volunteers try to anticipate the aim of experiments, and behave in ways that the experimenter 'wants' them to perform. Demand characteristics might have been especially important in the present case because we examined the process of deciding *whether* to engage in reward-guided behaviour. In Ch.III–IV we report two experiments that used rhesus macaques (*Macaca mulatta*), a species of old world monkey which diverged from hominids approximately 25 million years (*266*). Rhesus macaques are the most widely used non-human primate model in cognitive and systems neuroscience, and the homologies between human and macaque brains have been extensively characterised (*45*, *267–270*). This makes them ideally placed for characterising the function of conserved regions in the primate brain.

The experiments are reported in Ch.III–V as follows. In Ch.III, we show that macaque monkeys performing reward-seeking tasks exhibit discrete motivation-states that determine their baseline propensity for reward-pursuit. We show, moreover, that these motivation-states are related to the distribution of rewards in the environment, whereby animals adopt a high motivation-state when the environment is rich, and a low-motivation state when the environment is poor.

In Ch.IV, we show that DRN has a fundamental role in implementing these changes. We use fMRI recordings to demonstrate that patterns of brain activity in DRN – but no other neuromodulatory nucleus – signalled transitions between motivation-states and the fundamental aspects of the environment on which they depend. Disrupting DRN activity with

TUS reduced the frequency of motivation-state transitions and diminished the relationship between motivation-states and the richness of environment demonstrated in Ch.III. This suggests that DRN has a causal role in changing an animal's behaviour to match its environment. In addition, we take advantage of fMRI's whole-brain perspective to examine how DRN co-ordinated with other regions during reward-seeking behaviour. We show that in contrast to DRN's role in motivation change, activity in nearby subcortical nuclei like VTA and Hb tracked the level of an animal's motivation-state from one trial to the next. Finally, activity in AI signalled the richness of an animal's environment, which was an important organising principle for the motivation-states linked to subcortical nuclei. We then show patterns of functional connectivity between AI, and DRN, Hb and VTA which suggested they form a circuit for deciding whether to pursue reward opportunities.

In Ch.V we examine DRN function in the human brain. To do this, we developed a novel behavioural task featuring a limited set of reward options that are cyclically encountered, making it possible to track how and when a participant's behavioural policy toward each option changed. Similar to Ch.III, we show that option-specific policies in human behaviour were fundamentally shaped by the richness of the environment. We further show that activity in DRN – but no other neuromodulatory nucleus – signalled changes in behavioural policy that were driven by the environment, in addition to an evolving estimate of the environment's value. Finally, we show that multivariate patterns of activity in AI and dACC coded the relative value of each reward-option given the richness of the environment. This suggests that dACC and AI are important sources of information to DRN for reward-guided changes in behaviour.

Chapter III

A Hidden Markov Model General Linear Model identifies motivation-states in macaque monkeys

3.1: Introduction

Animals need rewards for their survival, and they need to obtain them as efficiently as possible. In many naturalistic scenarios, this means tracking general features of the surrounding environment: the foraging behaviour of many species, for example involves comparing the opportunity an animal is currently confronted with against the general richness and stochasticity of opportunities it has encountered in the recent past, which guide its expectations for the future (*16*, *17*). These are rational considerations for animals given the biological constraints that encumber them. Finding and pursuing rewards consumes precious metabolic resources that must later be replenished, and so it is critical that animals organise their reward-seeking activities in ways that exploit their external milieu – to press their advantage when things are good, and conserve their energy when things are bad.

In this chapter, we investigate this phenomenon in the behaviour of macaque monkeys. We gave animals a novel behavioural task with two important features. Firstly, each trial of the task involved an encounter with a single reward opportunity which animals could pursue in return for a probabilistic reward, or reject by withholding any behavioural response. This allowed us to study how animals became engaged in reward-guided action in the first instance, in contrast to canonical reward-guided decision-making paradigms with forced choices between different reward options (*12*, *13*). Secondly, we systematically manipulated the general richness and stochasticity of reward opportunities over time. This meant that we could

examine how an animal's broader environment influenced its propensity to pursue reward opportunities.

We predicted that an animal's motivation to pursue rewards would be affected by the richness and stochasticity of its environment given the considerations set forth in foraging theory (*16*, *17*). Our task design allowed us to directly test this by comparing behaviour between experimentally manipulated environments. In addition, we tested the hypothesis in a complementary way using a General Linear Model Hidden Markov Model (GLM-HMM) – a novel analytic approach for decoding temporal patterns in behaviour (*35*, *36*). The GLM–HMM extends the standard GLM by allowing a GLM's parameters to have different values depending on the state variable in a Hidden Markov Model (HMM). Conceptually, HMM states thus correspond to behavioural states in which there are systematic and temporally extended changes in the way an animal makes decisions. Here, we used the GLM–HMM framework to probe time-varying changes in an animal's baseline propensity to pursue reward, akin to motivation-states.

In brief, we found that animals were *ceteris paribus* more likely to pursue opportunities that occurred in rich environments relative to poor environments. In addition, we found evidence for two discrete motivation-states that followed the same logic: a high-motivation state was occupied in rich environments with abundant high-value opportunities, and a low-motivation state was occupied in poor environments where rewards were scarce. Taken together, these results provide converging evidence that an animal's motivation for rewards is influenced by the richness of the environment, and that this is explicable by environment-dependent motivation-states.

3.2: Results

3.2.1 Animals are more motivated to pursue rewards in rich environments

Four rhesus monkeys (*Macaca mulatta*) performed a simple decision-making task involving sequential encounters with reward opportunities that varied in reward-magnitude and reward-probability (fig. 3.1A). Upon each encounter, animals could either pursue the opportunity and incur a short temporal cost or let the opportunity pass and proceed to the next encounter. Each session comprised four blocks of 40-50 trials. We systematically controlled the distribution of reward-probabilities within blocks to engender different reward environments. These blocks varied on two levels across two dimensions: (i) *richness*, defined by the mean reward-probability of opportunities (i.e., the value of an average opportunity in a block), and where $\mu_{\text{reward-probability}}(\text{rich}) = 0.75$ and $\mu_{\text{reward-probability}}(\text{poor}) = 0.55$, and; (ii) *stochasticity*, defined as trial-to-trial variability in reward-probability, and implemented by changing the width of reward-probability distributions where $\sigma_{\text{reward-probability}}(\text{stochastic}) = 0.13$ and $\sigma_{\text{reward-probability}}(\text{predictable}) = 0.05$ (fig. 3.1B & 3.1C; see §3.4: Methods for details).

We first examined how behaviour changed between the experimentally defined blocks. We quantified this using a series of mixed-effect binomial general linear models (GLM) that accounted for animal-specific variation in behaviour. In these GLMs, motivation was conceived as the binomial likelihood of an animal pursuing a reward opportunity on a given trial.

Because rich and poor blocks were underpinned by trial-by-trial reward-probability values, it was not possible to enter reward-probability and block-richness regressors in the same GLM due to collinearity ($r = 0.80$). However, the reward-probability distributions for stochastic blocks were carefully chosen to ensure partial overlap between rich-stochastic and poor-

stochastic distributions. This provided a subset of opportunities that occurred in different contexts (rich-vs-poor) despite having the same reward values. Analysing data from overlapping trials indicated that animals were markedly more likely to pursue opportunities that occurred in rich blocks relative to poor blocks (GLM3.1, see §3.4: Methods for details and for all subsequent GLMs; $\beta_{\text{rich-vs-poor}}$=0.90, SE=0.14, p<.001; fig.3.2A). In contrast, there was no evidence that animals behaved differently in stochastic compared to predictable blocks (GLM3.2; $\beta_{\text{stochastic-vs-predicable}}$=-0.08, SE=0.11, p=.486; fig.3.2B).

The first analysis showed that animals were influenced by the environment as operationalised by the experimentally controlled blocks. We reasoned, however, that an animal's trial-by-trial understanding of the environment would likely depend on its recent reward history. We tested this hypothesis by quantifying the time-horizon of sensitivity to past rewards with a mixed-effects GLM that had individual terms for reward outcomes on trials t-1 to t-10, where t indicates the current trial (GLM3.3, see Methods). This suggested that behaviour was modulated by rewards as distant as five trials into the past, but not further (GLM3.3; fig. 3.2C). We therefore operationalised the richness of the environment as the average reward accumulated in the previous five trials, and the stochasticity of the environment as the standard deviation of the reward rate over the same five trial period. Importantly, changing the time-window used to operationalise environments did not impact any of the inferences drawn from the following analysis (see fig. 3.2D–ii).

Quantifying the influence of these factors revealed that animals were more likely to pursue opportunities as a function of increases in the richness and stochasticity of the environment (GLM3.4, see Methods; $\beta_{\text{environment-richness}}$=0.46, SE=0.08, p<.001; fig.3.D; $\beta_{\text{environment-stochasticity}}$=0.08, SE=0.03, p=.020; fig.3.2E). These effects were independent of trial-by-trial reward value – for example, opportunities with the same expectation-value were more likely to be pursued when the recent environment was compared to when it was poor, meaning

that the effect was not due to the correlation between rich environments and high-value reward opportunities on individual trials (fig.3.2D & 3.2E).

The fact that the richness of the environment predicted whether animals would pursue reward opportunities indicated a subtle but important feature of behaviour. Animals could only obtain rewards by pursuing reward opportunities, meaning that rich environments effectively corresponded to periods where many rewards had been pursued in the recent past. Insofar as receiving rewards in the past encouraged animals to pursue opportunities in the present, the richness of the environment effect therefore indicated that pursue-vs-reject decisions were autocorrelated over consecutive trials. We tested this hypothesis by predicting the pursue-vs-reject decision taken on each trial as a function of the pursue-vs-reject decision on the previous trial (behavioural history, hence). This confirmed that animals tended to repeat pursue-vs-reject decisions over consecutive trials, even though each trial featured a separate reward opportunity with distinct magnitude and probability parameters (GLM3.5; $\beta_{\text{behavioural-history}}$=0.64, SE=0.24, p<.001; fig.3.2G).

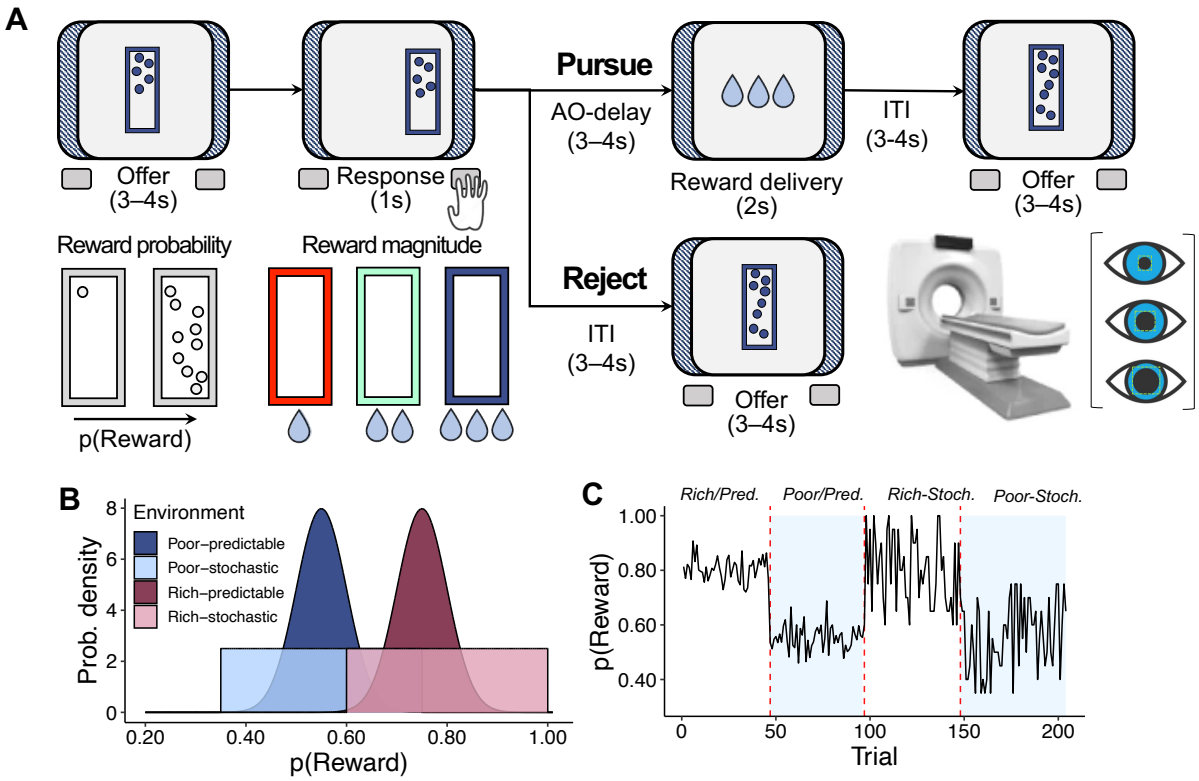


Figure 3.1. A behavioural task for controlling the richness and stochasticity of the environment. (A) Animals performed a behavioural task involving sequential encounters with reward opportunities that varied in reward-magnitude (stimulus-colour) and reward-probability (dots-per-stimulus). Animals performed the task while undergoing functional magnetic resonance imaging and pupillometry (see Ch.IV for fMRI). **(B)** The probability density of reward-probability in different reward environments. Rich environments [purple; mean(reward-probability)=.75] were characterised by higher mean reward-probability than poor environments [blue; mean(reward-probability)=.55]. Stochastic environments [light colours; sd(reward-probability)=.13] were characterised by uniform distributions and predictable environments by normal distributions [dark colours; sd(reward-probability)=.05]. **(C)** A depiction of changes in reward-probability as a function of the environment in an example session. Each session comprised four blocks covering each combination of richness and stochasticity.

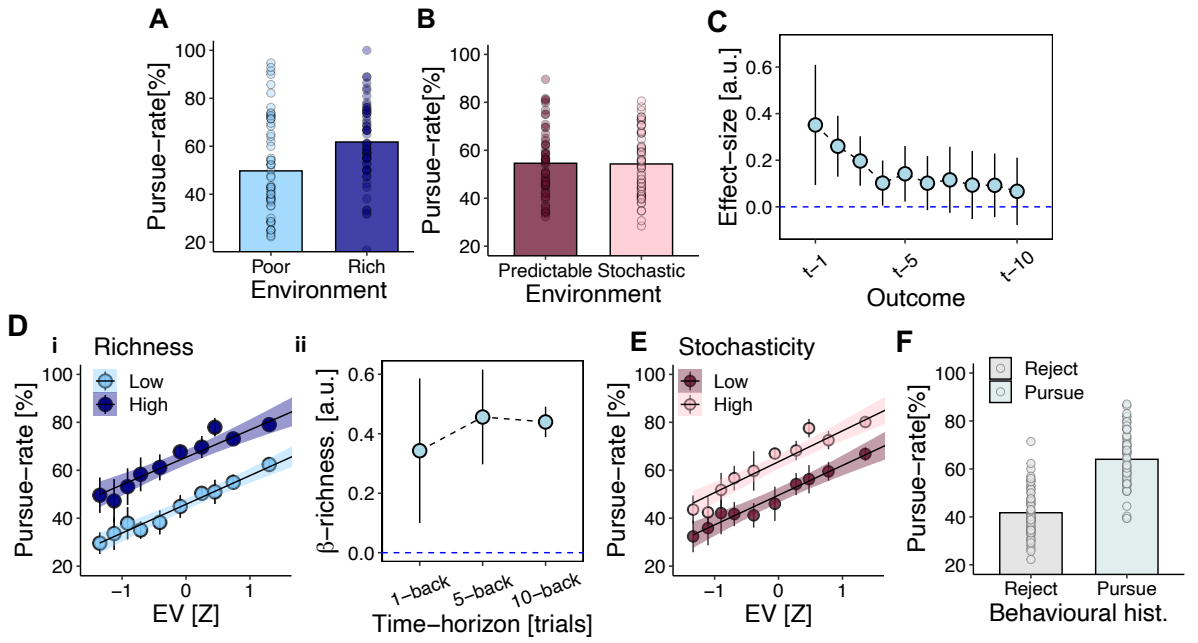


Figure 3.2. Animals are more likely to pursue reward opportunities in rich environments. (A) The reward-probability distributions for rich-volatile and poor-volatile blocks partly overlapped (fig.3.1B), which produced a subset of trials with the same reward value in different contexts (rich-vs-poor). Animals were more likely to pursue opportunities that occurred in rich blocks relative to poor blocks. Dots indicate mean rates of responding in each session and animal. **(B)** There was no evidence that animals behaved differently between predictable and stochastic blocks. Note that unlike (A), this analysis was performed over all trials because the block-stochasticity regressor was not collinear with other regressors. Dots indicate mean rates of responding in each session and animal. **(C)** The effect of reward outcomes from past trials on the pursue-vs-reject decision on the current trial in a mixed-effects logistic regression. Points and whiskers indicate effect-size (i.e. regression-weight) ± 95%CI for each predictor. Note that effects for outcomes past t-5 are not significant at .05 level. **(D–i)** When richness was operationalised relative to an animal's recent experience of reward outcomes, animals were more likely to pursue opportunities as the richness of the environment increased. X-axis indicates the expected value of the reward opportunity on each trial. Y-axis indicates rate of reward-pursuit. Colour-scale indicates mean-split according to richness of the environment [high = trials where richness(t) > μ(richness) and low = richness(t) ≤ μ(richness)]. Dots and whiskers indicate mean ± SEM of the pursue-rate in deciles of expected-value for each animal in each session. **(D–ii)** Importantly, the relationship shown in panel (D–i) was the same when the richness of the environment was operationalised using different time horizons of previous reward. Graph shows that

the effect-size for the richness of the environment in a GLM is similar regardless of the time-horizon used. We tested two additional time-horizons: (i) where the richness of the environment was operationalised as the reward outcome on the preceding trial [β(environment-richness)=0.34, SE=0.12, p=.006; GLM3.4], and; (ii) where the richness of the environment was operationalised as the average reward outcome on the previous ten trials [β(environment-richness)=0.44, SE=0.03, p<.001; GLM3.4]. Y-axis shows effect-size for the richness of the environment in GLMs. X-axis indicates the time horizon of previous reward outcomes used to operationalise the richness of the environment. Points and whiskers indicate effect-size (i.e. regression-weight) ± 95%CI for each predictor. **(E)** When stochasticity was operationalised relative to an animal's recent experience of reward outcomes, animals were more likely to pursue opportunities as the stochasticity of the environment increased. X-axis, Y-axis, dots, whiskers and follow conventions of panel D–i. Colour-scale indicates mean-split according to stochasticity of the environment [i.e. high corresponds to trials where stochasticity(t) > μ(stochasticity) and low to stochasticity(t) ≤ μ(stochasticity)]. **(F)** Animals were likely to repeat the pursue-vs-reject decision taken on the previous trial on the current trial. Dots indicate mean pursue-rate in each animal in each session.

3.2.2: A Hidden Markov Model identifies motivation-states in behaviour

We reasoned that the autocorrelations in animal behaviour evinced by the richness of the environment and behavioural history effects might reflect motivation-states – that is, changes in an animal's baseline predisposition to pursue-vs-reject rewards that were independent of the reward opportunity on each trial, and that persisted over many trials (fig.3.3A–B). It has recently been shown that latent time-varying dispositions in behaviour can be uncovered using a General Linear Model Hidden Markov Model (GLM-HMM) approach (*35*, *36*). We therefore tested the motivation-state hypothesis using a GLM-HMM.

GLM-HMMs are an extension of the classical GLM in which the parameter values of a GLM are allowed to change according to time-varying latent states. Transitions between latent states are governed by Markovian dynamics and are thus called 'HMM-states'. In simple terms, this means that if animals behave in fundamentally different ways – for example, if they adopt different motivation-states – over the course of the task, the GLM-HMM can detect these states, and specify how the relationship between the task and behaviour changes in each state via the weights in the GLM part of the model (fig.3.3B).

We formulated a GLM-HMM where the pursue-vs-reject decision on each trial was predicted with a binomial GLM that included the following terms: (i) a bias or 'intercept' term, (ii) a predictor for the expectation-value of the reward opportunity available on the current trial, and (iii) predictors for external environment cues that were presented throughout the task (fig.3.3B; see §3.4: Methods for details). Our hypothesis was that animals experienced changes in their baseline predisposition to pursue-vs-reject reward opportunities, over and above the influence of external factors like the value of the current opportunity. We therefore let only the bias parameter change between HMM-states and held the weights on other predictors constant.

Incorporating predictors for features of the task in the GLM part of the model ensured that patterns identified by the GLM-HMM were specifically due to shifts in an animal's baseline predisposition to pursue-vs-reject rewards, after taking into account external factors like the reward value on the current trial.

We used a previously reported model-selection procedure which involved testing models with different numbers of HMM states using 5-fold cross validation (*35*) (fig.3.4). We validated the procedure using a series of parameter recovery tests on simulated data (fig.3.4A–C). After confirming successful parameter recovery, we implemented it on behavioural data. This indicated that cross-validated log-likelihoods for 2-state GLM-HMMs were higher than 1-state GLM-HMMs for all animals (fig.3.4D & fig.3.5A), but that invoking additional states beyond this number was not warranted for any animal. We therefore used 2-state GLM-HMMs for all further analyses.

Each animal-specific 2-state GLM-HMM featured clearly distinct HMM-state-specific bias parameters (fig.3.5B) and profoundly autocorrelated HMM-state transition matrices, which were consistent with state-like fluctuations in baseline predispositions to pursue-vs-reject rewards (fig.3.5C). Furthermore, simulating data from fitted 2-state GLM-HMMs reproduced salient features of animal behaviour including qualitative patterns of autocorrelation in decision-making (fig.3.5D–E), and the richness of the environment and behavioural history effects described earlier (fig.3.5F–G). This demonstrates that the key behavioural effects described above – the richness of the environment effect, and the behavioural history effect – can be parsimoniously explained by a 2-state GLM-HMM in which there are changes in the bias term governing an animal's predisposition to pursue-vs-reject rewards. Further analyses demonstrated that HMM-states were not reducible to satiety, fatigue, or time-on-task (fig.3.6A–C).

We tested whether the statistical patterns identified by the GLM–HMM corresponded to biological motivation-states with a series of follow-up analyses. To do this, we first decoded the *maximum a posteriori* HMM-state on each trial using a well-established procedure called the Viterbi algorithm (see §3.4: Methods for details). We validated the Viterbi algorithm on simulated data, which demonstrated that it identified the correct HMM-state on 90% of trials, and identified HMM-state transitions within +/- 3 trials of a true transition in 85% of cases (fig.3.7A–C). We then examined how decoded HMM–states were related to behaviour. As expected, animals were markedly more likely to pursue reward opportunities in the putative high-motivation state compared to the low-motivation state, regardless of the reward value available (GLM3.6; $\beta_{\text{motivation-level}}$=2.66, *SE*=0.56, p<.001; fig.3.7D). Behaviour during transitions between HMM-states was characterised by abrupt changes in pursue-rates, suggesting that the GLM-HMM captured state-like differences in behaviour with a high degree of temporal precision (GLM3.7; $\beta_{\text{before-vs-after-transition(high-to-low)}}$=-1.48, SE=0.08, p<.001; $\beta_{\text{before-vs-after-transition(low-to-high)}}$=1.80, *SE*=0.30, p<.001, fig.3.7G). Finally, we tested the model's convergent validity by comparing HMM-states to: (i) trial-by-trial pupil-size, a well-validated indicator of physiological arousal (*271*), and (ii) trial-by-trial reaction times (RT) to pursued rewards, a simple way of quantifying vigour. Consistent with the link between HMM-states and motivation, the putative high-motivation state featured faster RTs when rewards were pursued (GLM3.8; $\beta_{\text{motivation-level}}$=-0.20, SE=0.03, p<.001; fig.3.7F) and increased pupil-diameter during decision-making (GLM3.9; $\beta_{\text{motivation-level}}$=0.20, SE=0.02, p<.001; fig.3.7E). The GLM-HMM, thus, provided quantitative evidence for discrete, persistent, and biologically meaningful motivation-states in monkey behaviour.

Finally, we tested the hypothesis that motivation-states were related to the external environment: was it the case, for example, that animals matched their motivation-state with the shifting distribution of reward opportunities in the task over time? We quantified the

availability of rewards by calculating, for each trial, the average expectation-value of reward opportunities in the preceding five trials. We did not use the *richness of the environment* metric introduced earlier because this was calculated from an animal's recent history of reward, and therefore was contingent on its behaviour – relating an animal's motivation-state to its recent history of behaviour, thus, would be a logically circular analysis. In contrast, the expectation value of reward opportunities was experimentally controlled via the reward distributions characterising the task (fig.3.1B) and was therefore not affected by an animal's decisions. This showed that animals were more likely to occupy high-motivation states as the average expected-value of recent reward opportunities increased (GLM3.10; $\beta_{Ave.EV}$=0.19, SE=0.08, *p*=.022; fig.3.7H). The GLM-HMM therefore demonstrated not just that animals experienced motivation-states that determined their predisposition to pursue-vs-reject rewards, but also that motivation-states were matched to the distribution of rewards in the environment.

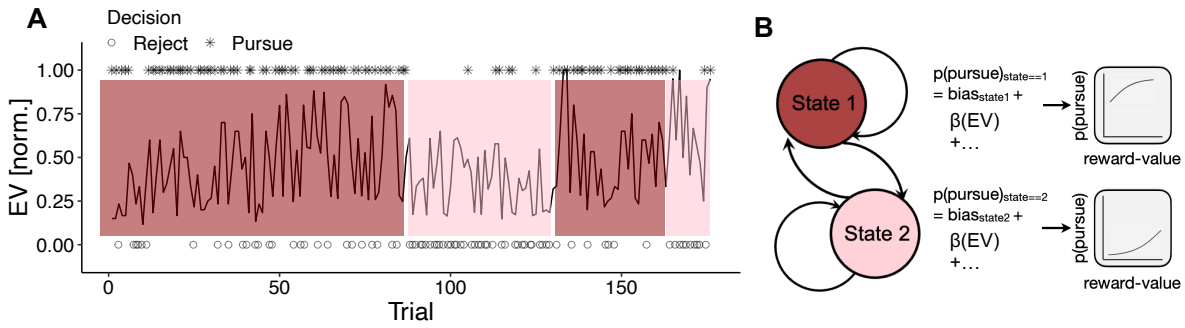


Figure 3.3. A General Linear Model Hidden Markov Model for explaining temporal patterns in behaviour. (A) Animals exhibited autocorrelated patterns of behaviour in which they were biased to pursue (red) or reject (pink) reward opportunities almost regardless of expectation-value. Data shows pursue-vs-reject decisions (stars at top indicate pursuits; dots at bottom indicate rejections) as a function of expectation-value (y-axis) over time (x-axis) in an example session. **(B)** We quantified these patterns using a GLM-HMM with HMM-state dependent bias parameters. This produced corresponding HMM-state dependent decision functions, which captured changes in the way animals made decisions over time.

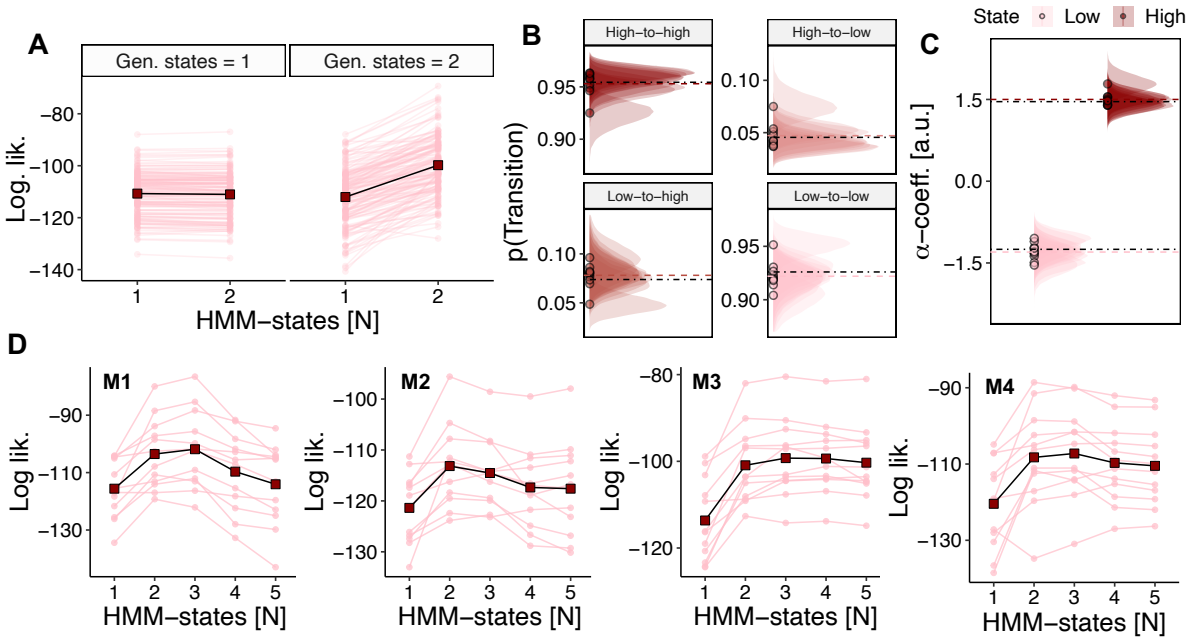


Figure 3.4. Validating and implementing a GLM–HMM fitting procedure using 5-fold cross validation. We implemented a previously reported fitting procedure for GLM-HMMs that involved 5-fold cross validation within each animal (*35*). In brief, we divided the data into five folds and iteratively fit the model to sessions in four folds (i.e. training folds) before testing performance on the held-out (i.e. test) fold. We assessed model performance using the log-likelihood of held-out sessions, which we computed using the forward algorithm (see §3.4: Methods). We constructed folds between-sessions instead of trials-within-sessions because between-session variance was greater than within-session variance. Between-session cross-validation was therefore a more demanding test of performance (*35*). For each fold, we tested models with $s \in \{1, 2, 3, 4, 5\}$ HMM-states and calculated the log-likelihood of test sessions in each case. **(A)** We validated the procedure by simulating 10 datasets from a fitted 2-state GLM-HMM. Each dataset was approximately the same size as the total dataset obtained from each individual animal (~16 sessions, ~3000 trials). We first tested whether the procedure recovered the true number of HMM-states when the data was generated by a 1-state (i.e. Binomial GLM; left panel) and/or 2-state GLM-HMM (right panel) models. 2-state GLM-HMMs fit better for data generated from a 2-state GLM–HMM, but not for data generated by a Binomial GLM. The number of generative states, thus, was successfully recovered. All subsequent data pertains to simulations with 2-state GLM-HMM generative models. **(B)** In addition, the fitting procedure successfully recovered transition-matrix parameters used to generate simulated data. Graphs show posterior distributions over transition-matrix components in fitted 2-state GLM-HMMs. Black

dashed lines indicate generative model parameters and coloured lines indicate mean parameters over all fitted models. **(C)** Finally, the fitting procedure successfully recovered HMM-state specific bias/intercept parameters used to generate simulated data. Graphs show posterior distributions over state-specific bias values. Dashed black lines indicate generative parameters and coloured lines indicate mean parameters over all fitted models. **(D)** Implementing the fitting procedure on behavioural data showed that all animals followed a similar pattern whereby 2-state GLM-HMMs performed better than Binomial GLMs (i.e. 1-state GLM–HMMs). Additional states, however, either failed to improve model performance, or impaired it. Graphs show relationship between log-likelihood of the held-out sessions as a function of HMM-states (foregrounded bold lines indicate mean log-likelihood of held-out sessions; background lines indicate log-likelihood of individual sessions).

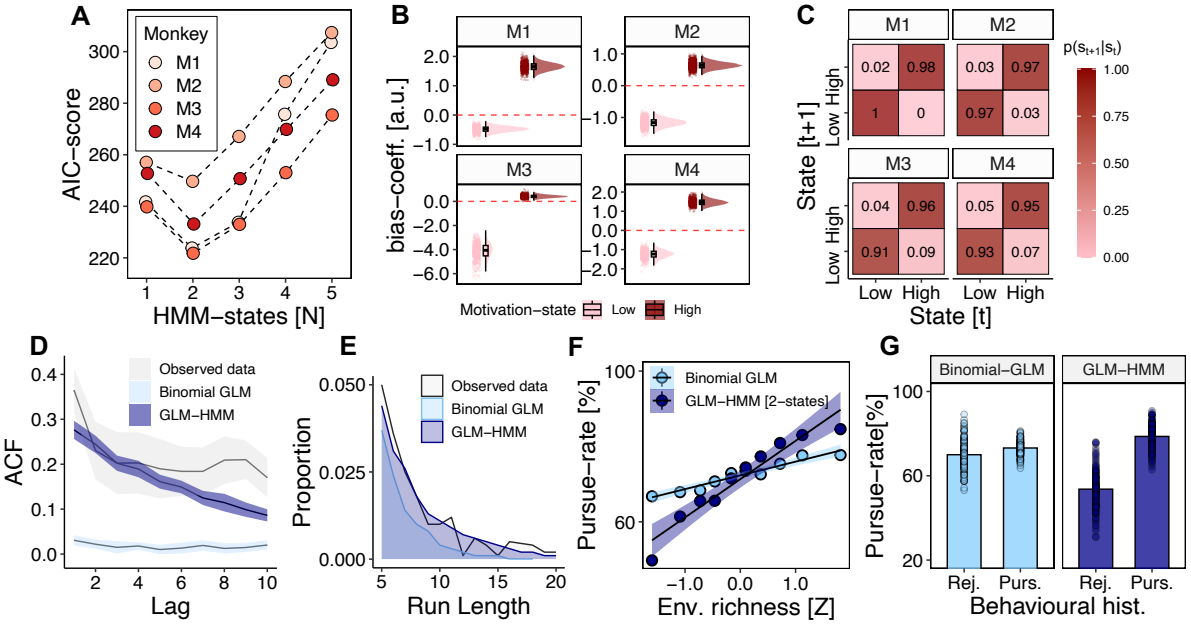


Figure 3.5. A 2-state GLM–HMM captures key quantitative and qualitative patterns of animal behaviour. (A) A model incorporating two HMM-states clearly improved model fit compared to alternative numbers of states. Here, this is shown using AIC – a form of log-likelihood that penalises model complexity (see supplementary fig.3.6 for full cross-validation results). Points show mean session-wise AIC for each animal. **(B)** The posterior probabilities of state-specific bias parameters in all animals (M1 – M4) indicate that HMM-states impact an animal's baseline predisposition to pursue-vs-reject rewards. **(C)** Transition matrices that describe the conditional probability of HMM-states from one trial to the next show strong autocorrelations in HMM-states in all animals (M1–M4). **(D)** Simulating data from a 2-state GLM-HMM reproduced trial-to-trial autocorrelations in behaviour, which were a key feature of animal behaviour. A binomial GLM was unable to reproduce this pattern. Y-axis shows autocorrelation function (ACF) between the decision on trial *t* and decisions on previous trials spanning *(t-1):(t-10)* for recorded behaviour (grey), the 2-state GLM-HMM (dark blue) and a Binomial GLM (light blue) for an example animal. Lines and shaded areas show the mean and SEM, respectively, of the ACF at each individual lag. ACFs for the behavioural data were obtained by calculating the ACF at each lag in each individual session (N=16) and then taking the mean and SE of ACFs at each level of lag. Simulated ACFs were obtained by simulating datasets from Binomial and 2-state GLM-HMM models that were fit to animal behaviour. We then calculated the mean and SE of the ACF at each lag, as per the behavioural data. **(E)** A concrete demonstration of autocorrelations in behaviour is

the number of times that the same pursue-vs-reject decision is repeated over consecutive trials. Here, we have called these *decision run-lengths*, where rejecting three consecutive reward opportunities would constitute a decision run-length of 3, for example. See also (*35*). The GLM-HMM closely tracks the tendency of animals to repeat their decisions over consecutive trials, especially for decision run lengths spanning five or more trials. A binomial GLM is unable to reproduce this pattern. Data from the same example animal and simulations as panel (E). **(F)** and **(G)** Implementing the GLMs (GLM 3.1–3.2) used to analyse behaviour in §3.2.1 on data simulated from a 2-state GLM-HMM recapitulated the richness of the environment effect (F) and behavioural history effect (G). The same patterns did not occur when data was simulated from a Binomial GLMs. In (F), dots indicate mean pursue-rate in deciles of richness of the environment across simulations, and shaded area indicates SEM around line-of-best-fit. In (G), dots indicate mean pursue-rates in individual simulations.

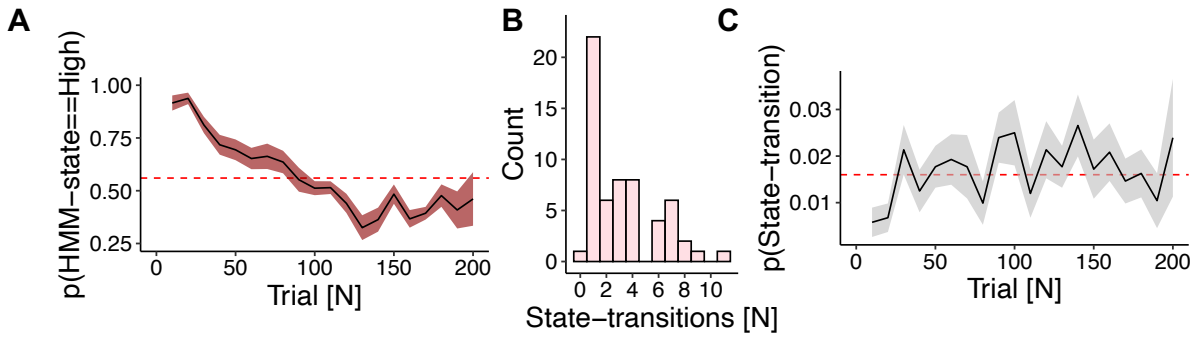


Figure 3.6. GLM–HMM states and state-transitions are not reducible to fatigue, satiety or time-on-task (A) The relative probability of motivation-states (y-axis) over time (x-axis). Animals are more likely to occupy the high-motivation state early in a session, and approximately equally likely to occupy the high-motivation and low-motivation states at the end of a session. The GLM–HMM states thus, are not reducible to satiety, fatigue, or time-on-task. Black line and shaded areas indicate mean and SEM, respectively, of state-occupancy as a function of time; dashed red line indicates mean probability of motivation-state occupancy, regardless of time. **(B)** Histogram of the number of GLM–HMM state transitions per session (x-axis). One session features no transitions. Data from all animals included. **(C)** The probability of GLM–HMM state transition (y-axis) as a function of time (x-axis). Although transitions are more likely to occur in later trials, there is no consistent and specific temporal pattern in transition events, which suggests that they are not reducible to the passage of time (per (**A)**). Line and shaded area indicate mean and SEM, respectively, of state transitions rates in quintile bins of trials. Dashed red line indicates overall mean probability of state-transition, regardless of time. Data from all animals included.

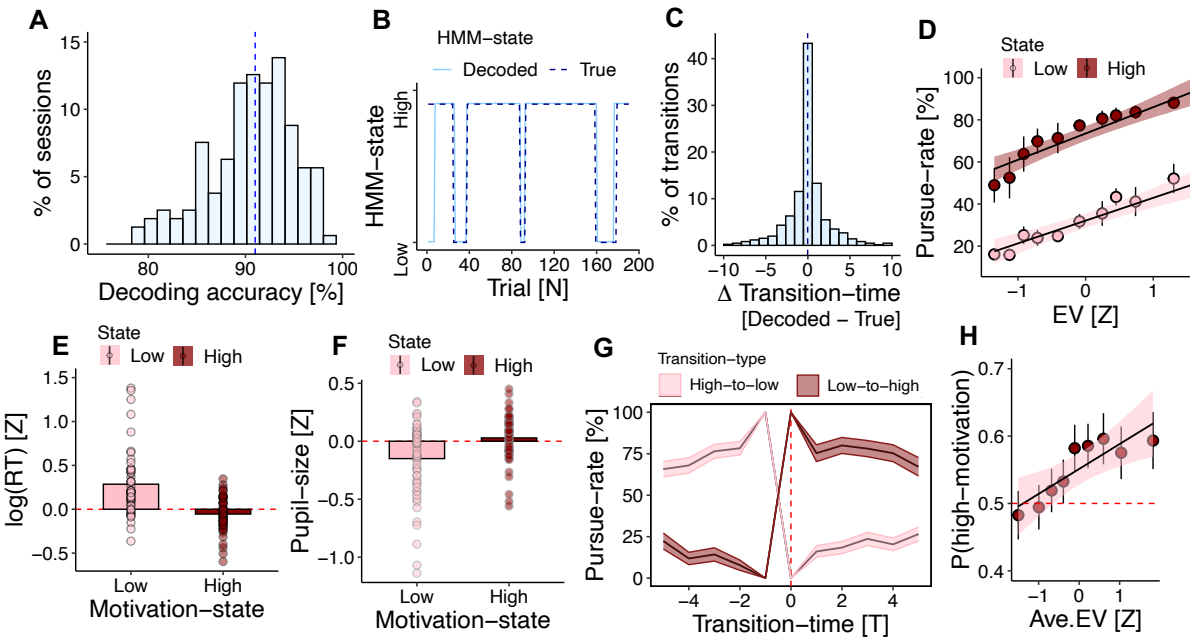


Figure 3.7. GLM–HMM shows distinct, persistent, and biologically meaningful motivation-states during behaviour. We further probed HMM-states by decoding the *maximum a posteriori* state on each trial and comparing it to features of an animal's behaviour and physiology. Decoding was accomplished using the Viterbi algorithm. **(A)** We established the accuracy of the Viterbi decoding algorithm by applying it to simulated datasets in which the true HMM-state was known. Viterbi decoding was successful on 90.68% (± 0.72%) of trials. Histogram shows distribution of session-level decoding accuracy (%) across 10 simulated datasets comprising 16 sessions each (i.e. 160 sessions total). Data simulated from a 2-state GLM–HMM fit to animal behaviour **(B)** True-HMM-state vs Viterbi-decoded-state in a representative example simulated session. **(C)** We assessed the Viterbi algorithm's ability to identify state-transitions by quantifying – for each decoded transition – the distance to the nearest true state-transition in simulated data (where distance was measured in number of trials). The majority of transition-points were exactly identified (mean-distance = -0.03 trials) and 84% of decoded transitions occurred within ±3 trial window of a true transition. Histogram shows proportion of decoded-vs-true transition differences. Same simulated data as (A). **(D)** Animals were more likely to pursue reward opportunities when occupying the high-motivation state relative to the low-motivation state regardless of the reward available on the current trial. Dots and whiskers indicate mean ± SEM pursue-rates in deciles of expected-value in each animal in each session. **(E)** Pupil-diameter – a common indicator of physiological arousal – when animals are making decisions is larger when animals occupy the high-motivation state compared to when they occupy

the low-motivation state. Dots indicate mean pupil-size in each animal in each session. **(F)** When rewards are pursued, reaction time – a common measure of vigour – is faster when animals occupy the high-motivation state. Dots indicate mean RT for each animal in each session. **(G)** Pursue-rates changed abruptly at decoded motivation-state transitions. For decoded low-to-high transitions, the pursue rate is lower in the five trials before the transition-trial than it is in the five trials after the transition trial. Similarly, for high-to-low transitions, the pursue-rate is higher in the five trials before the transition-trial than it is in the five trials after the transition trial. This suggests that the GLM-HMM identifies abrupt and state-like changes in behaviour, rather than gradual changes. Colour shades indicate transition direction. Line and shaded area indicates mean ± SEM of pursue-rate across all animals in all sessions **(H)** The likelihood that animals occupy the high-motivation state increases as a function of the expectation-value of recent reward opportunities (Ave. EV). Dots and whiskers indicate mean ± SEM of rate of motivation-state occupancy in deciles average expectation-value in all animals in all sessions.

3.3: Discussion

The constraints of biology dictate that animals should change their propensity for reward-seeking depending on the distribution of rewards in the environment. Here, we show with two complementary analytic approaches that macaque monkeys are more motivated, *ceteris paribus,* to pursue rewards in rich environments with many high-value opportunities.

We designed a task which featured experimentally controlled changes between rich and poor environments. Animals were markedly more likely to pursue opportunities that occurred in rich environments compared to poor environments, regardless of whether the environment was operationalised relative to experimentally controlled blocks, or an animal's recent empirical history of reward. We then implemented a novel GLM–HMM to show that these observations could be explained by state-like changes in an animal's baseline predisposition to pursue-vs-reject rewards. We found evidence for two motivation-states that were fundamentally related to the richness of the environment: a high motivation-state that was occupied when many high-value opportunities were available, and a low-motivation state that was occupied when low-value opportunities predominated.

Behavioural states are ubiquitous features of human and non-human animal behaviour. In systems neuroscience, they have traditionally been studied in simple organisms using stimulus-evoked paradigms, like avoidance states in response to predatory threats, aggression in response to social challenges, and courtship in response to reproductively salient stimuli (*32*, *272–274*). Here, we add to an emerging body of experiments showing that it is possible to decode states reflecting internal changes within an animal, like fluctuations in their motivation, attention, and arousal (*35*, *275*, *276*). It is likely that internal states of this kind arise from the biological constraints emphasised in behavioural ecology. For example, in the present case, animals needed to encode and interpret abstract visual stimuli, evaluate each stimulus given its

reward-probability and reward-magnitude, and enact swift and accurate movements to ensure that valuable opportunities were captured. The costs imposed by these processes were evinced in the fact that on approximately 50% of all trials, animals decided that the reward opportunity available was not worth engaging with.

What are the implications of biological constraints for reward-guided behaviour? In the present experiment, we found that animals were more motivated to pursue rewards that occurred in rich environments. This observation is striking because it appears to controvert prominent formulations of foraging theory. In brief, these theories suggest that animals should be less selective (i.e. more motivated) during periods of reward scarcity because they have homeostatic needs that must be met, regardless of how bad their opportunities are (*16–18*). A simplifying assumption of these models, however, is that animals know the background distribution of rewards in the environment. Although this is a reasonable assumption for species that occupy stable ecological niches, it is plausible that when animals lack this knowledge, they should instead concentrate their efforts on rich environments where they have the most to gain. The pattern observed in the present experiments may therefore ensure that reward intake is maximised in the manner that foraging theories envisage.

In support of this view, several other experiments in behavioural neuroscience have documented similar patterns of animal behaviour (*227*, *228*, *264*, *277*, *278*). For example, rodents sample rewards more frequently and vigorously when many high-value rewards are available, similar to the behavioural changes displayed by monkeys in the present experiment [reviewed in (*277*)]. Many of these rodent experiments involved operant self-paced paradigms in which the speed and vigour of responding determined an animal's reward intake. Although animals in the present task did not control the rate at which they encountered rewards in the same fashion, their behaviour may nevertheless have been shaped by a mechanism for invigorating reward-seeking in rich and opportune environments.

An alternative perspective is that the phenomenon derives from the so-called "spread of effect": a failure of credit assignment whereby rewards are not linked precisely to the choices which caused them (*227*, *279*). The spread of effect has a well-documented history in animal learning, and was recently advanced by Wittman and colleagues (*227*) to explain why 'global' average reward-rates (as opposed to option-specific reward rates) influence reversal learning in rhesus monkeys. Importantly, however, the reward-magnitude and reward-probability of opportunities in the present task were explicitly cued. This is fundamentally unlike learning tasks, where the critical aspects of reward options are uncertain and must, indeed, be learned. In principle, the fact that animals had complete knowledge about the value of each opportunity should have diminished the influence of credit assignment processes – and failures thereof – on behaviour.

Finally, it is important to touch on the stochasticity of the environment, which was the other experimentally controlled change in the task. Foraging theory implies that organisms should be more likely to pursue rewards in stochastic environments where future opportunities might be of little or no value (*280*). This is broadly consistent with the behavioural effect of stochasticity in the present work: although there was no evidence that behaviour differed between experimentally controlled stochastic and predictable blocks, animals were marginally more likely to pursue rewards as a function of stochasticity in their reward history. The size of this effect, however, paled in comparison to the richness of the environment. It is therefore possible that animals are, indeed, sensitive to the stochasticity of reward opportunities over time, but that the experimental manipulation was too subtle to elicit a meaningful signature of this in behaviour. In relation to this, it is important to note that although stochasticity was manipulated via the distribution of reward-probabilities, reward opportunities were characterised on both reward-magnitude and reward-probability dimensions. The relationship between manipulated stochasticity changes and the value of reward opportunities might

therefore have been outweighed when animals combined reward-magnitude with reward-probability into a scalar subjective-value. Future experiments could address this by operationalising environments using distributions over expectation-value instead of reward-probability. This would likely require finer quantisation of reward magnitude than the present experiment, because small gradations (e.g., *rw-mag* $\in \{1, 2, 3\}$, as used here) make it challenging to control expectation-value without inducing correlations between reward-probability and reward-magnitude dimensions.

3.4: Materials and Methods

3.4.1 Experimental model and subject details

The experiments were performed with four male rhesus macaques (*Macaca mulatta*). The sample size corresponded to those used in previous studies where it has been possible to identify significant and reliable whole-brain functional magnetic resonance (fMRI) and transcranial ultrasound stimulation (TUS) effects across multiple testing sessions but with the minimum number of animals (fMRI and TUS results are reported in Ch.IV). The animals were 10-12 years of age and weighed 14.30–17.09 kg. They lived in group housing with a 12hr light dark cycle and were afforded access to water for 12–16 hr on testing days and free access on non-testing days. All procedures were conducted under a license issued by the UK Home Office in accordance with the Animal (Scientific Procedures) Act 1986 and the European Union guidelines (EU Directive 2010/63/EU).

3.4.2 Behavioural training

All monkeys were fitted with magnetic resonance imaging (MRI) compatible cranial implants that facilitated head-fixation during testing and training (Rogue Research). Training was conducted in MRI compatible chairs designed to accommodate monkeys in a sphinx position, and took place in custom-built environments that replicated an MRI scanner. Animals were trained on simplified versions of the task in which they did not need to wait for go-cue and action-outcome delays, and were not exposed to controlled changes in their reward environment. Go-cue and action-outcome delays were then introduced and carefully increased over the course of training until they were sufficient for an fMRI experiment (see Ch.IV).

Changes in the reward environment were introduced after animals were accustomed to the final delay timings. Testing for both the fMRI and transcranial ultrasound stimulation (TUS) experiments began when the proportion of pursued trials per session was stationary over consecutive days. The behavioural data reported in this chapter was recorded during fMRI testing sessions.

3.4.3 Behavioural task

Animals performed a simple decision-making task involving sequential encounters with reward opportunities that appeared on a computer screen. Reward opportunities were presented in visual form via coloured boxes (size of the box: 8 x 26 cm) that were filled with dots. The colour of the stimulus indicated the number of juice drops that the opportunity was worth {red = 1; green = 2; blue = 3}. The number of dots comprising the stimulus indicated the probability that reward would be delivered if the opportunity was pursued. Reward probabilities ranged between (.05, 1), and dots indicated .05 linear increments of reward probability. The stimuli were designed so that dots gradually filled the box from the top downwards.

Opportunities first appeared in the centre of the screen before displacing either right or left (fig.3.1A). The displacement of the reward opportunity stimulus functioned as a go-cue which indicated that it was available for pursuit. Monkeys could pursue opportunities by manually responding to an infrared sensor corresponding to the opportunity's on-screen location – for example, if the opportunity displaced right, they needed to touch a sensor with their right-hand, and vice versa if it displaced left. The go-cue was designed to temporally dissociate neural activity related to pursue-vs-reject decisions from the motoric processes which translated decisions into movement. Similarly, the randomised right/left displacement of

the reward opportunity stimulus prevented monkeys from motor planning during the decision phase. The durations separating reward opportunity onsets and go-cues were drawn from *go-cue ~ uniform(3, 4)* and were optimised for a canonical macaque hemodynamic response function (HRF) of approximately 3-4s. Monkeys had 1s to pursue opportunities after the go-cue. If a reward opportunity was pursued, monkeys needed to wait an action-outcome delay drawn from the distribution *A-O-delay ~ uniform(3, 4)*. They then received any potential juice reward accompanied by visual reward feedback. The next trial then began after an inter-trial interval spanning *ITI ~ N(4, 1)*. If monkeys rejected a reward opportunity, they bypassed the action-outcome and reward-delivery delays and proceeded to the next trial. If monkeys pursued a reward opportunity before the go-cue – i.e. prematurely – they needed to wait the remaining duration of the reward opportunity presentation, in addition to go-cue, action-outcome and reward-feedback delays. The reward opportunity which elicited the premature response was then repeated on the next trial.

Each session of the task comprised four blocks of 40–50 trials. These blocks were used to engender different reward environments by systematically controlling the reward-probability dimension of reward opportunities. Environments varied on two dimensions: (i) *richness* was defined by the mean of the reward-probability distribution, where $\mu_{rich} = 0.75$ and $\mu_{poor} = 0.55$, and (ii) *stochasticity* was defined by the width of reward-probability distributions, where predictable environments were generated from Gaussian distributions with $\sigma_{predictable} = 0.05$ and unpredictable environments from random-uniform distributions with ranges of 0.40 such that $\sigma_{stochastic} = 0.13$. Blocks were characterised along both richness and stochasticity dimensions, which yielded four distinct environment-types – rich-predictable, rich-stochastic, poor-predictable, and poor-stochastic. Each session featured one of each environment-type, and the order of environments was counterbalanced along the richness dimension to avoid long periods of low-value offers (e.g., during consecutive poor environments) which were difficult for

monkeys to perform. Environment-types were indicated by visual cues which bordered the screen.

The behavioural data reported in this chapter was recorded during fMRI testing sessions. Three animals performed 16 sessions of the task and one animal performed 14 sessions.

The task was implemented in MATLAB v2019 (MathWorks) using Psychophysics Toolbox v3 (*281*) and presented on MRI-compatible screens (23in BOLD screen; Cambridge Research Systems) approximately 30cm away from the subjects. Juice rewards consisted of a solution of water, blackcurrant cordial and banana and each juice drop was approximately 1mL. Pupillometric data was obtained during fMRI sessions via an MR-compatible infrared EYElink 1000 eye-tracker device (SR Research Ltd.), which recorded pupil-diameter and gaze-direction along x and y planes at a 250-Hz sampling rate.

Pre-processing and analysis of pupillometric data was performed in *R* and entailed the following steps [adapted from (*282*)]. Samples reflecting eye blinks were identified with a detection algorithm native to the EYElink device. Artefacts were defined as consecutive samples with discrepancies of > 50 a.u. All samples within symmetric 25 sample (i.e., 0.1s) windows of artefacts or eye blinks were scrubbed and linearly interpolated, and interpolated time series were then low pass filtered with a 4Hz cut-off. These processes were performed for pupil-size, x-gaze-direction, and y-gaze-direction timeseries respectively. All timeseries were then z-scored, and we then regressed variance due to x-gaze-direction and y-gaze-direction from pupil-size via linear regression. All subsequent analysis was performed using the residuals.

To incorporate pupillometry in our behavioural analysis, we extracted pupil-size in specific epochs of the task, with a focus on periods after the presentation of reward

opportunities, when animals were making decisions. In particular, we quantified – for each trial – mean pupil-size 0.5–1.0s after stimulus presentation. We selected this epoch to avoid pupillometric responses reflecting sudden luminance changes, which occurred when the reward opportunity visual stimulus first appeared.

3.4.4 Behavioural analysis

Our initial behavioural analysis characterised which factors influenced the binomial pursue-vs-reject decision made on each trial. We accomplished this with mixed-effect binomial generalised linear models (GLMs) with subject-identity as a random variable – this means that all GLMs accounted for inter-subject variability for the effect of interest. We began by investigating the influence of experimentally defined block-types (rich-vs-poor, and stochastic-vs-predicable). We investigated the effect of block-richness with the following GLM:

GLM3.1

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 block\text{-}richness_t + \beta_3 trial\text{-}number_t + \mu_0 + \mu_1 block\text{-}richness_t + \varepsilon$$

where $\beta_{0\text{-}3}$ are fixed-effects, μ_0 is the by-subject random intercept, μ_1 is a by-subject random slope, and ε is an error term. GLM3.1 was fit to a subset of trials with overlapping reward-probabilities drawn from rich-stochastic and poor-stochastic blocks (see §3.2.1). There was therefore no term for block-stochasticity. We tested the effect of block-stochasticity with a separate GLM:

GLM3.2

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 reward\text{-}probability_t + \beta_3 block\text{-}stochasticity + \beta_4 trial\text{-}number_t + \mu_0 + \mu_1 block\text{-}stochasticity_t + \varepsilon$$

where $\beta_{0\text{-}4}$ are fixed-effects, μ_0 is the by-subject random intercept, μ_1 is a by-subject random slope, and ε is an error term.

We then analysed the effect of an animal's recent reward history, as distinct from the experimentally manipulated block-types. We began by investigating the time-horizon of sensitivity to past rewards with the following GLM:

GLM3.3

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 reward\text{-}probability_t + \beta_3 (reward\text{-}magnitude_t * reward\text{-}probability_t) + \beta_4 reward\text{-}outcome_{t-1} + \beta_5 reward\text{-}outcome_{t-2} + \beta_6 reward\text{-}outcome_{t-3} + \beta_7 reward\text{-}outcome_{t-4} + \beta_8 reward\text{-}outcome_{t-5} + \beta_9 reward\text{-}outcome_{t-6} + \beta_{10} reward\text{-}outcome_{t-7} + \beta_{11} reward\text{-}outcome_{t-8} + \beta_{12} reward\text{-}outcome_{t-9} + \beta_{13} reward\text{-}outcome_{t-10} + \mu_0 + \varepsilon$$

The small number of subjects precluded meant that it was not possible to fit GLM3.3 with the full random-effects structure – that is, with random effects for all predictors. To assess the significance of past reward outcomes whilst also taking inter-subject variability into account, we therefore fit an iterative series of GLMs in which one previous outcome predictor was given a random-slope (e.g. in the first GLM, $reward\text{-}outcome_{(t-1)}$ was given a random slope). We assessed the statistical significance of past outcome regressors using the random effect in the relevant GLM. Fig.3.2C shows the outcome of this analysis. Given the results from GLM 3.3, we operationalised the richness of an animal's reward environment as a moving-average with a retrospective five-trial window, and the stochasticity of its environment as the standard-deviation of the reward rate over the same period:

Richness of environment

$$environment\text{-}richness_t = mean(reward\text{-}received_{t-5}: reward\text{-}received_{t-1})$$

Stochasticity of environment

$$Environment\text{-}stochasticity_t = \mathrm{sd}(richness_{t-5}: richness_{t-1})$$

We implemented the following model to assess how richness and stochasticity influenced pursue/reject decisions:

GLM3.4

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 reward\text{-}probability_t + \beta_3 environment\text{-}richness_t + \beta_4 environment\text{-}stochasticity_t + \beta_5 (reward\text{-}magnitude_t * reward\text{-}probability_t) + \beta_6 trial\text{-}number_t + \mu_0 + \varepsilon$$

where $\beta_{0\text{-}6}$ are fixed-effects, μ_0 is the by-subject random intercept and ε is an error term. The small number of subjects precluded fitting GLM3.4 with the full random-effects structure. We therefore fit an iterative series of models in which the key predictors of interest – *environment-richness* and *environment-stochasticity* – were individually afforded random effects. To test whether the *environment-richness* effect was sensitive to the time horizon of past outcomes used to operationalise the richness of the environment, we repeated GLM3.4 for two additional cases: (i) a case where the richness of the environment was operationalised as the reward outcome on only the previous trial, and; (ii) a case where the richness of the environment was operationalised as the average reward outcome on the previous ten trials (fig.3.2D). Finally, we investigated the effect of behavioural history with the following GLM:

GLM3.5

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 reward\text{-}probability_t + \beta_3 behavioural\text{-}history_t + \beta_4 environment\text{-}stochasticity_t + \beta_5 trial\text{-}number_t + \mu_0 + \mu_1 behavioural\text{-}history + \varepsilon$$

where $\beta_{0\text{-}6}$ are fixed-effects, $\mu_{0\text{-}1}$ are random effects, and ε is an error term. *Behavioural-history* is a binary variable reflecting the pursue-vs-reject decision taken on the previous trial. We did not enter *behavioural-history* and *environment-richness* regressors into the same model because these effects were correlated ($r = 0.40$).

All mixed-effects GLMs were performed in *R* and models were fit via maximum likelihood estimation as implemented in the *lme4* toolbox.

3.4.5 General Linear Model Hidden Markov Model

We used a GLM-HMM to probe time-varying patterns in behaviour that might reflect motivation-states. A GLM-HMM is an extension of the GLM that consists of two parts. The HMM part of the model assumes that time series events arise from so-called hidden states, which produce observations according to state-dependent probability distributions (*35*, *36*). These states have Markov relations to one another, meaning that the state at time *t* is determined strictly by the state at time *t-1*, and the respective probabilities of transition between states. A simple HMM of animal behaviour in this context, for example, might comprise states in which responses arise from binomial distributions with state-dependent probability parameters that reflect changes in an animal's pursue-vs-reject decisions across time.

The GLM part of a GLM-HMM parameterises state-dependent probability distributions according to predictors (*35*). The weights afforded to each predictor can vary from state-to-state, which enables the model to capture time-varying differences in an animal's decision-making process. We implemented a GLM-HMM in which pursue-vs-reject decisions were parameterised by a bias term, in addition to a predictor for the expectation-value of the reward opportunity on the current trial, and binary environment cues indicating the experimentally controlled richness and stochasticity of the current block. We were interested in motivation-states – that is, changes in an animal's baseline predisposition to pursue-vs-reject reward opportunities. We therefore let only the bias parameter(s) change between states and constrained the GLM weight on expectation-value and environment-cue parameters to be the same across states. Formally:

$$p(y_t = pursue | z_t = k,\ \boldsymbol{x_t}, \boldsymbol{w_k}) = \frac{1}{1 + e^{-\boldsymbol{w_k} \cdot \boldsymbol{x_t}}}$$

Where y ∈ {pursue, reject}^{T} is the animal's decisions across *T* trials, z_t is the HMM state at trial *t* and $z_t \in \{1,..., K\}$ such that there are *K* states, $x_t \in \mathbb{R}^M$ is a matrix containing values trial-by-trial values for *M* predictors in the GLM part of the model, and $\boldsymbol{w}_k \in \mathbb{R}^M$ represents *M* GLM weights over predictors that are specific to state *k*. Transitions between the states unfold according to a transition matrix $A \in \mathbb{R}^{K \times K}$:

$$p(z_{t+1} = k | z_j, A) = A_{j,k}$$

where transition matrix is stationary. The joint probability of behaviours and HMM states is given by:

$$p(\mathbf{y}, \mathbf{z} | \{\mathbf{x}\}_{t=1}^{T}, \{\mathbf{w}\}_{k=1}^{K}, A, \tau) = p(z_1) \prod_{t=2}^{T} p(z_{t+1} | z_t) p(y_t, \boldsymbol{x}_t)$$

Where $\tau \in \mathbb{R}^K$ is the initial state distribution such that $\tau_k = p(z_1 = k)$. The likelihood function was fit to behavioural data via Markov Chain Monte Carlo optimisation implemented in STAN (*283*). A separate GLM-HMM was fit to each individual animal's behaviour.

We probed validity of the GLM–HMM framework by fitting a series of models with *K* ∈ {1, 2, 3, 4, 5} states. To begin, we performed a five-fold cross-validation protocol in which models were iteratively fit to 80% of sessions and tested on the remaining 20% of sessions (*35*). This process was performed within each animal. The best fitting model was determined on the basis of the log-likelihood of held-out sessions (see fig.3.4). In the main text, we further report session-wise AIC scores – a well-established and easily interpretable metric of model fit that penalises model complexity, and is a demanding test for GLM–HMMs with $K \geq 2$ because additional states increase complexity exponentially (fig.3.5A).

We further validated the GLM-HMM framework using simulated data. To begin, we tested whether the fitting procedure accurately recovered the parameters used to generate simulated datasets by taking a fitted Binomial GLM (i.e. *K* =1) and 2-state GLM–HMM for an

example animal and generating 10 datasets comprising 16 sessions from each model (i.e. the typical size of the behavioural data recorded from each animal). We then implemented the fitting procedure on simulated datasets and tested (i) whether the correct number of HMM-states was recovered (1-state vs 2-states), and (ii) whether the correct parameters were identified in the 2-state case (fig.3.4). We additionally examined simulated data for qualitative patterns that were reminiscent animal behaviour (fig.3.5). After model selection, we decoded the maximum a posteriori sequence of HMM-states in each session via the Viterbi algorithm (*283*):

$$\mathbf{z}_{\boldsymbol{s}}^{*} = \ argmax_{\boldsymbol{z}1:T} p(\boldsymbol{z}_{s,1:T} | \boldsymbol{y}_{s,1:T})$$

We used simulated data from a 2-state GLM HMM to validate the Viterbi decoding algorithm (fig. 3A–C).

Decoding HMM-states with the Viterbi algorithm allowed us to quantify their impact on behaviour, which we did using the following mixed-effects GLMs with subject-identity as the grouping variable:

GLM3.6

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 reward\text{-}probability_t + \beta_3 motivation\text{-}level_t + \beta_4\ reward\text{-}magnitude_t * reward\text{-}probability_t + \beta_5 trial\text{-}number_t + \mu_0 + \mu_1 motivation\text{-}level_t + \varepsilon$$

where $\beta_{0\text{-}5}$ are fixed-effects, $\mu_{0\text{-}1}$ are by-subject random effects for intercepts and motivation-level, respectively, and ε is an error term. *Motivation-level* $\in \{0,1\}$ is a binary variable reflecting the decoded motivation-state (low vs high) occupied on trial *t*. We identified motivation-state transitions by searching for cases where decoded motivation-states were different across consecutive trials (i.e., $motivation\text{-}level_t \neq motivation\text{-}level_{t-1}$). We compared behaviour before-vs-after motivation-state transitions with the following mixed-effects GLM:

GLM3.7

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 reward\text{-}probability_t + \beta_3 before\text{-}vs\text{-}after\text{-}transition_t + \beta_4\ reward\text{-}magnitude_t * reward\text{-}probability_t + \beta_5 trial\text{-}number_t\ \mu_0 + \mu_1 before\text{-}vs\text{-}after\text{-}transition_t\ + \varepsilon$$

where $\beta_{0\text{-}5}$ are fixed-effects, $\mu_{0\text{-}1}$ are by-subject random effects for intercepts and time-from-transition, respectively, and ε is an error term. *Before-vs-after-transition* was a binary variable that grouped trials into two categories: (i) 'before-transition' trials, defined as the five trials preceding a motivation-state transition, and (ii) 'after-transition' trials, defined as the five-trials in the immediate aftermath of a transition. We performed GLM3.7 separately for low-to-high and high-to-low motivation-state transitions.

We then characterised the GLM-HMM's convergent validity by comparing decoded states with variables that should change with motivation: (i) reaction-time, and (ii) pupil-size. To do this, we implemented the following mixed-effects GLMs:

GLM3.8

$$Pupil\text{-}size_t = \beta_o + \beta_1 reward\text{-}magnitude + \beta_2 reward\text{-}probability + \beta_3 HMM\text{-}state + \beta_4 trial\text{-}number + \mu_1 HMM\text{-}state + \varepsilon$$

GLM3.9

$$log(Reaction\text{-}time_t) = \beta_o + \beta_1 reward\text{-}magnitude + \beta_2 reward\text{-}probability + \beta_3 HMM\text{-}state + \beta_4 trial\text{-}number + \mu_1 HMM\text{-}state + \varepsilon$$

Where $\beta_{0\text{-}4}$ are fixed-effects and μ_1 is a random-effect within subjects. We did not implement random intercepts (i.e. μ_0) for either model because *pupil-size* and *reaction-time* were z-scored within-subjects during pre-processing. This meant that the central tendency of within-subject distributions for *reaction-time* and *pupil-size* was the same value (i.e., 0), and it was not necessary to add terms capturing inter-subject differences in the mean.

We used changes in decoded HMM-states to examine how motivation-states evolved over time. In particular, we probed the link motivation-states and aspects of an animal's

external environment, like the recent history of reward opportunities it had been presented with. We quantified the latter by calculating a moving average of the expectation-value of reward opportunities received in the previous five trials.

Availability of rewards (Ave. EV)

$$Ave.\ EV_t = mean(EV_{t-5}: EV_{-1})$$

We used Ave. EV to predict motivation-states via the following mixed-effects Binomial GLM:

GLM3.10

$$logit(motivation\text{-}state_t) \in \{0,1\} = \beta_o + \beta_1 EV + \beta_2 Ave.EV + \beta_3 trial\text{-}number + \mu_0 + \mu_1 Ave.EV + \varepsilon$$

Where $\beta_{0\text{-}3}$ are fixed-effects and $\mu_{0\text{-}1}$ are random-effects within each subject. We used EV – the expected value of the reward opportunity on the current trial – to control for the effects of the current reward opportunity on an animal's motivation-state. We used EV rather than the separate reward-magnitude and reward-probability dimensions of offers because Ave. EV and reward-probability were meaningfully correlated (r = 0.49). This was an inherent feature of the task because reward environments were engendered by manipulating the distribution of reward-probabilities over time, which therefore caused correlations between reward-probability and retrospective summary statistics of task events like the Ave.EV. Scaling reward-probabilities by reward-magnitude as in the expectation-value formula dampened this correlation (r = 0.23), meaning that EV and Ave. EV could be part of the same regression without collinearity.

Chapter IV

Dorsal raphe nucleus controls motivation-state transitions in monkeys

4.1 Introduction

In Ch.III, monkeys exhibited two motivation-states that shaped their baseline disposition to pursue-vs-reject rewards: a high motivation-state that was occupied in rich environments with many high-value opportunities, and a low motivation state that was occupied in poor environments with many low-value opportunities. In this chapter, we examine how the brain reconciles motivation for reward-seeking with the environment in this way. We argue for a critical role for the dorsal raphe nucleus (DRN) – a phylogenetically ancient part of the brainstem that is distinguished by its serotonergic innervation of the mammalian forebrain. Although fundamental aspects of DRN's function remain elusive, there are two discernible themes in previous experiments: (i) that DRN controls changes between opponent behavioural modes, like patience and impulsivity, or exploitation and exploration, and; (ii) that DRN responds to reward-related features of an animal's milieu, like the value, valence and uncertainty of recent outcomes (*202*, *204*, *208*, *221*, *227*, *228*, *284*, *285*).

We build on this previous work by arguing that DRN controls transitions between motivation-states that reconcile behaviour with the distribution of rewards in the environment. We take advantage of the whole-brain perspective afforded by functional magnetic resonance imaging (fMRI) to show that brain activity in DRN – but no other neuromodulatory nucleus – signals transitions in motivation-states. Such comparisons are critical for identifying the different functions of neuromodulatory nuclei (*286*). Interpreting the causal significance of

fMRI activation patterns is possible when the impact of disrupting the activity can be examined (*265*, *287*, *288*). We therefore conducted an additional experiment in which we perturbed neural activity with minimally invasive transcranial ultrasound stimulation. This demonstrated that DRN, but not ventral tegmental area (VTA) – a functionally related neuromodulatory nucleus – is causally involved in motivation-state transitions (*265*, *287*, *288*). We therefore provide the first demonstration that minimally invasive modulation of DRN is possible, and a new perspective on DRN's behavioural function.

4.2 Results

4.2.1 Brain activity in DRN signals the richness of the environment and changes in motivation state

To begin, we analysed fMRI recordings of brain activity that were taken while animals performed the task reported in Ch.III (see fig.4.1 for fMRI recording apparatus). Our analysis focused on *a priori* regions of interest (ROI; fig.4.2) comprising the ascending neuromodulatory systems (ANS) – an assemblage of phylogenetically ancient nuclei which includes the serotonergic DRN, in addition to the dopaminergic ventral tegmental area (VTA) and substantia nigra (SN), the cholinergic nucleus basalis (NB), and the noradrenergic locus coeruleus (LC). We also examined the habenula (Hb) – an epithalamic nucleus with diverse subcortical connections, including to several ANS nuclei (*245*, *289*). Broadening our analysis beyond DRN provided a comparative perspective on its function, which is valuable given that reward functions have been jointly attributed to different subcortical areas. For example, the richness of the environment has been linked to both VTA and DRN (*158*, *227*, *228*, *277*), while reward uncertainty has been ascribed in various guises to DRN, LC, and BF alike (*157*, *286*, *290–292*). We were able to address this by using fMRI's wholistic perspective to compare

signals between DRN and other ANS nuclei (*293*). This was accomplished with a novel suite of fMRI acquisition and pre-processing methods that optimised blood oxygen level dependent (BOLD) signal from subcortical regions and minimised artefacts and noise sources for midbrain and brainstem regions (see §4.4: Methods).

We first examined brain activity during the pursue-vs-reject decision made on each trial (GLM4.1 & GLM4.2; see §4.4: Methods for details, and for all subsequent GLMs). We focused on activity representing the richness and stochasticity of the reward environment at this time. Importantly, the richness of the environment was correlated with an animal's behavioural history (see Ch.III, §3.2.1). We therefore performed our analysis separately for trials in which the previous encounter was (i) pursued and (ii) rejected to ensure that any patterns of brain activity identified were specific to the richness of the environment, and not related to behavioural history. We note, however, that the key aspects of DRN's relationship with the richness of the environment hold regardless of whether the data are analysed separately in this way (fig.4.3A–B).

DRN activity was negatively correlated with the richness of the environment after rejection of opportunities (one-sample two-tailed t-test; GLM4.2; $t_{DRN;\ rejected}(58)=-2.80$, $p=.034$; $t_{DRN;\ pursued}(58)=-0.37$, $p=.713$ after Bonferroni correction for multiple comparisons, as are all subsequent t-tests; fig.4.3A & 4.3B; see also fig.4.3A top panel for effect of richness of the environment in all trials [i.e. regardless of behavioural history]; see fig.4.3 legend for further discussion of timing of BOLD signal). In contrast, VTA activity was prominent following opportunity pursuits while Hb exhibited aspects of both DRN and VTA patterns (fig.4.3A–B). Despite previous evidence linking DRN with reward uncertainty, we found no relationship between the stochasticity of the environment and DRN activity (GLM4.1; $t_{DRN}(58)=1.10$, $p=.824$) and no evidence that stochasticity was represented in any other ROI (GLM4.1; see fig.4.4).

After determining that DRN, VTA and Hb represented the richness of the environment, we tested whether they were linked to the motivation-states identified with the GLM-HMM. We first quantified the relationship between brain activity and motivation-state *level* (high vs. low). This showed that high motivation-states were signalled by increased activity in VTA and Hb, but not in DRN (GLM4.4; $t_{Hb;\ motivation\text{-}level}(58)=2.34$, $p=.045$; $t_{VTA;\ motivation\text{-}level}(58)=2.82$, $p=.019$; $t_{DRN;\ motivation\text{-}level}(58)=1.75$, $p=.084$; fig.4.3C–D). We next examined activity related to motivation-state *transitions* (change vs. no-change) by identifying transition-trials on which the *maximum a posteriori* motivation-state was different from the preceding trial. For each transition-trial, we defined a corresponding transition-period spanning a 7 trial window that was centred on the transition-trial (i.e. the transition trial +/- 3 trials) – this window was based on simulations showing that decoding of motivation-state transitions was accurate over similar timescales (see fig.3.7C, §3.2.2). Brain activity on these trials showed a striking dissociation whereby DRN – but neither Hb nor VTA – signalled transitions between motivation-states (GLM4.4; $t_{DRN;\ state\text{-}transition}(58)=-3.211$, $p=.006$; $t_{Hb;\ state\text{-}transition}(58)=-1.43$, $p=.160$; $t_{VTA;\ state\text{-}transition}(58)=-1.80$, $p=.152$; fig.4.3C–D).

To confirm that the motivation-state transition effect was specific to DRN, we analysed BOLD signal from adjacent anatomical features such as the medial raphe nucleus (MRN) and fourth ventricle. There was no evidence of corresponding effects in these areas, and the motivation-state effect in DRN was, furthermore, greater than the null effects in MRN and the fourth ventricle (GLM4.4; fig.4.3F–G; $t_{DRN}(58)=-3.21$, $p=.006$, $t_{MRN}(58)=1.01$, $p=.638$, $t_{Vent.}(58)=-0.007$, $p=.994$, $t_{DRN\text{-}vs\text{-}MRN}(58)=-3.13$, $p=.003$, $t_{DRN\text{-}vs\text{-}Vent}(58)=-2.21$, $p=.031$). We then probed the temporal specificity of the DRN activity pattern by varying the position of the transition-periods examined in the fMRI analysis relative to the actual transition-trials identified by the GLM-HMM. Activity changes in DRN were most prominent when the analysed period was closely aligned with the transition-trial decoded by the GLM-HMM,

suggesting that they were indeed specific to motivation-state transition events (fig.4.3H; fig.4.5 for further detail). Finally, we parsed motivation-state transitions according to their direction (i.e. high-to-low and low-to-high). This indicated that DRN activity signalled high-to-low transitions but not low-to-high transitions (GLM4.4 $t_{DRN;\ high\text{-}to\text{-}low\text{-}transitions}$ (58)=-2.51, p=.015; $t_{DRN;\ low\text{-}to\text{-}high\text{-}transitions}$ (58)=-0.91, p=.365; fig.4.3E). Taken together, these analyses suggest that DRN – and DRN specifically – implemented negative changes in an animal's motivation-state.

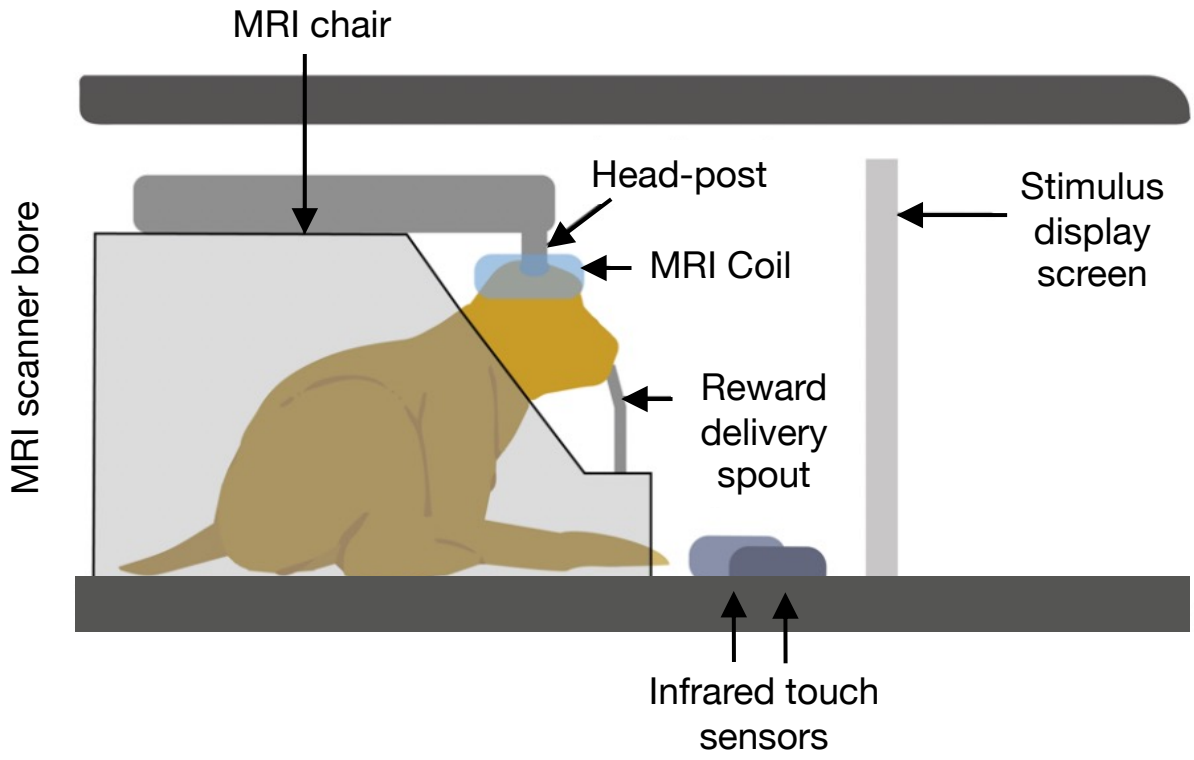


Figure 4.1. Diagram of fMRI scanning apparatus. Animals performed the task under fMRI. They were seated in MRI compatible behavioural testing chairs (Rogue Research) designed to accommodate the sphinx position, which can be occupied for long durations without discomfort. Animals were head-fixed using an extracranial head-post (Rogue Research) and positioned in front of a screen that displayed the behavioural task. Animals performed the behavioural task by making manual responses to lateralised infrared touch sensors, and received juice rewards via a delivery spout system. Pupil-size and gaze-direction data was collected with an Eyelink-1000 device positioned in front of the stimulus display screen (not pictured). Figure adapted from (*294*).

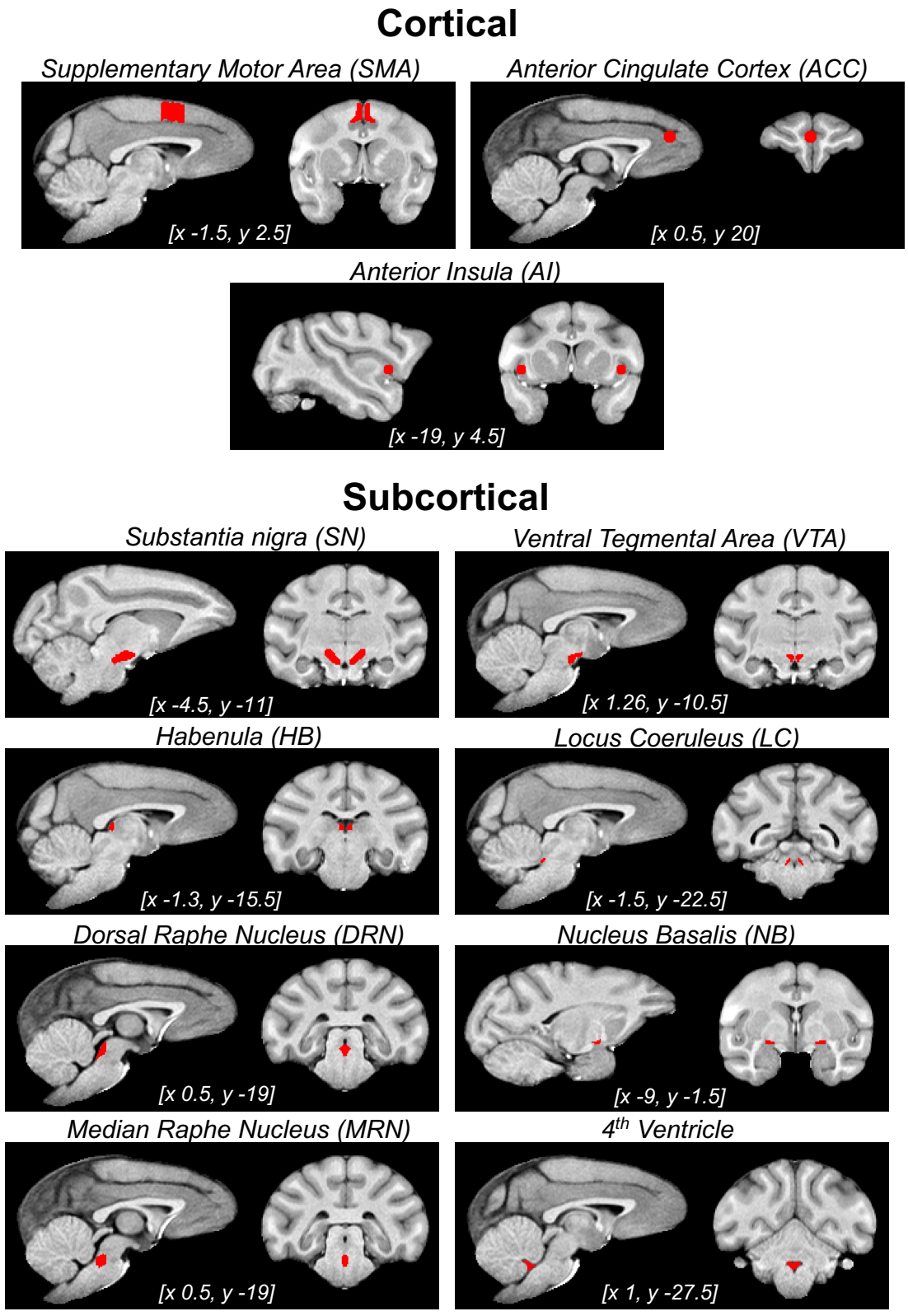


Figure 4.2. Cortical and subcortical regions of interest. An illustration of the ROIs implemented during fMRI analysis in CARET f99 macaque space. Subcortical and SMA ROIs consisted in anatomical masks that were drawn on a group structural template in CARET F99 macaque monkey space and then warped to individual structural and

functional spaces by nonlinear transformation. These masks were constructed separately by two-different assessors based on the Rhesus Monkey Brain Atlas (*69*) and then evaluated for convergence across assessors. ACC and AI ROIs were defined as 3mm spheres centred on the peak of functionally relevant activation contrasts obtained in previous studies (*227*, *264*). The medial raphe nucleus (MRN) and fourth ventricle were only used during control analyses to validate the effects observed in DRN.

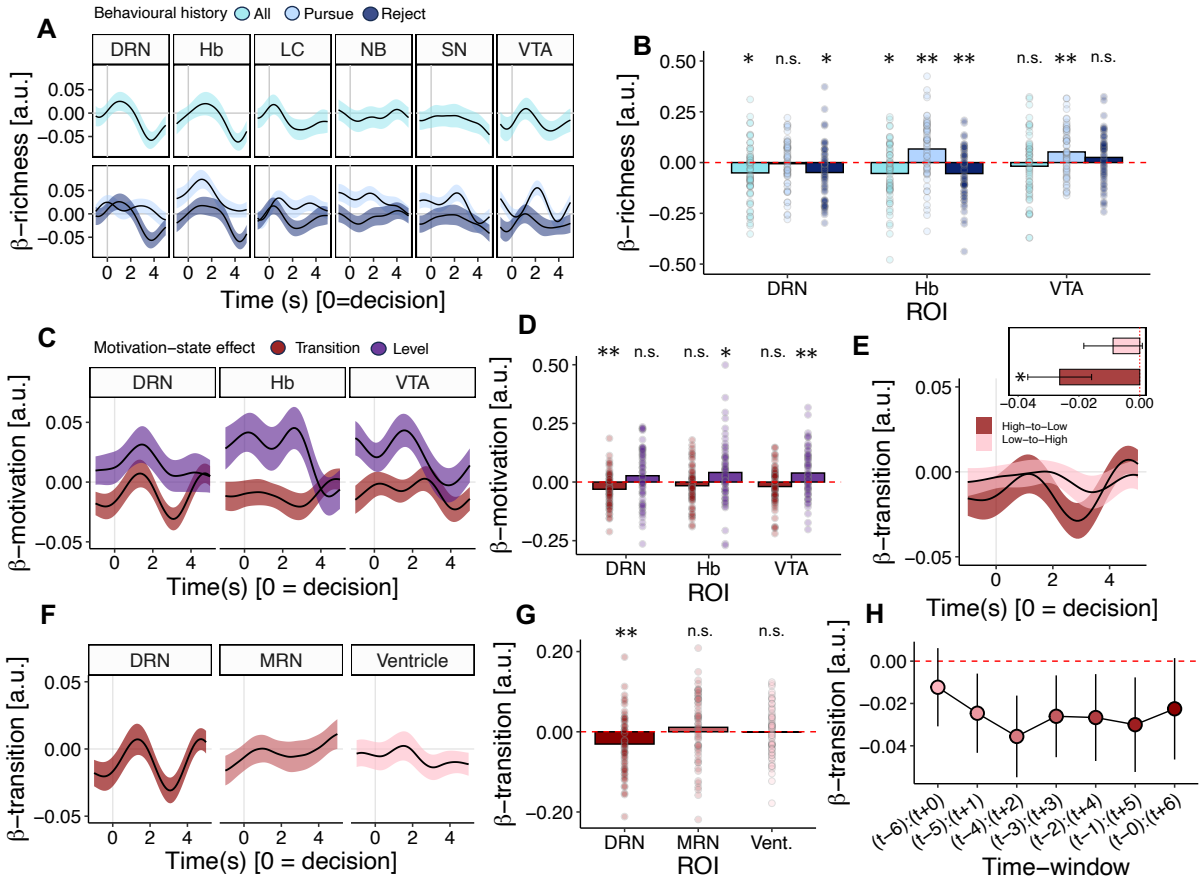


Figure 4.3 Brain activity in dorsal raphe nucleus represents the richness of the environment and changes in motivation-states. (A) Timecourse of the effect that the richness of the environment (β-richness) has on BOLD signal in subcortical ROIs. Time (x-axis) is depicted relative to the onset of reward-opportunities when pursue-vs-reject decisions are made (i.e. time=0 corresponds to decision-time). In macaque monkeys, BOLD signal reaches its maximum approximately 2-4s after the neural activity which caused it. The timing of peak β-weights in DRN, VTA and Hb is therefore consistent with brain activity that occurs at the time of decision-making. Top panel shows effects over all trials (GLM4.1) and bottom panel shows the analysis separately as a function of behavioural history (GLM4.2; see main text for explanation). Despite the partial collinearity between richness of the environment and behavioural history regressors, BOLD signal in both DRN and Hb represents richness of the environment across all trials (GLM4.1; $t_{DRN}(58)=-2.79$, $p=.036$; $t_{Hb}(58)=-2.86$, $p=.035$). **(B)** Distribution of peak effect sizes for richness of the environment (β-richness) on BOLD signal. Dots indicate peak effect sizes for individual sessions. Separating the data as a function of behavioural history revealed striking patterns in the direction of representations across regions. DRN represented the richness of the environment with negative sign when the previous reward opportunity was rejected (GLM4.2; $t_{DRN;\ rejected}(58)=-2.80$, $p=.034$; $t_{DRN;\ pursued}(58)=-$

0.37, p=.713); BOLD signal in VTA represented richness with positive sign when the previous reward opportunity was pursued (GLM4.2; $t_{VTA;\ pursued}(58)=3.30$, p=.009; $t_{VTA;\ rejected}(58)=1.60$, p=.350); finally, Hb recapitulated the sign and behavioural contingency of DRN and VTA with distinct positive and negative signals on trials after pursuits and rejections on the previous trial, respectively (GLM4.2; $t_{Hb;\ pursued}(58)=3.32$, p=.009; $t_{Hb;\ rejected}(58)=-3.37$, p=.008; $t_{Hb;\ rejected\text{-}vs\text{-}pursued}(58)=4.95$, p<.001). A two-way ANOVA followed by pairwise comparisons confirmed that signals in DRN and VTA after previous trial rejections and pursuits were different from one another but not from the corresponding Hb effect ($F_{ROI}(2, 348)=8.106$, p<.001, $F_{Behav.\text{-}history}(1, 348)=22.305$, p<.001; $F_{ROI\text{-}by\text{-}Behav.\text{-}history}(2, 348)=4.521$, p=.012; $t_{VTA\ vs\ DRN;\ pursued}(58)=2.501$, p=.015; $t_{DRN\ vs\ VTA;\ rejected}(58) =-3.403$, p<.001; $t_{VTA\ vs\ Hb;\ pursued}(58) =0.650$, p=.518; $t_{DRN\ vs\ Hb;\ rejected}(58)=-0.214$, p=.831). **(C)** Timecourse of the effect of motivation-state level (high-vs-low motivation state; purple) and motivation-state transitions (red) on BOLD activity in ROIs that represented the richness of the environment. Given haemodynamic delay (see (A)), the neural activity underlying the motivation-state transition signal in DRN is likely to occur in the ITIs preceding trials on which transitions occur, meaning that it is not an epiphenomenon of action or outcome related processes. **(D)** Distribution of peak effect sizes for motivation-state level (purple) and motivation-state transitions (red). Only DRN represents transitions between motivation-states. Dots indicate peak effect sizes for individual sessions. **(E)** The timecourse of the effect that low-to-high (red) and high-to-low (pink) motivation-state transitions have on DRN BOLD activity. The effect is strongest for high-to-low state transitions. Inset panel shows distribution of peak effect sizes. **(F)** and **(G)** Motivation-state transitions are signalled in BOLD activity in DRN but not in neighbouring areas like the median raphe nucleus (MRN) and 4th ventricle. This suggests that the pattern was DRN-specific **(H)** Shifting the position of transition-periods in the fMRI analysis relative to the transition points decoded in behaviour indicated that activity changes were specific to decoded transition-times. X-axis indicates position of transition-period relative to decoded transition time (where t=0 represents decoded transition trial). Dots and error bars indicate mean and 95% confidence interval of effect-sizes for each transition-window. See fig. 4.5 for further detail. In timecourse graphs (A, C, E and F), lines and shadings show the mean and standard error (SE) of the β weights across the sessions, respectively. Effect-size graphs (B, D, E inset, and G) show size of the peak regression weight in the timecourse selected using an unbiased leave-one-out procedure (see methods). Dots indicate peak regression weight from individual sessions. Error bars in E inset show SD of session-

wise regression weights. Dots and error bars in (H) show mean ± 95% confidence interval of regression weights for different transition-period windows.

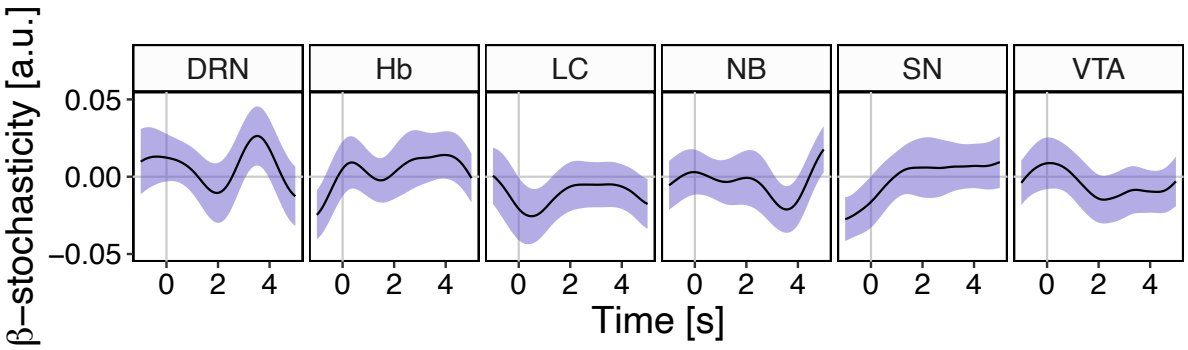


Figure 4.4. No evidence of environment-stochasticity representations in ROIs. The time-course of the effect of reward-stochasticity on BOLD activity in subcortical ROIs. Epochs are time-locked to decision-making. Reward-stochasticity had a modest influence on behaviour but there was no evidence that it was represented in the brain activity of ROIs (GLM4.1). Lines and shadings show the mean and standard error (SE) of the β weights across the sessions, respectively.

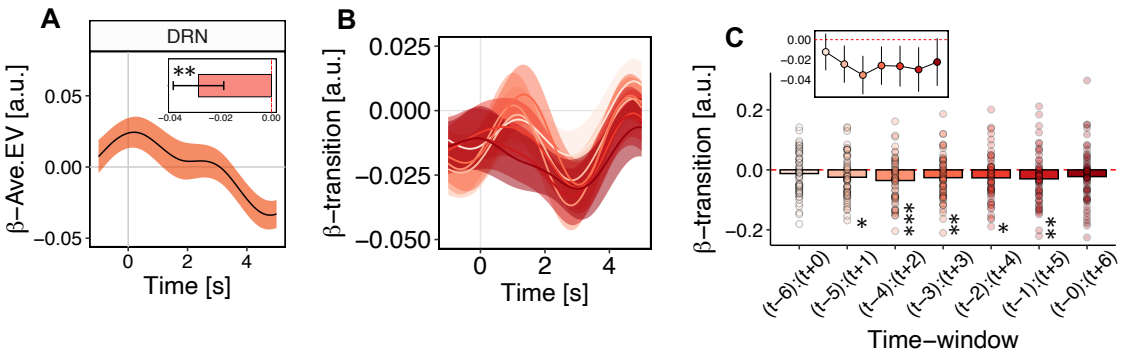


Figure 4.5. Further examination of motivation-state transitions in dorsal raphe nucleus. (A) Our analysis of behaviour indicated that animals were more likely to occupy high motivation-states as the average expected-value of recently encountered rewards increased (fig. 3.7H). Given that BOLD activity in DRN represented transitions between motivation-states, we tested whether DRN also encoded the average expected-value of recent reward opportunities – the idea being that DRN might control changes in motivation-states specifically in relation to the distribution of available rewards. This was, indeed, the case – the average expected-value of the preceding five reward opportunities exerted a negative effect on BOLD activity in DRN, comparable to the effect of the richness of the environment ($t_{EV}(58)$ =-2.89, p=.005; GLM4.5). Inset panel shows distribution of peak regression coefficients. **(B&C)** We initially tested the effect of motivation-state transitions on brain activity in a symmetric 7 trial window centred on the transition-trial identified by the GLM–HMM (i.e. the transition trial +/- 3 trials). This indicated that transitions were signalled by BOLD activity in DRN, as reported in the main text. We confirmed that this effect was not an artefact of our initial time-window selection by iterating the analysis over 7-trial windows with different positions relative to transitions trials identified with the GLM–HMM. This showed that the effect was robust to window position ($t_{(t-5):(t+1)}(58)$=-2.63, p=.011; $t_{(t-4):(t+2)}(58)$=-3.69, p<.001; $t_{(t-3):(t+3)}(58)$=-2.70, p=.009; $t_{(t-2):(t+4)}(58)$=-2.60, p=.012; $t_{(t-1):(t+5)}(58)$=-2.69, p=.009; $t_{(t-0):(t+6)}(58)$=-1.88, p=.065; GLM4.6). Note that the windows featuring the least overlap with decoded motivation-transitions show non-significant effects ($t_{(t-6):(t+0)}(58)$=-1.34, p=.180; $t_{(t-0):(t+6)}(58)$=-1.88, p=.065). Inset panel in C shows mean ± 95% confidence interval of transition effect across time-windows.

4.2.2 Non-invasive disruption of DRN perturbs motivation-state transitions

fMRI recordings linked DRN to behaviour in two complementary ways; (i) DRN represented the richness of the environment, which modulated an animal's pursue-vs-reject decisions for specific reward opportunities and (ii) DRN represented transitions between motivation-states, and therefore patterns of behaviour over multi-trial timescales. We tested DRN's causal contribution in these respects with a second experiment using transcranial ultrasound stimulation (TUS; fig.4.6A). TUS is a minimally invasive and reversible technique that disrupts brain activity via kinetic interactions between focused ultrasound waves and the neuronal and astrocyte membrane that, in turn, induce changes in adjacent synapses (*256*, *257*). Short TUS trains of the type used here are known to produce short term changes in neural activity by inducing N-methyl-d-aspartate (NMDA)-dependent plasticity (*258–262*). As a result, a region targeted with TUS alters its responsiveness to activity in interconnected areas, while non-targeted areas show no such change (*265*).

We used an offline TUS protocol that modulates neural activity for over an hour after sonication (*264*, *265*, *295*). We compared DRN-TUS to three control conditions: a sham condition (sham-TUS), an active cortical control condition (superior temporal sulcus; STS-TUS), and an additional active subcortical control condition (VTA). We chose VTA as a subcortical control region because it was implicated in motivation-states during fMRI recordings in distinct ways to DRN, thereby allowing us to test the specificity of DRN function. We first performed simulations of acoustic wave propagation under the TUS protocol, which established that selective perturbation of DRN and VTA was feasible despite their diminutive size and anatomical location (fig.4.6B, fig.4.7). We then empirically tested the TUS protocol's anatomical specificity by measuring changes in functional connectivity between DRN, VTA

and a series of key brain areas before-vs-after DRN-TUS and – as a control – before-vs-after VTA TUS. This showed striking and target-specific changes functional connectivity. After DRN-TUS, DRN's usual patterns of coactivation with key interconnected regions changed, but VTA connectivity did not (fig.4.6C). Conversely, after VTA-TUS, VTA-connectivity changed but DRN connectivity did not (fig.4.6C). In combination with the sonication simulations, this indicated that TUS induced, safe, efficacious, and highly localised disruption of DRN and VTA activity, analogous to previous investigations of TUS in cortical brain regions (*264*, *265*, *295*).

We tested the same monkeys from the fMRI experiment on five sessions of the behavioural task per TUS condition in a counterbalanced and pseudorandomised order (see §4.4: Methods). We first investigated how TUS modulated the influence of (i) richness of the environment, and (ii) behavioural history, which were the key factors driving the pursue-vs-reject decisions made on each trial of the fMRI experiment. DRN-TUS diminished the effect of richness of the environment relative to all other conditions (GLM4.9; $\beta_{\text{DRN-TUS vs sham-TUS*richness}}$=-0.18, SE=0.05, p<.001; $\beta_{\text{DRN-TUS vs STS-TUS*richness}}$=-0.12, SE=0.06, p=.025; $\beta_{\text{DRN-TUS vs VTA-TUS*richness}}$=-0.12, SE=0.06, p=.035; fig.4.6D). Similarly, DRN-TUS clearly reduced the effect of behavioural history relative to sham-TUS and STS-TUS, and produced a marginal difference relative to VTA-TUS (GLM4.10; $\beta_{\text{DRN-TUS vs sham-TUS*behave-history}}$=-0.42, SE=0.11, p<.001; $\beta_{\text{DRN-TUS vs STS-TUS*behave-history}}$=-0.24, SE=0.11, p=.027; $\beta_{\text{DRN-TUS vs VTA-TUS*behave-history}}$=-0.20, SE=0.11, p=.057; fig.4.6E). There were no differences between active control conditions (STS-TUS and VTA-TUS) and sham-TUS for the richness of the environment and behavioural history effects (GLM4.9 & fig.4.6D; $\beta_{\text{VTA-TUS vs sham-TUS*richness}}$=-0.06, SE=0.06, p=.282; $\beta_{\text{STS-TUS vs sham-TUS*richness}}$=-0.06, SE =0.05, p=.306; GLM4.10 & fig.4.6E; $\beta_{\text{VTA-TUS vs sham-TUS*behavioural-history}}$=-0.20, SE=0.10, p=.060; $\beta_{\text{STS-TUS vs sham-TUS*behavioural-history}}$=-0.18, SE=0.11, p=.090). DRN, therefore, was specifically and causally involved in the influence exerted by the richness of the environment and behavioural history on pursue-vs-reject decisions made in the present.

Next, we applied the GLM-HMM to behavioural data from the TUS experiment (see §4.4: Methods). Our analysis focused on motivation-state transitions, which were the distinguishing feature of DRN activity in fMRI recordings. These recordings suggested that DRN was specifically involved in high-to-low motivation-state transitions, but we first performed the analysis in a direction-neutral way due to the interdependence of time-series observations: because animals typically began the task in a high-motivation state (p(init-state=high)=0.70), reducing the likelihood of high-to-low transitions would necessarily reduce the frequency of low-to-high transitions in a fixed-length time series, even if the underlying low-to-high transition probability was unchanged. The analysis showed a striking pattern whereby DRN-TUS reduced the likelihood of motivation-state transitions relative to all other TUS conditions (GLM 4.11; $\beta_{\text{DRN-TUS vs sham-TUS}}$=-1.08, SE=0.27, p<.001; $\beta_{\text{DRN-TUS vs VTA-TUS}}$=-0.71, SE=0.29, p=.013; $\beta_{\text{DRN-TUS vs STS-TUS}}$=-0.57, SE=0.29, p=.051; fig.4.6F–G). There was no effect of VTA-TUS relative to sham-TUS (GLM4.11; $\beta_{\text{VTA-TUS vs sham-TUS}}$ = -0.37, SE = 0.22, p = .090). There was a moderate effect of STS-TUS compared to sham-TUS but this was outweighed by the fact that DRN-TUS reduced the likelihood of transitions even relative to STS-TUS (GLM4.11; $\beta_{\text{STS-TUS vs sham-TUS}}$=-0.51, SE=0.23, p=.022; $\beta_{\text{DRN-TUS vs STS-TUS}}$=-0.57, SE=0.29, p=.051). DRN-TUS was, moreover, the only condition that reduced high-to-low transitions specifically (GLM4.12; $\beta_{\text{DRN-TUS vs sham-TUS}}$=-0.77, SE=0.32, p=.016; $\beta_{\text{VTA-TUS vs sham-TUS}}$=-0.33, SE=0.28, p=.238; $\beta_{\text{STS-TUS vs sham-TUS}}$=-0.43, SE=0.29, p=.139). Finally, DRN-TUS caused animals to spend more time in the high-motivation state relative to sham-TUS (GLM4.13; $\beta_{\text{DRN-TUS vs sham-TUS}}$=0.28, SE=0.07, p<.001), consistent with the idea that DRN-TUS prevented animals from transitioning to low motivation states.

Finally, we sought to specify how DRN controlled transitions between motivation states. To do so, we returned to an earlier analysis demonstrating that motivation-states were matched to the availability of rewards: in brief, animals were more likely to occupy high-

motivation states when there were many high-value rewards available, analogous to the way that animals pursued more opportunities in rich environments (fig.3H). Given DRN's critical involvement in the latter phenomenon, we asked whether its control of motivational-state transitions was mediated by the availability of rewards. Consistent with this view, DRN-TUS diminished the relationship between the availability of rewards and motivation-states relative to all other TUS conditions (GLM4.13; fig.4.6H; $\beta_{\text{DRN-TUS vs sham-TUS*Ave. EV}}$=-0.28, SE=0.07, p<.001; $\beta_{\text{DRN-TUS vs STS-TUS*Ave. EV}}$=-0.29, SE=0.07, p<.001; $\beta_{\text{DRN-TUS vs VTA-TUS*Ave. EV}}$=-0.30, SE=0.07, p<.001). There were no effects of VTA-TUS or STS-TUS relative to sham-TUS ($\beta_{\text{VTA-TUS vs sham-TUS*Ave. EV}}$=-0.03, SE=0.06, p=.610; $\beta_{\text{STS-TUS vs sham-TUS*Ave. EV}}$=-0.05, SE=0.07, p=.499). Intriguingly, DRN-TUS's influence was more prominent for the link between low-value reward environments and low-motivation states, which dovetailed with fMRI signals implicating DRN specifically in high-to-low motivation-state transitions (GLM4.13; for Ave.EV< $\mu_{\text{Ave.EV}}$ $\beta_{\text{DRN-TUS*Ave. EV}}$=-0.52, SE=0.19, p=.006; for Ave.EV ≥ $\mu_{\text{Ave.EV}}$ $\beta_{\text{DRN-TUS*Ave. EV}}$=-0.13, SE=0.15, p=.403). Taken together, these results suggest that DRN has a fundamental role in ensuring that an animal's motivation-state is appropriate to the distribution of rewards in the environment.

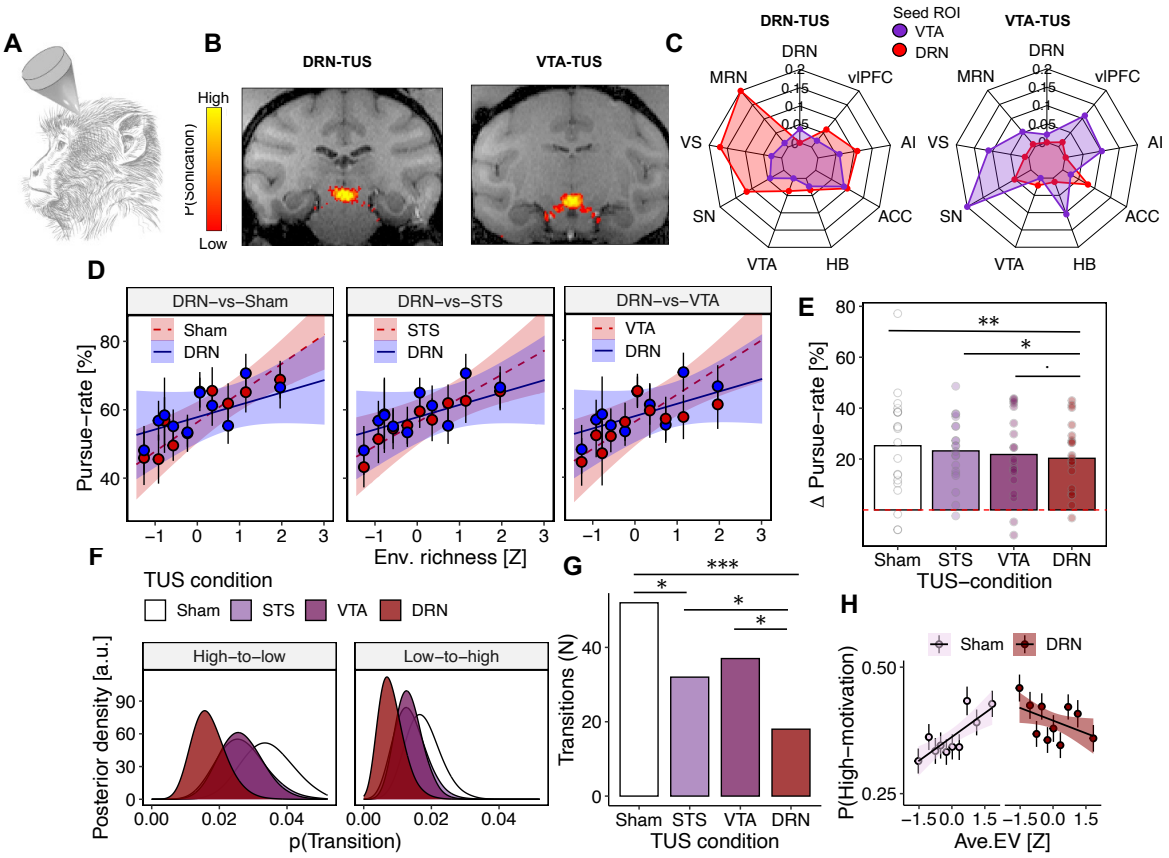


Figure 4.6. Non-invasive disruption of dorsal raphe nucleus shows causal role motivation-state transitions. (A) We investigated DRN's causal role in behaviour with transcranial ultrasound stimulation (TUS) – a minimally invasive form of brain stimulation suited to subcortical brain regions. **(B)** Simulating the propagation of acoustic waves under the TUS protocol suggested that sonication was target-specific after DRN-TUS and VTA-TUS trajectories. Colour map indicates probability of neuromodulation after bilateral TUS, calculated as the sum of stimulation intensity (Isppa) maps across two consecutive stimulations delivered over left and right hemispheres, respectively (see fig.4.7). Combined impact probability maps are overlaid on a standard F99 brain. **(C)** DRN-TUS (left) disrupted DRN (red) connectivity with key brain regions whilst leaving VTA (blue) connectivity intact. Conversely, VTA-TUS disrupted VTA connectivity whilst leaving DRN connectivity intact. Radial axis shows absolute-value of the difference in connectivity between seed and connectivity ROIs pre-vs-post TUS (see §4.4: Methods for details). ROIs were ventrolateral prefrontal cortex (vlPFC), anterior insular (AI), anterior cingulate cortex (ACC), habenula (Hb), ventral tegmental area (VTA), substantia nigra (SN), ventral striatum (VS), median raphe nucleus (MRN) and dorsal raphe nucleus (DRN). **(D)** DRN-TUS (left) attenuated the effect that the richness of an animal's environment had on behaviour relative to sham-TUS (left), STS-TUS (middle), and VTA-

TUS (right). Points and whiskers indicate mean ± SEM pursue-rate in decile bins of richness-of-the-environment in all animals. **(E)** DRN-TUS attenuated the effect of behavioural history on current behaviour relative to sham-TUS and STS-TUS, and marginally with respect to VTA-TUS. Y-axis indicates difference in pursue-rate as a function of behavioural history, coded as [pursue-rate(behav.history==pursue) – pursue-rate(behav.history==reject)]. Positive values therefore indicate that animals are more likely to pursue reward opportunities if they pursued the previous opportunity on the previous trial. Points indicate differences in mean pursue-rates for each animal in each session. **(F)** The posterior-probability of state-transitions in GLM-HMMs fit to data from each TUS condition. The central tendency of DRN-TUS posteriors is lower than for to STS-TUS, VTA-TUS and sham-TUS. **(G)** Fewer decoded motivation-state transitions occur in DRN-TUS sessions relative to sessions from other TUS conditions. Data are shown as sum of all motivation-state transitions occurring in each TUS-condition. **(H)** DRN-TUS diminishes the relationship between that the average value of recent opportunities and an animal's motivation-state level. Dots and whiskers indicate mean ± SEM of high-motivation state occupancy in deciles of average expected value (Ave. EV).

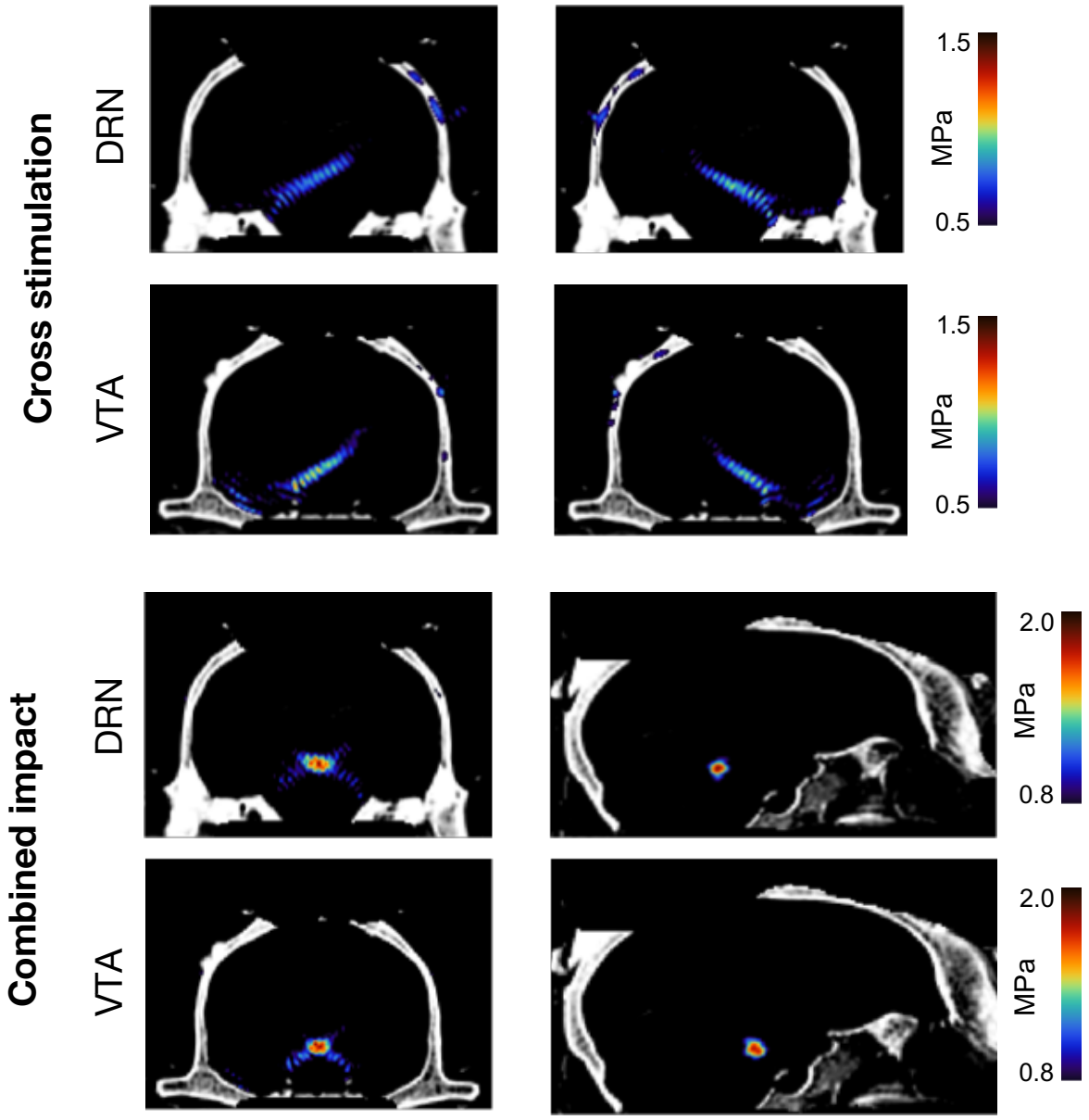


Figure 4.7. Simulation of the propagation of acoustic waves produced by the TUS protocol targeting DRN and VTA. Simulated focused ultrasound peak intensities and spatial distributions from successive left and right DRN / VTA stimulation. The skull was estimated from pseudo-CT images obtained from each monkey using a Black Bone MRI sequence (*296*). The maximum spatial-peak pulse-averaged intensity (Isppa) at the acoustic focus point was 24.4 W/cm2 (0.86 MPa) for the left DRN target, 21.5 W/cm2 (0.80 MPa) for the right DRN target, 30.3 W/cm2 (0.95 MPa) for the left VTA target and 19.9 W/cm2 (0.77 MPa) for the right VTA target. The combined impact is calculated as the sum of stimulation intensity map across the two consecutive stimulations delivered over the left and right hemisphere for DRN and VTA targets, respectively. The simulation

procedure is described in the §4.4: Methods and previous studies (*297*). The simulated data shown here is from monkey M2.

4.2.3 A cortico-subcortical circuit for reconciling behaviour with the environment

In a final analysis, we examined interactions between DRN and other brain regions related to its behavioural function. To do so, we first reverted to the analysis of brain activity and behaviour which did not rely on the GLM-HMM framework (fig.4.3A–B). Here, DRN, VTA and Hb represented an animal's environment with complementary patterns of activity, which suggested that they might form a circuit for reconciling decisions with the environment. To test the circuit hypothesis, we expanded our purview to functionally relevant cortical ROIs in supplementary motor area (SMA), anterior cingulate cortex (ACC) and anterior insula (AI) based on previous work implicating these regions in behavioural change (*227*, *264*, *298–300*). Only AI activity signalled the richness of the environment and with the distinctive contingency on behavioural history as seen in subcortical ROIs, and we therefore retained AI as a cortical ROI for connectivity analyses (see fig.4.8).

Next, we asked which ROIs coded the pursue-vs-reject decision on each trial – in other words, which ROIs might be the output stage of the decision-making circuit. This demonstrated that BOLD activity time-locked to decision-making in both Hb and AI represented the contrast between pursue-vs-reject decisions (GLM4.3; $t_{pursue;\ Hb}(58)=3.24$, $p=.011$; $t_{pursue;\ AI}(58)=2.78$, $p=.028$; fig.4.9). We then conducted psychophysiological interaction (PPI) analyses to probe changes in pairwise connectivity between regions as a function of the environment (*301*). PPIs mirrored the patterns observed earlier: connectivity between DRN and AI increased as a function of richness of the environment after *rejections* (GLM4.6; $t_{PPI(DRN\text{-}AI\ by\ richness);\ rejected}(58)=2.84$, $p=.006$; fig.4.10Ai) and increased between VTA and AI after *pursuits* (GLM 4.6; $t_{PPI(AI\text{-}VTA\ by\ richness);\ pursued}(58)=2.14$, $p=.036$; fig.4.10Aiii). Adopting the GLM-HMM

analysis approach showed a corresponding pattern of results (GLM4.7; figs.4.10A-i & 4.10A-iii). Connectivity between Hb and AI was not modulated by the richness of the environment but instead by the pursue/reject decision taken on each trial, consistent with a role in translating motivation to action (*84*) (GLM4.8 $t_{PPI(AI\text{-}Hb\ by\ action)}(58)=2.25$, $p=.028$; fig.4.10A-ii).

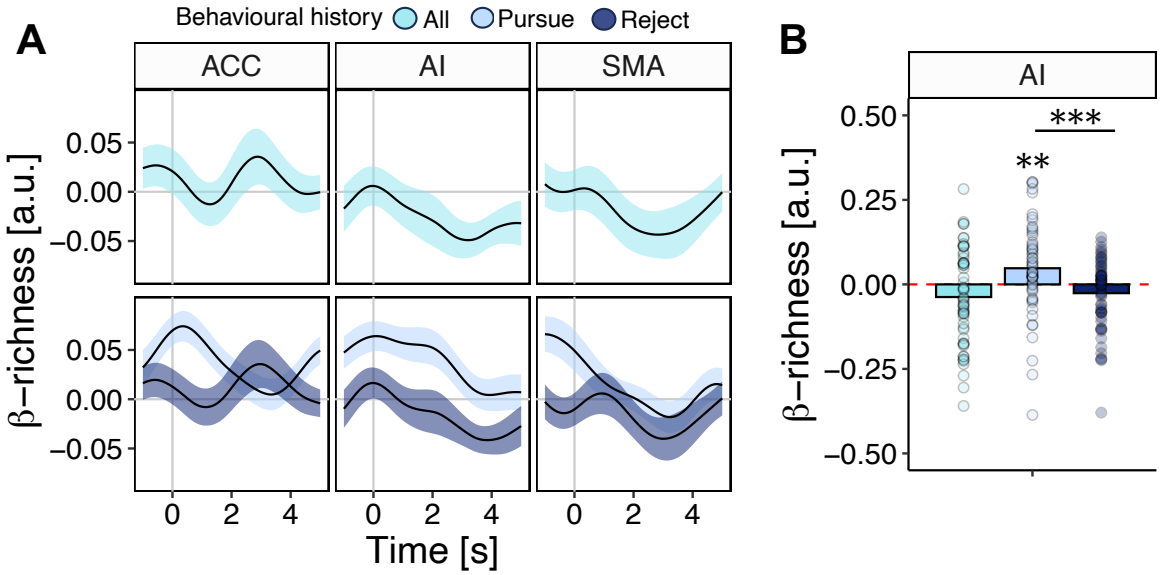


Figure 4.8. Anterior insula represents the richness of the environment with the same pattern as subcortical ROIs. The time-course (**A**) and peak-regression-coefficients (**B**) of the relationship between richness of the environment and BOLD signal in cortical ROIs. All epochs are time-locked to decision-making. AI is the only cortical ROI to represent an animal's environment as a function of behavioural history – i.e., the pursue-vs-reject decision taken on the previous trial (GLM4.2; $t_{AI;\ rejected}(58)=-1.94$, $p=.057$; $t_{AI;\ pursued}(58)=3.02$, $p=.004$; $t_{AI;\ pursued\text{-}vs\text{-}rejected}(58)=3.73$, $p<.001$; Holm-Bonferroni correction not applied).

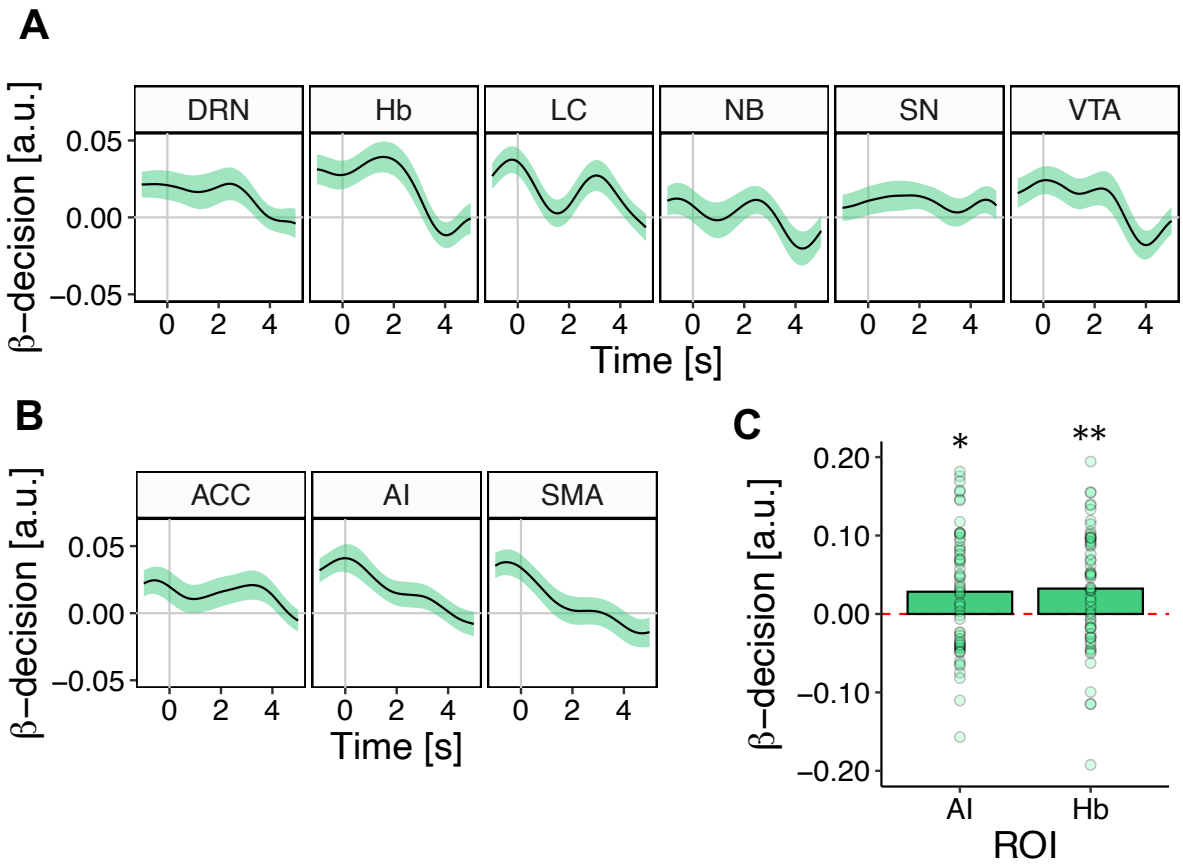


Figure 4.9. Habenula and anterior insula encode pursue-vs-reject decisions. The time-course and peak regression-coefficients of the relationship between pursue-vs-reject decisions and BOLD signal in subcortical **(A)** and cortical **(B)** ROIs. Epochs are time-locked to decision-making. **(C)** Pursue decisions were signalled by increased BOLD activity in Hb ($t_{HB}(58) = 3.25$, $p = .012$) and AI ($t_{AI}(58) = 2.78$, $p = .022$). No other ROIs represented pursuit decisions.

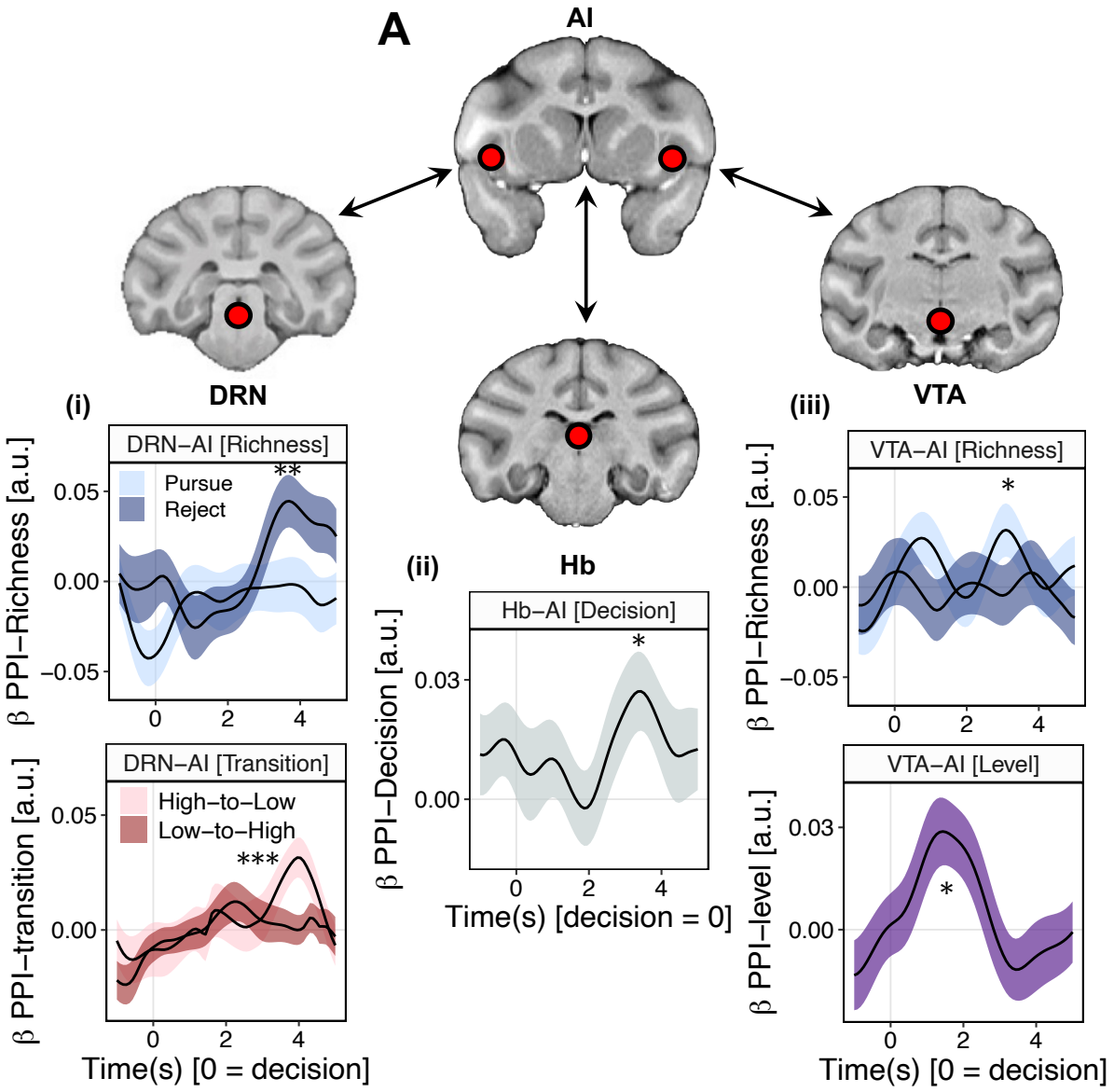


Figure 4.10. A circuit for reconciling motivation with the environment. (A) PPIs indicated the following patterns of cortico-subcortical connectivity. **(i)** Connectivity between DRN and AI increased as a function of the richness of the environment in the aftermath of rejections (GLM4.6; upper panel), and as a function of high-to-low motivation-state transitions (lower panel; GLM4.7; $t_{PPI(high\text{-}to\text{-}low)}$=3.64, p<.001), consistent with DRN's fMRI signals. **(ii)** Connectivity between AI and Hb increased as a function of the pursue-vs-reject decision taken on the current trial (GLM4.8). **(iii)** In contrast to DRN**,** connectivity between VTA and AI increased as a function of the richness of the environment in the aftermath of pursuits (GLM4.6; upper panel), and during high-motivation states (GLM3.7; $t_{motivation\text{-}level}$=2.47, p=.016).

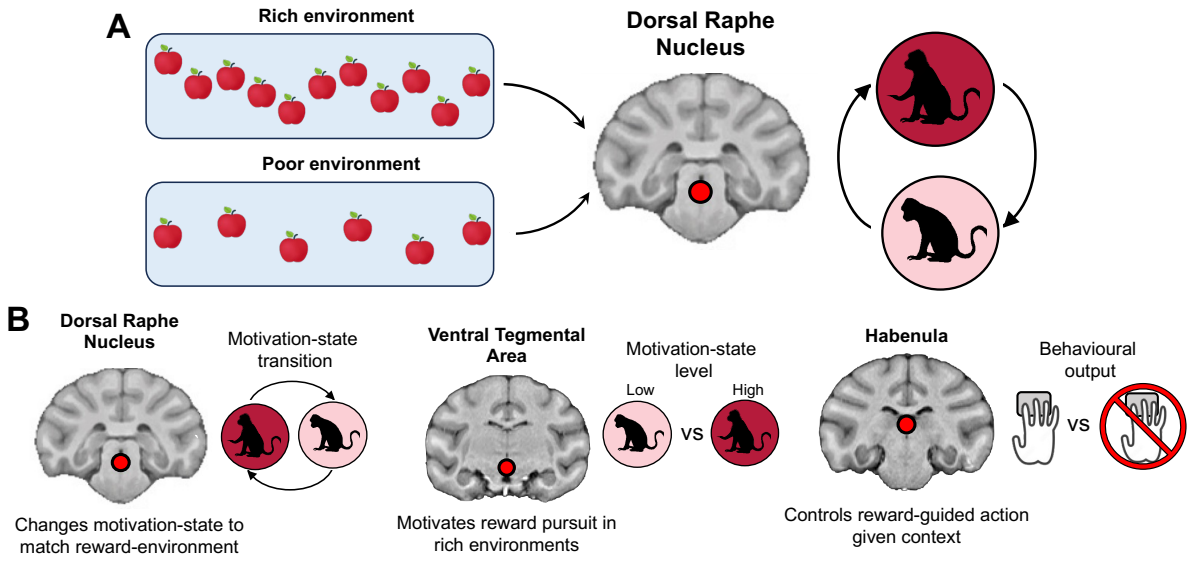


Figure 4.11. Dorsal raphe nucleus controls motivation-state transitions as a function of the environment. (A) We provide converging evidence from recording and disruption methods that DRN controls transitions between motivation-states, and that its control of these transitions reconcile an animal's motivation state with the richness of the environment. **(B)** We thereby distinguish DRN's function from interconnected subcortical nuclei which are also implicated in motivation. In combination with previous studies, we suggest a complementary scheme in which DRN (left) controls transitions between motivation states, VTA (middle) implements and maintains the changes in reward sensitivity that characterise motivation-states, and habenula (right) integrates an animal's internal and external context to control its behavioural output.

4.3 Discussion

Here, we provide converging evidence that DRN controls changes in motivation for reward-seeking. We used fMRI recordings to establish that DRN activity exhibits two distinctive patterns: (1) a pattern that represents the richness of the environment, particularly during periods of behavioural inactivity, and; (2) a pattern that represents transitions between temporally extended motivation-states, as identified with a GLM–HMM (see Ch.III). We probed the causal function of these patterns with first-of-its-kind minimally invasive DRN disruption using TUS. Disrupting DRN diminished the relationship between an animal's behaviour and the richness of the environment, and reduced the frequency of motivation-state transitions. This suggests that DRN causally controls changes in motivation for reward-seeking.

Our findings dovetail with two prevailing views on DRN function: the first view emphasises DRN's role in changes of behavioural strategy, for example between exploration and exploitation, or patience and impulsivity (*208*, *238*, *302*); the second view has linked DRN activity to statistical aspects of rewards, like the general value and/or uncertainty of the options characterising an environment (*202*, *219*, *221*, *227*, *228*, *231*). We observed similar relationships between brain activity and behaviour in this experiment, and the co-occurrence of both phenomena suggests that they are fundamentally related. In support of this interpretation, we found that disrupting DRN prevented animals from changing their motivation-state to match the recent distribution of rewards, which was an important motif in their usual behaviour. Taken together, these results suggest that previous observations regarding DRN function might in fact be different expressions of a common mechanism, whereby DRN implements behavioural change *specifically by* representing the distribution of rewards in the surrounding world (fig.4.11).

Several previous studies have linked DRN activity to uncertainty during learning paradigms, in which there is corresponding uncertainty about stimulus-outcome or action-outcome contingencies (*204*, *234*, *303*). We observed no evidence of uncertainty related DRN activity here, although this is likely attributable to features of the experimental design. Firstly, the reward-probability and reward-magnitude of the reward opportunity on each trial were directly cued, meaning that animals had complete knowledge of its expectation-value. This is fundamentally unlike standard learning paradigms, which require animals to infer reward values from experience. The closest approximation of uncertainty in the present task was controlled changes in the stochasticity of reward opportunities from one trial to the next. Stochasticity had a subtle influence on behaviour at best, befitting its null relationship with brain activity.

It is perhaps surprising that VTA was not more prominently involved in behaviour given its well-documented links with reward and motivation. However, many of these links pertain to reinforcement learning, and as described above, the task did not require reinforcement learning *per se*. It therefore obviated a cardinal function of the dopaminergic nuclei (*160*). The patterns of VTA activity we observed instead appeared to represent (i) the richness of the environment, particularly in the aftermath of reward-pursuits, and (ii) trial-by-trial fluctuations in motivation-state *level*, as distinct from the motivation-state *transitions* linked to DRN. These patterns are broadly consistent with auxiliary functions for the dopaminergic nuclei in signalling reward-value, and in motivating behaviour in proximity to specific rewards (fig.4.11) (*277*).

fMRI's wholistic perspective on brain activity enabled us to compare how DRN coordinated with other regions to control behaviour. These comparisons suggested that Hb is a key nucleus in this regard. Unlike other subcortical ROIs, Hb activity signalled the pursue-vs-reject decisions that animals ultimately made, in addition to the factors determining these

decisions like the richness of the environment. This is consistent with an emerging theory that Hb integrates motivationally-salient information from a diverse range of afferent sources to facilitate motor control via the basal ganglia (fig.4.11) (*84*, *254*, *304*, *305*). Intriguingly, although Hb is renowned for its inhibitory control over dopaminergic neurons in the midbrain, reward- and motivation-evoked Hb and VTA activity had the same sign in the present results. It is difficult to specify why this is the case. Previous recordings of Hb and VTA activity with fMRI show similar patterns, suggesting that it might be a feature of BOLD signal (*84*).

Finally, AI emerged as a critical cortical region for reward-seeking decisions, consistent with previous fMRI recordings in monkeys and human volunteers (*84*, *227*). AI has rich reciprocal connections with many subcortical regions, including to DRN, VTA, and Hb, specifically (*145*, *306*, *307*). Given the computational complexity of tracking rewards over multiple timepoints, AI is likely an important source of reward-related information to these nuclei (*84*, *145*, *227*, *306–308*). Although it is difficult to specify directions of influence from fMRI recordings, we observed patterns of functional connectivity between AI-DRN, AI-VTA, and AI-Hb that are consistent with this hypothesis. Cortico-subcortical interactions of this kind echo a common motif whereby information is first synthesised in the cortex and then transmitted to subcortical nuclei, which have wide-ranging efferent connections that can orchestrate brain-wide activity (*84*, *254*).

4.4 Methods

4.4.1 Experimental model and subject details

The experiments were performed with same four male rhesus macaques described in Ch.III, §3.4.1. All procedures were conducted under a license issued by the UK Home Office in accordance with the Animal (Scientific Procedures) Act 1986 and the European Union guidelines (EU Directive 2010/63/EU).

4.4.2 Behavioural training

Training and testing for the fMRI experiment is described in Ch.III, §3.4.2. Training for the transcranial ultrasound stimulation (TUS) experiment involved re-acclimatising animals to the behavioural task using a version with reduced go-cue and action-outcome delays. Delays were carefully increased until animals reached the same delay durations implemented in the fMRI experiment. Testing for the TUS experiment began when the proportion of pursued trials per session was stationary over consecutive sessions, similar to the criterion for beginning fMRI testing.

4.4.3 Behavioural task and analysis of behavioural data

The behavioural task used for both fMRI and TUS experiments is described in Ch.III, §3.4.3. The analysis of behavioural data for the fMRI experiment is also described in Ch.III,

§3.4.4–5. Analysis of behavioural data for the TUS experiment involved additional modelling described later in this chapter (§4.4.9)

4.4.4 Acquisition, reconstruction and pre-processing of MRI data

During collection of fMRI data, monkeys were head-fixed in a sphinx posture within an MRI compatible chair (Rogue Research). MR images were acquired with a horizontal bore clinical 3T scanner with a 15-channel non-human primate (NHP)-specific receive coil (RAPID Biomedical). Structural images were acquired during a previous experiment (*264*). Functional images were acquired via a CMMR multiband gradient-echo T2* echo planar imaging (EPI) sequence designed specifically to achieve high signal-to-noise (SNR) in subcortical structures (*309*, *310*) with the following parameters; 1.25mm isotropic voxels; repetition time (TR) = 1282ms; echo time (TE) = 25.40ms; multiband acceleration factor MB = 2; in-plane acceleration factor R = 2, and; flip angle = 63°.

Offline reconstruction of the raw functional data was performed following the dynamic off-resonance correction method developed by Shahdloo et al (*311*). In summary, standard Nyquist ghost correction and dynamic zeroth-order B0 correction were applied first. Then, the EPI reference navigator data acquired at every time-point was compared to navigator data from single-band references to estimate first-order dynamic off-resonance perturbations arising from the awake animal's body movements. Finally, the off-resonance estimates were used to correct the raw data prior to reconstruction.

Pre-processing of MR images was performed with a combination of FMRIB's software library, Advanced Normalization Tools, the Human Connectome Project Workbench and Magnetic Resonance Comparative Anatomy Toolbox (*312*). Although monkeys were head-fixed during MRI acquisition, incidental limb and body movements caused time-varying

distortions in the B_0 magnetic field and therefore non-linear motion artefacts along the phase encoding direction. To account for this, a low-noise EPI volume was identified for each session and then implemented as a reference to which other volumes were non-linearly registered slice-by-slice along the phase-encode direction. Aligned and distortion-corrected EPIs were then registered non-linearly first to monkey-specific high-resolution images, and then to a group template in CARET f99 macaque space. Further details of the group template construction are described elsewhere (*295*). Finally, the functional images were temporally filtered (high-pass temporal filtering, 3-dB cutoff of 100s) and spatially smoothed (Gaussian spatial smoothing, full-width half maximum of 2.5mm).

Three measures were used to detect artefacts in the data: (1) For each slice in each volume, we assessed the linear transform (in the y-plane) from that slice to the corresponding slice in the mean reference image; (2) we further assessed the normalised correlation between that slice and the corresponding slice in the mean reference image, and; (3) for each volume, we calculated the correlation between the volume (mean-filtered across z-slices) and the mean reference image after correction. Volumes were removed when they exceeded 2.5 SDs above the median of each measure. This threshold was chosen to keep the number of censored volumes below 10% of total volumes. We also added 13 PCA components describing, for each volume, the warping from that volume to the mean reference image when correcting motion artefacts as parametric regressors of non-interest that were not convolved in our general linear models.

4.4.5 fMRI data analysis

We focused our analysis of fMRI data on circumscribed *a priori* regions of interest (ROIs) comprising the dorsal raphe nucleus (DRN), ventral tegmental area (VTA), substantia

nigra (SN), nucleus basalis (NB), locus coeruleus (LC) and habenula (Hb) in the subcortex, and anterior cingulate cortex (ACC), anterior insula (AI) and supplementary motor area (SMA) in the cortex. Subcortical and SMA ROIs consisted in anatomical masks that were drawn on a group structural template in CARET F99 macaque monkey space and then warped to individual structural and functional spaces by nonlinear transformation. These masks were constructed separately by two-different assessors based on the Rhesus Monkey Brain Atlas (*69*) and then evaluated on for convergence across assessors. ACC and AI ROIs were defined as 3mm spheres centred on the peak of functionally relevant activation contrasts obtained in previous studies (*227*, *264*).

We extracted the filtered time series of BOLD signal from each ROI. The extracted signals were then averaged, normalised and up-sampled by a factor of 15 (*227*, *264*, *295*). The upsampled data was then epoched to 6s time windows spanning 1s before to 5s after the appearance of the reward-opportunity stimulus on each trial. We then examined the relationship between behaviour and brain activity with ordinary least squares (OLS) GLMs performed at each timepoint in each epoch.

Inferential statistics for time-course GLMs were performed using a leave-one-out cross-validation (LOOCV) procedure designed to estimate the peak regression-coefficient in each session without selection bias (*227*, *264*, *295*). For each session s ($N = 59$), we determined the timepoint t at which the largest absolute-value regression coefficient occurred in the remaining $N-1$ (i.e., 58) sessions. We restricted our search for t to a 4s window from 1s after to 5s after the decision-making time – that is, a four second window centred on the mean macaque HRF. We then calculated the regression coefficient in session s at time t. We repeated this iteratively for each session, which yielded a series of 59 regression coefficients. We performed significance testing on regression coefficients with one-sample two-tailed t-tests. To control for multiple comparisons, we applied the Bonferroni-Holm correction for any analysis

performed on more than one ROI. For each GLM, we repeated this for each regressor in each ROI. All time course analysis was conducted in MATLAB (MathWorks) using custom analysis scripts.

We implemented the following series of GLMs:

GLM4.1

$$BOLD_{ROI} = \beta_o + \beta_1 reward\text{-}magnitude + \beta_2 reward\text{-}probability + \beta_3 environment\text{-}richness + \beta_4 environment\text{-}stochasticity + \beta_5 behavioural\text{-}history + \beta_6 environment\text{-}richness * behavioural\text{-}history + \beta_7 environment\text{-}stochasticity * behavioural\text{-}history + \beta_8 pupil\text{-}size + \beta_9 trial\text{-}number$$

Where $BOLD_{ROI}$ indicates a *t*-by-*s* matrix containing time-series data for a given ROI (where *t* is trial, and *s* is time-sample). Pupil-size and trial-number were added as confound regressors to control for the effects of arousal and fatigue, respectively.

GLM4.2

$$BOLD_{ROI} = \beta_o + \beta_1 reward\text{-}magnitude + \beta_2 reward\text{-}probability + \beta_3 environment\text{-}richness + \beta_4 environment\text{-}stochasticity + \beta_5 pupil\text{-}size + \beta_6 trial\text{-}number$$

GLM4.2 was performed separately for subsets of trials in which animals (i) pursued and (ii) rejected the previous opportunity, respectively.

We tested which ROIs signalled the pursue-vs-reject decision made on each trial with the following GLM:

GLM4.3

$$BOLD_{ROI} = \beta_o + \beta_1 decision + \beta_2 pupil\text{-}size + \beta_3 trial\text{-}number$$

Where *decision* is the pursue-vs-reject decision made on a given trial.

We next tested which ROIs signalled features of the motivation-states identified with the GLM–HMM with the following GLMs:

GLM4.4

$$BOLD_{ROI} = \beta_o + \beta_1 transition\text{-}period + \beta_2\ motivation\text{-}state\text{-}identity + \beta_3 pupil\text{-}size + \beta_4 trial\text{-}number$$

Where *transition-period* is a dummy-coded variable covering symmetric 7-trial windows (*t-3:t+3*) around decoded motivation-state transitions and *motivation-state-identity* is the decoded motivation-state (high vs. low) for a given trial. We performed GLM4.4 separately for low-to-high and high-to-low transitions (see fig.4.3).We also performed it with 7-trial transition-periods aligned to from (t-6:t+0) to (t-0:t+6) relative to decoded transitions (where *t* indicates motivation-state transition trial decoded by the GLM-HMM). We did this to identify whether DRN's ostensive motivation-state transition representation was, indeed, specific to trials surrounding decoded motivation-state transitions (see fig.4.4 & fig.4.5).

We tested whether DRN coded the value of recently available rewards (Ave. EV) with the following GLM:

GLM4.5

$$BOLD_{DRN} = \beta_o + \beta_1 EV + \beta_2 Ave.\ EV + \beta_3 pupil\text{-}size + \beta_4 trial\text{-}number$$

GLM4.5 was performed on data from DRN alone, as it was designed to test a specific hypothesis arising from the combined neural and behavioural data. We used EV to control for the effects of the currently available reward opportunity instead of separate reward-magnitude and reward-probability regressors as in previous GLMs of BOLD signal. This is due to collinearity between Ave. EV and reward-probability regressors (see Ch.III, §3.4: Methods for details).

We probed changes in connectivity as a function of the richness of the environment using the following PPI-GLM.

GLM4.6

$$BOLD_{ROI} = \beta_o + \beta_1 BOLD_{seed} + \beta_2 PPI + \beta_3 environment\text{-}richness + \beta_4 reward\text{-}magnitude + \beta_5 reward\text{-}probability + \beta_6 environment\text{-}stochasticity + \beta_7 pupil\text{-}size + \beta_8 trial\text{-}number$$

Where $BOLD_{seed}$ *is* a t-by-s matrix containing timeseries data for seed regions in PPI analysis, and *PPI* is the interaction between $BOLD_{seed}$ and *environment-richness* regressors. GLM 4.6 was performed separately for subsets of trials in which an animals (i) pursued and (ii) rejected the previous reward opportunity, respectively. Analogously, we tested differences in functional connectivity with respect to (i) motivation-state level, and (ii) motivation-state transitions with the following PPI:

GLM4.7

$$BOLD_{ROI} = \beta_o + \beta_1 BOLD_{seed} + \beta_2 PPI + \beta_3 transition\text{-}period + \beta_4 motivation\text{-}state\text{-}identity + \beta_5 pupil\text{-}size + \beta_6 trial\text{-}number$$

Where $BOLD_{seed}$ *is* a t-by-s matrix containing time-series data for seed regions in PPI analysis. The analysis was performed twice – once where *PPI* was the interaction between $BOLD_{seed}$ and *transition-period* regressors, and once where *PPI* was the interaction between $BOLD_{seed}$ and *motivation-state-identity* regressors. Finally, we tested differences in functional connectivity with respect to pursue/reject decisions with the following PPI:

GLM4.8

$$BOLD_{ROI} = \beta_o + \beta_1 BOLD_{seed} + \beta_2 PPI + \beta_3 decision + \beta_4 pupil\text{-}size + \beta_5 trial\text{-}number$$

Where $BOLD_{seed}$ *is* a t-by-s matrix containing time-series data for seed regions in PPI analysis, and *PPI* is the interaction between $BOLD_{seed}$ and *decision* regressors.

4.4.6 Transcranial ultrasound stimulation (TUS)

TUS was performed using a four-element annular array transducer (NeuroFUS CTX-500, 64mm active diameter, Brainbox Ltd, Cardiff, UK) combined with a programmable amplifier (Sonic Concept Inc's Transducer Power Output System, TPO-105, Brainbox Ltd, Cardiff, UK). The transducer was paired with a transparent coupling cone filled with degassed water and sealed with a latex membrane. The water was degassed for 4-5 hr before each stimulation session and was replaced after each session. The resonance frequency of the ultrasonic wave was set to 500kHz. The stimulation protocol was based on previously established protocols in macaques (*263*). We used the following protocol: duty cycle 30%; pulse length 30ms; pulse repetition interval 100ms; total stimulation duration 30s. The pressure field from the transducer was measured in a water tank with a 75 μm diameter PVDF needle hydrophone (Precision Acoustics, Dorset UK) which had been calibrated at 500 kHz by the National Physical Laboratory (Teddington UK). The free-field spatial-peak pulse-average intensity (Isppa) at 60mm focal depth was 120 W/cm2, which was consistent with the output given by the transducer manufacturer.

At the beginning of each stimulation session the animal's skull was shaved and a conductive gel (SignaGel Electrode; Parker Laboratories Inc.) was applied to the skin. The water-filled coupling cone and the gel was used to ensure ultrasonic coupling between the transducer and the animal's head. Next, the ultrasound transducer / coupling cone was placed on the skull and a Brainsight Neuronavigation System (Rogue Research, Montreal, CA) was used to position the transducer so that the focal spot would be centred on the targeted brain region. There were four stimulation conditions: 1) DRN (target of interest); 2) VTA (subcortical active control condition); 2) superior-temporal-sulcus (STS; cortical active control

condition); 4) sham (passive control condition). DRN and VTA targets were approximately 50mm, and the STS 35mm, from the surface of the transducer (the exact focal distance depended on the subject). All targets were sonicated bilaterally for 60s in total, with 30s of stimulation applied from each hemisphere. Sonication of the midline targets (DRN and VTA) from one hemisphere was immediately followed by sonication of the same target from the contralateral hemisphere (cross-beam stimulation; (*264*)). Sonication of the STS in one hemisphere was immediately followed by sonication of a homologous target in the contralateral hemisphere. Hemispheres were sonicated in a pseudo-random order. After stimulation, monkeys were immediately moved to a testing room for behavioural data collection. The sham condition completely matched a typical stimulation session (setting, stimulation procedure, neuro-navigation, targeting, transducer preparation and timing of the bilateral application to the shaved skin on the head of the animal) except that sonication was not triggered. During the sham session, the montage was pseudo-randomly positioned to target DRN, VTA or STS. Each stimulation condition was repeated five times, on separate days, and the order of the stimulation sessions was pseudo-randomized for each animal. The stimulation was always performed at the same time of the day and there was always a 24-hour gap between each session, regardless of it being a real or sham stimulation session.

4.4.7 Acoustic modelling

We simulated the propagation of acoustic waves produced by the TUS protocol as described in (*297*). In brief, we used k-wave – a k-space pseudospectral solver (*313*) – and kArray tools to obtain estimates for the pressure amplitude, peak intensity and spatial distribution of TUS at steady-state. First, we simulated the acoustic wave propagation in water (free field) to characterise ultrasound beam for a target intensity of 120 W/cm^2 at 50mm focal

depth (focal depth of VTA and DRN from transducer). Next, we performed the simulation for DRN and VTA targets in the skull. The skull was estimated from pseudo-CT images obtained from each monkey using a Black Bone MRI sequence (*296*), and obtained by thresholding the pseudo-CT images at 1400-2100 Hounsfield Units (HU). We assumed a linear relationship between the pseudo-CT Hounsfield Units and the sound speed, as well as the density, and absorption coefficient as described elsewhere (*314*, *315*). The simulation grid size was set to the size of the T1-weighted MRI with a grid spacing of 0.5 mm which results in approximately 6 points per wavelength in water and tissue and up to 12.4 points per wavelength in bone.

4.4.8 Resting-state imaging data acquisition, pre-processing, and analysis

We further validated the TUS protocol by examining its effect on resting state coactivation patterns between VTA/DRN and key interconnected cortical and subcortical regions. Awake resting-state fMRI (rs-fMRI) data was acquired for all four monkeys (same animals as used in the fMRI experiment) pre-vs-post DRN-TUS and pre-vs-post VTA-TUS. Pre-processing and analysis of rs-fMRI data has been described elsewhere (*263*, *265*).

We characterised the effects of TUS on coactivation between DRN, VTA and key ROIs by comparing rs-fMRI data collected before DRN/VTA TUS with the rs-fMRI data collected immediately after DRN/VTA TUS. Pre-stimulation and post-stimulation rs-fMRI were collected on the same day. The impact of DRN/VTA TUS on coactivation patterns was quantified with seed-based connectivity analyses, which involved calculating a series of pairwise linear correlations in BOLD activity between a seed region (DRN or VTA) and the remaining ROIs. The resulting pre-stimulation connectivity fingerprints for DRN-TUS and VTA-TUS were then contrasted with post-stimulation connectivity fingerprints after DRN-TUS and VTA-TUS, respectively (see fig4.6C).

4.4.9 TUS data analysis

We characterised the behavioural effect of TUS by examining its two-way interactions with key predictors of decision-making in a series of Binomial GLMs. In these GLMs, the effect for TUS-condition was constructed to compare each individual active stimulation condition (STS, VTA, DRN) to sham-TUS as a reference. On some occasions, we compared DRN-TUS to one of the other active stimulation conditions – for example, DRN-TUS vs STS-TUS. We did this by performing GLMs on subsets of data that included only the TUS-conditions of interest. These analyses are specifically noted in the main text.

We used the following GLMs to examine how TUS modulates the effect of the richness of the environment, and the effect of behavioural history effects:

GLM4.9

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude + \beta_2 reward\text{-}probability + \beta_3 environment\text{-}richness + \beta_4 environment\text{-}stochasticity + \beta_5 TUS\text{-}condition + \beta_6 (reward\text{-}magnitude * reward\text{-}probability) + \beta_7 (environment\text{-}richness * TUS\text{-}condition) + \beta_8 (trial\text{-}number) + \mu_0 + \varepsilon$$

Where $\beta_{0\text{-}8}$ are fixed-effects, μ_0 is the by-subject random intercept and ε is an error term.

GLM4.10

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude + \beta_2 reward\text{-}probability + \beta_3 behavioural\text{-}history + \beta_4 TUS\text{-}condition + \beta_5 (reward\text{-}magnitude * reward\text{-}probability) + \beta_6 (behavioural\text{-}history * TUS\text{-}condition) + \beta_7 (trial\text{-}number) + \mu_0 + \varepsilon$$

Where $\beta_{0\text{-}7}$ are fixed-effects, μ_0 is the by-subject random intercept and ε is an error term. The paucity of data in each TUS condition for each subject prevented us from fitting models with more complex random-effects structures for these models. We examined the influence of TUS on the frequency of transitions between motivation-states with the following GLM:

GLM4.11

$$\text{logit}(\text{state-transition}_t) = \beta_0 + \beta_1\text{reward-magnitude} + \beta_2\text{reward-probability} + \beta_3\text{TUS-condition} + \beta_4\text{reward-magnitude*reward-probability} + \beta_5\text{trial-number} + \mu_0 + \varepsilon$$

Where *state-transition* ∈ {0,1} is a binary variable reflecting whether a motivation-state transition occurred, $\beta_{0\text{-}5}$ are fixed-effects, μ_0 is the by-subject random intercept and ε is an error term. We performed a similar GLM to examine the influence of TUS on high-to-low transitions specifically:

GLM4.12

$$\textit{logit}(\textit{state-decrease}_t) = \beta_o + \beta_1\textit{reward-magnitude} + \beta_2\textit{reward-probability} + \beta_3\textit{TUS-condition} + \beta_4\textit{reward-magnitude*reward-probability} + \beta_5\textit{trial-number} + \mu_0 + \varepsilon$$

Where *state-decrease* ∈ {0,1} is a binary variable reflecting whether a high-to-low motivation-state transition occurred, $\beta_{0\text{-}5}$ are fixed-effects, μ_0 is the by-subject random intercept and ε is an error term. We analysed the effect of TUS on the relationship between motivation-states and the availability of rewards (Ave. EV) with the following GLM:

GLM4.13

$$\textit{logit}(\textit{motivation-state}_t) = \beta_o + \beta_1\textit{EV} + \beta_2\textit{Ave.EV} + \beta_3\textit{TUS-condition} + \beta_4\textit{TUS-condition*Ave.EV} + \beta_5\textit{trial-number} + \mu_0 + \varepsilon$$

Where *motivation-state* ∈ {0,1} is a binary variable reflecting the decoded motivation-state on each trial and the high-motivation state =1 (i.e., positive coefficients correspond to increases in the likelihood of occupying the high-motivation state), $\beta_{0\text{-}5}$ are fixed-effects, μ_0 is the by-subject random intercept and ε is an error term. We performed GLM4.5 separately on behavioural data mean-split according to Ave.EV to determine whether DRN-TUS affected the relationship between low-value environments and low-motivation states., specifically.

Chapter V

Dorsal raphe nucleus signals behavioural policy changes in humans

5.1 Introduction

What makes a reward worth pursuing? For both humans and non-human animals, these judgments depend on the context in which they are made. A quintessential example is the prey selection dilemma set forth in behavioural ecology, in which a predator must decide whether to seize a foreground opportunity available in the present, or search the environment for better alternatives (*16*, *316*). According to normative accounts of this scenario, an opportunity should only be pursued if its value meets or exceeds the opportunity cost of obtaining it. In practice, this means that opportunities with the same value will evoke different behavioural policies depending on the context they occur in: for example, it is rational to pursue an objectively mediocre opportunity in a poor environment where everything on the horizon is similar or worse, and yet the same opportunity might be rejected in the context of a rich environment with plentiful high-value alternatives.

How does the brain adapt behavioural policies to the environment in this way? Here, we extend the hypothesis developed in Ch.III–IV and argue for the importance of a cortico-subcortical circuit comprised by dorsal raphe nucleus (DRN) – a brainstem nucleus that is the primary source of serotonin to the forebrain in mammals – , and anterior cingulate cortex (ACC) and anterior insular cortex in the cerebral cortex (AI) (*148*, *317*, *318*). This builds on previous studies in both human and non-human primates showing that activity in ACC and AI tracks reward-related statistics that are critical for adaptive decision-making, like the value,

valence and background-rate of recent reward outcomes (*24*, *26*, *84*, *227*). Similarly, it extends upon a rich vein of experiments in animal models linking DRN activity to learning and decision-making. A consistent theme in these studies is that DRN controls the balance between opponent modes of reward-driven behaviour, like patience and impulsivity, exploration and exploitation (*204*, *228*, *302*, *319*), and even motivation and demotivation for reward-seeking (Ch.III–IV). Here, we argue that these seemingly disparate functions involve a common process of setting a behavioural policy. We employ a novel behavioural task that allows us to identify when and how behavioural policies are set, and demonstrate for the first time that activity in human DRN is associated with policy changes.

We gave participants a novel, foraging-inspired behavioural task in which they encountered a limited set of reward options in a series of rich and poor environments. As expected, participants systematically changed their behavioural policy depending on the richness of the environment, and were especially responsive to poor environments which caused them to become less selective in their choices. We recorded brain activity while participants performed the task using an ultra-high field functional magnetic resonance imaging (fMRI) protocol specifically designed to maximise signal from brainstem and midbrain areas. These recordings demonstrated that DRN – but no other major neuromodulatory nucleus – exhibited distinctive univariate patterns of activity during behavioural-policy changes. In contrast, multivariate patterns of activity evoked by reward-options in dACC and AI paralleled the relative-value of each option given the environmental context, suggesting that these regions construct a representation of the environment in the first instance. Taken together, these results suggest that DRN, dACC and AI form a cortico-subcortical circuit that reconciles the brain's behavioural policy with the distribution of rewards in the environment.

5.2 Results

5.2.1 A behavioural task for manipulating the richness of the environment

Twenty-seven participants performed a simple decision-making task involving sequential encounters with reward opportunities (fig.5.1A–C). On each trial, participants were presented with a single reward opportunity drawn from a set of three possible options with known *a priori* values: a low-value option (5-points; bronze medallion), a middle-value option (10-points; silver medallion), and a high-value option (50 points; gold medallion). Upon each, encounter, participants could pursue opportunities by making a button press response or reject opportunities by withholding any response and waiting for the next trial to begin (see Methods).

Pursue a reward opportunity incurred a temporal opportunity cost equivalent to forgoing a single reward opportunity in the future. It was therefore critical to consider the prospective frequency of each reward-option when deciding whether an opportunity was worth pursuing. The frequency of reward-options was systematically manipulated by dividing the task into 4.5-minute blocks of trials with different reward-option frequency distributions (fig.5.1C). There were two types of block; (1) *rich* blocks, where the 50-point option was more frequent than the 10-point and 5-point options $P(5\text{-point}|Rich) = 0.16$, $P(10\text{-point}|Rich) = 0.33$, $P(50\text{-point}|Rich) = 0.50$), and; (2) *poor* blocks where the 5-point option was more frequent than the 10-point and 5-point options ($P(5\text{-point}|Poor) = 0.50$, $P(10\text{-point}|Poor) = 0.33$, $P(50\text{-point}|Poor) = 0.16$). These blocks allowed for experimental control of the richness of the environment throughout the task (fig.5.1C–D). Importantly, the frequency of the 10-point option was constant across both rich and poor blocks, meaning that changes in a participant's

behaviour with respect to the 10-point option could only arise from changes in the environmental context within which it occurred.

The parameters of the task were carefully chosen to ensure that the rational strategy was to accept the 10-point option in poor environments, and to reject the 10-point option in rich environments (fig.5.1E–F). Environments were not signalled to participants, and we therefore did not expect them to exhibit precisely this pattern. Nevertheless, the design ensured that there was a reward-maximising rationale for participants to change their behaviour toward the 10-point option. Accordingly, we focus predominantly trials featuring the 10-point option in the following behavioural analyses, although the same conclusions hold in omnibus analyses where all reward-options are considered together (fig.5.2).

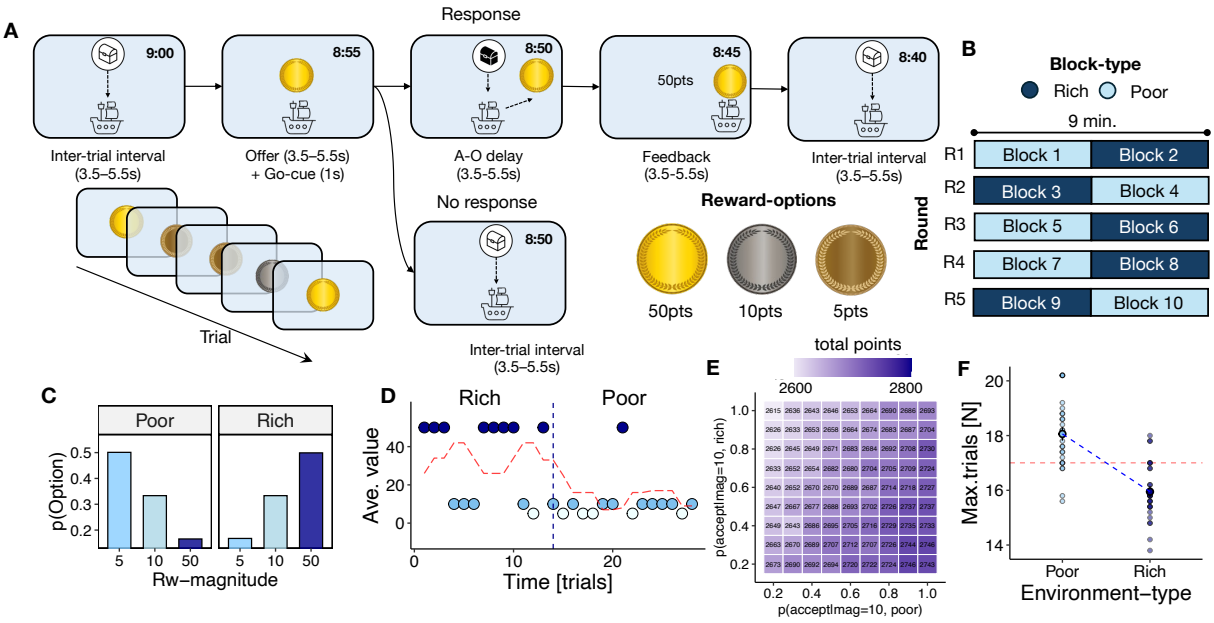


Figure 5.1. A behavioural task for controlling the richness of the environment. (A) Participants played the role of a treasure-seeking ship's captain in a behavioural task involving sequential encounters with reward-options. A single reward opportunity was encountered on each trial, and subjects needed to decide whether to pursue (button-press response) or reject (no response) that opportunity. Pursuing an opportunity incurred a temporal opportunity cost that was equivalent to giving up a future encounter, which was symbolised by a passing black treasure-chest during the action-outcome delay period. To earn as many points as possible, participants therefore needed to compare the reward-option available during each opportunity with the distribution of options they expected in future opportunities. Reward-options were symbolised by gold (50-points), silver (10-points) and bronze (5-points) medallions. The remaining time in the current round of the task was shown via a timer at the top left corner of the screen. **(B)** Diagram of the experiment's round and block structure in an example session. Each experimental session comprised five, 9-minute rounds of the behavioural task, and each round of the task involved two time-limited 4.5-minute blocks. Each round included one rich block and one poor block, which were randomised in order. **(C)** The frequency of 5-point, 10-point and 50-point reward options was systematically manipulated to engender differences in the richness of the environment. In rich blocks, the 50-point option was more frequent than 10-point and 5-point options, whereas in poor blocks the 5-point option was more frequent than 10-point and 50-point options. Note that the frequency of the 10-point option remained constant, meaning that behaviour toward the 10-point option could be compared across blocks without bias. **(D)** An example

sequence of rich and poor blocks illustrating key reward statistics and their evolution over time. Coloured dots indicate the specific reward-option available on each trial (dark-blue = 50-points, mid-blue = 10-points, light-blue = 5-points). Dashed red line indicates the mean value of rewards over the preceding five trials (i.e., *t*-1:*t*-5). **(E)** The temporal and reward-option value parameters of the task were carefully chosen to engender changes in behaviour toward the 10-point option. Tile-map and colour-scale shows the total points earned by a simulated agent performing the experiment as a function of different pursue-rates for the 10-point option in rich-blocks (y-axis) and poor blocks (x-axis). Note that environments were not signalled to participants, and so we did not expect them to display this pattern of behaviour immediately upon entering a new environment. Instead, we expected that they would approach the pattern as they accumulated experience within an environment. **(F)** Because blocks were time-limited, and the duration of each trial depended on whether the reward opportunity was pursued or rejected, the number of trials in each block was variable. Overall, participants performed more trials in poor blocks compared to rich blocks because they pursued fewer reward opportunities in poor blocks ($t_{\text{poor-vs-rich}}(26) = 15.52$, $p < .001$; see fig.5.2 for further details). Dots show participant-level mean number of trials in poor and rich blocks.

5.2.2 Reward-guided decisions are modulated by the richness of the environment

To begin, we tested the hypothesis that participants were more likely to pursue the 10-point option in poor environments compared to rich environments. All analyses in this section were implemented as mixed-effect generalised linear models (GLM) which account for subject-specific variation in behaviour. The dependent variable was the binary pursue-vs-reject decision on each trial.

We first tested the hypothesis by asking whether the experimentally manipulated environment-type (rich-vs-poor) influenced pursue/reject decisions. This showed that participants were, indeed, more likely to pursue the 10-point when it occurred in a poor environment compared to a rich environment (GLM5.1b; $\beta_{\text{environment-type}}$ = -0.19, SE = 0.10, p = .045; fig.5.2A–B). There was no evidence for similar effects in 5-point (GLM5.1b; $\beta_{\text{environment-type}}$ = -0.58, SE = 0.31, p = .062; fig.5.2A–B) or 50-point options (GLM5.1b; $\beta_{\text{environment-type}}$ = 0.64, SE = 0.51, p = .206; fig.5.2A–B). Analysing all options together in a single GLM indicated a two-way interaction between option-value and environment-type that was consistent with the pattern evinced by option-specific analyses (GLM5.1a; $\beta_{\text{environment-type*reward-option}}$ = -3.09, SE = 0.19, p < .001; fig.5.2A–B). Importantly, the difference in 10-point option pursue-rates between poor and rich environments [Δ(pursue-rate) = pursue-rate$_{\text{poor}}$ – pursue-rate$_{\text{rich}}$] was positively correlated with total points earned in the experiment ($r_{\Delta(\text{pursue-rate}) \text{ vs total-points}}$ = 0.40, p = .040; fig.5.2C). The degree to which participants changed their behaviour between environments determined thus their performance on the task, as intended by the experimental design.

Participants were not informed that there were distinct rich and poor environments, and they received no cues to this effect during the task. Environment-driven changes in behaviour therefore involved tracking the frequency of reward-options over time. To investigate whether the effect emerged over time in this way, we tested the two-way interaction between environment-type and time elapsed within an environment, where time was quantified in trials. This revealed that environment-driven changes in behaviour increased as a function of time-within-environment for the 10-point option (GLM5.2; $\beta_{\text{environment-type*trial-in-environment}}$= -0.20, SE = 0.06, p = .001; fig.5.2D), but that there were no such changes for the 5-point (GLM5.2; $\beta_{\text{environment-type*trial-in-environment}}$= -0.04, SE = 0.12, p = .693; fig.5.2D) and 50-point options (GLM5.2; $\beta_{\text{environment-type*trial-in-environment}}$= -0.43, SE = 0.25, p = .092; fig.5.2D). Simple main-effects analysis for the 10-point option indicated asymmetric sensitivity to environments, whereby participants were more likely to pursue the 10-point option as time elapsed in poor environments (GLM5.2; $\beta_{\text{trial-in-environment}}$= 0.27, SE = 0.08, p < .001), but not in rich environments (GLM5.2; $\beta_{\text{trial-in-environment}}$= -0.13, SE = 0.09, p = .165). This pattern of results was confirmed in a conceptually similar analysis in which the environment was operationalised as the average value of recently encountered reward-options, as distinct from experimentally defined environments (GLM5.3; fig.5.2E).

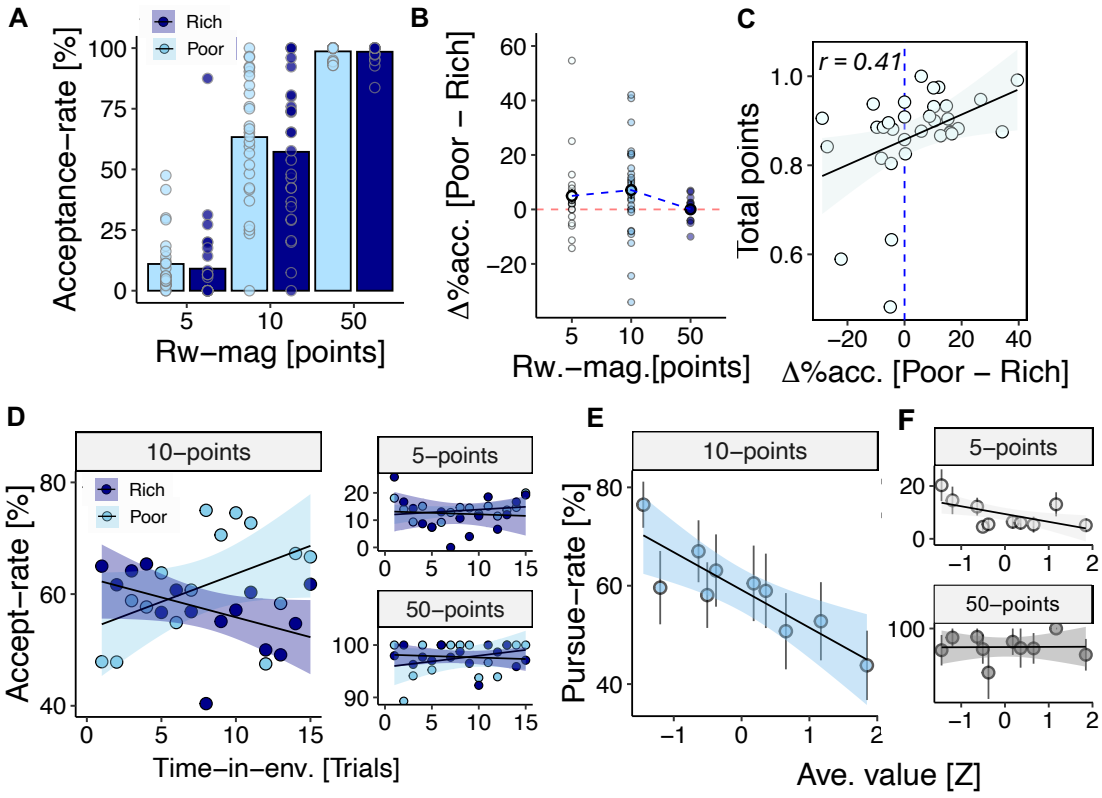


Figure 5.2 Pursue-vs-reject decisions are modulated by the richness of the environment. (A) & **(B)** A GLM predicting the probability of reward-pursuit indicated main effects of reward-magnitude (GLM5.1a; $\beta_{\text{reward-magnitude}}$= 6.12, SE = 0.25, p < .001) and environment-type (GLM5.1a; $\beta_{\text{environment-type}}$= -2.03, SE = 0.13, p < .001), in addition to a two-way interaction between these predictors (GLM5.1a; $\beta_{\text{reward-magnitude*environment-type}}$= -3.09, SE = 0.19 p < .001). Follow-up tests on the interaction indicated that the effect of environment-type was contingent on reward-magnitude: as per the main text, participants were more likely to pursue the 10-point option in poor environments relative to rich environments, but there was no evidence of changes for the 5-point or 50-point option as a function of environment-type. In **(A)**, bars and dots indicate sample-level and subject-level mean pursue-rates, respectively, for each reward option in each environment. In **(B)**, dots indicate participant-level difference-scores [$\Delta = \mu(\text{pursue})_{\text{poor}} - \mu(\text{pursue})_{\text{rich}}$] at each level of reward-magnitude. **(C)** The degree to which subjects changed their behaviour between environments predicted their success on the task. Y-axis denotes total-points earned in the experiment and x-axis denotes mean difference in 10-point option pursue-rates between poor environments and rich environments [poor – rich]. Dots represent data from individual subjects. **(D)** Environments were not explicitly signalled to subjects and needed to be inferred from the relative frequency of

each reward options over time. Accordingly, the difference between 10-point option pursue-rates (y-axis) for poor-environments and rich-environments increased as a function of time spent in an environment (x-axis). There were no such changes for 5-point and 50-point options. Dots and whiskers indicate mean ± SEM of pursue-rate in each level of trial-within-environment (x-axis). **(E)** Similarly, predicting the probability of reward pursuit as a function of the average-value of recent opportunities (where ave.-value$_t$ = mean(reward-magnitude$_{t-1}$: mean(reward-magnitude$_{t-5}$); see §5.4: Methods) indicated that participants were more likely to pursue the 10-point (GLM5.3; $\beta_{average\text{-}value}$ = -0.33, SE = 0.06, p = .003) and 5-point (GLM5.3; $\beta_{average\text{-}value}$ = -0.40, SE = 0.18, p = .021) options as average-value decreased. There was no evidence of this effect for the 50-point option (GLM5.3; $\beta_{average\text{-}value}$ = -0.14, SE = 0.47, p = .765). This is consistent with the hypothesis that participants changed their behaviour as a function of recently encountered reward opportunities. Dots and error bars indicate mean ± SEM pursue-rate in deciles of average-value.

5.2.3 Behaviour is organised in option-specific policies that are reconciled with the richness of the environment

Having shown that pursue-vs-reject decisions were influenced by the richness of the environment, we performed complementary analyses focussed on the temporal organisation of behaviour. Because the task featured only three reward-options, we reasoned that participants would form a behavioural policy for each option – that is, an option-specific pursue-vs-reject strategy that was consistent over successive encounters with the option in question (fig.5.3A). We validated this hypothesis by testing whether pursue-vs-reject decisions were autocorrelated across encounters with the same reward-option. This confirmed that participants tended to repeat their decisions across successive encounters, consistent with a temporally persistent behavioural policy (GLM5.4; $\beta_{\text{previous-policy}} = 1.53$, SE = 0.28, $p < .001$; fig.5.3B). Importantly, when the pursue-vs-reject decision on the preceding trial concerned a different option, it did not influence the pursue-vs-reject decision on the next trial, suggesting that policies were option-specific (GLM5.5; $\beta_{\text{previous-action}} = 0.05$, SE = 0.17, $p = .737$; fig.5.3B; see also fig.5.3A for explanation of distinction between switching in relation to a previous-action and switching in relation to a previous-policy).

Adopting the behavioural-policy perspective enabled us to precisely identify changes in behavioural policy (policy-switch, hence). We operationalised policy-switches as trials on which the pursue-vs-reject decision diverged from the policy exhibited the previous time the reward-option was encountered (fig.5.3A). Consistent with earlier analyses showing behavioural changes for the 10-point option, policy-switches were more frequent for the 10-point option compared to both its 5-point (GLM5.6; $\beta_{\text{middle-vs-low}} = 1.52$, SE = 0.39, $p < .001$; fig.5.3C) and 50-point counterparts (GLM5.6; $\beta_{\text{middle-vs-high}} = 3.63$, SE = 0.67, $p < .001$;

fig.5.3C). Subsequent analysis of policy-switches therefore focused on trials featuring the 10-point option.

Based on our analysis of pursue-vs-reject decisions, we predicted that policy-switches would be influenced by the richness of the environment. To test this, we categorised policy-switches into two-types depending on the direction of change: (i) *pursue-switches*, when the reward-option was pursued on the current trial after being rejected on the previous encounter, and (ii) *reject-switches*, when the reward-option was rejected on the current trial after being pursued on the previous encounter. Insofar as participants approximated the optimal task strategy, we expected that pursue-switches would be more likely in poor environments and reject changes would be more likely in rich environments.

When analysed in relation to the experimentally defined environments, pursue-switches followed the expected pattern: in poor environments, pursue-switches were more likely to occur than reject-switches (GLM5.7a; $\beta_{\text{pursue-switch vs reject-switch}}$= 0.89, SE = 0.19, p <.001; fig.5.3D-i), and pursue-switches in poor environments were also more likely to occur than pursue-switches in rich environments (GLM5.7b; $\beta_{\text{rich-vs-poor}}$= -0.36, SE = 0.12, p = .002; fig.5.3D-ii). Reject-switches, however, did not differ across environment types (GLM5.7c, $\beta_{\text{pursue-switch vs reject-switch}}$= -0.22, SE = 0.20, p =.274; GLM5.7d, $\beta_{\text{rich-vs-poor}}$= 0.07, SE = 0.12, p = .550; fig.5.3D). A similar pattern obtained when policy-switches were compared against the average-value of recent reward-options: pursue-switches were more likely as average-value decreased (GLM5.8a; $\beta_{\text{ave.-value}}$= -0.76, SE = 0.14, p <.001; fig.5.3E-i), but there was a null effect for reject-switches, which occurred at an approximately constant rate (GLM5.8b; $\beta_{\text{ave.-value}}$= 0.13, SE = 0.15, p = .386; fig.5.3E-i). In tandem with previous analyses, these results suggest that changes in behavioural policy approximated the rational task strategy, and that this pattern was more pronounced in poor environments.

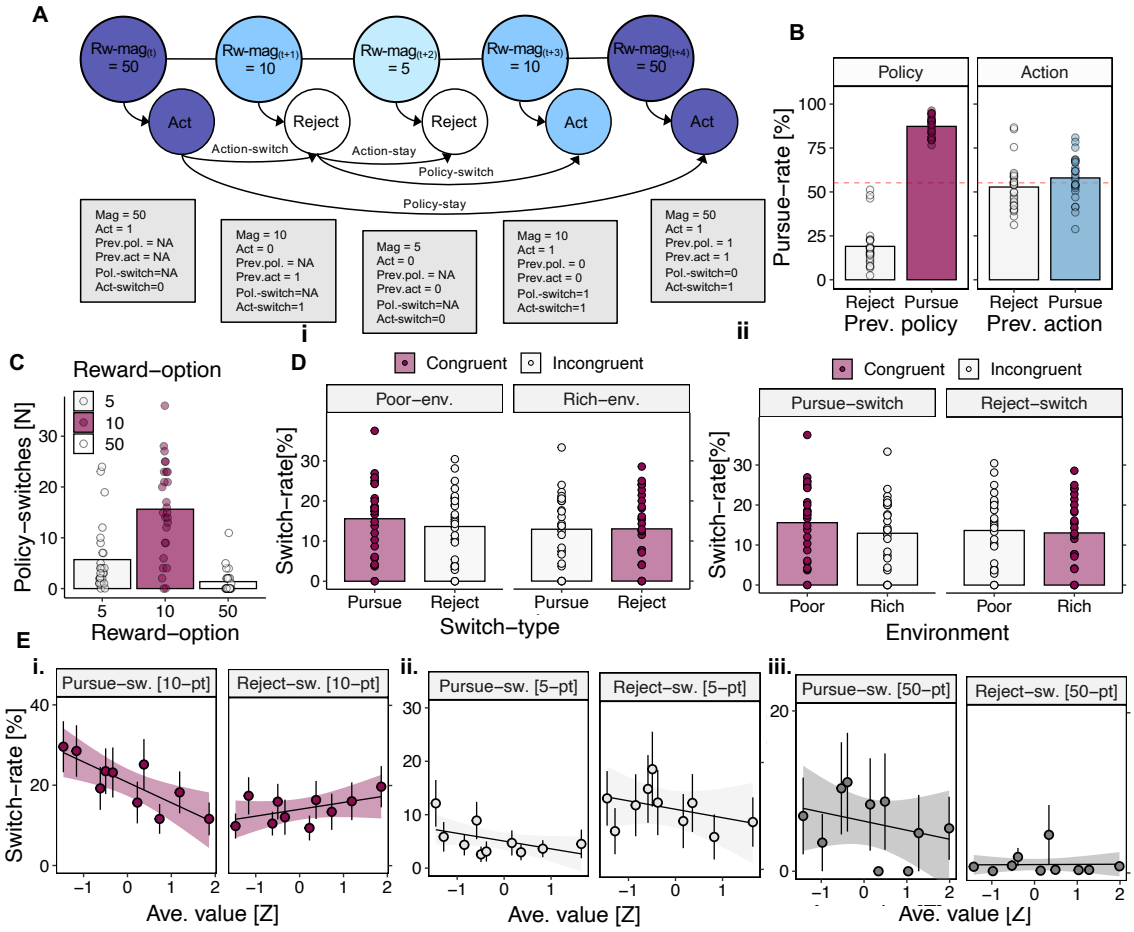


Figure 5.3 Participants exhibit consistent behavioural policies that are shaped by the richness of the environment. (A) Presenting a limited set of reward-options across the task allowed us to precisely track changes in option-specific behaviour. We categorised these changes along two dimensions: (1) a *policy* dimension reflecting the option-specific decision to pursue-vs-reject each of the three available reward-options, and (2) an *action* dimension reflecting the general decision to pursue-vs-reject opportunities regardless of the reward-option available. Within this framework, we then categorised trials as (i) *policy-stay* events if the pursue-vs-reject decision was consistent with the previous encounter, or (ii) *policy-switch* events if the pursue-vs-reject decision was different to the previous encounter. Correspondingly, we defined trials as (i) *action-stay* if the pursue-vs-reject decision was the same as the immediately preceding trial (regardless of reward-option), and (ii) action-switch if the pursue-vs-reject decision changed. **(B)** Participants were more likely to pursue a reward-option if they pursued it the previous time they encountered it and vice versa (previous-policy effect; left panel). Conversely, pursue-vs-reject decisions were not influenced by the action made on the previous trial (previous-action effect; right panel). Bars show

sample-mean pursue-rate in each condition. Dots indicate subject-specific pursue-rates. Dashed red line indicates overall mean pursue-rate, regardless of previous-policy and/or previous-action condition. **(C)** Policy-switches were more frequent for the 10-point option than for the 5-point and 50-point options. Bars show sample mean count of policy-switches per experimental session. Dots indicate subject-specific counts. **(D)** Policy switches approximated the rational task strategy. In **(i)**, pursue-switches (i.e. switching from reject-to-pursue) are more likely to occur than reject-switches in poor environments, consistent with the rational strategy for performing the task. There was no corresponding difference between pursue-switches and reject-switches in the rich-environments. Similarly, in **(ii)** pursue-switches were more likely to occur in poor environments compared to rich environments but there was no difference in the frequency of reject-switches between environment types. Colour-scale denotes congruency of switch-direction relative to environment-type – i.e., the switch-direction that subjects should display if following the rational strategy. Data shown for 10-point option only. Bars indicate overall mean rate of switch-events and dots indicate subject-specific rates of switch events. **(E)** Probability of pursue-switch and reject-switch events when the richness of the environment is operationalised as the average value of recent reward-options. Similar to (D), for the 10-point option **(i)**, pursue-switches are more likely to occur as the richness of the environment decreases (i.e. in poor environments) but reject-switches follow no systematic pattern. There was no evidence of a relationship between the richness of the environment and switches for 5-point **(ii)** and 50-point **(iii)** options. Points and error bars indicate mean ± SEM switch-rates in decile bins of average-value.

5.2.4 Brain activity in dorsal raphe nucleus represents the richness of the environment, and environment-driven changes in behaviour

Participants performed the behavioural task while undergoing ultra-high field fMRI recordings of blood oxygen level dependent (BOLD) signal. We used these recordings to investigate brain activity related to tracking the richness of the environment, and to environment-driven changes in behavioural policy. We focused on *a priori* regions of interest (ROIs) comprising the ascending neuromodulatory systems (ANS): an assemblage of phylogenetically ancient nuclei in the basal forebrain, midbrain, and brainstem that release fundamental neuromodulatory chemicals via diffuse, whole-brain connections (*89*, *90*, *320*). According to a prominent theory, the ANS broadcast salient low-dimensional signals that orchestrate activity across multiple brain regions at once (*320*). The ANS, thus, are ideally placed to implement changes in behavioural policy.

We extracted BOLD signal from an anatomical ROI for the dorsal raphe nucleus (DRN), in addition to a combined ROI covering the substantia nigra and ventral tegmental area (midbrain dopaminergic nuclei; MBD), the locus coeruleus (LC), the nucleus basalis (NB), and the habenula – an epithalamic nucleus characterised by reciprocal interactions with subcomponents of the ANS (*245*, *289*) (see fig.5.4A for anatomical illustration of ROIs and §5.4: Methods for details of ROI construction). Although our key hypotheses concerned DRN, expanding our purview allowed us to compare DRN activity with patterns in other ANS nuclei, which is critical for identifying functional specialisations in different brain regions (*286*).

We first examined brain activity related to the pursue-vs-reject decision made on each trial, with a focus on activity related to the richness of the environment. Although there were no differences in activity between experimentally manipulated rich-vs-poor environments

(fig.5.4B–C), DRN activity was modulated by the two-way interaction between environment-type and time-within-environment in a manner that paralleled the gradual adjustment of behaviour to the environment illustrated in fig.5.2D (GLM5.11; $t_{DRN;\ environment\text{-}type*trial\text{-}in\text{-}environment}(26)=2.84$, $p=.043$; fig.5.4B–C). No other ROI exhibited this pattern. It is, perhaps, surprising that MBD did not evince environment-related activity given its widely reported role in reward-value. Further examination of MBD activity, however, demonstrated two patterns of activity that dovetail with the broader literature on MBD: (i) activity related to the pursue-vs-reject decision made on each trial, and reminiscent of an action initiation signal, and; (ii) activity conveying the value-difference between the reward opportunity on the current trial and the average value of the recent environment, akin to a reward prediction-error response (fig.5.5) (*254*, *321–323*).

We next examined brain activity related to policy-switches. We did this by contrasting brain activity on policy-switch trials with brain activity on policy-stay trials – i.e., we asked whether specific patterns of brain activity occurred when a policy-switch was revealed in behaviour. This indicated that activity in DRN – but no other subcortical ROI – signalled policy-switch trials (GLM5.12; $t_{DRN;\ policy\text{-}change}(26)=2.84$, $p=.043$; fig.5.4D–E). We confirmed that this effect was specific to behavioural-policy changes by testing DRN's relationship with other forms of behavioural change, like action-switches where both the pursue-vs-reject decision and option-identity changed across consecutive trials such that the change in behaviour did not constitute change in option-specific behavioural policy (fig.5.3A). Action-switches had a null effect on DRN activity (GLM5.12; $t_{DRN;\ action\text{-}change}(26)=-1.69$, $p=.110$; fig.5.4D–E), and the policy-switch signal reported earlier was, moreover, significantly different to the null action-switch signal (GLM5.12; $t_{DRN;\ action\text{-}switch\ vs\ policy\text{-}switch}(26)=3.18$, $p=.002$; fig.5.4D–E).

After linking DRN activity to policy-switches in general, we next tested whether DRN was involved in environment-driven policy-switches specifically. We therefore examined DRN activity during policy-switches for the 10-point option that were congruent with the environment – i.e., pursue-switches in poor environments, and reject-switches in rich environments. DRN activity signalled both congruent pursue-switches and congruent reject-switches, albeit with oppositely signed patterns of activity (GLM5.13a; $t_{DRN;\ pursue\text{-}change}(26)=2.79$, $p=.009$; GLM5.13c; $t_{DRN;\ reject\text{-}change}(26)=-2.81$, $p=.009$; fig.5.4F-i). Because pursue-switches were correlated with action and reject-switches were correlated with inaction, we ensured the effects were not artefacts of motor preparation by comparing DRN activity for congruent-vs-incongruent pursue-changes, and congruent-vs-incongruent reject-switches: in other words, we compared events with the same motor-output that occurred in different environmental contexts. Unlike congruent-pursue switches (GLM5.13a; $t_{DRN;\ congruent\ pursue}(26)=2.38$, $p=.024$; fig.5.4F-ii), incongruent-pursue switches (GLM5.13b; $t_{DRN;\ incongruent\ pursue}(26)=-1.10$, $p=.280$; fig.5.4F-ii) had a null effect on DRN activity, and the congruent-pursue switch signal was stronger than its incongruent-pursue counterpart ($t_{DRN;\ congruent\text{-}vs\text{-}incongruent\ pursue}(26)=2.60$, $p=.014$; fig.5.4F-ii). Similarly, incongruent-reject switches had a null effect on DRN activity (GLM5.13d; $t_{DRN;\ incongruent\ reject}(26)=-1.35$, $p=.189$; fig.5.4F-iii) in contrast to the significant effect of congruent-reject switches (GLM5.13c; $t_{DRN;\ congruent\ reject}(26)=-2.08$, $p=.046$; fig.5.4F-iii), although the difference between these patterns was not significant ($t_{DRN;\ congruent\text{-}vs\text{-}incongruent\ reject}(26)=-1.30$, $p=.204$; fig.5.4F-iii). In this way, DRN policy-switch representations paralleled behavioural adjustment to the environment (fig.5.3D–E): they were only significant for policy-switches that were appropriate for the richness of the environment, and they were especially prominent for pursue-switches engendered by poor environments (fig.5.3D–E).

As a final test of DRN policy-switch signals, we compared congruent policy-switches for the 10-point and 5-point options. Behavioural analysis (fig.5.3) demonstrated that policy-switches for the 10-point option were modulated by the richness of the environment, whereas policy-switches for the 5-point option instead appeared to be random or exploratory in nature. Comparing the two events therefore allowed us to further isolate policy-switches driven by the richness of the environment from behavioural switches with other origins. Unlike the significant effect of congruent 10-point option pursue-switches (GLM5.13a; $t_{DRN;\ 10pt\text{-}option\ pursue}(26)=2.38$, $p=.024$; fig.5.4F-iv), 5-point option pursue-switches in poor environments had a null relationship with DRN activity (GLM5.13a; $t_{DRN;\ 5pt\text{-}option\ pursue}(26)=-1.23$, $p=.229$; fig.5.4F-iv). These signals were significantly different from one another ($t_{DRN;\ 5pt\text{-}vs\text{-}10pt\ pursue}(26)=2.52$, $p=.020$; fig. 5.4F-iv). Similarly, reject-switches for the 5-point option in rich environments had a null relationship with DRN activity (GLM5.13c; $t_{DRN;\ 5pt\text{-}option\ reject}(26)=-0.73$, $p=.471$; fig.5.4F-v) in contrast to the significant effect of middle-option reject-switches (GLM5.13c; $t_{DRN;\ 10pt\text{-}option\ reject}(26)=-2.08$, $p=.046$; fig.5.4F-v) although the difference between signals was not significant in this case ($t_{DRN;\ 5pt\text{-}vs\text{-}10pt\ reject}(26)= 1.77$, $p=.090$; fig.5.4F-v). This provides further evidence that DRN activity was specifically linked to environment-driven policy-switches.

Taken together, these analyses suggest that activity in DRN – but no other ANS nucleus – signalled changes in the richness of the environment and changes in behavioural policy that aligned behaviour with the environment. The policy-switch effect was specifically linked to environment-driven changes in behaviour and was consistently more prominent for policy-switches engendered by poor environments. In this way, policy-switch representations in DRN parallelled behavioural sensitivity to poor environments (fig.5.4Fi–v & fig.5.3D–E).

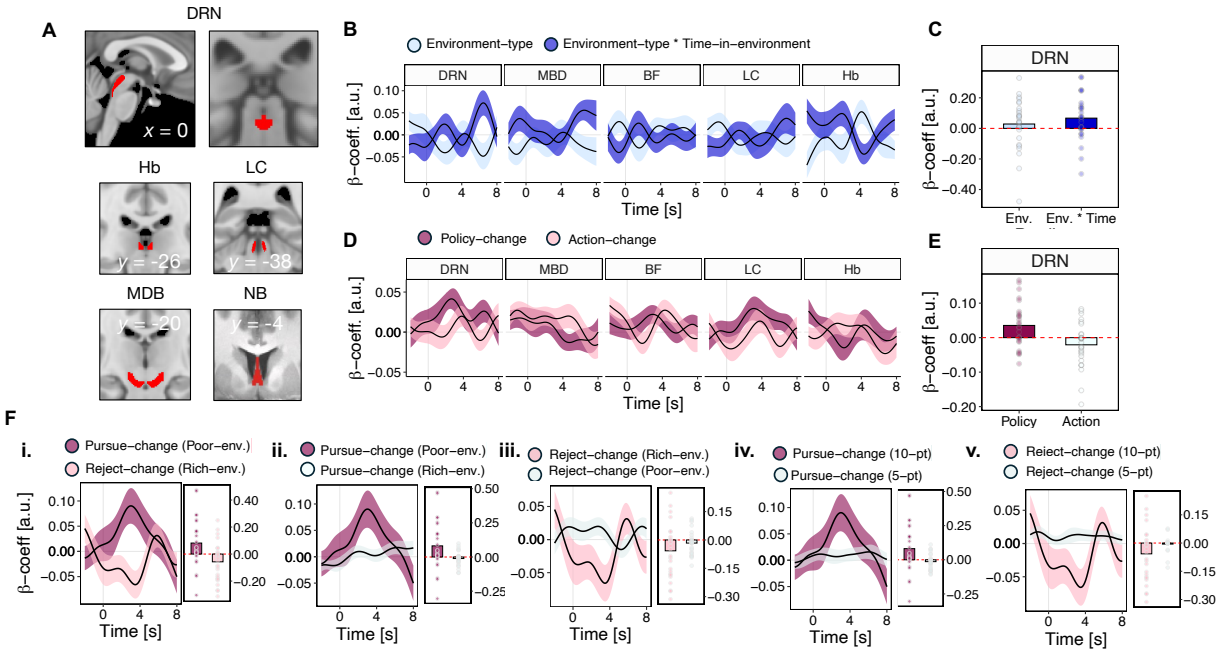


Figure 5.4 Brain activity in dorsal raphe nucleus represents the richness of the environment and environment-driven changes in behavioural policy. (A) fMRI analysis focussed on anatomical regions of interest (ROIs) in the ascending neuromodulatory systems. ROIs comprised the dorsal raphe nucleus (DRN), habenula (Hb), locus coeruleus (LC), substantia nigra and ventral tegmental area (MDB), and nucleus basalis (NB). See §5.4: Methods for details of mask construction. **(B)** Timecourse of the relationship between BOLD signal and (i) experimentally-manipulated environment-type (i.e., rich-vs-poor; light-blue), and (ii) interaction between environment-type and time-within-environment (dark-blue). Time=0 on the x-axis corresponds to the onset of the reward-opportunity stimulus. Given haemodynamic delay of approximately 4-6s, neural activity related to the reward-opportunity stimulus should therefore appear 4-6s after onset in BOLD signal (i.e. at 4-6s on the x-axis). **(C)** Distribution of peak regression-weights for environment-type and environment-type*time-in-environment predictors in DRN. Brain activity in DRN – but no other ROI– was modulated by environment-type, but only as time within the environment increased, parallelling behavioural adaptations to the environment as per fig.5.2D. **(D)** Timecourse of the relationship between BOLD signal and: (i) policy-switch events (dark-pink), and; (ii) action-switch (light-pink) events. Per (B), time=0 on the x-axis corresponds to the onset of the reward-opportunity stimulus. Given haemodynamic delay, it is likely that policy-switch neural activity in DRN occurs in ITI periods, when subjects are integrating reward feedback from the previous trial and planning their future behaviour. **(E)** Distribution of peak regression-weights for policy-switch and action-switch events in

DRN. Policy-switches – but not action-switches – modulate DRN activity, suggesting that DRN is involved in option-specific forms of behavioural change. **(F)** Timecourse and peak regression-weights (inset panels) for additional analyses linking DRN to environment-driven policy switches. Time=0 on the x-axis corresponds to the onset of the reward-opportunity stimulus in all timecourse graphs. **(i)** indicates DRN activity during policy-switches that are congruent to the environment – i.e., pursue-switch events in poor environments (dark-pink) and reject-switch events in rich environments (light-pink). **(ii)** shows DRN activity during pursue-switch events in poor environments (dark-pink) and rich-environments (grey) – i.e., events with the same motor output, but that differ in environment-congruence. **(iii)** shows DRN activity during reject-switch events in rich environments (light-pink) and poor-environments (grey) – i.e., events matched on motor output, but that differ in environment-congruence. **(iv)** shows DRN activity during pursue-switch events for the 10-point option (dark-pink) and 5-point option (grey). In behaviour, only 10-point policy-switches are modulated by the environment (fig.5.3). **(v)** shows DRN activity during reject-switch events for the 10-point option (light-pink) and 5-point option (grey). In timecourse graphs (B, D, and F), lines and shadings show the mean and standard error (SE) of the β weights across participants. In effect-size graphs (C, E, F inset panels), bars show sample-mean magnitude of the peak regression weight in the timecourse according to an unbiased leave-one-out procedure (see §5.4: Methods). Dots indicate peak regression weights for individual participants.

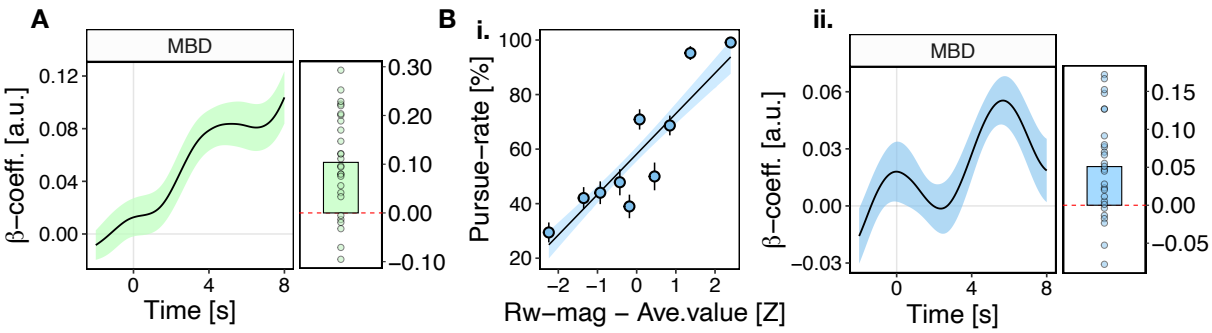


Figure 5.5. Activity in the midbrain dopaminergic nuclei represents pursue-vs-reject decisions, and the value-difference between current and recent reward opportunities. (A) BOLD signal in the MBD ROI was strongly modulated by the pursue-vs-reject decision made on each trial, consistent with the link between dopaminergic nuclei and action initiation (GLM5.15; $t_{MBD;\ pursue\text{-}vs\text{-}reject}(26)=5.21$, $p<.001$). **(B-i)** Although BOLD signal in the MBD ROI did not covary with experimentally controlled block-type or average-value, it was modulated by the value-difference between the reward opportunity on the current trial and the average-value of recent opportunities [i.e. value-difference$_t$ = reward-magnitude$_t$ – average-value$_t$]. This response is reminiscent of a reward-prediction error in the sense that the average-value of recent trials constitutes a simple prediction for the reward-value of the current trials (GLM5.17; $t_{MBD;\ value\text{-}difference}(26)=3.96$, $p<.001$). **(B-ii)** Importantly, participants were also more likely to pursue reward-opportunities as a function of value-difference (GLM5.16; $\beta_{value\text{-}difference}= 0.90$, SE = 0.07, $p <.001$). In this way, the value-difference signal in MBD is consistent with the hypothesis that it invigorates pursuit of specific reward opportunities. In **(B-i)**, dots and error bars indicate mean ± SEM pursue-rate in deciles of value-difference. In **(A)** and **(B-ii)**, lines and shadings of the time course graph show mean ± SEM of β weights across participants. Inset panels show effect-size graphs, where bar indicates mean peak regression weight in the timecourse according to an unbiased leave-one-out procedure (see). Dots indicate peak regression weights for individual participants.

5.2.5 Multivariate representations of value in frontal cortex are modulated by the richness of the environment

Having shown that DRN activity signalled the richness of the environment and environment-driven policy-switches, we expanded our fMRI analysis to cortical areas that might construct representations of the environment in the first instance. We hypothesised that in regions encoding the task structure, the distance between neural representations of each specific reward-option would change in tandem with the relative-value of reward options given the environment (fig.5.7A). We tested this using representational similarity analysis (RSA) on a series of ROIs deriving from functional fMRI analyses (*324*). First, we searched for relevant ROIs in our own data using two whole-brain fMRI analyses that searched for regions coding: (i) the pursue-vs-reject decision made on each trial (GLM5.9; fig.5.6A), and (ii) the reward-magnitude of the opportunity on each trial (GLM5.10; fig.5.6B). This identified clusters of activity related to the pursue-vs-reject contrast in anterior insular cortex (AI), dorsal anterior cingulate cortex (dACC), and lateral frontopolar cortex (lFPC), and a cluster in subgenual anterior cingulate cortex (sgACC) related to the reward-magnitude contrast (fig.5.6A–B; see appendix Tables 1 and 2 for full results of whole-brain analyses; see §5.4: Methods for details of ROI construction from functional activations). In addition, we investigated ROIs in dorsomedial prefrontal cortex (dmPFC) and pregenual anterior cingulate cortex (pgACC) based on activations observed in previous studies (*325*, *326*) (fig.5.7B).

Our analysis focussed on four representational distances; (i) d_1, the distance between 5-point and 10-point options in poor environments; (ii) d_2, the distance between 10 and 50-point options in poor environments; (iii) d_3, the distance between 5-point and 10-point options in rich environments, and; (iv) d_4, the distance between 10-point and 50-point options in rich

environments (fig.5.7A). Insofar as the neural representation of reward-options tracked the relative value of reward-options given the environment, we made the following predictions: (1) H1: $d_1 - d_3 > 0$ – i.e., that the 10-point option's representation is closer to the 5-point option's representation in rich environments compared to poor environments, and; (2) H2: $d_4 - d_2 > 0$ – i.e., that the 10-point option's representation is closer to the 50-point option's representation option in poor environments compared to rich environments (fig.5.7A). Importantly, not all participants followed the rational strategy of pursuing the 10-point option more frequently in poor environments (fig.5.2). We therefore anticipated that the magnitude of representational distance changes would be correlated with participant-specific levels of behavioural change between environments.

Comparing neural representations with behaviour demonstrated a striking effect in dACC and AI whereby the magnitude of H2 ($d_4 - d_2$) was positively correlated with behavioural change between environments ($r_{ACC\text{-}vs\text{-}behav.}$=0.60, $t(23)$=3.46, p=.002; $r_{AI\text{-}vs\text{-}behav.}$=0.65, $t(23)$=3.90, p<.001; fig.5.7E; correlation statistics reported after familywise error control using Bonferroni-Holm method) – in other words, in participants who adopted the rational strategy, neural representations of the 10-point option were more similar to the 50-point option in poor environments compared to in rich environments. We confirmed this interpretation by contrasting $d_4 - d_2$ in participants who were sensitive and insensitive to task environments (see §5.4: Methods for details). This indicated that the $d_4 - d_2$ difference was greater in the environment-sensitive group than in the environment-insensitive group ($t_{ACC}(11) = 3.36$, p = .006; $t_{AI}(11) = 3.15$, p = .012). There was no evidence of these patterns in any other ROI. Similarly, there was no evidence that the distance between 10-point and 5-point options (i.e. H1, or d_3-vs-d_1) was related to behaviour (fig.5.7D-i), and no evidence that H2 (d_4-vs-d_2), or H1 (d_3-vs-d_1) were different from 0 in and of themselves (fig.5.7D-ii). This suggests that participants represented the structure of the task in terms of the 50-point and 10-point options,

and that such representations were stronger in participants who adopted the rational task strategy.

Having shown that multivariate patterns of activity in dACC and AI captured critical aspects of the task, we conducted univariate timecourse analysis in these regions to relate them to subcortical ROIs. In contrast to multivariate patterns, there was no evidence of univariate relationships between dACC or AI activity and experimentally manipulated environments ($t_{dACC;\ environment\text{-}type}(26)= 1.52$, $p=.282$; $t_{AI;\ environment\text{-}type}(26)= 0.48$, $p=.636$). There was, similarly, no evidence that dACC or AI represented environment-driven policy-switches in the manner of DRN ($t_{dACC;\ pursue\text{-}change}(26)=-1.81$, $p=.081$; $t_{dACC;\ reject\text{-}change}(26)= -2.14$, $p=.085$; $t_{AI;\ pursue\text{-}change}(26)=0.96$, $p=.346$; $t_{AI;\ reject\text{-}change}(26)=-2.13$, $p=.085$; fig.5.8). In combination with earlier analyses of DRN activity, we therefore reasoned that DRN and dACC/AI might constitute a circuit in which cortical components represented salient aspects of the task structure like the relative value of each reward-option contingent on the richness of the environment, while DRN implemented the changes in behavioural policy consequent on these values. We tested this hypothesis with a series of psychophysiological interaction (PPI) analyses, which quantified changes in DRN–dACC/AI connectivity that might reflect communication between regions as a function of the environment. Consistent with this hypothesis, there was a PPI between dACC and DRN whereby coactivation was stronger in poor environments, in which environment-driven behavioural changes and neural signals were most prominent ($t_{PPI;\ dACC\text{-}DRN}(26) = -2.45$, $p = .021$; fig.5.9B). There was no evidence of the same pattern between DRN and AI ($t_{PPI;\ AI\text{-}DRN}(23) = -0.86$, $p = .400$).

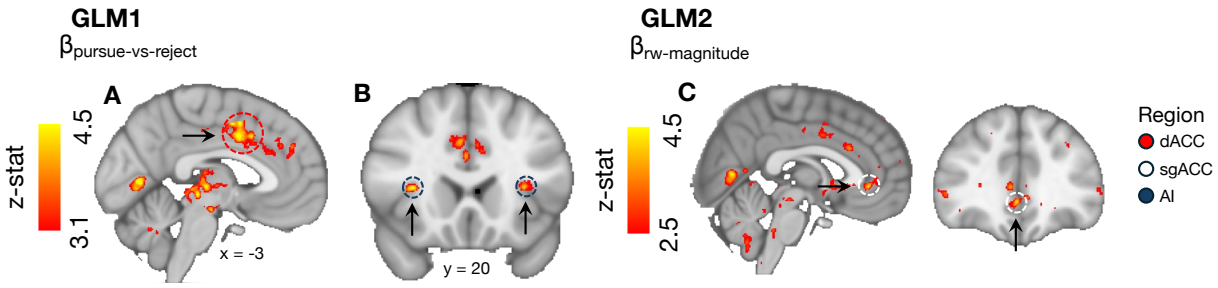


Figure 5.6. Whole-brain fMRI analysis identifies regions of interest for representational similarity analysis. We conducted representational similarity analysis (RSA) to probe patterns of brain activity that captured the value of a reward options relative to the richness of the environment. We identified regions of interest (ROIs) for RSA partly using whole-brain fMRI GLMs. **(A)** In the first whole-brain analysis (GLM5.9), we searched for activity related to the pursue-vs-reject decision made on each trial. This indicated a prominent series of clusters in dorsal anterior cingulate cortex (dACC) centred on the cingulate sulcus and extending dorsally into supplementary and presupplementary motor areas (*z*-max = 4.95; MNI-coordinates = [-1, 5, 38]). **(B)** In addition there were bilateral clusters in anterior insular cortex (AI) at the border with the frontal operculum (left z-max = 3.89, MNI-coordinates = [-32, 10, 10]; right z-max = 4.36, MNI-coordinates = [43, 9, 3]). **(C)** In the second whole-brain analysis, we searched for activity related to the factors which influenced pursue-vs-reject decisions. This identified a cluster of activity in subgenual anterior cingulate cortex (sgACC) related to the reward-magnitude of the opportunity on each trial (*z*-max=3.87, MNI coordinates = [2, 36, -3]). The z-score for the cluster was below the conventional cluster-correction threshold of *z*=3.1. However, we retained it as an ROI because sgACC and neighbouring areas like perigenual anterior cingulate cortex (pgACC) have been linked to the neural representation of subjective value, which is a closely related concept to relative-value (*325*).

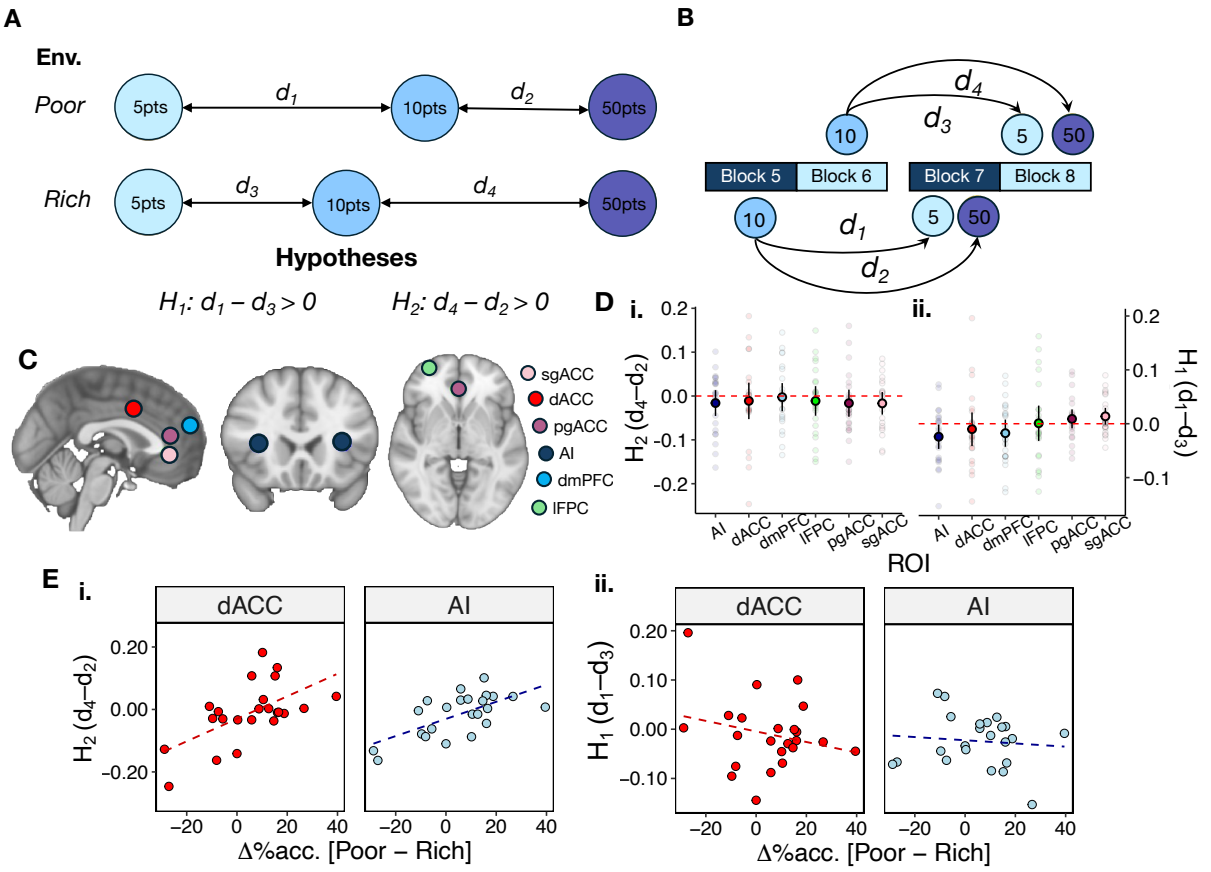


Figure 5.7 Multivariate representations of value are modulated by the environment in dorsal anterior cingulate cortex and anterior insular cortex. (A) We used representational similarity analysis (RSA) to examine how the distance between neural representations of reward options changed between rich and poor environments. Our hypothesis was that in regions representing the task's structure, the distance between representations would mirror changes in the relative value of the 10-point option. We therefore formulated two key hypotheses: (H_1) that the 10-point option would be closer to the 5-point option in rich environments (d_3) compared to poor environments (d_1), i.e., $d_3 - d_1 > 0$, and; (H_2) that the 10-point option would be closer to the 50-point option in poor environments (d_4) compared to rich environments (d_2), i.e., $d_4 - d_2 > 0$. **(B)** Diagram of method for calculating representational distances for RSA. The RSA approach involved comparing the representational distance between pairs of reward options across rich and poor environments (i.e. blocks). We calculated representational distances by comparing the similarity of multi-voxel reward-evoked patterns of activity. To avoid similarity artefacts arising from temporal autocorrelations in BOLD signal, these calculations used multi-voxel patterns from non-contiguous blocks (see diagram). For example, to estimate the distance between 5-point and 10-point options during poor blocks, we might compare the similarity of 10-point evoked patterns of activity in Block

5 (a poor block) to 5-point evoked patterns of activity in Block 7 (also a poor block). See §5.4: Methods for further details. **(C)** We tested H1 and H2 in six cortical ROIs: subgenual anterior cingulate cortex (sgACC), dorsal anterior cingulate cortex (dACC), pregenual anterior cingulate cortex (pgACC), anterior insular cortex (AI), dorsomedial prefrontal cortex (dmPFC) and lateral frontopolar cortex (lFPC). Colour scale corresponds to ROI identity. **(D)** In **(i)**, there was no evidence that representations of 10-point and 5-point options were different between poor and rich blocks in any ROI. Similarly, in **(ii)** there was no evidence that representations of 10-point and 50-point options were different between poor and rich blocks. Together with the result in fig.5.7E, this suggests that the neural representation of reward-options only obeyed changes in the relative value of reward-options in participants who adopted the rational task strategy. Bold points and error bars indicate mean ± 95%CI H1 & H2 scores in each ROI (n.b. CI not corrected for multiple comparisons). Dots indicate H1 and H2 values for individual participants. **(E)** In **(i)** dACC and AI, H2 [$d_4 - d_2$] (i) was positively correlated with the degree of behavioural change between environments – i.e., the more neurally similar the 10-point option was to the 50-point in poor environments, the more frequently subjects pursued the 10-point option in poor environments (and vice versa in rich environments). In contrast **(ii)**, H1 [$d_1 - d_3$] was not correlated with behaviour in dACC or AI ($t_{dACC}(21) = -1.12$, $p=.271$; $t_{AI}(21) = -0.45$, $p=.650$), nor in any other ROI. Points indicate individual participants. Dashed lines indicate line-of-best-fit.

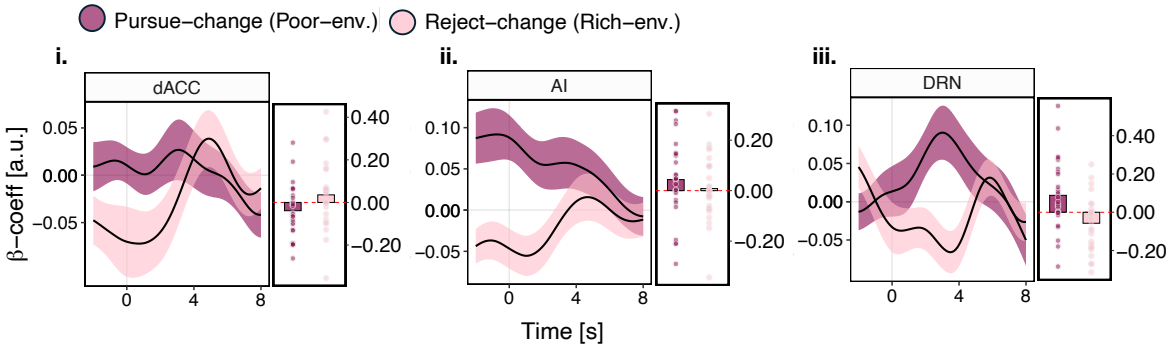


Figure 5.8 Univariate activity in dorsal anterior cingulate cortex and anterior insular cortex does not represent policy-switches. In contrast to DRN **(iii)**, univariate activity in dACC **(i)** and AI **(ii)** did not covary with environment-driven changes in behavioural policy (see fig.5.4). This suggests that although dACC and AI represent changes in the relative value of reward-options given the environment, they do not implement changes in behaviour that arise from those values (see also fig.5.9). Lines and shadings show the mean and standard error (SE) of β weights across participants. Inset panels show distribution of peak regression weights, where bar indicates mean value and dots indicate participant-specific regression weights.

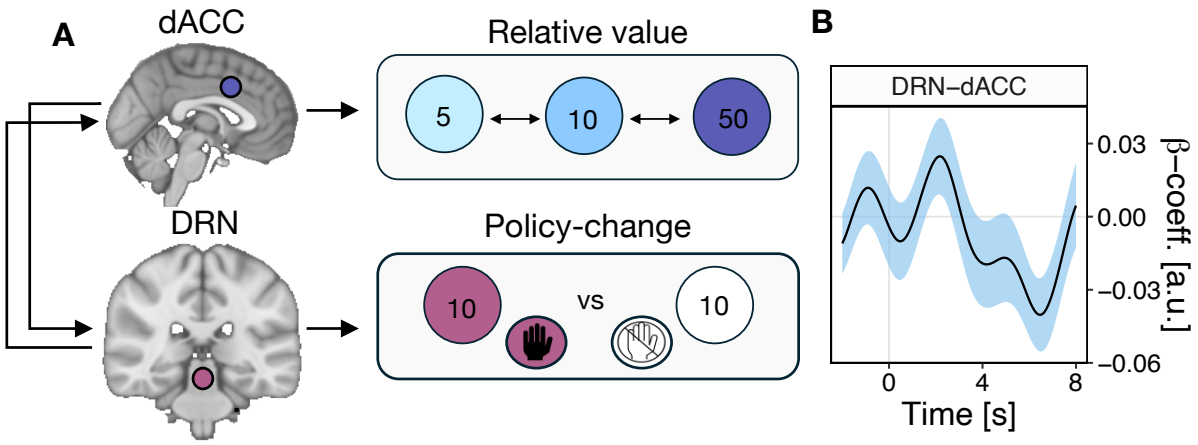


Figure 5.9. A cortico-subcortical circuit for reconciling behavioural policies with the environment. (A) We demonstrate complementary patterns of brain activity in the cortex and dorsal raphe nucleus (DRN) during reward-guided decision-making. In tandem with previous studies, we propose a circuit in which: (i) regions of cortex such as dACC and AI compute the relative value of reward-options given the distribution of recent opportunities and outcomes, and; (ii) DRN implements and/or facilitates changes in behavioural policy that arise from relative value computations. **(B)** Consistent with this hypothesis, we observed a PPI between dACC and DRN as a function of environment-type. The sign of the PPI indicates that there was stronger coactivation between DRN and dACC in poor environments, where behavioural changes were most prominent.

5.3 Discussion

Normative theories of reward-guided behaviour suggest that reward opportunities should be evaluated relative to the context they occur in (*17*). Here, we show that human behaviour obeys this principle. We gave participants a novel behavioural task involving repeated encounters with the same set of reward options, and found that participants were *ceteris paribus* more likely to pursue opportunities in poor environments with few high-value alternatives. fMRI recordings of brain activity linked successful task performance to a distributed neural circuit comprising DRN in the brainstem, and dACC and AI in the cerebral cortex. Within this group, DRN activity signalled changes in option-specific behavioural policies, but only insofar as those switches were appropriate to the richness of the environment. In contrast, multivariate activity in dACC and AI paralleled changes in the relative value of reward opportunities given the environment, suggesting that these regions represented the task structure. Co-activation between DRN and cortical regions like dACC changed as a function of the environment, and was particularly strong during poor environments where changes in behaviour and brain activity were most prominent. We therefore suggest that DRN, dACC and AI constitute a circuit for reconciling behaviour with the richness of the environment – a circuit in which cortical components calculate the relative value of reward opportunities given the current context, and DRN implements changes in behavioural policy consequent on these values.

The present work adds to a growing corpus of experiments inspired by behavioural ecology (*26*, *28*, *84*, *254*). These studies have shown that human participants are keenly sensitive to background reward-rates during foraging-style scenarios, where their behaviour approximates the predictions of optimal foraging theory. Here, we provide further evidence to

this effect by showing that participants pursued a moderately-valuable reward option – the 10-point option – more frequently in poor environments compared to rich environments. A novel feature of our task design was that participants repeatedly encountered the same reward options many times over, and that changes in the environment were engendered by simply manipulating the frequency of each individual option. This meant that we could explicitly track option-specific behavioural policies and their relationship with the richness of the environment. Policy switches were consistent with the behavioural patterns described above, whereby participants switched toward pursuing the 10-point option in poor environments. This is unsurprising because the policy switch framework was a complementary perspective on pursue-vs-reject decisions, rather than a separate behaviour *per se*.

Importantly, pinpointing the trials when policy switches occurred allowed us to uncover changes in DRN activity that signalled these switches. This pattern was specific to DRN, and was specific to switches that were appropriate to the richness of the environment. It was further accompanied by activity representing the richness of the environment itself, which emerged gradually over time within in the same way as environment-related patterns of decision-making. The joint observation of these neural signals is consistent with animal experiments implicating DRN in environment-driven changes in behaviour. This has now been demonstrated for many kinds of behavioural change, including changes between patience and impulsivity during intertemporal choice (*204*, *233*), exploitation and exploration during reversal learning (*208*, *222*, *231*, *302*), and motivation and demotivation during reward seeking (Ch.III-IV). The present results provide the first evidence of this phenomenon in the human brain, and indicate that DRN has a conserved function of matching the general policy governing behaviour with the general reward statistics of the environment.

Intriguingly, environment-related modulation of both behaviour and DRN activity was more prominent in poor environments compared to rich environments. Given that DRN is the

main source of serotonin to the mammalian forebrain, it is tempting to relate this asymmetry to classic theories implicating serotonin in aversive processing (*158*, *211*, *327*). The present results do not implicate DRN activity in aversion *per se* – for example, one of DRN's critical activity patterns related to *pursuits* of the 10-point option in poor environments – but they are broadly consistent with the notion that DRN facilitates behavioural adaptation to adverse circumstances. Whether this function arises specifically from serotonin neurons is, however, unclear from the present data, as many DRN neurons are non-serotonergic, and it is not possible to discriminate serotonergic from non-serotonergic populations in fMRI recordings (*327*, *328*). Future experiments could address this using genetically specific recording and manipulation techniques in rodents, which should be capable of the basic behavioural paradigm developed here.

In contrast to DRN's role in policy switches, dACC and AI were linked to changes in the external task environment. This accords with previous evidence that these regions track rewards over multiple timesteps and, in dACC's case, that it evaluates rewards relative to the counterfactual alternatives of the current context (*24*, *65*, *84*, *227*). We probed this role here by testing whether the representational distance between reward options changed across rich and poor environments. This demonstrated that individual differences in the distance between 10-point and 50-point options paralleled the degree of behavioural adaptation to rich and poor environments, suggesting that successful performance of the task was fundamentally linked to warps in the neural representations of value – warps which brought the 10-point and 50-point options closer together when they evoked the same behavioural policy, and pushed them further apart when behaviour diverged. Strikingly similar patterns have been observed in experiments applying RSA to other cognitive domains like navigation, where neural representations of spatial location appear to warp according to an agent's current and goal locations (*329*). To our

knowledge, the presents results are the first evidence of the same phenomenon in value-space, and suggest that warping is a common neural motif of context-dependent computation.

Notably, dACC and AI were the only ROIs with task-related changes in representational distances, and they were selected as ROIs in the first instance because they showed univariate increases in activity during reward-pursuit decisions. It is important to highlight, therefore, that we measured the distance between multivariate patterns of reward-evoked activity with cosine dissimilarity – a method that obviates the influence of univariate changes on representational distance. This means, for example, that environment-driven changes in the representational distance between 10-point and 50-point options is not reducible to the fact that: (i) both 10-point and 50-point options are pursued in poor environments, and therefore evoke similar levels of univariate activity, and (ii) the 50-point option is pursued more than the 10-point option in rich environments, leading to divergent in univariate activity.

5.4 Methods

5.4.1 Participants

Twenty-nine participants performed the experiment. All were between 18-40 years of age, reported normal or corrected-to-normal vision, and reported no current diagnosis or treatment for psychiatric or neurological disorder. Participants received £40 for completing the experiment and could earn an additional payment of up to £10 depending on their performance in the decision-making task. All relevant ethical regulations governing research with human participants were observed, and participants provided written informed consent at the beginning of each session. Ethical approval was given by the Oxford University Central University Research Ethics Committee (CUREC) (Ref-Number: MSD-IDREC-R55856/RE001). Two participants were excluded due to excessive head motion during fMRI scans, which prevented accurate registration (motion outliers > 15% of total fMRI volumes).

5.4.2 Behavioural task

Before performing the experiment in the MRI scanner, participants received written instructions and completed a short practice version of the task. The experiment proper involved five rounds of the behavioural task, and each round was approximately 9 minutes (fig.5.1B). Participants were given a 60s break between rounds, meaning that the total duration was approximately 50 minutes.

The behavioural task involved a series of encounters with sequential reward opportunities (fig.5.1A). The aim was to earn as many points as possible in a fixed amount of time (45 minutes), and participants were incentivised to do this because their payment was partly determined by their performance. In brief, participants were presented with a single reward opportunity on each trial and needed to decide whether to pursue (i.e. accept) or reject it. Pursuing an opportunity incurred a temporal opportunity cost. Given the time-limited nature of the experiment, it was therefore critical to track the frequency of the reward options over time – i.e., to anticipate what kind of reward opportunity they might get in future trials – when deciding whether the opportunity cost was worth incurring. To make the task intuitive, it was presented as a treasure-hunt scenario in which the participant played the role of a ship's captain searching for treasure opportunities (fig.5.1A). Upon encountering a treasure opportunity, the captain needed to decide whether it was worth the time and effort to sail toward it (pursue-decision), or to continue to explore the seas elsewhere (reject-decision).

Reward opportunities were drawn from a set of three, discrete options represented by visual stimuli: (1) a low-value option (5 points; bronze medallion) (2) a moderate-value option (10 points; silver medallion), and (3) and a high-value option (50 points; gold medallion). The value of options was deterministic, constant over time, and known *a priori* to participants before they performed the experiment. The task, thus, did not involve learning about stimulus-reward relationships, or uncertainty about the payout of an option if/when it was pursued. Although option values were constant, the frequency of each option was systematically manipulated to create differences in the richness of the environment. Environments were operationalised via blocks of trials that were time-limited to 4.5 minutes. In rich blocks, the 50-point option was more frequent than the 10-point and 5-point options [fig.5.1B; $P(5pt|Rich) = 0.16$, $P(10pt|Rich) = 0.33$, $P(50pt|Rich) = 0.50$]. In poor blocks, the 5-point option was more frequent than the 10-point and 50-point counterparts [fig.5.1B; $P(5pt|Poor) = 0.50$,

P(10pt|Poor) = 0.33, *P(50pt|Poor)* = 0.16]. There were no environment-related cues during the task, and participants were not informed that the task was divided into discrete rich and poor blocks – all they were told was that the frequency of the options might vary over time. As such, participants needed to learn the richness of the environment by tracking the frequency with which specific options occurred.

Per above, the experiment consisted of five, 9-minute rounds and each round consisted of two 4.5-minute blocks (fig.5.1B). The amount of time remaining in each round was signalled using a timer in the top right-hand side of the screen. Each round featured one rich block and one poor block, and the order of rich/poor blocks within each round was random (fig.5.1B). The precise number of trials completed in each block – and therefore each round – varied depending on how many options were pursued during the allotted time (see supplementary fig.5.1F; $\mu(\text{trials})_{\text{rich}} = 15.95 \pm 1.06$, $\mu(\text{trials})_{\text{poor}} = 18.06 \pm 1.27$, $t(26)_{\text{rich-vs-poor}} = 15.52$, $p < .001$). Each trial began with an inter-trial-interval (ITI; *ITI ~TruncExp*(μ *= 4.5, min=3.5, max=5.5*), during which a treasure chest visual stimulus representing an upcoming reward opportunity descended from the top to the centre of the screen. A picture of a medallion replaced the treasure chest when the chest reached the centre of the screen, thereby revealing the reward-value of the opportunity on the current trial. Participants needed to consider the reward-option for a short duration (*Opportunity ~TruncExp*(μ *= 4.5, min=3.5, max=5.5*) before it displaced either left- or right-ward from the centre. The displacement functioned as a go-cue which indicated that the opportunity was eligible for pursuit. Imposing the go-cue delay created a temporal dissociation between brain activity related to pursue-vs-reject decisions, and the motor processes undertaken to execute the decision via button press. Participants could pursue options by making a button press response within 1s of the go-cue. Different responses were required for leftward and rightward pursuits to prevent anticipatory planning of movements during the

go-cue delay. To reject an opportunity, participants needed to withhold any button-press response and let the opportunity pass.

Pursuing an opportunity incurred a temporal opportunity cost via an action-outcome delay (*Action-outcome ~ TruncExp*(μ *= 4.5, min=3.5, max=5.5*) and a reward feedback delay (*Rw-feedback ~ TruncExp*(μ *= 4.5, min=3.5, max=5.5*). The duration of these delays was equivalent to the time needed for one future encounter – in other words, pursuing a reward opportunity on the current trial meant forgoing a reward opportunity in the future. The magnitude of this cost was explained to participants during the instructions, and visually indicated during the task via a passing treasure-chest stimulus during the action-outcome delay, which represented the opportunity forgone.

The parameters of the task were carefully chosen to ensure that the optimal strategy was to always pursue the 10-point option in poor blocks, and to never pursue the 10-point option in rich blocks. Per above, blocks were not signalled to participants, and we therefore did not expect them to display the optimal pattern of behaviour immediately. Instead, we expected that they would move towards the optimal pattern after accumulating experience of each environment over time. Designing the task in this way ensured that there was a reward-maximising rationale for participants to change their behavioural policy toward the 10-point option.

The task was implemented in MATLAB R2022b (Mathworks Inc.) using the Psychophysics Toolbox extension (*281*).

5.4.3 Behavioural analysis

Our initial analysis of behaviour quantified the factors influencing the pursue-vs-reject decision made on each trial. We accomplished this with a series of mixed-effect binomial

generalised linear models (GLMs) with participant-identity as a random variable. All GLMs therefore accounted for inter-subject variability in effects of interest.

We first investigated whether pursue-vs-reject decisions were influenced by the richness of the environment, where richness of the environment was operationalised as the experimentally controlled environment-type (i.e. rich-vs-poor block) on a given trial:

GLM5.1a

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 environment\text{-}type_t + \beta_3(reward\text{-}magnitude_t * environment\text{-}type_t) + \mu_0 + \mu_1 reward\text{-}magnitude_t + \mu_2 environment\text{-}type_t + \mu_3(reward\text{-}magnitude_t * environment\text{-}type_t) + \varepsilon$$

Where $\beta_{0\text{-}3}$ are fixed-effects and $\mu_{0\text{-}3}$ are random-effects within each participant.

We followed this with a GLM that tested the effect of the environment separately for each reward option:

GLM5.1b

$$logit(decision_t) = \beta_o + \beta_1 environment\text{-}type_t + \mu_0 + \mu_1 environment\text{-}type_t + \varepsilon$$

Where $\beta_{0\text{-}1}$ are fixed-effects, $\mu_{0\text{-}1}$ are random-effects within each participant. Per above, GLM5.1b was fit separately to subsets of trials featuring the 5-point, 10-point and 50-point options, respectively.

We then investigated whether the effect of environment-type on pursue-vs-reject decisions was modulated by the amount of time elapsed within a block (*trial-within-block*). This interaction is consistent with the hypothesis that participants incrementally learned the richness of the environment with experience.

GLM5.2

$$logit(decision_t) = \beta_o + \beta_1 environment\text{-}type_t + \beta_2\ trial\text{-}within\text{-}block_t + \beta_3(environment\text{-}type_t * trial\text{-}within\text{-}block_t) + \mu_0 + \mu_1 environment\text{-}type_t + \varepsilon$$

Where $\beta_{0\text{-}3}$ are fixed-effects and $\mu_{0\text{-}1}$ are random-effects within each participant. GLM5.2 was fit separately to subsets of trials featuring the 5-point, 10-point and 50-point options, respectively.

We then tested the same hypothesis from a complementary perspective by operationalising the richness of the environment as the average-value of the reward-options encountered on the previous five-trials (*average-value*):

Average-value

$$\text{average-value}_t = \text{mean}(\text{reward-magnitude}_{t-1}\text{: reward-magnitude}_{t-5})$$

Operationalising the richness of the environment in this way captured a participant's experience of the task, instead of relying on the task's underlying generative structure. We investigated the effect of average-value on pursue-reject decisions with the following GLM:

GLM5.3

$$logit(decision_t) = \beta_o + \beta_1 average\text{-}value_t + \mu_0 + \mu_1 average\text{-}value_t + \varepsilon$$

Where $\beta_{0\text{-}1}$ are fixed-effects and $\mu_{0\text{-}1}$ are random-effects within each subject. GLM5.3 was fit separately to subsets of trials featuring the 5-point, 10-point and 50-point options, respectively.

Because the task featured three discrete reward-options that occurred many times over, we were able to precisely identify changes in a participant's behavioural policy toward each option (where behavioural policy indicates an option-specific pursue-vs-reject strategy; fig.5.3A). We first tested the hypothesis that participants formed option-specific policies by examining whether pursue-vs-reject decisions were autocorrelated over consecutive encounters with a particular reward-option:

GLM5.4

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 previous\text{-}policy_t + \mu_0 + \mu_1 reward\text{-}magnitude_t + \mu_2 previous\text{-}policy_t + \varepsilon$$

Where $\beta_{0\text{-}2}$ are fixed-effects and $\mu_{0\text{-}2}$ are random-effects within each subject. Here, previous-policy indicates the pursue-vs-reject decision the last time that the reward option on trial *t* was encountered.

We confirmed that behavioural policies were option-specific by testing the effect of the pursue-vs-reject decision taken on the preceding trial (i.e. *t*-1) regardless of which reward option occurred:

GLM5.5

$$logit(decision_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \beta_2 previous\text{-}action_t + \mu_0 + \mu_1 reward\text{-}magnitude_t + \mu_2 previous\text{-}action_t + \varepsilon$$

Where $\beta_{0\text{-}2}$ are fixed-effects and $\mu_{0\text{-}2}$ are random-effects within each subject. Together, GLM5.4 and GLM5.5 allowed us to test whether temporal structure in behaviour was option-specific, or whether participants also had autocorrelated predispositions to pursue-vs-reject options in general, regardless of reward-option identity.

We then operationalised policy-switches as trials on which participants changed their pursue-vs-reject decisions across successive encounters with a particular option – for example, if a participant rejected the 10-point option on trial *t=1* and subsequently pursued the 10-point option on trial *t=5*, trial *t*-5 was therefore coded as policy-switch event (fig. 5.3A). We also defined trials along an action dimension, where an action-switch occurred when participants changed their pursue-vs-reject decision across consecutive trials (regardless of which reward-option appeared). This allowed us to compare option-specific tendencies (i.e. the policy dimension) in purse-vs-reject decisions with general tendencies (i.e. the action dimension) in pursue-vs-reject decisions.

After defining each trial along policy-switch and action-switch dimensions, we conducted a series of analyses to identify the factors that predicted policy-switches. Because the policy-switch status of each trial was binary (switch=1, stay =0), we used mixed-effect binomial generalised linear models (GLMs) where the dependent variable was the policy-switch status on each trial, and participant-identity was a random variable.

We first investigated whether the probability of policy-switches differed as a function of reward-option:

GLM5.6

$$logit(policy\text{-}switch_t) = \beta_o + \beta_1 reward\text{-}magnitude_t + \mu_0 + \mu_1 reward\text{-}magnitude_t + \varepsilon$$

Where $\beta_{0\text{-}1}$ are fixed-effects and $\mu_{0\text{-}1}$ are random-effects within each participant.

We then investigated whether the direction of policy-switches for the 10-point option was modulated by environment-type. To do this, we separated policy-switches into two different types: (1) *reject-switches*, when subjects rejected a given reward-option after pursuing it on the preceding encounter, and (2) *pursue-switches,* where subjects pursued a given reward-option after rejecting it on the previous encounter. The optimal strategy for performing the task involved pursuing the 10-point option in poor environments and rejecting the 10-point option in rich environments. Insofar as behaviour approximated this strategy, we expected that pursue-switches would occur more frequently in poor environments and reject-switches would occur more frequently in rich environments. We tested these hypotheses with the following GLMs:

GLM5.7a

$$logit(reject\text{-}switch_t) = \beta_0 + \beta_1 environment\text{-}type_t + \mu_0 + \mu_1 environment\text{-}type_t + \varepsilon$$

GLM5.7b

$$logit(pursue\text{-}switch_t) = \beta_0 + \beta_1 environment\text{-}typet + \mu_0 + \mu_1 environment\text{-}type_t + \varepsilon$$

Where $\beta_{0\text{-}1}$ are fixed-effects and $\mu_{0\text{-}1}$ are random-effects within each participant. GLM5.7a and GLM5.7b were fit separately to subsets of trials featuring the 5-point, 10-point and 50-point options respectively. We then tested the same hypotheses in an analysis where the richness of the environment was operationalised as the average-value of the reward-options encountered on the previous five-trials:

GLM5.8a

$$logit(pursue\text{-}switcht) = \beta_o + \beta_1 average\text{-}value_t + \mu_0 + \mu_1 environment\text{-}type_t + \varepsilon$$

GLM5.8b

$$logit(reject\text{-}switch_t) = \beta_o + \beta_1 averaget\text{-}value_t + \mu_0 + \mu_1 environment\text{-}type_t + \varepsilon$$

Where $\beta_{0\text{-}1}$ are fixed-effects and $\mu_{0\text{-}1}$ are random-effects within each participant. GLM5.8a and GLM5.8b were fit separately to subsets of trials featuring the 5-point, 10-point and 50-point options respectively. All behavioural analyses were performed in *R* (R Core Team, 2023) using the *lme4* implementation of mixed-effects GLMs (*330*).

5.4.4 Imaging data acquisition

Structural and functional MRI data was collected with a Siemens 7 Tesla MRI scanner. High-resolution functional data were acquired with a multiband gradient-echo T2* echo planar imaging sequence with 1.5mm isotropic voxels, multiband acceleration factor 2, repetition time (TR) = 1.775s, echo time (TE) = 17.8ms, flip angle = 66°, and GRAPPA acceleration factor 2. The parameters were selected to maximise signal-to-noise ratio in subcortical areas. To accommodate the high temporal and spatial resolution of the protocol, functional scans had a limited field of view (FOV) oriented at 30 degrees with respect to the AC-PC line (66 slices

with a coverage of 99 mm). The FOV captured all regions of interest in the midbrain, brainstem and cortex. Before acquiring the task-related functional scan, we acquired a pre-saturation single-measurement, whole-brain functional scan with the same orientation. The pre-saturation scan was used to facilitate registration of the limited-FOV task-related functional scan to the whole brain.

Structural data were acquired using a T1-wieghted MP-RAGE sequence with 0.7mm isotropic voxels, GRAPPA acceleration factor 2, TR = 2200ms, TE = 3.02ms, and inversion time (TI) = 1050ms. To correct distortions arising from inhomogeneities in the magnetic field, a fieldmap sequence was acquired with 2mm isotropic voxels, TR = 620ms, TE1 = 4.08ms, and TE2 = 5.1ms. To account for the effects of physiological noise on functional MRI data, participants were fitted with a pulse oximeter and respiratory bellows that acquired cardiac and respiratory timeseries at 50Hz using a BioPac MP160 device (BIOPAC Systems Inc., USA).

5.4.5 fMRI data preprocessing

Preprocessing of fMRI data was performed with the FMRIB Software Library (*312*, *331*). The Brain Extraction Tool (*332*) was used to separate brain from non-brain matter in structural and functional images. Functional images were normalised, spatially smoothed (Gaussian kernel with a 3mm full-width half-maximum) and temporally high-pass filtered (3 dB cut-off = 100s), and artefacts arising from head motion were removed using MCFLIRT (*333*). Registration of task-related functional images to Montreal Neurological Institute (MNI)-space was performed in three stages:

1. The task-related limited-FOV EPI was registered to the pre-saturation whole-brain EPI using FMRIB's Linear Image Registration Tool with 6 degrees of freedom transformation.

2. The whole-brain EPI was registered to the subject-specific structural image using Boundary-Based Registration (BBR) incorporating fieldmap correction (*334*).
3. The subject-specific structural image was registered to a 1mm resolution Standard MNI template with FMRIB's Non-linear Registration Tool (FNIRT; (*312*)).

5.4.6 Whole-brain fMRI data analysis

Statistical analysis of whole-brain functional data was performed at two levels using FMRIB's Expert Analysis Tool (FEAT; (*312*)). In the first level, a univariate general linear model was used to compute parameter estimates for each regressor in each participant (*335*). Contrast and variance estimates for each parameter in each participant were subsequently combined in a mixed-effects analysis conducted at the second level (FLAME 1 + 2), where participant-identity was a random effect (*336*). Significance testing was performed with cluster-correction, a cluster significance threshold of $p = .001$, and a voxel inclusion threshold of $z = 3.1$. Data were pre-whitened before analysis to account for temporal autocorrelations in BOLD signal.

We performed two-whole brain analyses. The first GLM identified voxels where BOLD signal represented the pursue-vs-reject decision made on each trial. Note that the pursue-vs-reject decision was regressed at two different timepoints: (i) at the time of opportunity onset, i.e., when participants made a decision about whether to pursue the reward opportunity, and; (ii) at go-cue onset, i.e., at the time when participants implemented their decision in motor output. These events were separated in time by a jittered interval drawn from $Opportunity \sim TruncExp(\mu = 4.5, min=3.5, max=5.5)$, meaning that neural activity at each time was temporally decorrelated. The second GLM identified voxels where BOLD signal represented the reward-magnitude of the reward-option on each trial. We could not include both pursue-vs-

reject and reward-magnitude regressors in the same GLM because they were highly correlated (r = 0.64). The formulas for GLMs were:

GLM5.9

$$BOLD_{v,t} = \beta_1 \text{opportunity-onset} + \beta_2 \text{trial-number} + \beta_3 \text{decision(opportunity-onset)} + \beta_4 \text{go-cue-onset} + \beta_5 \text{ decision(go-cue-onset)} + \beta_6 \text{feedback-onset} + \beta_7 \text{reward-feedback}$$

GLM5.10

$$BOLD_{v,t} = \beta_1 \text{opportunity-onset} + \beta_2 \text{trial-number} + \beta_3 \text{reward-magnitude} + \beta_4 \text{go-cue-onset} + \beta_5 \text{decision(go-cue-onset)} + \beta_6 \text{feedback-onset} + \beta_7 \text{reward-feedback}$$

Where $BOLD_{v,t}$ is a t x v matrix containing the BOLD signal measurements at time t in voxel v. The regressors are as follows: *opportunity-onset* is an unmodulated regressor representing the occurrence of the reward-option stimulus on each trial; *trial-number* is a parametric regressor representing the number of trials completed in the experiment as a proxy for the passage of time, time-locked to opportunity onset; *decision(opportunity-onset)* is a binary regressor reflecting the pursue-vs-reject decision on each trial, timelocked to opportunity onset; *reward-magnitude* is a parametric regressor reflecting the value $\in \{5, 10, 50\}$ of the reward-option timelocked to opportunity onset; *go-cue-onset* is an unmodulated regressor representing the occurrence of the go-cue stimulus on each trial; *decision(go-cue-onset)* is a binary regressor reflecting the pursue-vs-reject decision on each trial, timelocked to the onset of the go-cue stimulus; *feedback-onset* is an unmodulated regressor representing the occurrence of reward-feedback received on trials when the opportunity was pursued; *reward-feedback* is a parametric regressor with three levels (5, 10, 50) representing the value of the reward-option, and time-locked to feedback presentation on trials where opportunities were accepted.

All regressors were boxcar functions with a constant duration of 0.1s and convolved with a double-gamma hemodynamic response function. Further non-task confound regressors

were added to reduce noise in BOLD signal, including: (1) head motion parameters estimated using MCFLIRT during pre-processing (*333*); (2) regressors for voxel-wise estimates of physiological noise arising from cardiac and respiratory activity, estimated using FSL's Physiological Noise Monitoring (PNM) tool (*337*), and; (3) regressors for motion outliers, indicating volumes with head motion that could not be corrected with linear methods.

5.4.7 ROI timecourse fMRI analysis

Additional analysis of brain activity was conducted on epoched timeseries extracted from regions of interest (ROIs). Anatomical ROIs were constructed for a series of brainstem and midbrain nuclei previously implicated in reward-guided decision-making including dorsal raphe nucleus (DRN), Locus Coeruleus (LC), Nucleus Basalis (NB), a combined mask covering the Substantia Nigra and Ventral Tegmental Area (Midbrain Dopaminergic Nuclei; MDB), and Habenula (Hb). The LC mask derived from a brain atlas developed by Pauli and colleagues (*338*). DRN, MDB, Hb and NB masks were manually drawn in MNI-space based on an atlas of the human brain (*44*). The DRN mask was drawn in consultation with the anatomical guidelines of Kranz and colleagues (*339*) to ensure specificity to the dorsal portion of raphe nucleus. Standard-space masks were transformed to subject-specific structural space by applying a standard-to-structural warp field. To ensure conformation between anatomical masks and subject-specific neuroanatomy, subject-specific masks were manually checked and edited by two experimenters and evaluated for inter-rater reliability by a third experimenter. Structural-space masks were then transformed from structural to functional space by applying a structural-to-functional affine matrix, and were subsequently thresholded, binarized and dilated by one voxel.

To prepare ROI data for analysis, the filtered time-series of BOLD signal from each voxel was averaged, normalised and up-sampled by a factor of 20 with spline interpolation (*340*). Up-sampled timeseries were then epoched in 10s windows starting 2s before and ending 8s after the reward-opportunity onset on each trial. Timeseries data was analysed by fitting a GLM with ordinary least squares (OLS) at each timepoint in each epoch. The statistical significance of parameters in GLMs was assessed using a leave-one-out procedure that avoided temporal biases in the selection of peak-effects (*341*). The procedure involved estimating, for each regressor *r* in each participant *s*, the time *t* at which the mean effect of *r* was greatest in the remaining *N*-1 participants. The effect of *r* in *s* was then estimated by calculating the effect size for *r* at time *t* in participant *s*. This was repeated iteratively for each participant, leaving an *N*x1 vector of parameter estimates that were not biased by the timecourse of data of individual participants. We assessed statistical significance of participant -specific parameter estimates by running a single-sample t-test on the vector of parameter estimates. All t-tests were two-sided. Whenever analyses involved three or more ROIs, we maintained the family-wise error rates at α=.05 using the Holm-Bonferroni procedure for multiple comparisons (*342*).

We tested which ROIs represented the richness of the environment with the following GLM:

GLM5.11

$$BOLD_{i,\,t,\,r} = \beta_0 \text{constant} + \beta_1 \text{environment-type} + \beta_2 \text{trial-within-block} + \beta_3 (\text{environment-type})^{*} (\text{trial-within-block}) + \beta_4 \text{trial-number}$$

Where: $BOLDi,_t$ is an *i* x *t* x *r* (*i* = timepoint, *t* = trial, *r* = ROI) matrix containing the average BOLD signal at each timepoint in each epoch in ROI *r*; *constant* is an unmodulated constant regressor; *environment-type* is the experimentally manipulated environment-type (rich-vs-poor) on trial *t*; *trial-within-block* is the trial-number within the block at time *t* as a proxy for

time-elapsed within the environment, and; *trial-number* is number of trials passed from the beginning of the experiment to trial *t*.

We then tested which ROIs represented policy-switches with the following GLM:

GLM5.12

$$BOLD_{i,\,t,\,r} = \beta_0 \text{constant} + \beta_1 \text{policy-switch} + \beta_2 \text{action-switch} + \beta_3 \text{trial-number}$$

Where *policy-switch* is a binary variable indicating whether a policy-switch occurred on trial *t*, and *action-switch* is a binary variable indicating whether an action-switch occurred on trial *t* (see fig.5.3A for definitions).

To investigate brain activity related specifically to environment-driven policy-switches, we classified policy-switch events as either *congruent* or *incongruent* in relation to the environment they occurred in. Recall that the optimal strategy for preforming the task was to pursue the 10-point opportunity in poor environments and reject it in rich environments. Accordingly, we coded pursue-switches in poor environments as congruent and pursue-switches in rich environments as incongruent. Correspondingly, reject-switches in rich environments were congruent and reject-switches in poor environments were incongruent. We used these four event types as regressors in the following series of GLMs:

GLM5.13a

$$BOLD_{i,\,t,\,r} = \beta_0 \text{constant} + \beta_1 \text{congruent-switch-poor} + \beta_3 \text{trial-number}$$

GLM5.13b

$$BOLD_{i,\,t,\,r} = \beta_0 \text{constant} + \beta_1 \text{incongruent-switch-poor} + \beta_3 \text{trial-number}$$

GLM 5.13c

$$BOLD_{i,\,t,\,r} = \beta_0 \text{constant} + \beta_1 \text{congruent-switch-rich} + \beta_3 \text{trial-number}$$

GLM 5.13d

$$BOLD_{i,\,t,\,r} = \beta_0 \text{constant} + \beta_1 \text{incongruent-switch-rich} + \beta_3 \text{trial-number}$$

Where *congruent-switch-poor* is a binary variable indicating whether a congruent policy-switch in a poor environment occurred on trial *t*, *incongruent-switch-poor* is a binary variable indicating whether an incongruent policy-switch in a poor environment occurred on trial *t*, *congruent-switch-rich* is a binary variable indicating whether a congruent policy-switch in a rich environment occurred on trial *t*, and *incongruent-switch-rich* is a binary variable indicating whether an incongruent policy-switch in a rich environment occurred on trial *t*. We tested differences in the brain activity evoked by these event-types by performing paired-sample t-tests on the effect-sizes for different event types.

Per above, the optimal strategy in the task involved changing behavioural-policy toward the 10-point option depending on the environment. Conversely, there was no environment-driven reason for changing behavioural-policy toward the 5-point option, which should always have been rejected. To further test for brain activity related to environment-driven policy-switches, we separately analysed the effect of policy-switches on trials featuring the 10-point and 5-point options. We did this by performing the analyses in GLM5.13a and GLM5.13c for subsets of data containing 10-point option trials and the 5-point option trials, respectively.

5.4.8 Representational similarity analysis

We further investigated neural representations of the task environment using representational similarity analysis [RSA; (*324*)]. RSA was performed on the following cortical ROIs: anterior insular cortex (AI), dorsal anterior cingulate cortex (dACC), lateral frontopolar cortex, subgenual anterior cingulate cortex (sgACC), pregenual anterior cingulate cortex (pgACC) and dorsomedial prefrontal cortex (dmPFC). AI, dACC and lFPC were selected

because they exhibited clusters of activity related to the pursue-vs-reject contrast in whole-brain fMRI analysis (GLM5.9; see appendix table 1). Similarly, sgACC was selected because it exhibited activity related to the reward-magnitude contrast in a whole-brain fMRI analysis (GLM5.10; see appendix table 2). dmPFC and pgACC were selected based on functional activations in previous studies (*325*, *326*). Each ROI was constructed as a sphere with 7mm radius centred on the voxel with the highest cluster-corrected β-weight.

We used RSA to compare differences in the neural representation of reward-options between rich and poor environments. We predicted that the distance between option-specific neural representations would mirror changes in the relative-value of reward options given the environment – i.e., the 10-point would be closer to the 50-point option in poor environments compared to in rich environments, and closer to the 5-point option in rich environments compared to poor environments (fig.5.7A). The analysis consisted of two stages: (i) quantifying the multi-voxel patterns of activity evoked by 5-point, 10-point and 50-point reward options in rich and poor blocks of the experiment, and; (ii) computing the distance between multi-voxel representations, and quantifying how distances changed as a function of environment-type.

We measured multi-voxel representations of reward-options using a GLM implemented in FSL FEAT. fMRI data pre-processing followed the pipeline described in *fMRI data preprocessing* above except that spatial smoothing was omitted to maximise voxel-specific information. Omitting spatial smoothing prevented accurate registration for two participants, who were therefore excluded from the analysis (N_{RSA} = 25). The GLM quantified the multi-voxel patterns of activity evoked by each reward-option with separate binary predictors indicating the onset of each option in each block (i.e., 30 predictors in total). It had the following formula:

GLM5.14

$BOLD_{v,t,r}$ =

$\beta_1$op.-onset(5pt; block-1) + $\beta_2$op.-onset(10pt; block-1) + $\beta_3$op.-onset(50pt; block-1) +
$\beta_4$op.-onset(5pt; block-2) + $\beta_5$op.-onset(10pt; block-2) + $\beta_6$op.-onset(50pt; block-2) +
$\beta_7$op.-onset(5pt; block-3) + $\beta_8$op.-onset(10pt; block-3) + $\beta_9$op.-onset(50pt; block-3) +
β_{10}op.-onset(5pt; block-4) + β_{11}op.-onset(10pt; block-4) + β_{12}op.-onset(50pt; block-4) +
β_{13}op.-onset(5pt; block-5) + β_{14}op.-onset(10pt; block-5) + β_{15}op.-onset(50pt; block-5) +
β_{16}op.-onset(5pt; block-6) + β_{17}op.-onset(10pt; block-6) + β_{18}op.-onset(50pt; block-6) +
β_{19}op.-onset(5pt; block-7) + β_{20}op.-onset(10pt; block-7) + β_{21}op.-onset(50pt; block-7) +
β_{22}op.-onset(5pt; block-8) + β_{23}op.-onset(10pt; block-8) + β_{24}op.-onset(50pt; block-8) +
β_{25}op.-onset(5pt; block-9) + β_{26}op.-onset(10pt; block-9) + β_{27}op.-onset(50pt; block-9) +
β_{28}op.-onset(5pt; block-10) + β_{29}op.-onset(10pt; block-9) + β_{30}op.-onset(50pt; block-10)

Where $BOLD_{v,t,r}$ is a t x v x r matrix containing the BOLD signal measurements at time t in voxel v in ROI r, $\beta_{1\text{-}30}$ are unmodulated regressors representing the occurrence of a given reward-option in a given block (e.g. option-onset(5pt; block-1) indicates onsets of the 5-point reward option in the first block of the experiment). All option-onset regressors were implemented as boxcar functions with a constant duration of 0.1s and convolved with a double-gamma hemodynamic response function. Confound regressors for physiological noise and motion artefacts were included, as per whole-brain fMRI analysis.

For each ROI r, and each predictor I, in $\beta_{1\text{-}30}$, GLM5.14 produced a vector of v regression weights captured the effect of β_i at each voxel v in r. Each vector of regression weights constituted a multivariate neural representation of a reward-option. We then quantified pairwise distances between multivariate representations using cosine dissimilarity – i.e., the complement of the cosine-angle between vectors. We used cosine dissimilarity as opposed to standard parametric (e.g. Pearson's R) and non-parametric (e.g. Kendall's Tau) methods to avoid artefacts arising from univariate differences in activity (see §5.3: Discussion for further details). Our analysis focussed on four representational distances; (i) d_1, the distance between 5-point and 10-point options in poor environments; (ii) d_2, the distance between 10-point and 50-point options in poor environments; (iii) d_3, the distance between 5-point and 10-point

options in rich environments, and; (iv) d_4, the distance between 10-point and 50-point options in rich environments.

To avoid biases arising from the temporal autocorrelation of BOLD signal, distance calculations were performed by comparing option-specific representations between non-contiguous blocks of the same environment-type (fig.5.7B). For example, to calculate d_1 – the distance between 5-point and 10-point options in poor environments – we might compare β_{11}op.-onset(10pt; block-4) to β_{16}opportunity-onset(5pt; block-6), where blocks 4 and 6 are poor environments that are 1-block apart (block-space, hence). Because the analysis was performed to compare changes in representational distance between environments, we excluded representational distance scores if their block-space was not present for both rich and poor environments. To avoid overrepresentation of specific block-space values in representational distance scores, we calculated the mean representational distance at each value of block-space for each environment. For final representational distances, we calculated the mean representational distance over all levels of block-space in each environment. This process was performed separately for each ROI in each participant.

Once subject-specific d_1, d_2, d_3, and d_4, scores were calculated, we tested our hypotheses by taking the difference between environment-specific distance scores (e.g. H1= d_1-d_3; H2 = d_4 – d_2). We tested the relationship between representational distances and behaviour by performing Pearson's *R* correlations between H1 and H2 scores and the degree of behavioural change for the 10-point option between environments [Δ(pursue-rate = pursue-rate$_{poor}$ – pursue-rate$_{rich}$]. We supplemented this by dividing participants into two groups based on their degree of behavioural change: (i) a sensitive group, where (pursue-rate$_{poor}$ – pursue-rate$_{rich}$ ≥ 0), and; (ii) an insensitive group, where (pursue-rate$_{poor}$ – pursue-rate$_{rich}$ < 0). We then tested whether H2 scores were different between groups using a two-sample t-test.

In addition, we investigated differences in representational distance regardless of behaviour by testing H1 and H2 against 0 using single-sample t-tests. These analyses were performed separately for each ROI. The family-wise error rate was controlled separately for t-test and correlation analyses using the Bonferroni-Holm correction for multiple comparisons (*342*).

5.4.9 Analysis of midbrain dopaminergic function

Finally, we conducted additional analyses of brain activity and behaviour to probe the role of the midbrain dopaminergic nuclei (MDB) – a combined ROI including SN and VTA. We tested the hypothesis that MDB activity signalled pursue-vs-reject decisions with the following GLM:

GLM5.15

$$BOLD_{i,\,t,} = \beta_0 \text{constant} + \beta_1 \text{pursue-vs-reject} + \beta_3 \text{trial-number}$$

Where *pursue-vs-reject* was the binary pursue-vs-reject decision taken on each trial (pursue=1; reject = 0). We then probed whether MDB activity might signal the difference between the present reward opportunity and the recent average value of reward opportunities (referred to as value-difference, hence) in the manner of a reward prediction error. Value difference was calculated as:

Value difference

$$\text{Value-difference} = \text{reward-magnitude}_t - \text{average-value}_t$$

See above formal calculation for *average-value*. We first tested whether value-difference was related to behaviour with the following GLM:

GLM5.16

$$logit(decision_t) = \beta_o + \beta_1 value\text{-}difference + \mu_0 + \mu_1\ value\text{-}difference_t + \varepsilon$$

Note that we did not enter *reward-magnitude*, *average-value* or *environment-type* as additional regressors because they were highly correlated with *value-difference*. We then probed the relationship between MBD activity and value-difference with the following GLM:

GLM5.17

$$BOLD_{i,\ t,\ r} = \beta_0 constant + \beta_1 value\text{-}difference + \beta_3 trial\text{-}number$$

Chapter VI

General discussion

This thesis presents two experiments showing that DRN controls environment-driven changes in behaviour. We developed two novel, foraging-inspired behavioural paradigms to examine how human and non-human primates decide whether to pursue reward opportunities, and how these decisions are modulated by the richness of the environment. Careful experimental design and state-of-the-art analytic techniques enabled us to identify temporally persistent patterns of behaviour and points of behavioural change – in the monkey case, between general states of motivation for reward-seeking (Ch.III), and in the human case between option-specific policies of pursuit and rejection (Ch.V). DRN's relationship with behaviour was strikingly convergent across species. A combination of fMRI recording and TUS disruption experiments in monkeys showed that DRN was causally involved in motivation-state transitions, and that animals were unable to match their motivation state to the richness of the environment when DRN activity was perturbed (Ch.IV). In humans, fMRI recordings showed that DRN activity signalled changes in option-specific behavioural policies, but only insofar as changes were appropriate to the richness of the environment (Ch.V). Taken together, these results suggest that DRN has a fundamental and general role in reconciling behaviour with the reward statistics of the surrounding world.

The connection between DRN and behavioural change accords with previous experiments in myriad species, including non-human primates, rodents, and zebrafish (*204*, *208*, *222*, *233*, *238*, *302*, *343*). Together, these experiments have implicated DRN activity and/or DRN-derived serotonin in transitions between diverse forms of behaviour including patience-to-impulsivity in intertemporal choice, and exploration-to-exploitation during

reversal learning. It is striking, thus, that DRN's link with behavioural change has rarely been explicitly drawn, with most reports instead focussing on how DRN activity represents aspects of reward outcomes (*231*, *232*, *234*, *235*). Our results are broadly consistent with the view that DRN activity represents rewards, but suggest that it might not relate to specific reward events *per se*. Like previous monkey experiments, we show that DRN activity tracks the general richness of the environment over many individual reward outcomes (*228*, *344*) – a quantity that was fundamentally related to policy-switches and motivation-state transitions in behaviour. It is difficult to specify the mechanistic link between these phenomena based on fMRI data alone. One possibility pertains to the spiking properties of serotonergic DRN neurons, which are driven by both the intensity and temporal derivative of synaptic inputs (*345*). Although speculative, it is plausible that the patterns of activity we recorded arise from a mechanism where: (i) changes in afferent reward information over time causes activation of DRN serotonin neurons, which; (ii) causes behavioural change by serotonergic neuromodulation of downstream target regions.

DRN activity was strongest in monkeys when animals transitioned from high- to low-motivation states, which typically occurred in poor environments. Similarly, DRN activity in the human brain was more prominent when participants switched from rejecting to pursuing the 10-point reward opportunity, which was adaptive when the richness of the environment deteriorated. DRN activity was thus linked in both species to poor environments predominated by low-value opportunities, which evokes classic theories implicating DRN and DRN-derived serotonin in aversive processing (*98*, *158*, *211*). Although it is unlikely that DRN activity is specific to aversive events as envisaged by this view, our results suggest that DRN might be particularly important for adapting behaviour to adverse environments. This is not *a priori* inconsistent with evidence that rewards positively modulate DRN activity (*201*). If, for example, DRN facilitates behavioural-change by indicating when the reward-rate ebbs below

a critical level, then positive coding of rewards might be crucial to the timing and incidence of behaviour-change signals.

The discussion so far has frequently referred to DRN's distinctive serotonergic properties. It is therefore important to acknowledge that although DRN is the main source of forebrain serotonin, non-serotonergic neurons make up at least 50% of DRN's total volume, and they are spatially intermixed with serotonin neurons throughout the region (*145*, *147*, *149*). It is thus not possible to discriminate serotonergic from non-serotonergic DRN activity in fMRI recordings, and not possible to determine whether the patterns reported were serotonin specific. Future experiments could address this by recording and/or manipulating activity using genetically specific recording and perturbation techniques. Although these methods have been intractable in primates, they are widely used in lower mammals like rodents, which are capable of foraging-based behaviours studied here.

Very few studies have investigated DRN function in the primate brain due to its challenging anatomical location (although see (*219*, *229*, *303*, *346*)). Indeed, the present experiments are perhaps the first to compare DRN function across the monkey and human brain, and certainly the first to demonstrate that primate DRN can be non-invasively disrupted. Despite the feasibility advantages of small model organisms, it remains important to probe DRN function at the primate level. Many aspects of DRN function likely arise from its interactions with the cortex, which is characterised by a number of primate-specific phylogenetic innovations that have no rodent homologue (*45*). In the present experiments, for example, we observed task-related coactivations between DRN and AI, which is preferentially expanded in primates compared to rodents, and further expanded in humans compared to old world monkeys (*67*). Whether the same interactions occur in rodents is an empirical question, although it is perhaps doubtful given the degree of phylogenetic development that AI has clearly undergone.

The hypothesis of DRN function put forward here is striking for its ostensive similarity to widely held theories of SN/VTA function, which are deeply entwined with motivated behaviour. Why would the brain devote two systems to the same purpose? Firstly, it is important to be specific about the details of our hypothesis. SN and VTA neurons predominantly show transient phasic responses to individual events, like the initiation of *specific* reward-guided actions, or learning updates from *specific* reward outcomes (*160*, *189*). In contrast, we have argued that DRN reconciles *global* states of reward and behaviour, like ensuring that an animal's general preparedness to pursue rewards is appropriate for the richness of the environment. In addition, it is worth highlighting that although SN and VTA have compelling links with reward-guided behaviour, most of the corroborating experiments examined SN and/or VTA neural populations alone and are therefore silent on the comparative role of other areas. Most experiments on elements of the ANS share a similarly singular focus, which has produced a confusing literature in which multiple functions are attributed to single pathways and multiple pathways are linked to single functions. Some of this overlap might reflect biological reality – for example, the selection advantages of functional redundancy imply that a degree of functional similarity between neural circuits is adaptive, even though the brain is characterised by specialisation on the whole (*347*). In either the case, the present experiments addressed this by directly comparing the function of the ANS nuclei by recording from all regions simultaneously with fMRI. This demonstrated a clear and specific role for DRN that was not recapitulated in SN, VTA or any other region.

Despite converging links between DRN activity and behaviour across experiments, behaviour itself showed puzzling cross-species differences. Human participants in Ch.V approximated the predictions of optimal foraging theory and were more likely to pursue rewards in poor environments. This is unsurprising insofar as humans have a well-documented ability to identify and implement optimal strategies in foraging scenarios (*28*, *348*). Macaque

monkeys, however, displayed the opposite pattern: they were more likely to pursue rewards during periods of abundance, and their high-motivation states were synchronised to environments with many high-value opportunities. Intriguingly, rhesus monkeys in natural environments display the basic features of optimal foraging: for example, they actively seek unpreferred food items when their favoured foods are scarce (*349*). This suggests that they capable in principle of evaluating foreground rewards against their background reward-context, even if they did not do so in the present task.

What might explain this? One possibility concerns the biological origins of optimal foraging behaviour. Foraging theory emphasises that animals have homeostatic requirements which entail a baseline level of reward intake no matter how bad the options are – hence, the distinctive pattern whereby rewards of constant value are more favourably evaluated in poor environments. Importantly, it is unlikely that homeostatic concerns were a factor for animals in the present experiments: task sessions were short (approximately 40-minutes) and reward opportunities were frequent (approximately every 12-seconds), making it improbable that any individual reward was pursued out of the need to survive. Animals may therefore have preferentially engaged with rich environments because the demands imposed by the task made them unable – or more accurately, unwilling – to perform it *ad infinitum*, and they therefore concentrated their efforts on periods where they had the most to gain. It is unlikely that human participants in Ch.V were motivated by homeostatic concerns either. However, even though they did not need the rewards in the task *per se*, many participants were able to grasp the reward-maximising strategy in a way that animals were not. The cross-species difference in behaviour might therefore reflect the better ability of human participants to both: (i) understand the abstract goal of the behavioural task – in this case, reward-maximisation – and; (ii) extrapolate from their experiences with similarly structured tasks, where the reward consequences were salient enough to induce learning of an adaptive strategy.

Finally, it is worth briefly addressing how the present results relate to affective disorders. Affective disorders are marked by an inability to adapt behaviour to negative outcomes, and the relationship between DRN and poor environments reported here is thus broadly consistent with the hypothesis that they are caused by serotonergic dysfunction (*98*, *350*, *351*). It is important to emphasise, however, that the relationship between serotonin and affect is poorly understood, as are the cognitive mechanisms of serotonin-specific reuptake inhibitors (SSRI), which inspired the serotonin hypothesis in the first instance (*352*). The present experiments do little to resolve these controversies. However, the focus on high-level patterns like motivation-states and behavioural-policies opens dimensions of behaviour that have not, to our knowledge, been explored in psychopharmacology or computational psychiatry. These might be fruitful avenues for future experiments.

Appendices

Table 1: fMRI cluster locations from GLM5.9

Clusters for $\beta_3$decision(opportunity-onset) contrast in GLM 5.9

Anatomical location	x	y	z	*z*-max	*p*	Size (# voxels)
Left Thalamus	-15	-21	7	5.28	9.61e-32	7132
Postcentral Gyrus	-56	-18	44	6.18	9.54e-31	6811
Anterior Cingulate Gyrus	-1	5	38	4.95	1.6e-30	6739
Right Cerebellum	18	-50	-20	6.04	2.03e-29	6391
Occipital Pole	16	-95	4	5.15	1.46e-19	3570
Insular/Central Opperculum	-39	-4	16	5.65	6.35e-16	2675
Right Cerebellum	17	-58	-46	4.72	1.58e-14	2355
Left Cerebellum	-23	-59	-19	4.63	3.42e-11	1648
Precentral Gyrus	-55	6	36	5.12	5.71e-11	1604
Lingual Gyrus	-3	-71	4	4.79	5.96e-08	1068
Posterior Cingulate Gyrus/Precentral Gyrus	-8	-24	44	5.16	8.34e-07	855
Left Putamen	-32	-8	-2	4.44	7.45e-06	709
Occipital Pole/Superior Lateral Occipital Cortex	-22	-93	9	4.65	1.08e-05	685
Inferior Lateral Occipital Cortex	-41	-85	9	4.54	0.000365	471
Lateral Frontal Pole (Right)	30	59	0	4.14	0.000528	450
Postcentral Gyrus/Anterior Supramarginal Gyrus	40	-31	41	4.31	0.00144	395
Lateral Frontal Pole (Left)	-24	63	2	3.89	0.00155	391
Insular/Central Opperculum (Right)	44	2	11	3.91	0.00289	358
Anterior Insular/Frontal Opperculum (Left)	-32	19	10	4.29	0.00461	334
Cerebral White Matter (Right)	24	-23	-4	4.56	0.00789	307
Posterior Cingulate Gyrus/Precentral Gyrus (Right)	16	-31	39	4.01	0.0269	248
Central Opperculum/Insular	43	9	3	4.36	0.0334	238
Anterior Insular/ Frontal Opperculum/	33	21	9	4.69	0.0484	221

Table 2: fMRI cluster locations from GLM5.10

Clusters for $\beta_3$reward-magnitude contrast in GLM5.10

Anatomical location	x	y	z	*z*-max	*p*	Size (# voxels)
Right Cerebellum	20	-50	-20	4.35	9.48e-20	197
Occipital Pole	15	-92	3	4.6	4.9e-15	135
Postcentral Gyrus	-56	-19	45	4.28	6.16e-13	110
Insular/Central Opercular Cortex	-40	-3	15	4.41	1.42e-10	84
Postcentral/Supramarginal Gurys	-58	-23	23	3.98	1.19e-07	55
Lingual Gyrus	-4	-74	-2	4.15	1.31e-06	46
Anterior Cingulate Cortex/Cingulate Gyrus	- 6	6	43	4.05	5.13e-06	41
Right Nucleus Accumbens	6	11	-5	4.71	2.12e-05	36
Supramarginal Gyrus	-59	-34	31	4.01	2.83e-05	35

References

1. G. E. M. Anscombe, *Intention* (Harvard University Press, 2000).

2. P. Godfrey-Smith, Individuality, subjectivity, and minimal cognition. *Biol Philos* **31**, 775–796 (2016).

3. J. I. Gold, M. N. Shadlen, The Neural Basis of Decision Making. *Annual Review of Neuroscience* **30**, 535–574 (2007).

4. C. Lillesaar, The serotonergic system in fish. *Journal of Chemical Neuroanatomy* **41**, 294–308 (2011).

5. T. Fujita, N. Aoki, C. Mori, K. J. Homma, S. Yamaguchi, Molecular biology of serotonergic systems in avian brains. *Front Mol Neurosci* **16**, 1226645 (2023).

6. A. Parent, Comparative anatomy of the serotoninergic systems. *J Physiol (Paris)* **77**, 147–156 (1981).

7. F. Ayala-Guerrero, S. Huitrón-Reséndiz, R. Mancilla, Characterization of the raphe nuclei of the reptile Ctenosaura pectinata. *Physiol Behav* **50**, 717–722 (1991).

8. I. G. Woods, P. D. Kelly, F. Chu, P. Ngo-Hazelett, Y.-L. Yan, H. Huang, J. H. Postlethwait, W. S. Talbot, A Comparative Map of the Zebrafish Genome. *Genome Research* **10**, 1903 (2000).

9. P. Haggard, The Neurocognitive Bases of Human Volition. *Annual Review of Psychology* **70**, 9–28 (2019).

10. R. E. Passingham, S. L. Bengtsson, H. C. Lau, Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences* **14**, 16–21 (2010).

11. B. Libet, C. A. Gleason, E. W. Wright, D. K. Pearl, Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* **106 (Pt 3)**, 623–642 (1983).

12. C. Summerfield, K. Tsetsos, Building Bridges between Perceptual and Economic Decision-Making: Neural and Computational Mechanisms. *Front Neurosci* **6**, 70 (2012).

13. P. W. Glimcher, *Neuroeconomics: Decision Making and the Brain* (Academic Press, 2013).

14. P. A. Samuelson, A Note on the Pure Theory of Consumer's Behaviour. *Economica* **5**, 61–71 (1938).

15. J. von Neumann, O. Morgenstern, A. Rubinstein, *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)* (Princeton University Press, 1944; https://www.jstor.org/stable/j.ctt1r2gkx).

16. J. R. Krebs, J. T. Erichsen, M. I. Webber, E. L. Charnov, Optimal prey selection in the great tit (Parus major). *Animal Behaviour* **25**, 30–38 (1977).

17. D. W. Stephens, J. R. Krebs, *Foraging Theory* (Princeton University Press, 1986).

18. N. B. Davies, J. R. Krebs, *An Introduction to Behavioural Ecology* (John Wiley & Sons, 2009).

19. L. B. Crowder, Optimal foraging and feeding mode shifts in fishes. *Environ Biol Fish* **12**, 57–62 (1985).

20. D. Goulson, Foraging strategies of insects for gathering nectar and pollen, and implications for plant ecology and evolution. *Perspectives in Plant Ecology, Evolution and Systematics* **2**, 185–209 (1999).

21. K. Jensen, D. Mayntz, S. Toft, F. J. Clissold, J. Hunt, D. Raubenheimer, S. J. Simpson, Optimal foraging for specific nutrients in predatory beetles. *Proc Biol Sci* **279**, 2212–2218 (2012).

22. G. H. Pyke, C. K. Starr, "Optimal Foraging Theory" in *Encyclopedia of Social Insects*, C. K. Starr, Ed. (Springer International Publishing, Cham, 2021; https://doi.org/10.1007/978-3-030-28102-1_90), pp. 677–685.

23. C. R. Townsend, I. J. Winfield, "The Application of Optimal Foraging Theory to Feeding Behaviour in Fish" in *Fish Energetics: New Perspectives*, P. Tytler, P. Calow, Eds. (Springer Netherlands, Dordrecht, 1985; https://doi.org/10.1007/978-94-011-7918-8_3), pp. 67–98.

24. B. Y. Hayden, J. M. Pearson, M. L. Platt, Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience* **14**, 933 (2011).

25. B. Y. Hayden, M. E. Walton, Neuroscience of foraging. *Front. Neurosci.* **8** (2014).

26. N. Kolling, T. E. J. Behrens, R. B. Mars, M. F. S. Rushworth, Neural Mechanisms of Foraging. *Science* **336**, 95–98 (2012).

27. N. Kolling, M. K. Wittmann, T. E. J. Behrens, E. D. Boorman, R. B. Mars, M. F. S. Rushworth, Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience* **19**, 1280–1285 (2016).

28. S. M. Constantino, N. D. Daw, Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective, & Behavioral Neuroscience* **15**, 837–853 (2015).

29. T. C. Blanchard, B. Y. Hayden, Neurons in Dorsal Anterior Cingulate Cortex Signal Postdecisional Variables in a Foraging Task. *Journal of Neuroscience* **34**, 646–655 (2014).

30. T. C. Blanchard, B. Y. Hayden, Monkeys are more patient in a foraging task than in a standard intertemporal choice task. *PloS one* **10**, e0117057 (2015).

31. R. C. Bolles, *Theory of Motivation* (Harper & Row, 1975).

32. S. W. Flavell, N. Gogolla, M. Lovett-Barron, M. Zelikowsky, The emergence and influence of internal states. *Neuron* **110**, 2545–2570 (2022).

33. N. Tinbergen, *The Study of Instinct* (Pygmalion Press, an imprint of Plunkett Lake Press, 2020).

34. P. Tovote, J. P. Fadok, A. Lüthi, Neuronal circuits for fear and anxiety. *Nat Rev Neurosci* **16**, 317–331 (2015).

35. Z. C. Ashwood, N. A. Roy, I. R. Stone, A. E. Urai, A. K. Churchland, A. Pouget, J. W. Pillow, Mice alternate between discrete strategies during perceptual decision-making. *Nat Neurosci* **25**, 201–212 (2022).

36. W. Zucchini, *Hidden Markov Models for Time Series: An Introduction Using R [Electronic Resource]* (CRC Press/Taylor & Francis Group, Boca Raton, 2009)*Monographs on statistics and applied probability (Series)*.

37. T. I. B. Laboratory, V. Aguillon-Rodriguez, D. Angelaki, H. Bayer, N. Bonacchi, M. Carandini, F. Cazettes, G. Chapuis, A. K. Churchland, Y. Dan, E. Dewitt, M. Faulkner, H. Forrest, L. Haetzel, M. Häusser, S. B. Hofer, F. Hu, A. Khanal, C. Krasniak, I. Laranjeira, Z. F. Mainen, G. Meijer, N. J. Miska, T. D. Mrsic-Flogel, M. Murakami, J.-P. Noel, A. Pan-Vazquez, C. Rossant, J. Sanders, K. Socha, R. Terry, A. E. Urai, H. Vergara, M. Wells, C. J. Wilson, I. B. Witten, L. E. Wool, A. M. Zador, Standardized and reproducible measurement of decision-making in mice, *eLife* (2021). https://doi.org/10.7554/eLife.63711.

38. O. Odoemene, S. Pisupati, H. Nguyen, A. K. Churchland, Visual Evidence Accumulation Guides Decision-Making in Unrestrained Mice. *J. Neurosci.* **38**, 10143–10155 (2018).

39. A. E. Urai, A. Braun, T. H. Donner, Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nat Commun* **8**, 14637 (2017).

40. G. E. Alexander, M. R. DeLong, P. L. Strick, Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* **9**, 357–381 (1986).

41. S. N. Haber, Corticostriatal circuitry. *Dialogues Clin Neurosci* **18**, 7–21 (2016).

42. O. Hikosaka, H. F. Kim, M. Yasuda, S. Yamamoto, Basal Ganglia Circuits for Reward Value–Guided Behavior. *Annual Review of Neuroscience* **37**, 289–306 (2014).

43. A. Klaus, J. A. da Silva, R. M. Costa, What, If, and When to Move: Basal Ganglia Circuits and Self-Paced Action Initiation. *Annual Review of Neuroscience* **42**, null (2019).

44. J. K. Mai, M. Majtanik, G. Paxinos, *Atlas of the Human Brain* (Academic Press, Amsterdam, 4th edition., 2016).

45. R. E. Passingham, S. P. Wise, R. E. Passingham, S. P. Wise, *The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight* (Oxford University Press, Oxford, New York, 2012)*Oxford Psychology Series*.

46. B. A. Vogt, *Cingulate Neurobiology and Disease* (Oxford University Press, 2023).

47. B. A. Vogt, D. N. Pandya, D. L. Rosene, Cingulate cortex of the rhesus monkey: I. Cytoarchitecture and thalamic afferents. *Journal of Comparative Neurology* **262**, 256–270 (1987).

48. B. A. Vogt, E. A. Nimchinsky, L. J. Vogt, P. R. Hof, Human cingulate cortex: Surface features, flat maps, and cytoarchitecture. *Journal of Comparative Neurology* **359**, 490–506 (1995).

49. H. Barbas, D. N. Pandya, Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey, *Journal of Comparative Neurology* (1989). https://doi.org/10.1002/cne.902860306.

50. R. J. Morecraft, G. W. Van Hoesen, Convergence of Limbic Input to the Cingulate Motor Cortex in the Rhesus Monkey. *Brain Research Bulletin* **45**, 209–232 (1998).

51. G. W. Van Hoesen, R. J. Morecraft, B. A. Vogt, "Connections of the Monkey Cingulate Cortex" in *Neurobiology of Cingulate Cortex and Limbic Thalamus: A*

Comprehensive Handbook, B. A. Vogt, M. Gabriel, Eds. (Birkhäuser, Boston, MA, 1993; https://doi.org/10.1007/978-1-4899-6704-6_9), pp. 249–284.

52. B. A. Vogt, D. N. Pandya, Cingulate cortex of the rhesus monkey: II. Cortical afferents. *Journal of Comparative Neurology* **262**, 271–289 (1987).

53. T. Paus, Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nat Rev Neurosci* **2**, 417–424 (2001).

54. M. F. S. Rushworth, M. E. Walton, S. W. Kennerley, D. M. Bannerman, Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences* **8**, 410–417 (2004).

55. S. R. Heilbronner, B. Y. Hayden, Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annu. Rev. Neurosci.* **39**, 149–170 (2016).

56. M. F. S. Rushworth, M. P. Noonan, E. D. Boorman, M. E. Walton, T. E. Behrens, Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron* **70**, 1054–1069 (2011).

57. C. Amiez, J.-P. Joseph, E. Procyk, Reward encoding in the monkey anterior cingulate cortex. *Cereb Cortex* **16**, 1040–1055 (2006).

58. K. A. Hadland, M. f. s. Rushworth, D. Gaffan, R. E. Passingham, The Anterior Cingulate and Reward-Guided Selection of Actions. *Journal of Neurophysiology* **89**, 1161–1164 (2003).

59. S. W. Kennerley, M. E. Walton, T. E. J. Behrens, M. J. Buckley, M. F. S. Rushworth, Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience* **9**, 940–947 (2006).

60. P. H. Rudebeck, T. E. Behrens, S. W. Kennerley, M. G. Baxter, M. J. Buckley, M. E. Walton, M. F. S. Rushworth, Frontal Cortex Subregions Play Distinct Roles in Choices between Actions and Stimuli. *J. Neurosci.* **28**, 13775–13785 (2008).

61. A. U. Turken, D. Swick, Response selection in the human anterior cingulate cortex. *Nat Neurosci* **2**, 920–924 (1999).

62. M. E. Walton, D. M. Bannerman, M. F. S. Rushworth, The Role of Rat Medial Frontal Cortex in Effort-Based Decision Making. *The Journal of Neuroscience* **22**, 10996 (2002).

63. M. E. Walton, D. M. Bannerman, K. Alterescu, M. F. S. Rushworth, Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *J Neurosci* **23**, 6475–6479 (2003).

64. E. D. Boorman, M. F. Rushworth, T. E. Behrens, Ventromedial Prefrontal and Anterior Cingulate Cortex Adopt Choice and Default Reference Frames during Sequential Multi-Alternative Choice. *J. Neurosci.* **33**, 2242–2253 (2013).

65. E. F. Fouragnan, B. K. H. Chau, D. Folloni, N. Kolling, L. Verhagen, M. Klein-Flügge, L. Tankelevitch, G. K. Papageorgiou, J.-F. Aubry, J. Sallet, M. F. S. Rushworth, The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. *Nat Neurosci* **22**, 797–808 (2019).

66. B. Y. Hayden, J. M. Pearson, M. L. Platt, Neuronal basis of sequential foraging decisions in a patchy environment. *Nature neuroscience* **14**, 933–939 (2011).

67. H. C. Evrard, The Organization of the Primate Insular Cortex. *Front Neuroanat* **13**, 43 (2019).

68. H. C. Evrard, T. Forro, N. K. Logothetis, Von Economo Neurons in the Anterior Insula of the Macaque Monkey. *Neuron* **74**, 482–489 (2012).

69. G. Paxinos, *The Rhesus Monkey Brain in Stereotaxic Coordinates* (Academic, San Diego ;, 2000).

70. J. P. Aggleton, M. J. Burton, R. E. Passingham, Cortical and subcortical afferents to the amygdala of the rhesus monkey (*Macaca mulatta*). *Brain Research* **190**, 347–368 (1980).

71. X. An, R. Bandler, D. Öngür, J. l. Price, Prefrontal cortical projections to longitudinal columns in the midbrain periaqueductal gray in Macaque monkeys. *Journal of Comparative Neurology* **401**, 455–479 (1998).

72. S. T. Carmichael, J. L. Price, Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology* **363**, 642–664 (1995).

73. S. T. Carmichael, J. L. Price, Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology* **363**, 615–641 (1995).

74. M. Chikama, N. R. McFarland, D. G. Amaral, S. N. Haber, Insular Cortical Projections to Functional Regions of the Striatum Correlate with Cortical Cytoarchitectonic Organization in the Primate. *J. Neurosci.* **17**, 9686–9705 (1997).

75. A. T. Ferry, D. Öngür, X. An, J. L. Price, Prefrontal cortical projections to the striatum in macaque monkeys: Evidence for an organization related to prefrontal networks. *Journal of Comparative Neurology* **425**, 447–470 (2000).

76. E. J. Mufson, M.-M. Mesulam, Insula of the old world monkey. II: Afferent cortical input and comments on the claustrum. *Journal of Comparative Neurology* **212**, 23–37 (1982).

77. D. Öngür, X. An, J. l. Price, Prefrontal cortical projections to the hypothalamus in Macaque monkeys. *Journal of Comparative Neurology* **401**, 480–505 (1998).

78. C.-J. Shi, M. d. Cassell, Cortical, thalamic, and amygdaloid connections of the anterior and posterior insular cortices. *Journal of Comparative Neurology* **399**, 440–468 (1998).

79. M. J. C. Showers, E. W. Lauer, Somatovisceral motor patterns in the insula. *Journal of Comparative Neurology* **117**, 107–115 (1961).

80. S. M. Oppenheimer, A. Gelb, J. P. Girvin, V. C. Hachinski, Cardiovascular effects of human insular cortex stimulation. *Neurology* **42**, 1727–1727 (1992).

81. W. PENFIELD, M. E. FAULK JR., THE INSULA: FURTHER OBSERVATIONS ON ITS FUNCTION1. *Brain* **78**, 445–470 (1955).

82. (Bud) Craig, A. D, How do you feel — now? The anterior insula and human awareness. *Nat Rev Neurosci* **10**, 59–70 (2009).

83. A. D. Craig, How do you feel? Interoception: the sense of the physiological condition of the body. *Nat Rev Neurosci* **3**, 655–666 (2002).

84. N. Khalighinejad, N. Garrett, L. Priestley, P. Lockwood, M. F. S. Rushworth, A habenula-insular circuit encodes the willingness to act. *Nat Commun* **12**, 6329 (2021).

85. V. Menon, L. Q. Uddin, Saliency, switching, attention and control: a network model of insula function. *Brain structure & function* **214**, 655 (2010).

86. W. W. Seeley, The Salience Network: A Neural System for Perceiving and Responding to Homeostatic Demands. *J. Neurosci.* **39**, 9878–9882 (2019).

87. X. Wang, Q. Wu, L. Egan, X. Gu, P. Liu, H. Gu, Y. Yang, J. Luo, Y. Wu, Z. Gao, J. Fan, Anterior insular cortex plays a critical role in interoceptive attention. *eLife* **8** (2019).

88. O. Hikosaka, Y. Takikawa, R. Kawagoe, Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.* **80**, 953–978 (2000).

89. M. Stephenson-Jones, A. A. Kardamakis, B. Robertson, S. Grillner, Independent circuits in the basal ganglia for the evaluation and selection of actions. *PNAS* **110**, E3670–E3679 (2013).

90. M. Stephenson-Jones, K. Yu, S. Ahrens, J. M. Tucciarone, A. N. van Huijstee, L. A. Mejia, M. A. Penzo, L.-H. Tai, L. Wilbrecht, B. Li, A basal ganglia circuit for evaluating action outcomes. *Nature* **539**, 289–293 (2016).

91. E. R. Kandel, J. H. Schwartz, T. M. Jessell, *Principles of Neural Science* (Elsevier, 1991).

92. M. Burrows, "Neurotransmitters, neuromodulators and neurohormones" in *The Neurobiology of an Insect Brain*, M. Burrows, Ed. (Oxford University Press, 1996; https://doi.org/10.1093/acprof:oso/9780198523444.003.0005), p. 0.

93. A. Thiele, M. A. Bellgrove, Neuromodulation of Attention. *Neuron* **97**, 769–785 (2018).

94. Ö. D. Özçete, A. Banerjee, P. S. Kaeser, Mechanisms of neuromodulatory volume transmission. *Mol Psychiatry*, 1–14 (2024).

95. J. R. Cooper, F. E. Bloom, R. H. Roth, *The Biochemical Basis of Neuropharmacology* (Oxford University Press, 2003).

96. P. A. Getting, Emerging Principles Governing the Operation of Neural Networks. *Annual Review of Neuroscience* **12**, 185–204 (1989).

97. R. M. Harris-Warrick, E. Marder, Modulation of Neural Networks for Behavior. *Annual Review of Neuroscience* **14**, 39–57 (1991).

98. P. Dayan, Q. J. M. Huys, Serotonin in affective control. *Annu Rev Neurosci* **32**, 95–126 (2009).

99. B. L. Jacobs, E. C. Azmitia, Structure and function of the brain serotonin system. *Physiological Reviews* **72**, 165–229 (1992).

100. G. R. Poe, S. Foote, O. Eschenko, J. P. Johansen, S. Bouret, G. Aston-Jones, C. W. Harley, D. Manahan-Vaughan, D. Weinshenker, R. Valentino, C. Berridge, D. J. Chandler, B. Waterhouse, S. J. Sara, Locus coeruleus: a new look at the blue spot. *Nat Rev Neurosci* **21**, 644–659 (2020).

101. P. R. Mouton, B. Pakkenberg, H. J. G. Gundersen, D. L. Price, Absolute number and size of pigmented locus coeruleus neurons in young and aged individuals. *Journal of Chemical Neuroanatomy* **7**, 185–190 (1994).

102. E. J. Nestler, P. J. Kenny, S. J. Russo, A. Schaefer, "Widely Projecting Systems: Monoamines, Acetylcholine, and Orexin" in *Nestler, Hyman & Malenka's Molecular*

Neuropharmacology: A Foundation for Clinical Neuroscience (McGraw-Hill, New York, NY, ed. 4, 2020; neurology.mhmedical.com/content.aspx?aid=1174973672).

103. K. L. Agster, C. A. Mejias-Aponte, B. D. Clark, B. D. Waterhouse, Evidence for a regional specificity in the density and distribution of noradrenergic varicosities in rat cortex. *Journal of Comparative Neurology* **521**, 2195–2207 (2013).

104. V. Breton-Provencher, M. Sur, Active control of arousal by a locus coeruleus GABAergic circuit. *Nat Neurosci* **22**, 218–228 (2019).

105. D. A. Lewis, J. H. Morrison, Noradrenergic innervation of monkey prefrontal cortex: A dopamine-β-hydroxylase immunohistochemical study. *Journal of Comparative Neurology* **282**, 317–330 (1989).

106. P.-H. Luppi, G. Aston-Jones, H. Akaoka, G. Chouvet, M. Jouvet, Afferent projections to the rat locus coeruleus demonstrated by retrograde and anterograde tracing with cholera-toxin B subunit and *Phaseolus vulgaris* leucoagglutinin. *Neuroscience* **65**, 119–160 (1995).

107. J. H. Morrison, S. L. Foote, Noradrenergic and serotoninergic innervation of cortical, thalamic, and tectal visual structures in old and new world monkeys. *Journal of Comparative Neurology* **243**, 117–138 (1986).

108. L. A. Schwarz, K. Miyamichi, X. J. Gao, K. T. Beier, B. Weissbourd, K. E. DeLoach, J. Ren, S. Ibanes, R. C. Malenka, E. J. Kremer, L. Luo, Viral-genetic tracing of the input–output organization of a central noradrenaline circuit. *Nature* **524**, 88–92 (2015).

109. M. R. Ananth, P. Rajebhosale, R. Kim, D. A. Talmage, L. W. Role, Basal forebrain cholinergic signalling: development, connectivity and roles in cognition. *Nat Rev Neurosci* **24**, 233–251 (2023).

110. J. Mena-Segovia, J. P. Bolam, Rethinking the Pedunculopontine Nucleus: From Cellular Organization to Function. *Neuron* **94**, 7–18 (2017).

111. H.-L. Wang, M. Morales, Pedunculopontine and laterodorsal tegmental nuclei contain distinct populations of cholinergic, glutamatergic and GABAergic neurons in the rat. *The European journal of neuroscience* **29** (2009).

112. N. J. Woolf, L. L. Butcher, Cholinergic systems in the rat brain: III. Projections from the pontomesencephalic tegmentum to the thalamus, tectum, basal ganglia, and basal forebrain. *Brain Research Bulletin* **16**, 603–637 (1986).

113. N. J. Woolf, L. L. Butcher, Cholinergic systems in the rat brain: IV. descending projections of the pontomesencephalic tegmentum. *Brain Research Bulletin* **23**, 519–540 (1989).

114. T. Aitta-aho, Y. A. Hay, B. U. Phillips, L. M. Saksida, T. J. Bussey, O. Paulsen, J. Apergis-Schoute, Basal Forebrain and Brainstem Cholinergic Neurons Differentially Impact Amygdala Circuits and Learning-Related Behavior. *Current Biology* **28**, 2557-2569.e4 (2018).

115. B. Bloem, L. Schoppink, D. C. Rotaru, A. Faiz, P. Hendriks, H. D. Mansvelder, W. D. J. van de Berg, F. G. Wouterlood, Topographic Mapping between Basal Forebrain Cholinergic Neurons and the Medial Prefrontal Cortex in Mice. *J. Neurosci.* **34**, 16234–16246 (2014).

116. R. B. Crouse, K. Kim, H. M. Batchelor, E. M. Girardi, R. Kamaletdinova, J. Chan, P. Rajebhosale, S. T. Pittenger, L. W. Role, D. A. Talmage, M. Jing, Y. Li, X.-B. Gao, Y. S. Mineur, M. R. Picciotto, Acetylcholine is released in the basolateral amygdala in response to predictors of reward and enhances the learning of cue-reward contingency. *eLife* **9**, e57335 (2020).

117. J. P. Do, M. Xu, S.-H. Lee, W.-C. Chang, S. Zhang, S. Chung, T. J. Yung, J. L. Fan, K. Miyamichi, L. Luo, Y. Dan, Cell type-specific long-range connections of basal forebrain circuit. *eLife* **5**, e13214 (2016).

118. M. R. Gielow, L. Zaborszky, The Input-Output Relationship of the Cholinergic Basal Forebrain. *Cell Rep* **18**, 1817–1830 (2017).

119. I. Gritti, P. Henny, F. Galloni, L. Mainville, M. Mariotti, B. E. Jones, Stereological estimates of the basal forebrain cell population in the rat, including neurons containing choline acetyltransferase, glutamic acid decarboxylase or phosphate-activated glutaminase and colocalizing vesicular glutamate transporters. *Neuroscience* **143**, 1051–1064 (2006).

120. H. J. Gritton, W. M. Howe, C. S. Mallory, V. L. Hetrick, J. D. Berke, M. Sarter, Cortical cholinergic signaling controls the detection of cues. *Proceedings of the National Academy of Sciences* **113**, E1089–E1097 (2016).

121. F. Huppé-Gourgues, K. Jegouic, E. Vaucher, Topographic Organization of Cholinergic Innervation From the Basal Forebrain to the Visual Cortex in the Rat. *Front. Neural Circuits* **12** (2018).

122. H. Kondo, L. Zaborszky, Topographic organization of the basal forebrain projections to the perirhinal, postrhinal, and entorhinal cortex in rats. *Journal of Comparative Neurology* **524**, 2503–2515 (2016).

123. T. Laszlovszky, D. Schlingloff, P. Hegedüs, T. F. Freund, A. Gulyás, A. Kepecs, B. Hangya, Distinct synchronization, cortical coupling and behavioral function of two basal forebrain cholinergic neuron types. *Nat Neurosci* **23**, 992–1003 (2020).

124. X. Li, H. Yu, B. Zhang, L. Li, W. Chen, Q. Yu, X. Huang, X. Ke, Y. Wang, W. Jing, H. Du, H. Li, T. Zhang, L. Liu, L.-Q. Zhu, Y. Lu, Molecularly defined and functionally distinct cholinergic subnetworks. *Neuron* **110**, 3774-3788.e7 (2022).

125. J. F. Muller, F. Mascagni, V. Zaric, A. J. McDonald, Muscarinic cholinergic receptor M1 in the rat basolateral amygdala: Ultrastructural localization and synaptic relationships to cholinergic axons. *Journal of Comparative Neurology* **521**, 1743–1759 (2013).

126. C. B. Saper, Organization of cerebral cortical afferent systems in the rat. II. Magnocellular basal nucleus. *Journal of Comparative Neurology* **222**, 313–342 (1984).

127. L. Slomianka, F. A. Geneser, Distribution of acetylcholinesterase in the hippocampal region of the mouse: I. Entorhinal area, parasubiculum, retrosplenial area, and Presubiculum. *Journal of Comparative Neurology* **303**, 339–354 (1991).

128. L. Slomianka, F. A. Geneser, Distribution of acetylcholinesterase in the hippocampal region of the mouse: II. Subiculum and hippocampus. *Journal of Comparative Neurology* **312**, 525–536 (1991).

129. L. Teles-Grilo Ruivo, J. Mellor, Cholinergic modulation of hippocampal network function. *Front. Synaptic Neurosci.* **5** (2013).

130. P. Turrini, M. A. Casu, T. P. Wong, Y. D. Koninck, A. Ribeiro-da-Silva, A. C. Cuello, Cholinergic nerve terminals establish classical synapses in the rat cerebral cortex: synaptic pattern and age-related atrophy. *Neuroscience* **105**, 277–285 (2001).

131. D. Umbriaco, K. C. Watkins, L. Descarries, C. Cozzari, B. K. Hartman, Ultrastructural and morphometric features of the acetylcholine innervation in adult rat parietal cortex: An electron microscopic study in serial sections. *Journal of Comparative Neurology* **348**, 351–373 (1994).

132. L. Zaborszky, A. Csordas, K. Mosca, J. Kim, M. R. Gielow, C. Vadasz, Z. Nadasdy, Neurons in the Basal Forebrain Project to the Cortex in a Complex Topographic Organization that Reflects Corticocortical Connectivity Patterns: An Experimental Study Based on Retrograde Tracing and 3D Reconstruction. *Cereb Cortex* **25**, 118–137 (2015).

133. S. C. Gantz, C. P. Ford, H. Morikawa, J. T. Williams, The Evolving Understanding of Dopamine Neurons in the Substantia Nigra and Ventral Tegmental Area. *Annual Review of Physiology* **80**, 219–241 (2018).

134. S. N. Haber, The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy* **26**, 317–330 (2003).

135. B. Berger, P. Gaspar, C. Verney, Dopaminergic innervation of the cerebral cortex: unexpected differences between rodents and primates. *Trends in Neurosciences* **14**, 21–27 (1991).

136. L. J. Porrino, P. S. Goldman-Rakic, Brainstem innervation of prefrontal and anterior cingulate cortex in the rhesus monkey revealed by retrograde transport of HRP. *J Comp Neurol* **205**, 63–76 (1982).

137. S. M. Williams, P. S. Goldman-Rakic, Widespread origin of the primate mesofrontal dopamine system. *Cereb Cortex* **8**, 321–345 (1998).

138. K. T. Beier, E. E. Steinberg, K. E. DeLoach, S. Xie, K. Miyamichi, L. Schwarz, X. J. Gao, E. J. Kremer, R. C. Malenka, L. Luo, Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping. *Cell* **162**, 622–634 (2015).

139. K. T. Beier, X. J. Gao, S. Xie, K. E. DeLoach, R. C. Malenka, L. Luo, Topological Organization of Ventral Tegmental Area Connectivity Revealed by Viral-Genetic Dissection of Input-Output Relations. *Cell Reports* **26**, 159-167.e6 (2019).

140. S. Ikemoto, Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens–olfactory tubercle complex. *Brain Research Reviews* **56**, 27–78 (2007).

141. V. M. Pickel, A. C. Towle, T. H. Joh, J. Chan, Gamma-aminobutyric acid in the medial rat nucleus accumbens: Ultrastructural localization in neurons receiving monosynaptic input from catecholaminergic afferents. *Journal of Comparative Neurology* **272**, 1–14 (1988).

142. T. N. Lerner, C. Shilyansky, T. J. Davidson, K. E. Evans, K. T. Beier, K. A. Zalocusky, A. K. Crow, R. C. Malenka, L. Luo, R. Tomer, K. Deisseroth, Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* **162**, 635–647 (2015).

143. Y. Zhang, K. M.-H. Larcher, B. Misic, A. Dagher, Anatomical and functional organization of the human substantia nigra and its connections. *eLife* **6**, e26653 (2017).

144. W. Tang, O. Kochubey, M. Kintscher, R. Schneggenburger, A VTA to Basal Amygdala Dopamine Projection Contributes to Signal Salient Somatosensory Events during Fear Learning. *J. Neurosci.* **40**, 3969–3980 (2020).

145. R. P. Vertes, S. B. Linley, "Efferent and afferent connections of the dorsal and median raphe nuclei in the rat" in *Serotonin and Sleep: Molecular, Functional and Clinical Aspects*, J. M. Monti, S. R. Pandi-Perumal, B. L. Jacobs, D. J. Nutt, Eds. (Birkhäuser Basel, Basel, 2008; http://link.springer.com/10.1007/978-3-7643-8561-3_3), pp. 69–102.

146. E. C. Azmitia, M. Segal, An autoradiographic analysis of the differential ascending projections of the dorsal and median raphe nuclei in the rat. *Journal of Comparative Neurology* **179**, 641–667 (1978).

147. R. P. Vertes, A. M. Crane, Distribution, quantification, and morphological characteristics of serotonin-immunoreactive cells of the supralemniscal nucleus (B9) and pontomesencephalic reticular formation in the rat. *J Comp Neurol* **378**, 411–424 (1997).

148. K. W. Huang, N. E. Ochandarena, A. C. Philson, M. Hyun, J. E. Birnbaum, M. Cicconet, B. L. Sabatini, Molecular and anatomical organization of the dorsal raphe nucleus. *eLife* **8**, e46464 (2019).

149. W. Fu, E. Le Maître, V. Fabre, J.-F. Bernard, Z.-Q. David Xu, T. Hökfelt, Chemical neuroanatomy of the dorsal raphe nucleus and adjacent structures of the mouse brain. *Journal of Comparative Neurology* **518**, 3464–3494 (2010).

150. R. P. Vertes, A PHA-L analysis of ascending projections of the dorsal raphe nucleus in the rat. *Journal of Comparative Neurology* **313**, 643–668 (1991).

151. R. P. Vertes, Brainstem afferents to the basal forebrain in the rat. *Neuroscience* **24**, 907–935 (1988).

152. L. P. Morin, E. L. Meyer-Bernstein, The ascending serotonergic system in the hamster: comparison with projections of the dorsal and median raphe nuclei. *Neuroscience* **91**, 81–105 (1999).

153. A. Gasbarri, A. Sulli, C. Pacitti, J. L. McGaugh, Serotonergic input to cholinergic neurons in the substantia innominata and nucleus basalis magnocellularis in the rat. *Neuroscience* **91**, 1129–1142 (1999).

154. V. Di Matteo, M. Cacchio, C. Di Giulio, E. Esposito, Role of serotonin2C receptors in the control of brain dopaminergic function. *Pharmacology Biochemistry and Behavior* **71**, 727–734 (2002).

155. R. P. Vertes, G. F. Martin, Autoradiographic analysis of ascending projections from the pontine and mesencephalic reticular formation and the median raphe nucleus in the rat. *J Comp Neurol* **275**, 511–541 (1988).

156. R. P. Vertes, W. J. Fortin, A. M. Crane, Projections of the median raphe nucleus in the rat. *J Comp Neurol* **407**, 555–582 (1999).

157. A. J. Yu, P. Dayan, Uncertainty, Neuromodulation, and Attention. *Neuron* **46**, 681–692 (2005).

158. N. D. Daw, S. Kakade, P. Dayan, Opponent interactions between serotonin and dopamine. *Neural Netw* **15**, 603–616 (2002).

159. R. Cools, K. Nakamura, N. D. Daw, Serotonin and Dopamine: Unifying Affective, Activational and Decision Functions. *Neuropsychopharmacology* **36**, 98–113 (2011).

160. W. Schultz, P. Dayan, P. R. Montague, A Neural Substrate of Prediction and Reward. *Science* **275**, 1593–1599 (1997).

161. R. S. Sutton, *Reinforcement Learning: An Introduction* (The MIT Press, Cambridge, Massachusetts, Second edition., 2018)*Adaptive computation and machine learning*.

162. D. Marr, *Vision [Electronic Resource]: A Computational Investigation into the Human Representation and Processing of Visual Information* (MIT Press, Cambridge, MA, 2010; https://ezproxy-prd.bodleian.ox.ac.uk/login?url=http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=3339148)*Ebook central*.

163. J. Y. Cohen, S. Haesler, L. Vong, B. B. Lowell, N. Uchida, Neuron-type specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).

164. N. D. Daw, S. J. Gershman, B. Seymour, P. Dayan, R. J. Dolan, Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron* **69**, 1204–1215 (2011).

165. N. Eshel, J. Tian, M. Bukwich, N. Uchida, Dopamine neurons share common response function for reward prediction error. *Nat Neurosci* **19**, 479–486 (2016).

166. A. A. Hamid, J. R. Pettibone, O. S. Mabrouk, V. L. Hetrick, R. Schmidt, C. M. Vander Weele, R. T. Kennedy, B. J. Aragona, J. D. Berke, Mesolimbic dopamine signals the value of work. *Nature Neuroscience* **19**, 117–126 (2016).

167. M. R. Roesch, D. J. Calu, G. Schoenbaum, Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci* **10**, 1615–1624 (2007).

168. E. E. Steinberg, R. Keiflin, J. R. Boivin, I. B. Witten, K. Deisseroth, P. H. Janak, A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* **16**, 966–973 (2013).

169. K. A. Zaghloul, J. A. Blanco, C. T. Weidemann, K. McGill, J. L. Jaggi, G. H. Baltuch, M. J. Kahana, Human substantia nigra neurons encode unexpected financial rewards. *Science* **323**, 1496–1499 (2009).

170. M. W. Howe, P. L. Tierney, S. G. Sandberg, P. E. M. Phillips, A. M. Graybiel, Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* **500**, 575–579 (2013).

171. T. A. Krausz, A. E. Comrie, A. E. Kahn, L. M. Frank, N. D. Daw, J. D. Berke, Dual credit assignment processes underlie dopamine signals in a complex spatial environment. *Neuron* **111**, 3465-3478.e7 (2023).

172. Y. Niv, Dopamine ramps up. *Nature* **500**, 533–535 (2013).

173. S. J. Gershman, Dopamine Ramps Are a Consequence of Reward Prediction Errors. *Neural Computation* **26**, 467–471 (2014).

174. A. A. Hamid, M. J. Frank, C. I. Moore, Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* **184**, 2733-2749.e16 (2021).

175. H. R. Kim, A. N. Malik, J. G. Mikhael, P. Bech, I. Tsutsui-Kimura, F. Sun, Y. Zhang, Y. Li, M. Watabe-Uchida, S. J. Gershman, N. Uchida, A Unified Framework for Dopamine Signals across Timescales. *Cell* **183**, 1600-1616.e25 (2020).

176. U. Ungerstedt, Adipsia and Aphagia after 6-Hydroxydopamine Induced Degeneration of the Nigro-striatal Dopamine System. *Acta Physiologica Scandinavica* **82**, 95–122 (1971).

177. S. Hostiuc, E. Drima, O. Buda, Shake the Disease. Georges Marinesco, Paul Blocq and the Pathogenesis of Parkinsonism, 1893. *Frontiers in Neuroanatomy* **10** (2016).

178. W. Birkmayer, O. Hornykiewicz, The effect of l-3,4-dihydroxyphenylalanine (=DOPA) on akinesia in parkinsonism1. *Parkinsonism & Related Disorders* **4**, 59–60 (1998).

179. A. Carlsson, The occurrence, distribution and physiological role of catecholamines in the nervous system. *Pharmacol Rev* **11**, 490–493 (1959).

180. K. G. Lloyd, L. Davidson, O. Hornykiewicz, The neurochemistry of Parkinson's disease: effect of L-dopa therapy. *J Pharmacol Exp Ther* **195**, 453–464 (1975).

181. I. Sano, T. Gamo, Y. Kakimoto, K. Taniguchi, M. Takesada, K. Nishinuma, Distribution of catechol compounds in human brain. *Biochim Biophys Acta* **32**, 586–587 (1959).

182. J. A. da Silva, F. Tecuapetla, V. Paixão, R. M. Costa, Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* **554**, 244–248 (2018).

183. P. D. Dodson, J. K. Dreyer, K. A. Jennings, E. C. J. Syed, R. Wade-Martins, S. J. Cragg, J. P. Bolam, P. J. Magill, Representation of spontaneous movement by

dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *PNAS* **113**, E2180–E2188 (2016).

184. R. N. Hughes, G. D. R. Watson, E. A. Petter, N. Kim, K. I. Bakhurin, H. H. Yin, Precise Coordination of Three-Dimensional Rotational Kinematics by Ventral Tegmental Area GABAergic Neurons. *Current Biology* **29**, 3244-3255.e4 (2019).

185. B. J. Hunnicutt, B. C. Jongbloets, W. T. Birdsong, K. J. Gertz, H. Zhong, T. Mao, A comprehensive excitatory input map of the striatum reveals novel functional organization. *eLife* **5**, e19103 (2016).

186. M. Howe, D. Dombeck, Rapid signaling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**, 505–510 (2016).

187. N. F. Parker, C. M. Cameron, J. P. Taliaferro, J. Lee, J. Y. Choi, T. J. Davidson, N. D. Daw, I. B. Witten, Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat Neurosci* **19**, 845–854 (2016).

188. J.-F. Poulin, G. Caronia, C. Hofer, Q. Cui, B. Helm, C. Ramakrishnan, C. S. Chan, D. A. Dombeck, K. Deisseroth, R. Awatramani, Mapping projections of molecularly defined dopamine neuron subtypes using intersectional genetic approaches. *Nat Neurosci* **21**, 1260–1271 (2018).

189. J. D. Berke, What does dopamine mean? *Nature Neuroscience* **21**, 787 (2018).

190. J. D. Salamone, M. Correa, The Mysterious Motivational Functions of Mesolimbic Dopamine. *Neuron* **76**, 470–485 (2012).

191. W. Schultz, Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *Journal of Neurophysiology* **56**, 1439–1461 (1986).

192. W. Schultz, R. Romo, Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioral reactions. *Journal of Neurophysiology* **63**, 607–624 (1990).

193. C. M. Portas, B. Bjorvatn, R. Ursin, Serotonin and the sleep/wake cycle: special emphasis on microdialysis studies. *Progress in Neurobiology* **60**, 13–35 (2000).

194. D. R. Loyd, M. A. Henry, K. M. Hargreaves, Serotonergic Neuromodulation of Peripheral Nociceptors. *Semin Cell Dev Biol* **24**, 51–57 (2013).

195. G. Curzon, Serotonin and appetite. *Ann N Y Acad Sci* **600**, 521–30; discussion 530-1 (1990).

196. J. Neumann, B. Hofmann, S. Dhein, U. Gergs, Cardiac Roles of Serotonin (5-HT) and 5-HT-Receptors in Health and Disease. *Int J Mol Sci* **24**, 4765 (2023).

197. G. Hilaire, N. Voituron, C. Menuet, R. M. Ichiyama, H. H. Subramanian, M. Dutschmann, THE ROLE OF SEROTONIN IN RESPIRATORY FUNCTION AND DYSFUNCTION. *Respir Physiol Neurobiol* **174**, 76–88 (2010).

198. R. Li, G. Tang, J. Yang, S. Gao, Y. Wang, X. Wu, Y. Bai, J. Liu, The avBNSTGABA–VTA and avBNSTGABA–DRN pathways are respectively involved in the regulation of anxiety-like behaviors in parkinsonian rats. *Neurochemistry International* **175**, 105720 (2024).

199. C. A. Marcinkiewcz, C. M. Mazzone, G. D'Agostino, L. R. Halladay, J. A. Hardaway, J. F. DiBerto, M. Navarro, N. Burnham, C. Cristiano, C. E. Dorrier, G. J. Tipton, C. Ramakrishnan, T. Kozicz, K. Deisseroth, T. E. Thiele, Z. A. McElligott, A. Holmes, L.

K. Heisler, T. L. Kash, Serotonin engages an anxiety and fear-promoting circuit in the extended amygdala. *Nature* **537**, 97–101 (2016).

200. R. Cools, A. C. Roberts, T. W. Robbins, Serotoninergic regulation of emotional and behavioural control processes. *Trends in Cognitive Sciences* **12**, 31–40 (2008).

201. Y. Li, W. Zhong, D. Wang, Q. Feng, Z. Liu, J. Zhou, C. Jia, F. Hu, J. Zeng, Q. Guo, L. Fu, M. Luo, Serotonin neurons in the dorsal raphe nucleus encode reward signals. *Nat Commun* **7**, 10503 (2016).

202. K. Nakamura, M. Matsumoto, O. Hikosaka, Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus. *J Neurosci* **28**, 5331–5343 (2008).

203. S. F. Maier, L. R. Watkins, Stressor controllability and learned helplessness: The roles of the dorsal raphe nucleus, serotonin, and corticotropin-releasing factor. *Neuroscience & Biobehavioral Reviews* **29**, 829–841 (2005).

204. K. W. Miyazaki, K. Miyazaki, K. F. Tanaka, A. Yamanaka, A. Takahashi, S. Tabuchi, K. Doya, Optogenetic Activation of Dorsal Raphe Serotonin Neurons Enhances Patience for Future Rewards. *Current Biology* **24**, 2033–2040 (2014).

205. Y. Worbe, G. Savulich, V. Voon, E. Fernandez-Egea, T. W. Robbins, Serotonin depletion induces “waiting impulsivity” on the human four-choice serial reaction time task: cross-species translational significance. *Neuropsychopharmacology* **39**, 1519–1526 (2014).

206. K. Mitsui, A. Takahashi, Aggression modulator: Understanding the multifaceted role of the dorsal raphe nucleus. *BioEssays* **46**, 2300213 (2024).

207. A. Takahashi, R. X. Lee, T. Iwasato, S. Itohara, H. Arima, B. Bettler, K. A. Miczek, T. Koide, Glutamate Input in the Dorsal Raphe Nucleus As a Determinant of Escalated Aggression in Male Mice. *J. Neurosci.* **35**, 6452–6463 (2015).

208. H. F. Clarke, J. W. Dalley, H. S. Crofts, T. W. Robbins, A. C. Roberts, Cognitive Inflexibility After Prefrontal Serotonin Depletion. *Science* **304**, 878–880 (2004).

209. E. Lottem, D. Banerjee, P. Vertechi, D. Sarra, M. oude Lohuis, Z. F. Mainen, Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nat Commun* **9**, 1000 (2018).

210. B. L. Jacobs, C. A. Fornal, Activity of serotonergic neurons in behaving animals. *Neuropsychopharmacology* **21**, 9S-15S (1999).

211. P. Soubrié, Reconciling the role of central serotonin neurons in human and animal behavior. *Behavioral and Brain Sciences* **9**, 319–335 (1986).

212. C. J. Carter, C. J. Pycock, Differential effects of central serotonin manipulation on hyperactive and stereotyped behaviour. *Life Sciences* **23**, 953–960 (1978).

213. E. Courtiol, E. C. Menezes, C. M. Teixeira, Serotonergic regulation of the dopaminergic system: Implications for reward-related functions. *Neuroscience & Biobehavioral Reviews* **128**, 282–293 (2021).

214. P. J. Fletcher, K. M. Korth, Activation of 5-HT1B receptors in the nucleus accumbens reduces amphetamine-induced enhancement of responding for conditioned reward. *Psychopharmacology (Berl)* **142**, 165–174 (1999).

215. J. F. W. Deakin, F. G. Graeff, 5-HT and mechanisms of defence. *J Psychopharmacol* **5**, 305–315 (1991).

216. F. G. Graeff, F. S. Guimarães, T. G. C. S. De Andrade, J. F. W. Deakin, Role of 5-HT in stress, anxiety, and depression. *Pharmacology Biochemistry and Behavior* **54**, 129–141 (1996).

217. N. C. Tye, B. J. Everitt, S. D. Iversen, 5-Hydroxytryptamine and punishment. *Nature* **268**, 741–743 (1977).

218. S. Auerbach, C. Fornal, B. L. Jacobs, Response of serotonin-containing neurons in nucleus raphe magnus to morphine, noxious stimuli, and periaqueductal gray stimulation in freely moving cats. *Experimental Neurology* **88**, 609–628 (1985).

219. E. S. Bromberg-Martin, O. Hikosaka, K. Nakamura, Coding of task reward value in the dorsal raphe nucleus. *J Neurosci* **30**, 6262–6272 (2010).

220. J. Y. Cohen, M. W. Amoroso, N. Uchida, Serotonergic neurons signal reward and punishment on multiple timescales. *eLife* **4**, e06346 (2015).

221. K. Hayashi, K. Nakao, K. Nakamura, Appetitive and Aversive Information Coding in the Primate Dorsal Raphé Nucleus. *J. Neurosci.* **35**, 6195–6208 (2015).

222. S. Matias, E. Lottem, G. P. Dugué, Z. F. Mainen, Activity patterns of serotonin neurons underlying cognitive flexibility. *eLife* **6**, e20552 (2017).

223. S. P. Ranade, Z. F. Mainen, Transient Firing of Dorsal Raphe Neurons Encodes Diverse and Specific Sensory, Motor, and Reward Events. *Journal of Neurophysiology* **102**, 3026–3037 (2009).

224. J. Ren, D. Friedmann, J. Xiong, C. D. Liu, B. R. Ferguson, T. Weerakkody, K. E. DeLoach, C. Ran, A. Pun, Y. Sun, B. Weissbourd, R. L. Neve, J. Huguenard, M. A.

Horowitz, L. Luo, Anatomically Defined and Functionally Distinct Dorsal Raphe Serotonin Sub-systems. *Cell* **175**, 472-487.e20 (2018).

225. H.-L. Wang, S. Zhang, J. Qi, H. Wang, R. Cachope, C. A. Mejias-Aponte, J. A. Gomez, G. E. Mateo-Semidey, G. M. J. Beaudoin, C. A. Paladini, J. F. Cheer, M. Morales, Dorsal Raphe Dual Serotonin-Glutamate Neurons Drive Reward by Establishing Excitatory Synapses on VTA Mesoaccumbens Dopamine Neurons. *Cell Rep* **26**, 1128-1142.e7 (2019).

226. M. Luo, Y. Li, W. Zhong, Do dorsal raphe 5-HT neurons encode "beneficialness"? *Neurobiology of Learning and Memory* **135**, 40–49 (2016).

227. M. K. Wittmann, E. Fouragnan, D. Folloni, M. C. Klein-Flügge, B. K. H. Chau, M. Khamassi, M. F. S. Rushworth, Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. *Nat Commun* **11**, 3771 (2020).

228. N. Khalighinejad, S. Manohar, M. Husain, M. F. S. Rushworth, Complementary roles of serotonergic and cholinergic systems in decisions about when to act. *Curr Biol* **32**, 1150-1162.e7 (2022).

229. K. Inaba, T. Mizuhiki, T. Setogawa, K. Toda, B. J. Richmond, M. Shidara, Neurons in Monkey Dorsal Raphe Nucleus Code Beginning and Progress of Step-by-Step Schedule, Reward Expectation, and Amount of Reward Outcome in the Reward Schedule Task. *J Neurosci* **33**, 3477–3491 (2013).

230. E. F. Harkin, C. D. Grossman, J. Y. Cohen, J.-C. Béïque, R. Naud, Serotonin predictively encodes value. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.09.19.558526.

231. C. D. Grossman, B. A. Bari, J. Y. Cohen, Serotonin neurons modulate learning rate through uncertainty. *Current Biology* **32**, 586-599.e7 (2022).

232. K. Iigaya, M. S. Fonseca, M. Murakami, Z. F. Mainen, P. Dayan, An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nat Commun* **9**, 2477 (2018).

233. K. Miyazaki, K. W. Miyazaki, K. Doya, Activation of Dorsal Raphe Serotonin Neurons Underlies Waiting for Delayed Rewards. *J. Neurosci.* **31**, 469–479 (2011).

234. K. Miyazaki, K. W. Miyazaki, A. Yamanaka, T. Tokuda, K. F. Tanaka, K. Doya, Reward probability and timing uncertainty alter the effect of dorsal raphe serotonin neurons on patience. *Nature Communications* **9**, 2048 (2018).

235. K. Miyazaki, K. W. Miyazaki, G. Sivori, A. Yamanaka, K. F. Tanaka, K. Doya, Serotonergic projections to the orbitofrontal and medial prefrontal cortices differentially modulate waiting for future rewards. *Sci Adv* **6**, eabc7246 (2020).

236. P. A. Correia, E. Lottem, D. Banerjee, A. S. Machado, M. R. Carey, Z. F. Mainen, Transient inhibition and long-term facilitation of locomotion by phasic optogenetic activation of serotonin neurons. *eLife* **6**, e20975 (2017).

237. M. S. Fonseca, M. Murakami, Z. F. Mainen, Activation of Dorsal Raphe Serotonergic Neurons Promotes Waiting but Is Not Reinforcing. *Current Biology* **25**, 306–315 (2015).

238. H. F. Clarke, S. C. Walker, H. S. Crofts, J. W. Dalley, T. W. Robbins, A. C. Roberts, Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. *J Neurosci* **25**, 532–538 (2005).

239. O. Hikosaka, The habenula: from stress evasion to value-based decision-making. *Nat Rev Neurosci* **11**, 503–513 (2010).

240. O. Hikosaka, S. R. Sesack, L. Lecourtier, P. D. Shepard, Habenula: Crossroad between the Basal Ganglia and the Limbic System. *Journal of Neuroscience* **28**, 11825–11829 (2008).

241. G. K. Aghajanian, R. Y. Wang, Habenular and other midbrain raphe afferents demonstrated by a modified retrograde tracing technique. *Brain Res* **122**, 229–242 (1977).

242. M. Herkenham, W. J. H. Nauta, Efferent connections of the habenular nuclei in the rat. *Journal of Comparative Neurology* **187**, 19–47 (1979).

243. H. Ji, P. D. Shepard, Lateral Habenula Stimulation Inhibits Rat Midbrain Dopamine Neurons through a GABAA Receptor-Mediated Mechanism. *J. Neurosci.* **27**, 6923–6930 (2007).

244. M. Matsumoto, O. Hikosaka, Negative motivational control of saccadic eye movement by the lateral habenula. *Prog Brain Res* **171**, 399–402 (2008).

245. M. Matsumoto, O. Hikosaka, Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* **447**, 1111–1115 (2007).

246. P. L. Brown, H. Palacorolla, D. Brady, K. Riegger, G. I. Elmer, P. D. Shepard, Habenula-Induced Inhibition of Midbrain Dopamine Neurons Is Diminished by Lesions of the Rostromedial Tegmental Nucleus. *J Neurosci* **37**, 217–225 (2017).

247. S. Hong, T. C. Jhou, M. Smith, K. S. Saleem, O. Hikosaka, Negative Reward Signals from the Lateral Habenula to Dopamine Neurons Are Mediated by Rostromedial Tegmental Nucleus in Primates. *J. Neurosci.* **31**, 11457–11471 (2011).

248. H. Hu, Y. Cui, Y. Yang, Circuits and functions of the lateral habenula in health and in disease. *Nat Rev Neurosci* **21**, 277–295 (2020).

249. V. Varga, B. Kocsis, T. Sharp, Electrophysiological evidence for convergence of inputs from the medial prefrontal cortex and lateral habenula on single neurons in the dorsal raphe nucleus. *European Journal of Neuroscience* **17**, 280–286 (2003).

250. M. R. Park, Monosynaptic inhibitory postsynaptic potentials from lateral habenula recorded in dorsal raphe neurons. *Brain Res Bull* **19**, 581–586 (1987).

251. W. C. Stern, A. Johnson, J. D. Bronzino, P. J. Morgane, Effects of electrical stimulation of the lateral habenula on single-unit activity of raphe neurons. *Experimental Neurology* **65**, 326–342 (1979).

252. M. Herkenham, W. J. Nauta, Afferent connections of the habenular nuclei in the rat. A horseradish peroxidase study, with a note on the fiber-of-passage problem. *J Comp Neurol* **173**, 123–146 (1977).

253. S. A. Huettel, A. W. Song, G. McCarthy, *Functional Magnetic Resonance Imaging* (Oxford University Press, Incorporated, 2009).

254. N. Khalighinejad, L. Priestley, S. Jbabdi, M. F. S. Rushworth, Human decisions about when to act originate within a basal forebrain–nigral circuit. *Proceedings of the National Academy of Sciences* **117**, 11799–11810 (2020).

255. K. J. Friston, C. Buechel, G. R. Fink, J. Morris, E. Rolls, R. J. Dolan, Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* **6**, 218–229 (1997).

256. S. Yoo, D. R. Mittelstein, R. C. Hurt, J. Lacroix, M. G. Shapiro, Focused ultrasound excites cortical neurons via mechanosensitive calcium accumulation and ion channel amplification. *Nat Commun* **13**, 493 (2022).

257. D. G. Blackmore, D. Razansky, J. Götz, Ultrasound as a versatile tool for short- and long-term improvement and monitoring of brain function. *Neuron* **111**, 1174–1190 (2023).

258. Y. Shamli Oghli, T. Grippe, T. Arora, T. Hoque, G. Darmani, R. Chen, Mechanisms of theta burst transcranial ultrasound induced plasticity in the human motor cortex. *Brain Stimulation* **16**, 1135–1143 (2023).

259. N. Samuel, K. Zeng, I. E. Harmsen, M. Y. R. Ding, G. Darmani, C. Sarica, B. Santyr, A. Vetkas, A. Pancholi, A. Fomenko, V. Milano, K. Yamamoto, U. Saha, R. Wennberg, N. C. Rowland, R. Chen, A. M. Lozano, Multi-modal investigation of transcranial ultrasound-induced neuroplasticity of the human motor cortex. *Brain Stimulation* **15**, 1337–1347 (2022).

260. G. Darmani, T. O. Bergmann, K. Butts Pauly, C. F. Caskey, L. de Lecea, A. Fomenko, E. Fouragnan, W. Legon, K. R. Murphy, T. Nandi, M. A. Phipps, G. Pinton, H. Ramezanpour, J. Sallet, S. N. Yaakub, S. S. Yoo, R. Chen, Non-invasive transcranial ultrasound stimulation for neuromodulation. *Clinical Neurophysiology* **135**, 51–73 (2022).

261. K. R. Murphy, J. S. Farrell, J. L. Gomez, Q. G. Stedman, N. Li, S. A. Leung, C. H. Good, Z. Qiu, K. Firouzi, K. Butts Pauly, B. P. T. Khuri-Yakub, M. Michaelides, I. Soltesz, L. de Lecea, A tool for monitoring cell type–specific focused ultrasound neuromodulation and control of chronic epilepsy. *Proceedings of the National Academy of Sciences* **119**, e2206828119 (2022).

262. T. Di Ianni, K. P. Morrison, B. Yu, K. R. Murphy, L. de Lecea, R. D. Airan, High-throughput ultrasound neuromodulation in awake and freely behaving rats. *Brain Stimulation* **16**, 1743–1752 (2023).

263. L. Verhagen, C. Gallea, D. Folloni, C. Constans, D. E. Jensen, H. Ahnine, L. Roumazeilles, M. Santin, B. Ahmed, S. Lehericy, M. C. Klein-Flügge, K. Krug, R. B. Mars, M. F. Rushworth, P. Pouget, J.-F. Aubry, J. Sallet, Offline impact of transcranial focused ultrasound on cortical activation in primates. *Elife* **8**, e40541 (2019).

264. N. Khalighinejad, A. Bongioanni, L. Verhagen, D. Folloni, D. Attali, J.-F. Aubry, J. Sallet, M. F. S. Rushworth, A Basal Forebrain-Cingulate Circuit in Macaques Decides It Is Time to Act. *Neuron* **105**, 370-384.e8 (2020).

265. D. Folloni, L. Verhagen, R. B. Mars, E. Fouragnan, C. Constans, J.-F. Aubry, M. F. S. Rushworth, J. Sallet, Manipulation of Subcortical and Deep Cortical Activity in the Primate Brain Using Transcranial Focused Ultrasound Stimulation. *Neuron* **101**, 1109-1116.e5 (2019).

266. P. Perelman, W. E. Johnson, C. Roos, H. N. Seuánez, J. E. Horvath, M. A. M. Moreira, B. Kessing, J. Pontius, M. Roelke, Y. Rumpler, M. P. C. Schneider, A. Silva, S. J. O'Brien, J. Pecon-Slattery, A Molecular Phylogeny of Living Primates. *PLOS Genetics* **7**, e1001342 (2011).

267. S. P. Wise, *Cortical Evolution in Primates: What Primates Are, What Primates Were, and Why the Cortex Changed* (Oxford University Press, 2024).

268. F.-X. Neubert, R. B. Mars, J. Sallet, M. F. S. Rushworth, Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proceedings of the National Academy of Sciences* **112**, E2695–E2704 (2015).

269. F.-X. Neubert, R. B. Mars, A. G. Thomas, J. Sallet, M. F. S. Rushworth, Comparison of Human Ventral Frontal Cortex Areas for Cognitive Control and Language with Areas in Monkey Frontal Cortex. *Neuron* **81**, 700–713 (2014).

270. J. Sallet, R. B. Mars, M. P. Noonan, F.-X. Neubert, S. Jbabdi, J. X. O'Reilly, N. Filippini, A. G. Thomas, M. F. Rushworth, The Organization of Dorsal Frontal Cortex in Humans and Macaques. *J. Neurosci.* **33**, 12255–12274 (2013).

271. S. Mathôt, Pupillometry: Psychology, Physiology, and Function. *J Cogn* **1**, 16.

272. R. Remedios, A. Kennedy, M. Zelikowsky, B. F. Grewe, M. J. Schnitzer, D. J. Anderson, Social behaviour shapes hypothalamic neural ensemble representations of conspecific sex. *Nature* **550**, 388–392 (2017).

273. S. X. Zhang, L. E. Miner, C. L. Boutros, D. Rogulja, M. A. Crickmore, Motivation, Perception, and Chance Converge to Make a Binary Decision. *Neuron* **99**, 376-388.e6 (2018).

274. E. D. Hoopfer, Y. Jung, H. K. Inagaki, G. M. Rubin, D. J. Anderson, P1 interneurons promote a persistent internal state that enhances inter-male aggression in Drosophila. *eLife* **4**, e11346 (2015).

275. S. S. Bolkan, I. R. Stone, L. Pinto, Z. C. Ashwood, J. M. Iravedra Garcia, A. L. Herman, P. Singh, A. Bandi, J. Cox, C. A. Zimmerman, J. R. Cho, B. Engelhard, J. W. Pillow, I. B. Witten, Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. *Nat Neurosci* **25**, 345–357 (2022).

276. A. J. Calhoun, J. W. Pillow, M. Murthy, Unsupervised identification of the internal states that shape natural behavior. *Nat Neurosci* **22**, 2040–2049 (2019).

277. Y. Niv, N. D. Daw, D. Joel, P. Dayan, Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* **191**, 507–520 (2007).

278. Y. Niv, N. Daw, P. Dayan, "How fast to work: Response vigor, motivation and tonic dopamine" in *Advances in Neural Information Processing Systems* (MIT Press, 2005; https://proceedings.neurips.cc/paper/2005/hash/3812f9a59b634c2a9c574610eaba5bed-Abstract.html)vol. 18.

279. E. L. Thorndike, A proof of the law of effect. *Science* **77**, 173–175 (1933).

280. P. Anselme, O. Güntürkün, How foraging works: Uncertainty magnifies food-seeking motivation. *Behavioral and Brain Sciences* **42** (2019).

281. M. Kleiner, D. Brainard, D. Pelli, A. Ingling, R. Murray, C. Broussard, What's new in psychtoolbox-3. *Perception.* **36**, 1–16 (2007).

282. N. Tardiff, J. D. Medaglia, D. S. Bassett, S. L. Thompson-Schill, The modulation of brain network integration and arousal during exploration. *Neuroimage* **240**, 118369 (2021).

283. Stan Development Team, Stan Modeling Language Users Guide and Reference Manual, *stan-dev.github.io* (2022). //mc-stan.org/.

284. H. E. M. den Ouden, N. D. Daw, G. Fernandez, J. A. Elshout, M. Rijpkema, M. Hoogman, B. Franke, R. Cools, Dissociable Effects of Dopamine and Serotonin on Reversal Learning. *Neuron* **80**, 1090–1100 (2013).

285. E. S. Bromberg-Martin, O. Hikosaka, K. Nakamura, Coding of task reward value in the dorsal raphe nucleus. *J Neurosci* **30**, 6262–6272 (2010).

286. L. Marshall, C. Mathys, D. Ruge, A. O. de Berker, P. Dayan, K. E. Stephan, S. Bestmann, Pharmacological Fingerprints of Contextual Uncertainty. *PLOS Biology* **14**, e1002575 (2016).

287. J. Kubanek, J. Brown, P. Ye, K. B. Pauly, T. Moore, W. Newsome, Remote, brain region–specific control of choice behavior with ultrasonic waves. *Science Advances* **6**, eaaz4193 (2020).

288. T. Deffieux, Y. Younan, N. Wattiez, M. Tanter, P. Pouget, J.-F. Aubry, Low-intensity focused ultrasound modulates monkey visuomotor behavior. *Curr Biol* **23**, 2430–2433 (2013).

289. O. Hikosaka, The habenula: from stress evasion to value-based decision-making. *Nature Reviews Neuroscience* **11**, 503–513 (2010).

290. I. E. Monosov, D. A. Leopold, O. Hikosaka, Neurons in the Primate Medial Basal Forebrain Signal Combined Information about Reward Uncertainty, Value, and Punishment Anticipation. *J. Neurosci.* **35**, 7443–7459 (2015).

291. N. M. Ledbetter, C. D. Chen, I. E. Monosov, Multiple Mechanisms for Processing Reward Uncertainty in the Primate Basal Forebrain. *J. Neurosci.* **36**, 7852–7864 (2016).

292. K. Zhang, C. D. Chen, I. E. Monosov, Novelty, Salience, and Surprise Timing Are Signaled by Neurons in the Basal Forebrain. *Current Biology* **29**, 134-142.e3 (2019).

293. K. Miyamoto, T. Osada, R. Setsuie, M. Takeda, K. Tamura, Y. Adachi, Y. Miyashita, Causal neural network of metamemory for retrospection in primates. *Science* **355**, 188–193 (2017).

294. B. R. Cottereau, A. T. Smith, S. Rima, D. Fize, Y. Héjja-Brichard, L. Renaud, C. Lejards, N. Vayssière, Y. Trotter, J.-B. Durand, Processing of Egomotion-Consistent Optic Flow in the Rhesus Macaque Cortex. *Cereb Cortex* **27**, 330–343 (2017).

295. A. Bongioanni, D. Folloni, L. Verhagen, J. Sallet, M. C. Klein-Flügge, M. F. S. Rushworth, Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature* **591**, 270–274 (2021).

296. K. A. Eley, S. R. Watt-Smith, F. Sheerin, S. J. Golding, "Black Bone" MRI: a potential alternative to CT with three-dimensional reconstruction of the craniofacial skeleton in the diagnosis of craniosynostosis. *Eur Radiol* **24**, 2417–2426 (2014).

297. S. N. Yaakub, T. A. White, J. Roberts, L. Verhagen, C. J. Stagg, S. Hall, E. F. Fouragnan, Transcranial focused ultrasound-mediated neurochemical and functional connectivity changes in deep cortical regions in humans. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.01.20.524869.

298. D. G. R. Tervo, E. Kuleshova, M. Manakov, M. Proskurin, M. Karlsson, A. Lustig, R. Behnam, A. Y. Karpova, The anterior cingulate cortex directs exploration of alternative strategies. *Neuron* **109**, 1876-1887.e6 (2021).

299. R. Quilodran, M. Rothé, E. Procyk, Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* **57**, 314–325 (2008).

300. F. M. Stoll, V. Fontanier, E. Procyk, Specific frontal neural dynamics contribute to decisions to check. *Nat Commun* **7**, 11990 (2016).

301. K. J. Friston, C. Buechel, G. R. Fink, J. Morris, E. Rolls, R. J. Dolan, Psychophysiological and Modulatory Interactions in Neuroimaging. *NeuroImage* **6**, 218–229 (1997).

302. E. Lottem, D. Banerjee, P. Vertechi, D. Sarra, M. oude Lohuis, Z. F. Mainen, Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature Communications* **9**, 1000 (2018).

303. Y.-Y. Feng, E. S. Bromberg-Martin, I. E. Monosov, Dorsal raphe neurons integrate the values of reward amount, delay, and uncertainty in multi-attribute decision-making. *Cell Reports* **43**, 114341 (2024).

304. O. Hikosaka, S. R. Sesack, L. Lecourtier, P. D. Shepard, Habenula: Crossroad between the Basal Ganglia and the Limbic System. *Journal of Neuroscience* **28**, 11825–11829 (2008).

305. E. S. Bromberg-Martin, Y.-Y. Feng, T. Ogasawara, J. K. White, K. Zhang, I. E. Monosov, A neural mechanism for conserved value computations integrating information and rewards. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.08.14.503903.

306. P. L. A. Gabbott, T. A. Warner, P. R. L. Jays, P. Salway, S. J. Busby, Prefrontal cortex in the rat: Projections to subcortical autonomic, motor, and limbic centers. *Journal of Comparative Neurology* **492**, 145–177 (2005).

307. L. Yetnikoff, A. Y. Cheng, H. N. Lavezzi, K. P. Parsley, D. S. Zahm, Sources of input to the rostromedial tegmental nucleus, ventral tegmental area, and lateral habenula compared: A study in rat. *Journal of Comparative Neurology* **523**, 2426–2456 (2015).

308. D. Folloni, E. Fouragnan, M. K. Wittmann, L. Roumazeilles, L. Tankelevitch, L. Verhagen, D. Attali, J.-F. Aubry, J. Sallet, M. F. S. Rushworth, Ultrasound modulation of macaque prefrontal cortex selectively alters credit assignment–related activity and behavior. *Science Advances* **7**, eabg7700 (2021).

309. S. Moeller, E. Yacoub, C. A. Olman, E. Auerbach, J. Strupp, N. Harel, K. Uğurbil, Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med* **63**, 1144–1153 (2010).

310. K. Setsompop, B. A. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, L. L. Wald, Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine* **67**, 1210–1224 (2012).

311. M. Shahdloo, U. Schüffelgen, D. Papp, K. L. Miller, M. Chiew, Model-based dynamic off-resonance correction for improved accelerated fMRI in awake behaving nonhuman primates. *Magnetic Resonance in Medicine* **87**, 2922–2932 (2022).

312. M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, S. M. Smith, FSL. *NeuroImage* **62**, 782–790 (2012).

313. B. T. Cox, S. Kara, S. R. Arridge, P. C. Beard, k-space propagation models for acoustically heterogeneous media: Application to biomedical photoacoustics. *The Journal of the Acoustical Society of America* **121**, 3453–3464 (2007).

314. L. Marsac, D. Chauvet, R. La Greca, A.-L. Boch, K. Chaumoitre, M. Tanter, J.-F. Aubry, Ex vivo optimisation of a heterogeneous speed of sound model of the human skull for non-invasive transcranial focused ultrasound at 1 MHz. *Int J Hyperthermia* **33**, 635–645 (2017).

315. J. K. Mueller, L. Ai, P. Bansal, W. Legon, Numerical evaluation of the skull for human neuromodulation with transcranial focused ultrasound. *J Neural Eng* **14**, 066012 (2017).

316. E. L. Charnov, Optimal foraging, the marginal value theorem. *Theoretical Population Biology* **9**, 129–136 (1976).

317. A. Charara, A. Parent, Chemoarchitecture of the primate dorsal raphe nucleus. *Journal of Chemical Neuroanatomy* **15**, 111–127 (1998).

318. J.-P. Hornung, The human raphe nuclei and the serotonergic system. *J Chem Neuroanat* **26**, 331–343 (2003).

319. L. Priestley, M. Chiew, M. Shahdloo, A. Mahmoodi, X. Cheng, R. Cleveland, M. Rushworth, N. Khalighinejad, Dorsal Raphe Nucleus Controls Motivational State Transitions in Monkeys. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.02.13.580224.

320. P. Dayan, Twenty-Five Lessons from Computational Neuromodulation. *Neuron* **76**, 240–256 (2012).

321. M. Guitart-Masip, E. Duzel, R. Dolan, P. Dayan, Action versus valence in decision making. *Trends Cogn Sci* **18**, 194–202 (2014).

322. W. Schultz, A dopamine mechanism for reward maximization. *Proc Natl Acad Sci U S A* **121**, e2316658121 (2024).

323. P. Dayan, Dopamine, Reinforcement Learning, and Addiction. *Pharmacopsychiatry* **42**, S56–S65 (2009).

324. N. Kriegeskorte, M. Mur, P. A. Bandettini, Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2** (2008).

325. J. W. Kable, P. W. Glimcher, The neural correlates of subjective value during intertemporal choice. *Nat Neurosci* **10**, 1625–1633 (2007).

326. M. C. Klein-Flügge, A. Bongioanni, M. F. S. Rushworth, Medial and orbital frontal cortex in decision-making and flexible behavior. *Neuron* **110**, 2743–2770 (2022).

327. B. L. Jacobs, E. C. Azmitia, Structure and function of the brain serotonin system. *Physiological Reviews* **72**, 165–229 (1992).

328. W. Fu, E. Le Maître, V. Fabre, J.-F. Bernard, Z.-Q. David Xu, T. Hökfelt, Chemical neuroanatomy of the dorsal raphe nucleus and adjacent structures of the mouse brain. *Journal of Comparative Neurology* **518**, 3464–3494 (2010).

329. P. S. Muhle-Karbe, H. Sheahan, G. Pezzulo, H. J. Spiers, S. Chien, N. W. Schuck, C. Summerfield, Goal-seeking compresses neural codes for space in the human hippocampus and orbitofrontal cortex. *Neuron* **111**, 3885-3899.e6 (2023).

330. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48 (2015).

331. S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, P. M. Matthews, Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, S208–S219 (2004).

332. S. M. Smith, Fast robust automated brain extraction. *Human Brain Mapping* **17**, 143–155 (2002).

333. M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).

334. D. N. Greve, B. Fischl, Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).

335. M. W. Woolrich, B. D. Ripley, M. Brady, S. M. Smith, Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage* **14**, 1370–1386 (2001).

336. M. W. Woolrich, T. E. J. Behrens, C. F. Beckmann, M. Jenkinson, S. M. Smith, Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage* **21**, 1732–1747 (2004).

337. J. C. W. Brooks, C. F. Beckmann, K. L. Miller, R. G. Wise, C. A. Porro, I. Tracey, M. Jenkinson, Physiological noise modelling for spinal functional magnetic resonance imaging studies. *Neuroimage* **39**, 680–692 (2008).

338. W. M. Pauli, A. N. Nili, J. M. Tyszka, A high-resolution probabilistic *in vivo* atlas of human subcortical brain nuclei. *Scientific Data* **5**, 180063 (2018).

339. G. S. Kranz, A. Hahn, M. Savli, R. Lanzenberger, Challenges in the differentiation of midbrain raphe nuclei in neuroimaging research. *Proceedings of the National Academy of Sciences* **109**, E2000–E2000 (2012).

340. T. E. Behrens, L. T. Hunt, M. W. Woolrich, M. F. Rushworth, Associative learning of social value. *Nature* **456**, 245–249 (2008).

341. N. Kolling, J. Scholl, A. Chekroud, H. A. Trier, M. F. S. Rushworth, Prospection, Perseverance, and Insight in Sequential Behavior. *Neuron* **99**, 1069-1082.e7 (2018).

342. S. Holm, A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**, 65–70 (1979).

343. J. C. Marques, M. Li, D. Schaak, D. N. Robson, J. M. Li, Internal state dynamics shape brainwide activity and foraging behaviour. *Nature* **577**, 239–243 (2020).

344. M. K. Wittmann, P. L. Lockwood, M. F. S. Rushworth, Neural Mechanisms of Social Cognition in Primates. *Annual Review of Neuroscience* **41**, 99–118 (2018).

345. E. F. Harkin, M. B. Lynn, A. Payeur, J.-F. Boucher, L. Caya-Bissonnette, D. Cyr, C. Stewart, A. Longtin, R. Naud, J.-C. Béïque, Temporal derivative computation in the dorsal raphe network revealed by an experimentally driven augmented integrate-and-fire modeling framework. *eLife* **12**, e72951 (2023).

346. K. Nakamura, M. Matsumoto, O. Hikosaka, Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus. *J Neurosci* **28**, 5331–5343 (2008).

347. G. Tononi, O. Sporns, G. M. Edelman, Measures of degeneracy and redundancy in biological networks. *Proceedings of the National Academy of Sciences* **96**, 3257–3262 (1999).

348. N. Garrett, N. D. Daw, Biased belief updating and suboptimal choice in foraging decisions. *Nat Commun* **11**, 3417 (2020).

349. K. Zhang, F. Karim, Z. Jin, H. Xiao, Y. Yao, Q. Ni, B. Li, W. Pu-Cuo, Z. Huang, H. Xu, Diet and feeding behavior of a group of high-altitude rhesus macaques: high adaptation to food shortages and seasonal fluctuations. *Current Zoology* **69**, 304–314 (2023).

350. J. Moncrieff, R. E. Cooper, T. Stockmann, S. Amendola, M. P. Hengartner, M. A. Horowitz, The serotonin theory of depression: a systematic umbrella review of the evidence. *Mol Psychiatry* **28**, 3243–3256 (2023).

351. R. Cools, A. C. Roberts, T. W. Robbins, Serotoninergic regulation of emotional and behavioural control processes. *Trends in Cognitive Sciences* **12**, 31–40 (2008).

352. P. J. Cowen, M. Browning, What has serotonin to do with depression? *World Psychiatry* **14**, 158–160 (2015).