The computational neurology of active vision

Thomas Parr

A thesis submitted for the degree of Doctor of Philosophy

Wellcome Centre for Human Neuroimaging

Supervised by:

Professor Karl Friston

Professor Geraint Rees

I, Thomas Parr confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Abstract

In this thesis, we appeal to recent developments in theoretical neurobiology – namely, active inference – to understand the active visual system and its disorders. Chapter 1 reviews the neurobiology of active vision. This introduces some of the key conceptual themes around attention and inference that recur through subsequent chapters. Chapter 2 provides a technical overview of active inference, and its interpretation in terms of message passing between populations of neurons. Chapter 3 applies the material in Chapter 2 to provide a computational characterisation of the oculomotor system. This deals with two key challenges in active vision: deciding where to look, and working out how to look there. The homology between this message passing and the brain networks solving these inference problems provide a basis for *in silico* lesion experiments, and an account of the aberrant neural computations that give rise to clinical oculomotor signs (including *internuclear ophthalmoplegia*). Chapter 4 picks up on the role of uncertainty resolution in deciding where to look, and examines the role of beliefs about the quality (or precision) of data in perceptual inference. We illustrate how abnormal prior beliefs influence inferences about uncertainty and give rise to neuromodulatory changes and visual hallucinatory phenomena (of the sort associated with *synucleinopathies*). We then demonstrate how synthetic pharmacological perturbations that alter these neuromodulatory systems give rise to the oculomotor changes associated with drugs acting upon these systems. Chapter 5 develops a model of *visual neglect*, using an oculomotor version of a line cancellation task. We then test a prediction of this model using magnetoencephalography and dynamic causal modelling. Chapter 6 concludes by situating the work in this thesis in the context of computational neurology. This illustrates how the variational principles used here to characterise the active visual system may be generalised to other sensorimotor systems and their disorders.

Impact statement

Given that the nervous system is engaged in computation, it follows that neurological disorders may be thought of as disorders of computation. The work presented in this thesis illustrates a first principles approach that may be used to characterise the computations performed by the active visual system and the ways in which these may be compromised. The utility of this is that it offers a formal approach to understanding the processes by which a healthy brain infers the causes of its sensations and the appropriate course of action to take, and enables simulation of the consequences of computational pathology. Through this approach, it is possible to provide precise accounts of the functional disconnections that could underwrite neurological disorders, and the consequence of such lesions to the brain as a whole (i.e. the network level changes or *diaschisis* that result from disrupting one part of an interconnected system), and to behaviour (e.g. oculomotor behaviour). This has potential applications in phenotyping of patient populations in terms of the functional deficits giving rise to clinical syndromes. This has implications for personalised approaches to medicine, where the phenotypic

characteristics of a given patient may be used to inform decisions about their treatment or selection for clinical trials. In addition, the pharmacological interventions simulated in this thesis offer a simple proof-of-principle that could allow for monitoring of therapeutic interventions in terms of the computational manifestations of those therapies. In principle, the combination of computational phenotyping and synthetic therapeutic interventions could be used to predict the response of an individual to different sorts of therapy through simulating their trajectories into the future under different interventions. Another application of the approach outlined here is in understanding the functional anatomy of the active visual system from first principles. In brief, the (variational) theoretical approach we have pursued implies that the inferential problems the brain must solve to engage in active vision mandate a specific architecture that supports the passing of inferential messages. Drawing from clinical (neuropsychological) data, the consequences of simulated lesions that disrupt these messages constrain how they map to known neuroanatomy. This enables the elaboration of specific neuroanatomical hypotheses in terms of changes in the coupling between brain regions. Our penultimate chapter provides an example of the development of a model informed by neuropsychology and the evaluation of a hypothesis implied by this model using neuroimaging. Finally, while our focus has been on active vision and the oculomotor system, the methods applied here could also be used to characterise other sensorimotor systems. We hope that this work will aid in the understanding of these systems and, ultimately, in informing treatment of their disorders.

Acknowledgements

I am very grateful to all the friends, family, and colleagues who have supported me during my PhD. To name a few:

I am very grateful to Karl Friston for his supervision and have benefited immensely from his guidance and mentorship. Thanks also to Geraint Rees, for his supervision and input.

The hard work of the support staff at both the FIL and the Queen's Larder has been invaluable during my time at Queen Square.

During the last few years, I have been fortunate to have the opportunity to spend time at academic institutions around the world and am grateful to those who made this possible. Specifically, I would like to thank Jakob Hohwy for my time at Monash University; Michael Halassa, for inviting me to visit the Massachusetts Institute of Technology; and Giovanni Pezzulo, for my visit to Istituto di Scienze e Tecnologie della Cognizione.

I would like to thank those who have helped me, directly or indirectly, with the work in this thesis. These include the people who worked with me on the research on which this thesis is based: David Benrimoh, Hayriye Cagnan, Stefan Kiebel, Dimitrije Markovic, Berk Mirza, and Peter Vincent; and others including (but not limited to): Rick Adams, Micah Allen, Anjali Bhat, Jelle Bruineberg, Emma Holmes, Michael Kirchoff, Jakub Limanowski, Maxwell Ramstead, Richard Rosch, and Peter Zeidman.

Finally, I am grateful to my parents, brothers, and to Hugo Vine, Prateek Yadav, and Aimee Goel for their moral support.

This work was made possible through the support of the Rosetrees Trust.

Contents

Glossary of mathematical notation and key variables

Notation	Description
y	Continuous sensory data
x	Continuous hidden state
v	Continuous hidden cause
a	Continuous action[1]
o	Categorical sensory data (observation)
s	Categorical hidden state
π	Categorical policy (sequence of actions)
u	Categorical action
ε	Prediction error (continuous)
$\boldsymbol{\varepsilon}$	Prediction error (categorical)
$\tilde{x}$	Trajectory (of hidden states)
$\dot{x}$	Rate of change (of hidden states)
F	Free energy
$\mathbf{F}$	Vector of free energies for each policy
G	Expected free energy
$\mathbf{G}$	Vector of expected free energies for each policy
$\mathcal{N}(\mu, C^{-1})$	Normal distribution (mode μ, covariance C)
$Cat(\mathbf{s})$	Categorical distribution (expectation $\mathbf{s}$)
$Dir(\mathbf{b})$	Dirichlet distribution (parameters $\mathbf{b}$)
$\Gamma(\alpha,\beta)$	Gamma distribution (shape α, rate β)
$\psi(\cdot)$	Digamma function (derivative of gamma function)
$\sigma(\cdot)$	Softmax (normalised exponential) function
$E_P[\cdot]$	Expected (averaged) under the distribution P
$H[\cdot]$	Shannon entropy (negative expected log probability)
$D_{KL}[\cdot \Vert \cdot]$	Kullback-Leibler divergence (expected log probability ratio)
$P(o, s, \pi)$, $p(v, x, y)$	Generative model (categorical, continuous)
$Q(s, \pi)$, $q(v, x)$	Variational density (categorical, continuous)

[1] In Chapter 5, a is used to indicate prior Dirichlet parameters

1 - Active vision and the oculomotor system

Introduction

Although our experience of the visual world seems temporally and spatially continuous, the sensations we derive it from are not. Saccadic eye movements constitute a series of discrete fixations, interspersed with rapid movements. Little meaningful visual information is obtained as the eyes sweep from one fixation to the next (Bridgeman et al. 1975) and, at any moment, the proportion of the visual field from which any high resolution information is sampled is tiny. These observations, seemingly so contrary to perceptual experience, can be reconciled under the metaphor of perception as hypothesis testing (Friston et al. 2012a; Gregory 1980). By forming hypotheses about a continuous world, saccades can be deployed as experiments to adjudicate among alternatives. This licenses a description of active vision as if it were a scientific process (with certain qualifications c.f. (Bruineberg et al. 2016)).

This view implies perception of space is fundamentally tied to motor representations, as visual input at a point in space is the consequence of a motor experiment (saccade to that location) (Zimmermann and Lappe 2016). This enactivist take on perceptual synthesis means that objects in the visual field become hypotheses or explanations for 'what would I see if I looked there?' In this chapter[2], we describe the neuronal apparatus used to perform these experiments – and thereby implement active vision (Andreopoulos and Tsotsos 2013; Mirza et al. 2016; Ognibene and Baldassarre 2014; Wurtz et al. 2011). This functional anatomy comprises the brainstem network that gives rise to the nerves to the extraocular muscles. The superior colliculus is an important structure in this network, receiving input from both subcortical and cortical regions. Particular focus will be afforded structures that determine the choice of saccade target, and the mechanisms by which the data from previous saccades are combined, accumulated or assimilated to construct a seamless temporal experience (Marchetti 2014). These mechanisms can fail in the damaged brain, and a common syndrome resulting from this failure is visual neglect. Patients suffering from this fail to attend to one side (typically the left) of visual space (Halligan and Marshall 1998). One manifestation of this attentional deficit is a decreased frequency of saccadic sampling in the neglected half of space relative to the other (Karnath and Rorden 2012) despite intact early visual processing of stimuli on the neglected side; as evidenced by electrophysiology (Di Russo et al. 2007) and neuroimaging (Rees et al. 2000). We will address some of the links between the neurobiology of visual scene

[2] This chapter is adapted from (Parr and Friston 2017a)

construction, and the consequences of its disruption. A number of theoretical concepts recur throughout this chapter. These include consideration of the mnemonic processes required for scene construction, the relationship between eye movements and attention, and the inferential (Bayesian) foundation of these processes.

Brainstem oculomotor control

All forms of eye movement depend upon the connections from the cranial nerve nuclei in the midbrain (CN III), and the pons (CN IV, VI) to the extraocular muscles. Saccadic eye movements depend specifically upon the connections to these nuclei from the paramedian pontine reticular formation (PPRF) and the rostral interstitial nucleus of the medial longitudinal fasciculus (riMLF). The former generates horizontal saccades (Cohen et al. 1968; Henn 1992), and the latter vertical (Büttner-Ennever and Büttner 1978). Other important influences come from the vestibular system (Baker and Highstein 1978), and the cerebellum (Berretta et al. 1993), but are outside the scope of this chapter. A subset of neurons within the PPRF monosynaptically target the ipsilateral abducens (CN VI) nucleus (Strassman et al. 1986). From the abducens nucleus, some neurons have axons which first decussate, then ascend as part of the MLF, to the oculomotor (CN III) nucleus in the midbrain (Sparks 2002). The PPRF can use this pathway to initiate conjugate eye movements in the ipsilateral direction. An additional anatomical pathway allows the PPRF to influence the riMLF (Büttner-Ennever and Büttner 1978), ensuring it can generate saccades with a vertical directional component (see Figure 1.1 for a summary of this anatomy).

Saccadic movements are rapid movements that occur between short periods of fixation. In order to maintain fixation between saccades, PPRF ‘burst’ neurons are tonically inhibited by ‘omnipause’ neurons, located in the nucleus raphe interpositus (RIP) (Büttner-Ennever et al. 1988). These cells cease firing immediately before a burst of firing in the PPRF cells, but resume before the saccade is complete. ‘Omnipause’ neurons may have a role in synchronising different directional components of saccade generation, as the RIP also projects to the riMLF (Büttner-Ennever and Büttner 1978). The electrophysiological correlates of the fixation and saccadic phases suggest the brain treats saccadic eye movements as a series of discrete events, consistent with the view that attentional processes are both serial and discrete (Buschman and Miller 2010a).

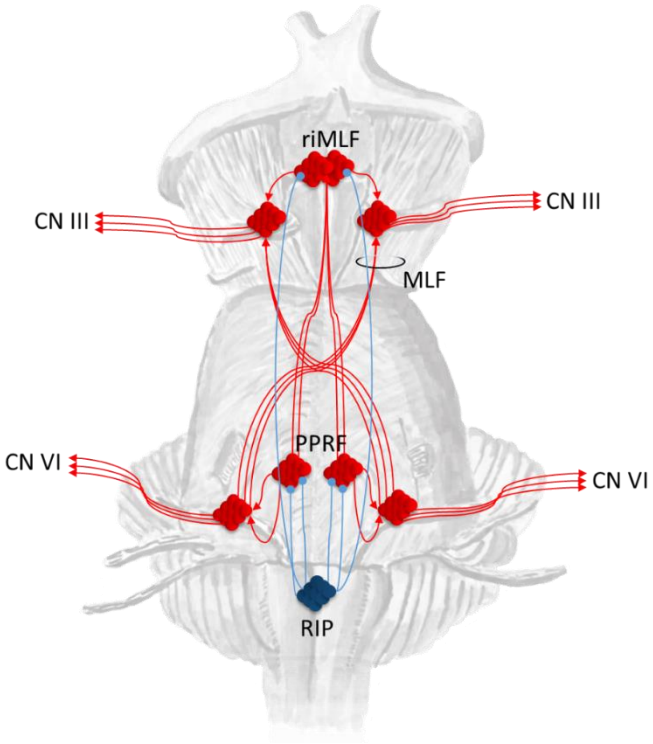

Figure 1.1 – Brainstem control of saccadic movements. This schematic shows some of the brainstem nuclei involved in the generation and control of saccadic eye movements. The paramedian pontine reticular formation (PPRF) is responsible for the generation of horizontal saccades, through its influence on the ipsilateral abducens nucleus, which gives rise to cranial nerve (CN) VI. A subset of neurons in the abducens nucleus projects to the contralateral oculomotor (CN III) nucleus in the midbrain, via the medial longitudinal fasciculus (MLF), ensuring conjugate eye movements occur. The PPRF additionally projects to the rostral interstitial nucleus of the MLF (riMLF), which generates vertical saccades. 'Omnipause' neurons in the nucleus raphe interpositus (RIP) synchronise the onset of vertical and horizontal components of saccades. The superior colliculus (not shown) influences both the PPRF and RIP. Excitatory connections are shown in red, while inhibitory connections are shown in blue.

The superior colliculus

An important input to the PPRF and the RIP is the superior colliculus (Raybourn and Keller 1977). This is a midbrain structure, found at the same level as the oculomotor (CN III) nucleus. The superior colliculus represents visual space according to several integrated topographic maps. Superficially, it contains a retinotopic map, making use of the input it receives directly from the optic nerve (Schiller and Stryker 1972). Intermediate layers are thought to house a motor map, with each location corresponding to a potential saccadic target (Sparks 1986).

Deeper layers have maps that exhibit multisensory features, including somatosensation (Peck et al. 1993; Stein et al. 1989). Some accounts of collicular function propose that it contains a saliency map (Veale et al. 2017; Zelinsky and Bisley 2015), and mediates attention to salient locations. Attention here refers to planned or performed eye movements leading to foveation of the 'attended' location. This is a distinct process to attention as 'gain control' (Feldman and Friston 2010; Hillyard et al. 1998) of sensory streams (that does not necessarily depend upon oculomotor contingencies) (Parr and Friston 2019a). The colliculus receives an input from cortical layer V (Fries 1984). This layer-specific input is shared with other structures with a role in salience computations, including the basal ganglia (Shipp 2007) and the pulvinar nucleus of the thalamus (Shipp 2003). It is interesting that many of the areas implicated in attentional selection and salience conform to this laminar input pattern.

Neurons in the superior colliculus can be classified according to distinct electrophysiological profiles. Three broad categories of neurons are identifiable in this way. These are the collicular 'burst' neurons, the 'fixation' neurons, and the 'build-up' neurons (Ma et al. 1991; Munoz and Wurtz 1995a). The first of the three are found more dorsally, while the latter two are more ventral within the colliculus. 'Fixation' neurons are active during fixation, and are found at the rostral pole of the colliculus. These synapse on the 'omnipause' neurons of the nucleus raphe interpositus (Gandhi and Keller 1997), so that decreases in 'fixation' neuron activity causes a disinhibition of the PPRF 'burst' neurons, resulting in a saccade. The 'burst' neurons discharge immediately before a saccade, and the target location of the saccade corresponds to the location of these neurons in the colliculus. 'Build-up' neurons have a slowly increasing activity that terminates when a saccade occurs, although this activity is not always followed by a saccade. This observation is important in the context of the premotor theory of attention (Rizzolatti et al. 1987), as this theory suggests that covert attention may correspond to a planned saccade which does not take place. 'Build-up' neurons, as a population, have the interesting property that the activity across the population appears to travel as a 'hill' across the colliculus towards the rostral pole, which represents the foveal location (Munoz and Wurtz 1995b).

The notion of a travelling 'hill' of excitation corresponds well to a set of theoretical constructs known as attractor networks. Representations of states that evolve in metric space have been extensively modelled using continuous attractor networks (Zhang et al. 2008). One reason for emphasising this point is that, due to the serial nature of saccadic sampling, the apparent temporal continuity of visual experience requires explanation. The constraints placed upon a 'hill' of activity in a continuous attractor network mean that changing representation of one location in a metric space to another requires the transient representation of all intermediate locations. This enforces a form of memory, as the proximal future and past are heavily constrained by one another. This represents an imposition of prior beliefs on the interpretation

of sensory data, providing a simple example of a form of Bayesian inference (Pouget et al. 2013).

If the superior colliculus is unilaterally damaged, or pharmacologically inactivated, the frequency of saccades to the contralateral side of space is reduced (Schiller et al. 1987; Schiller et al. 1980). However, in the presence of intact frontal eye fields, collicular ablation does not permanently prevent the generation of voluntary saccades (Albano and Wurtz 1982). While this suggests that the frontal eye fields can make use of brainstem projections, which bypass the colliculus, reversible inactivation experiments indicate that the collicular route is the pathway used in structurally normal brains (Hikosaka and Wurtz 1985). The deficits following these pharmacological lesions resemble those observed in visual neglect, as one side of space appears to be neglected by the lesioned animals, in terms of both saccadic sampling, and covert attention (Lovejoy and Krauzlis 2010). The superior colliculus is rarely involved in lesions giving rise to neglect, but it is plausible that it is a component of the networks damaged in this syndrome – in the sense of a functional lesion or diaschisis (Corbetta and Shulman 2011; Price et al. 2001). This brings us to consider the nature of the inputs to the colliculus.

The basal ganglia

The substantia nigra pars reticulata (SNr) is an output nucleus of the basal ganglia located in the midbrain. It has a direct inhibitory, GABAergic, connection to the superior colliculus (Hikosaka and Wurtz 1983). This can be seen as a gate on the many direct cortical inputs to the colliculus, each of which identifies a different potential saccadic target. Consistent with this view is the observation that disruption of the SNr (Hikosaka and Wurtz 1985b), or its projections to the colliculus (Hikosaka and Wurtz 1985), increases the frequency of spontaneous saccades. The SNr receives a glutamatergic input from the subthalamic nucleus, a component of the indirect and hyperdirect pathways through the basal ganglia (Nambu 2004), and a GABAergic input from the D1 receptor expressing medium spiny neurons (MSNs) in the striatum, as part of the direct pathway. The striatum also contributes to the indirect pathway, as D2 receptor expressing MSNs inhibit the external part of the globus pallidus, thereby disinhibiting the subthalamic nucleus. The balance between the activity in the direct and indirect pathways is modulated by dopaminergic projections from the midbrain (Moss and Bolam 2008), which act to bias this balance in favour of the direct pathway. Activity in the direct pathway disinhibits the targets of the basal ganglia output nuclei, while the indirect pathway increases this inhibition (Freeze et al. 2013).

While visual neglect is often caused by cortical lesions, several subcortical regions have also been associated with the syndrome. The putamen, pulvinar, and caudate nucleus have all been associated with neglect (Karnath et al. 2002). These all communicate with cortical regions, such as the superior temporal gyrus, which, when damaged, can result in neglect. Changes in these regions have been observed following basal ganglia strokes which cause neglect (Karnath et al. 2005b). In addition to these observational data, animal studies have demonstrated that a neglect-like syndrome can be induced through manipulations at the level of the striatum. Unilateral infusions of MPTP, which is toxic to dopaminergic axons, have been shown to bias memory guided (Kori et al. 1995) and spontaneous (Kato et al. 1995) saccades towards the ipsilateral visual field. As the dopaminergic input to the striatum is also affected in Parkinson's disease, involvement of the basal ganglia plausibly explains the 'directional hypokinesia' component described in some forms of neglect (Mattingley et al. 1992). This is an impairment in initiating contralesional movements, more classically (but non-directionally) associated with Parkinson's disease. In neglect patients who have anterior or subcortical lesions, 'directional bradykinesia' has additionally been observed.

The cortical attention networks

The cortical regions that project directly to the superior colliculus include both frontal (Künzle and Akert 1977) and parietal (Gaymard et al. 2003) areas associated with the 'dorsal attentional network' (Corbetta et al. 2000; Szczepanski et al. 2013). This is a set of cortical regions which have been defined, using fMRI, on the basis of their signal changes during attentional tasks (Corbetta and Shulman 2002). The activity in these areas is largely bilateral (Hopfinger et al. 2000; Kastner et al. 1999), but asymmetries have been found for some tasks (Corbetta et al. 2002; Szczepanski et al. 2010). Interhemispheric differences in regions of the dorsal network have also been elicited through causal manipulations, including transcranial magnetic stimulation (Szczepanski and Kastner 2013), although the network as a whole was found to be approximately symmetrical. As might be expected for a region involved in directing eye movements, greater responses were found in the hemisphere contralateral to the visual field which was attended. These regions are connected by a white matter tract called the superior longitudinal fasciculus (SLF). The SLF is made up of three branches (Makris et al. 2004), and it is the first of these which connects the dorsal network of frontoparietal areas (Thiebaut de Schotten et al. 2011) (see Figure 3). The other two branches connect the regions of the 'ventral attention network' to each other, and connect the dorsal and ventral networks to one another.

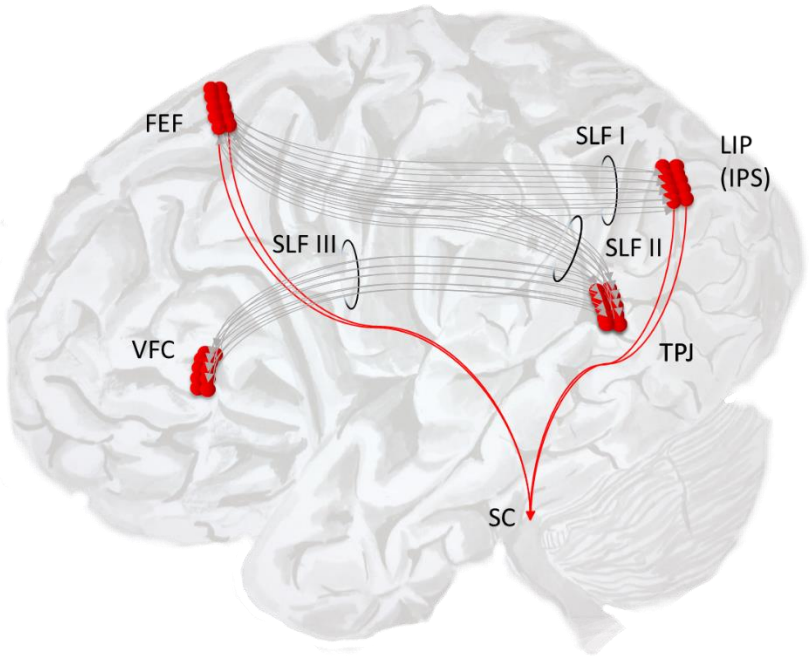

Figure 1.3 – The dorsal and ventral attentional networks. The dorsal and ventral networks each involve both frontal and parietal regions. The dorsal areas – including those in the region of the frontal eye fields (FEF), the lateral intraparietal (LIP) area and the intraparietal sulcus (IPS) – project to the superior colliculus (SC), suggesting a direct involvement of these areas in the control of eye movements. Note that these parietal areas are sometimes referred to as the parietal eye fields (Shipp 2004). These areas are connected by the first branch of the superior longitudinal fasciculus (SLF I). The ventral network is made up of areas in the ventral frontal cortex (VFC) and areas close to the temporoparietal junction (TPJ). These are connected by the third branch of the SLF (SLF III). SLF II connects the parietal part of the ventral network to the frontal part of the dorsal network. This schematic is based on the descriptions in (Corbetta and Shulman 2002) and in (Thiebaut de Schotten et al. 2011).

The premotor theory

The premotor theory of attention (Rizzolatti et al. 1987) draws evidence from these anatomical observations, as 'attentional' networks overlap substantially with those involved in eye movement control (Büchel et al. 1998; Corbetta et al. 1998; Nobre et al. 2000). The premise of this theory is that the allocation of (overt) attention to a given location is equivalent to making a saccade to that location. Attention can also be covertly directed to a location by

planning a saccade to it, even if this saccade is not performed. The behavioural evidence for this theory comes from eye tracking studies in which the deployment of covert attention has been shown to systematically alter the trajectory of saccades (Sheliga et al. 1994; Sheliga et al. 1995). Psychophysical measures are consistent with this, as stimulus discrimination is enhanced at saccade target locations compared to other visual field locations (Deubel and Schneider 1996). Further evidence comes from patients with palsies of the abducens (CN VI) nerve (see Figure 1.1). These injuries result in an inability to abduct the eye on the affected side. In a detection task, consistent with the premotor theory, these patients do not show the reduced reaction time characteristic of covert attention when the stimulus is placed in a location which is impossible for them to perform a saccade to (Craighero et al. 2001). Physiological evidence in favour of the theory is compelling. By stimulating frontal eye field neurons in the monkey, it is possible to cause saccadic eye movements. Subthreshold stimulation of these same cells increases detection performance of stimuli presented at the saccadic target location of those neurons (Moore and Fallah 2001). While not uncontroversial (Smith and Schenk 2012), the premotor theory highlights the important relationship between attention and eye movements, and the anatomical structures common to both.

Active inference

The question of how a salient location is selected as a (covert or overt) saccadic target has stimulated much theoretical study. Bayesian frameworks have been extensively employed to address this question, including definitions of salience, and surprise, in terms of information theoretic quantities (Itti and Baldi 2006; Itti and Koch 2000). More recently, this question has been formulated in terms of Active Inference (Friston et al. 2012a; Mirza et al. 2016). This is a theory derived from the principle that adaptive (living) systems must minimise the dispersion of their states in order to continue to exist in a meaningful way (Friston et al. 2006). A consequence of this theory is that organisms should sample (e.g. by performing a saccade to) the parts of the sensory environment that resolve most uncertainty about the causes of their sensations. In order to select the locations that best serve this process, they are equipped with a probabilistic model of how sensory data are generated, which includes beliefs about their own actions (Friston et al. 2012b). This is used to generate predictions about the sensations they will encounter. By performing an approximate Bayesian inversion of this model, given sensory data, organisms are able to infer their own optimal policy (sequence of actions). Optimal in this context means the active sampling of sensations that afford the greatest reduction in uncertainty or, equivalently, the greatest information gain. This is also known as

intrinsic value and, mathematically, is the expected Bayesian surprise that underwrites salience in the earlier formulations above (Itti and Baldi 2006; Itti and Koch 2000). A set of classical reflex arcs can then fulfil the predictions made under the implicit generative model (Friston et al. 2017a). A key aspect of this Bayes optimal, epistemic, uncertainty resolving formulation implies that the best saccade is selected from representations of all possible saccades, according to their salience or epistemic value. In turn, this implies the existence of a salience map; where the epistemic values of all possible saccade locations are evaluated. This may provide a complementary perspective on the attractor dynamics discussed above as models of activity in the deep layers of the superior colliculus; namely, an encoding of salience.

The anatomy of visual neglect

In visual neglect patients, cortical lesions can induce a lateral bias in the saccadic sampling of a scene. Typically, the frequency of saccades to the right side of space is increased, compared to the left. This appears to be related to the selection of saccadic targets, rather than an impairment in the production of saccades to the neglected hemifield (Bartolomeo and Chokron 2002). Intuitively, one might expect the lesion sites to correspond to the dorsal frontal and parietal regions directly involved in saccadic control. However, although cortical lesions associated with neglect can occur in both frontal and parietal regions, they are typically more ventral than the frontal eye fields or the intraparietal sulcus (Corbetta and Shulman 2002). A neglect-like syndrome can be elicited by lesioning the frontal eye fields (Latto and Cowey 1971), but this is only temporary. Additionally, as noted above, the ‘dorsal attentional network’ is symmetrically distributed. This contrasts with the observation that spatial hemineglect is much more common following a right hemispheric lesion. While the behavioural correlates render it unlikely that cortically driven neglect precludes no dysfunction of the dorsal network, the above observations indicate that this is likely to be secondary to the disruption of other structures.

The lateral biasing of saccadic movements in neglect can be reconciled with the fact that the cortical inputs to the superior colliculus are often preserved. The more ventral frontoparietal regions which are associated with neglect overlap with the ‘ventral attentional network’ (Corbetta and Shulman 2002; Corbetta and Shulman 2011). In contrast to the dorsal network, the ventral network is more prominent in the right hemisphere, consistent with the greater frequency of spatial neglect following right hemispheric lesions. These regions are connected by the third branch of the SLF, which is known to have a greater volume in the right hemisphere (Thiebaut de Schotten et al. 2011). The ventral parietal regions of this network are

connected to the frontal regions of the dorsal network by the second branch of the SLF. This means that the ventral network directly influences the cortical sites that project to the saccade generating areas of the brainstem.

The second branch of the SLF has been associated with some interesting lateralised behavioural correlates. In normal subjects, under certain conditions, a 'pseudo-neglect' can be elicited (Bowers and Heilman 1980; Jewell and McCourt 2000). This has been shown for a line bisection task, also used to assess hemineglect, in which a subject marks what they believe to be the midpoint of a horizontal line. While hemineglect patients typically mark to the right of the midline, small deviations to the left can occur in healthy subjects. The degree to which this 'pseudo-neglect' occurs is related to the volume of the right SLF II. The larger this is, the greater the leftward deviation (Thiebaut de Schotten et al. 2011). It has been proposed that neglect represents a disconnection syndrome, in which the frontoparietal interactions mediated by the SLF have been disrupted (Bartolomeo et al. 2007; He et al. 2007). This structurally motivated hypothesis complements the functionally motivated suggestion that an interaction between the dorsal and ventral networks is necessary for normal attentional function (Corbetta and Shulman 2002). There is some evidence for this from lesion studies. For example, one study looking at lesion overlaps between patients found maximal subcortical overlaps in the SLF (Doricchi and Tomaiuolo 2003). Case reports (Ciaraffa et al. 2013) endorse this finding, which is further strengthened by the observation that SLF II damage is a good predictor of hemineglect (Lunven et al. 2015; Thiebaut de Schotten et al. 2014). In addition to this, inactivation of the right SLF by electrical stimulation during surgery caused a temporary rightward deviation in the line bisection task (Thiebaut de Schotten et al. 2005).

Dorsal and ventral

The distinction between the dorsal and ventral networks mirrors the distinction between the dorsal and ventral visual pathways (Goodale and Milner 1992). These are often referred to as the 'what' and 'where' visual pathways, as the former appears to represent stimulus identity, while the latter represents stimulus location (Ungerleider and Haxby 1994). Given that an object retains its identity, regardless of its position in space, the brain appears to have treated these as independent factors. In probabilistic inference, this is referred to as a 'mean field approximation' (Friston and Buzsáki 2016). If the dorsal and ventral attention networks represent a similar factorisation, this could provide an intuitive explanation for the lateralisation of the latter network, and the symmetry of the former. Each hemisphere is thought to contain maps of the contralateral side of space (Wandell et al. 2007). It is

unsurprising then that more dorsal regions, associated with the 'where' pathway, are relatively symmetrical. However, stimulus identity does not require representation in a specific location, due to the factorisation of these variables. As such, a unilateral representation is sufficient for the 'what' stream. This is consistent with clinical neuropsychological observations, as lesions to regions in the right ventral visual pathway are can give rise to disorders of object recognition (Warrington and James 1967; Warrington and James 1988; Warrington and Taylor 1973), while the homologous regions on the left are more likely to be associated with difficulty naming objects (Kirshner 2003). This could explain the lateralisation of the ventral network and, given its influence over the dorsal network, is consistent with the higher prevalence of spatial neglect among patients with right hemispheric lesions. The connection between the two networks would be mandated by the need to direct the eyes to different locations to resolve uncertainty about a stimulus or scene identity. Note that a popular alternative explanation for this pathological asymmetry is that the right hemisphere represents both left and right sides of space, while the left represents only the right side (Mesulam 1999).

Working memory and temporal continuity

As has been emphasised above, saccadic eye movements involve sampling of locations in a serial and discrete fashion. The frequency of spontaneous saccades is about 2-3 Hz (Büttner and Büttner-Ennever 2006), but clearly we do not reset our beliefs about a visual scene at this frequency. In order to construct a temporally continuous representation of the visual world, it is clear that some form of short term memory must be involved, so that the information obtained at one fixation carries over – or is assimilated – into the next. Broadly, there are two mechanisms that allow the temporary storage of information in the brain. These are sustained neuronal activity (Goldman-Rakic 1995), and short term changes in synaptic efficacy (Mongillo et al. 2008). In Bayesian approaches to understanding brain function, these two mechanisms correspond inference and learning respectively; namely, updating beliefs (approximate posterior distributions) about hidden states of the world, and parameters (generative model) that describe the probabilistic relationships between hidden states (Friston et al. 2016b).

Memory as sustained neuronal activity

Sustained neuronal activity has been extensively studied in the context of 'delay-period' activity (Goldman-Rakic 1995). This is the increase in firing rate observed in some neurons, which persists even after the stimulus that evoked the increase is no longer present. 'Delay-period' working memory tasks during single unit recordings have been used to demonstrate this phenomenon (Funahashi 2015). An example of such a task is an oculomotor delay task, in which an animal fixates a location on a screen. A stimulus is presented which indicates a saccadic target. During a delay, in which no stimulus is present, the animal must remember the target location. When instructed, they should perform a saccade to that location. From the presentation of the stimulus, until the performance of the saccade, neurons in the principal sulcus of the prefrontal cortex remain persistently active (Funahashi et al. 1989). Among these neurons, many are tuned to the eventual saccade direction. Other parts of the frontal cortex have been shown to contain populations of neurons that exhibit similar properties for other planned actions (Cisek and Kalaska 2005). The relationship between these forms of memory and planned actions have prompted some authors (Frank et al. 2001; Hikosaka et al. 2000) to suggest that the raison d'être of working memory is in evaluating future actions. This complements work on decision processes in the field of artificial intelligence (Kaelbling et al. 1998), in which memory serves a similar purpose. There is an attractive circularity to the notion that the temporal continuity of visual experience is due to the use of memories from past saccades to evaluate potential future saccades.

Single unit recordings have demonstrated that there are neurons with responses limited to the duration of a stimulus presentation (Hubel and Wiesel 1959), and also those which have responses that transcend this time scale (Funahashi et al. 1989). This speaks to a temporal hierarchy (Cocchi et al. 2016; Hasson et al. 2008; Kiebel et al. 2008; Murray et al. 2014) in the brain, with different neurons representing different rates of environmental change. Temporal responses in different areas of the brain have been shown (Hasson et al. 2015; Hasson et al. 2008; Honey et al. 2012; Murray et al. 2014) to map closely to the hierarchical structure of the cortex as derived from studies of laminar connectivity (Felleman and Van Essen 1991; Zeki and Shipp 1988). This is consistent with the idea that the brain contains a hierarchical generative model (Friston 2008) of a temporally structured environment, and allows for slowly changing contexts to inform the evolution of states which change over a faster time scale. Under this view, working memory, in the form of persistent neuronal activity, corresponds to a process of evidence accumulation over multiple timescales.

As mentioned above in the context of the superior colliculus, sustained activity patterns have been extensively modelled using continuous attractor networks. Working memory has not escaped this treatment (Compte 2006; Wimmer et al. 2014). While many accounts of working memory focus on prefrontal regions, such networks have been used to model activity in many

different brain regions, including those for brainstem oculomotor control (Seung 1998), navigational regions (Redish et al. 1996; Zhang 1996), and motor planning (Georgopoulos et al. 1982; Lukashin et al. 1996). Given the computational nature of these architectures, all could be described as implementing a form of working memory. All involve a sustained representation, which is updated as new observations are made. However, these memories have different temporal properties, depending on the rate of change of what they represent, and so may not be sustained over the time course associated with the classical notion of working memory. Notably, it is areas considered high in the anatomical (and consequently temporal) hierarchy (Felleman and Van Essen 1991), such as the dorsolateral prefrontal cortex (Goldman-Rakic 1987; Kojima et al. 1982) and hippocampus (Chadwick et al. 2010; Squire et al. 2004), which are often thought to perform working memory functions.

Memory as short-term plasticity

For some situations requiring working memory, persistent activation of neurons is an inefficient way to store temporary information. This is due to the number of dimensions required for some memories, and the metabolic constraints (Lennie 2003) on the number of neurons required to represent these. To build some intuition for this point, consider the example (shown graphically in Figure 4) of a cancellation task. Variants of these tasks are frequently used both clinically (Albert 1973; Fullerton et al. 1986) and experimentally (Husain et al. 2001; Malhotra et al. 2004; Mannan et al. 2005) to assess spatial neglect. Subjects are shown an array of targets, and are asked to cancel each target once, and only once. Cancellation may involve marking the target with a pencil, or clicking on it in a computer display. In the latter set up, there need not be a visible marker alerting the subject that they have previously cancelled it. Despite this, there is a relatively low rate of re-cancellation of a stimulus in healthy subjects (Mannan et al. 2005), showing that cancelled locations are remembered. If this task were performed using a set of possible locations on an 8x8 grid, there would be 64 possible target locations. For each of these, there are 3 possible states: no target, target, and cancelled target. To be able to represent beliefs about the state at each location as persistent activity in populations of neurons, it would be necessary to employ 64 x 3 = 192 computational units, and to maintain activity patterns across all of these simultaneously. In many natural scenes, the number of locations, and possible stimuli at each location, is clearly much greater than this, and would require huge numbers of neurons if remembered in this manner.

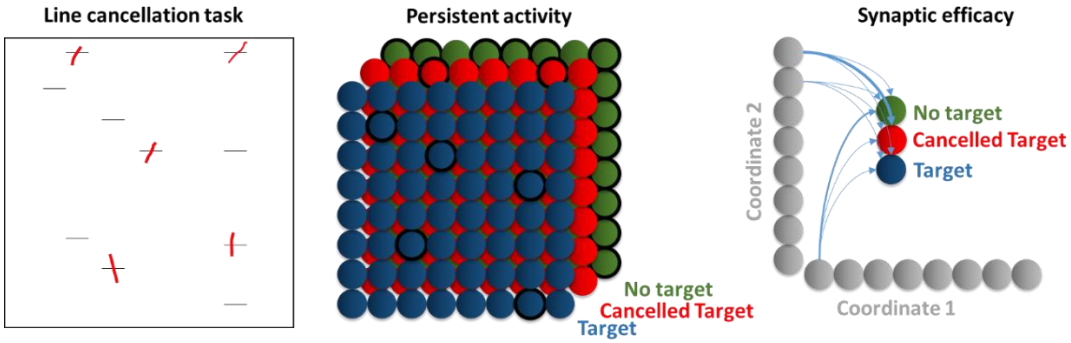

Figure 1.4 – Mechanisms of memory. *On the left,* an example of a line cancellation task is shown. The subject is presented with a sheet of paper with a set of horizontal lines and is asked to cancel (red marks) each of these lines. *The middle panel* shows the set of 192 neurons which would be required to represent the subject's beliefs about where the lines are, and whether they have cancelled them, if the memory of previously visited locations were stored in terms of persistent activity in a neuronal population. The currently active neurons are represented by a black outline. *The panel on the right* shows a more efficient way to represent this information, in terms of a mapping from a representation of space to representations of each of the possible observations that could be made on visiting a particular location. Clearly it is more efficient to make use of synaptic efficacy when storing temporary, high dimensional, memories. In short, synaptic efficacy represents probabilistic mappings (i.e., 'if I were to look there, I would see that') as opposed to beliefs about the current state of the world (i.e., 'I am looking there' or 'seeing that') encoded by synaptic activity.

Contrast this with a memory system in which information is stored in the interactions between different neurons (i.e. synaptically). In this case, it is only necessary to employ 3 computational units to represent the state at each location. Each location unit can then represent its current state as an interaction between itself and the three alternative states. For example, on viewing a target for the first time, the synapses between the unit representing the location and that representing the presence of a target can be potentiated. This reduces the need for 192 neuronal populations to 67; a number which can be further reduced to 19 using a factorised representation of location (i.e. a coordinate system) in place of explicit representations of each location. This simple example demonstrates that, while low dimensional memories can be stored as persistent activity, synaptic updates are a much more efficient way to store higher dimensional representations. This might explain why some working memory tasks have failed to show a clear relationship between working memory deficits and re-cancellation rates in neglect patients (Wansard et al. 2014). There may be impairment in (short-term) synaptic plasticity, which would not be detected by probing with a delay-period type task. We will

return to this idea in Chapter 5, where we will illustrate how impairments in synaptic plasticity (e.g. due to disconnection of the neurons either side of that synapse) alter the optimal, uncertainty resolving, saccadic policies one could engage in.

Short term plasticity may be due to several mechanisms, but calcium dependent processes clearly play a substantial role. In a presynaptic neuron, an increase in calcium ion concentration, as a result of an action potential, triggers vesicular release. With repeated action potentials, intracellular calcium buffers can become saturated (Blatow et al. 2003; Deng and Klyachko 2011), ensuring that the increase in calcium at the next action potential will be greater. This means that the synapse is temporarily potentiated. Pre and postsynaptic mechanisms have been used to explain the opposite phenomenon, in which there is a temporary depression of the synapse. Changes in plasticity over very short time scales, such as these, have been described in neurons in the prefrontal cortex (Hempel et al. 2000; Wang et al. 2006). Computational studies (Barak et al. 2010; Mongillo et al. 2008) have demonstrated that dynamics such as these could account for some working memory phenomena.

Spatial neglect provides some clues as to the anatomical regions that may be involved in this kind of short term plasticity for spatial memories (Mannan et al. 2005). For patients with lesions of the intraparietal sulcus, the probability of re-cancellation of a target increases with time. In contrast, lesions of the inferior frontal regions give a constant increased re-cancellation probability. Although both regions are related to the attentional networks, these results suggest distinct mechanisms of neglect following each lesion. The former appears to be memory dependent, while the latter does not. This hints at the importance of axons in the region of intraparietal sulcus. These connections could furnish the candidate synapses that store spatial memories through short term plastic changes. Consistent with this, patients with neglect who have a more severe spatial working memory deficit have been reported to have parietal white matter lesions not found in those with who have neglect but relatively intact spatial working memory (Malhotra et al. 2005).

Conclusion

The neuroanatomical system which supports the interrogation of a visual scene includes a complex network of brainstem areas under the influence of cortical and subcortical structures. Damage to almost any component of this system can cause a neglect syndrome, emphasising their important roles in visual experience. The mnemonic properties of many of these components have been highlighted, as these allow information from the past to be integrated

into representations of the present and future. In other words, posterior beliefs following one observation become prior beliefs about the causes of the next. The updating of this form of working memory on the basis of new observations is necessarily a Bayesian (belief updating) process, likely involving a factorisation of variables, such that ‘what’ and ‘where’ are represented independently. This is consistent with the dorsal and ventral streams hypothesis, and the anatomy of the attentional networks, which provide a cortical influence over eye movements. In doing so, hypotheses derived from past experience are combined with new sensory data to construct visual percepts.

2 – Neuronal message passing and active inference

Introduction

Recent advances in theoretical neurobiology rest upon the idea that the brain uses an internal (generative) model of its environment to try to explain the causes of its sensations. This involves combining prior beliefs about the world with beliefs about how sensations are generated (Doya 2007; Knill and Pouget 2004). The process of computing the most probable explanation for the data at hand is known as Bayesian inference, and optimises a quantity known as Bayesian model evidence (also known as the *marginal likelihood*, or *negative surprisal*). Model evidence quantifies how probable data are under a particular model of how they were generated. The drive to maximise model evidence, sometimes referred to as 'self-evidencing' (Hohwy 2016), may be motivated in a number of ways (Friston 2013). Perhaps the simplest, from a physiological perspective, is to see this as a generalisation of the principle of homeostasis (Cannon 1929). While normally applied to interoceptive data (temperature, blood pressure, acidity, etc.), homeostasis expresses the idea that there is an allowable distribution within which physiological parameters should be maintained. Deviations from these distributions elicit corrective actions (autonomic reflexes) to reverse the deviation. Interpreting a distribution over interoceptive data as representing the probability of those data under some model, we can think of homeostasis as expressing the imperative to maximise the evidence for this model. In the next few sections, we unpack this idea, extending it to proprioceptive and exteroceptive modalities. We outline the architectures of generative models the nervous system could employ, and the implications self-evidencing has for the structure and function of neuronal networks.

This chapter[3] outlines the methods used in subsequent chapters, which each employ different sorts of generative model, while appealing to the same (variational) principles. In preparation for Chapter 3, we outline the generic form of the discrete state-space models that we will use to simulate active visual sampling to reduce uncertainty. We highlight the inference scheme obtained through minimising a free energy functional (used throughout this thesis), and compare this numerically to other established schemes. We additionally illustrate the form of the continuous state-space models that underwrite predictive coding (Friston and Kiebel 2009; Rao and Ballard 1999) and will be leveraged in understanding brainstem oculomotor control and its pathologies. These are extended in Chapter 4, to address neuromodulatory mechanisms.

[3] Some of the material used in this chapter was initially published in (Parr et al. 2019b)

The same mathematical tools are used in Chapter 5 in a more pragmatic way, to draw inferences about the networks giving rise to measured magnetoencephalography data using a dynamic causal modelling approach (Friston et al. 2017d).

Free energy

As outlined above, we can think of action as a process of evidence maximisation, regulating our environment by correcting deviations in sensor values (y) from some optimal distribution. As these values will depend upon things that cannot directly be observed (e.g. the pattern of photoreceptor activity on the retina depends upon the position of a light source). These variables are referred to as *hidden states* (x), and must be integrated (or marginalised) out from a joint probability distribution to obtain the model evidence:

$$\ln p(y) = \ln \int p(y, x) dx \tag{2.1}$$

In most practical contexts, this integration is either computationally or analytically intractable. However, it is possible to express a lower bound on the evidence by introducing an arbitrary probability density (q) that expresses beliefs (in the sense of Bayesian belief updating) about the hidden variables. This is referred to as a *variational distribution* or an *approximate posterior* (for reasons that will become clear below). This lets us write down a free energy (F) functional of these beliefs that is always greater than or equal to the negative log evidence (Beal 2003; Dayan et al. 1995):

$$\begin{aligned} F[q, y] &= E_q[\ln q(x) - \ln p(y, x)] \\ &= D_{KL}[q(x) \,||\, p(x \mid y)] - \ln p(y) \\ &\geq -\ln p(y) \end{aligned} \tag{2.2}$$

The KL-Divergence (expected difference between two log probability densities) in the second line is always greater than or equal to zero, with equality when both densities are equal. This means the bound on the evidence becomes tighter as the variational distribution becomes a better approximation to the posterior density ($p(x \mid y)$). In summary, once free energy has been

minimised by changing the variational distribution, acting to change sensory inputs such that reducing free energy is equivalent to self-evidencing. Active Inference can then be succinctly articulated as the process of minimising free energy through optimising beliefs (perception) and sensory data (action). Note that this process depends upon the form of the joint distribution ($p(y, x)$), referred to as a *generative model*. The next sections will outline the generic structure of the generative models that will be employed throughout the rest of this thesis.

Generative models

This thesis employs two broad types of generative model (Friston et al. 2017c). The first of these is defined in continuous time (Friston et al. 2010a), while the latter is in discrete time (Friston et al. 2012b). Continuous time models are important in interfacing with the physical world, where sensory receptors communicate continuous data (e.g. luminance, pressure, temperature) and effectors (muscles or glands) cause continuous changes (muscle length, chemical concentration). A general way of expressing the dynamics of a continuous time model is through a stochastic differential equation describing the flow of a hidden (unobserved) variable. This is equipped with a second stochastic equation expressing the generation of data from the hidden variables:

$$\begin{aligned} \tilde{y} &= \tilde{g}(\tilde{x}, \tilde{v}) + \tilde{\omega}_y \\ D\tilde{x} &= \tilde{f}(\tilde{x}, \tilde{v}) + \tilde{\omega}_x \\ &\Rightarrow \\ p(\tilde{y} \mid \tilde{x}, \tilde{v}) &= \mathcal{N}(\tilde{g}, \tilde{\Pi}_y) \\ p(\tilde{x} \mid \tilde{v}) &= \mathcal{N}(D \cdot \tilde{f}, \tilde{\Pi}_x) \\ p(\tilde{v}) &= \mathcal{N}(\tilde{\eta}, \tilde{\Pi}_v) \end{aligned} \tag{2.3}$$

The first two equalities indicate how hidden states (x) generate data (y), and how hidden states evolve over time. These depend (respectively) upon functions (g and f) that take into account a second hidden variable (v); sometimes referred to as a 'hidden cause'. In hierarchical generative models, this variable links together different levels (which normally operate over different time scales). In this setting, the hidden causes are generated by higher levels of a model just as the data are generated by the single level considered here. The tilde (~) symbol

indicates a trajectory, here expressed in terms of generalised coordinates of motion[4]. These are vectors whose first element is the position of the state, second element is its velocity, third is acceleration, and so on. These can be used to reconstruct a trajectory by treating them as the coefficients of a Taylor series expansion[5] of $x(\tau)$. The D matrix is a derivative operator, expressed in terms of a matrix with ones above the leading diagonal. This shifts the elements of the generalised motion vectors up by one, such that each element of the resulting vector is the first temporal derivative of the original element. When transposed (note that we use the dot-product notation $a \cdot b \triangleq a^T b$), this derivative operator instead shifts all elements down by one. The random fluctuations (ω) have a Gaussian form, enabling the expression of the stochastic equations in terms of their associated probability densities. The precisions (inverse variances) of these generalised quantities are constructed through use of an autocorrelation function that determines the smoothness of the fluctuations (Cox and Miller 1965). This means the generative models described using these equations are not constrained by Weiner assumptions about the structure of the fluctuations. Combining the distributions above, we can express the joint distribution of the variables in the generative model:

$$p(\tilde{y}, \tilde{x}, \tilde{v}) = p(\tilde{y} \mid \tilde{x}, \tilde{v})\, p(\tilde{x} \mid \tilde{v})\, p(\tilde{v}) \qquad \text{(2.4}$$

Note the factorisation of this density into a likelihood, a dynamical term, and a prior. Figure 2.1 illustrates this generative model in factor-graph form (Dauwels 2007; de Vries and Friston 2017; Forney Jr and Vontobel 2011; Laar and Vries 2016; Loeliger 2004; Loeliger et al. 2007). The lower part of this figure also depicts the inferential message passing implied by free energy minimisation for this generative model. This will be unpacked in detail in the next section.

[4] $\tilde{x} \triangleq [x \quad x' \quad x'' \quad \cdots]^T = [x^{[0]} \quad x^{[1]} \quad x^{[2]} \quad \cdots]^T$

[5] $x(\tau) = x_0 + \tau x_0' + \frac{1}{2}\tau^2 x_0'' + \ldots$

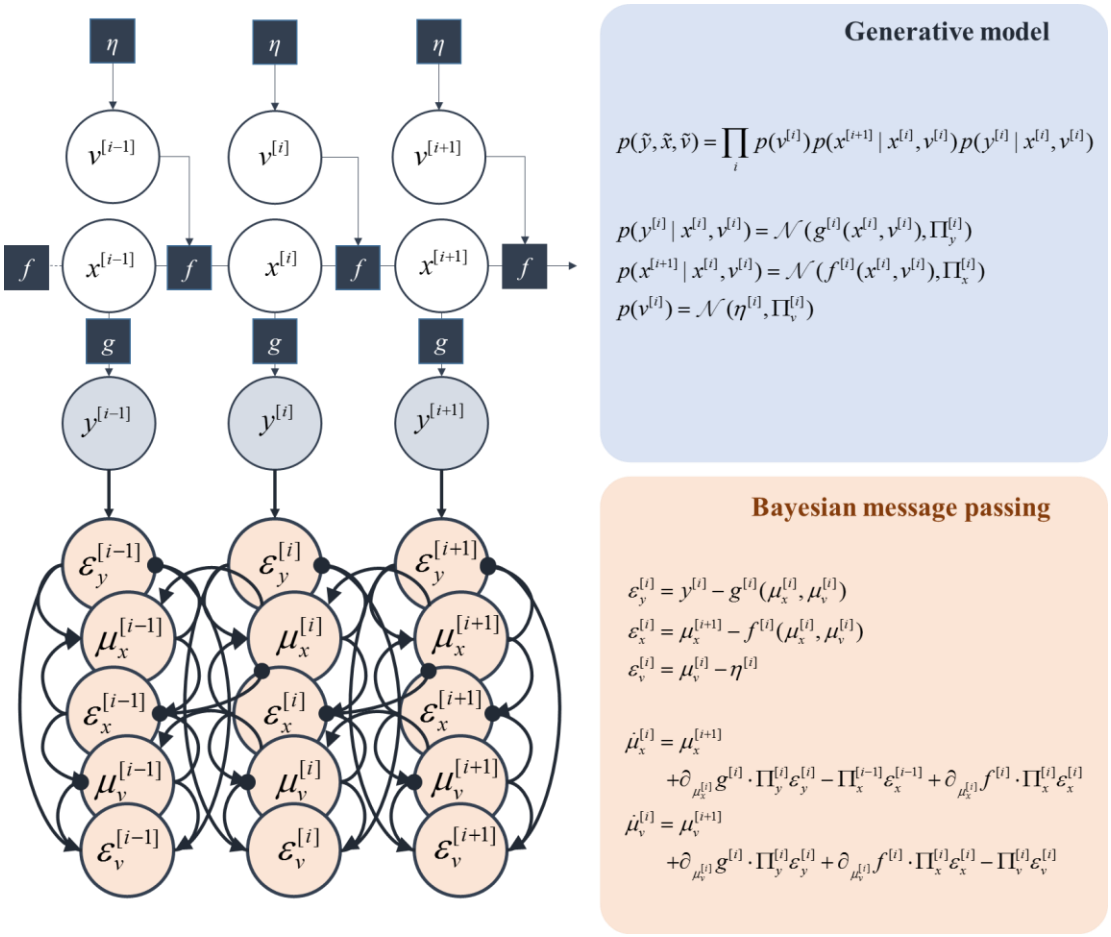

Figure 2.1 – Continuous state-space generative model. *The upper (blue) part* of this figure specifies a continuous-time dynamic model both graphically and in terms of its associated probability distributions (blue panel). Circles indicate random variables, with filled blue circles representing observable data. Arrows from one circle to another indicate that the variable in the second circle is conditionally dependent upon the first. The square nodes indicate the probability distributions (specified in the blue panel) that mediate these dependencies. *The lower (pink) part* of this figure shows the Bayesian message passing (or filtering) scheme that can be used to draw inferences about the variables in the generative model above. This is expressed in terms of prediction errors (ε) and expectations or modes (μ). While these equations look a little complicated, they are obtained simply by setting the rate of change of the posterior mode to be the negative free energy gradient with respect to this mode.

Before outlining the inversion of these models to explain sensory data, we first outline the analogous generative model used to account for discrete-time (i.e. sequential) dynamics, of the sort associated with planning. This is a (partially observed) Markov decision process (MDP), and can be expressed in terms of a set of categorical probability distributions:

$$
\begin{aligned}
P(o_\tau^m \mid s_\tau) &= Cat(\mathbf{A}^m) \\
P(s_{\tau+1}^n \mid s_\tau^n, \pi) &= Cat(\mathbf{B}_\pi^n) \\
P(o_\tau^m) &= Cat(\mathbf{C}^m) \\
P(s_1^n) &= Cat(\mathbf{D}^n) \\
P(\pi) &= Cat(\boldsymbol{\pi}_0)
\end{aligned}
\tag{2.5}
$$

To distinguish the categorical variables used here from the continuous variables (*x*, *y*, *v*) above, we use *o* to indicate an observable outcome (i.e. data), *s* to indicate hidden states, and π to indicate a policy (sequence of actions). The superscripts *m* and *n* indicate different outcome modalities (e.g. vision and proprioception) or hidden state factors (e.g. what and where) respectively. The bold letters are the sufficient statistics (vectors and matrices) of the categorical distributions. For example, a column of the **A**-matrix represents the probability, given the state indicated by that column, of each possible outcome (row). Similarly, **B** specifies, for each policy, the probability of each state transition (from column to row). The remaining distributions have sufficient statistics that are all vectors; where each element provides the probability of an alternative value for the associated random variable. The probability distribution for policies is a little different to the others, as $\boldsymbol{\pi}_0$ must be computed very carefully. This form is specified in Figure 2.2, and we will unpack this in full in the *Active perception* section in this chapter. As in Equation 2.4, we can express the joint distribution over these variables simply by multiplying together these factors:

$$
P(\tilde{o}, \tilde{s}, \pi) = P(\tilde{o} \mid \tilde{s}) P(\tilde{s} \mid \pi) P(\pi) \tag{2.6}
$$

As before, the tilde (~) indicates a trajectory. However, this is now expressed more simply as a sequence of values through time. Figure 2.2 illustrates the MDP model outlined above in graphical form, showing a sequence of states evolving through time, depending upon the policy selected. The states at each time generate an observation. This has a very similar form to the model expressed in Figure 2.1. Note that one key difference is that the same policy variable is used to influence transitions over multiple time-steps. This emphasises the importance of MDP models in planning trajectories through time.

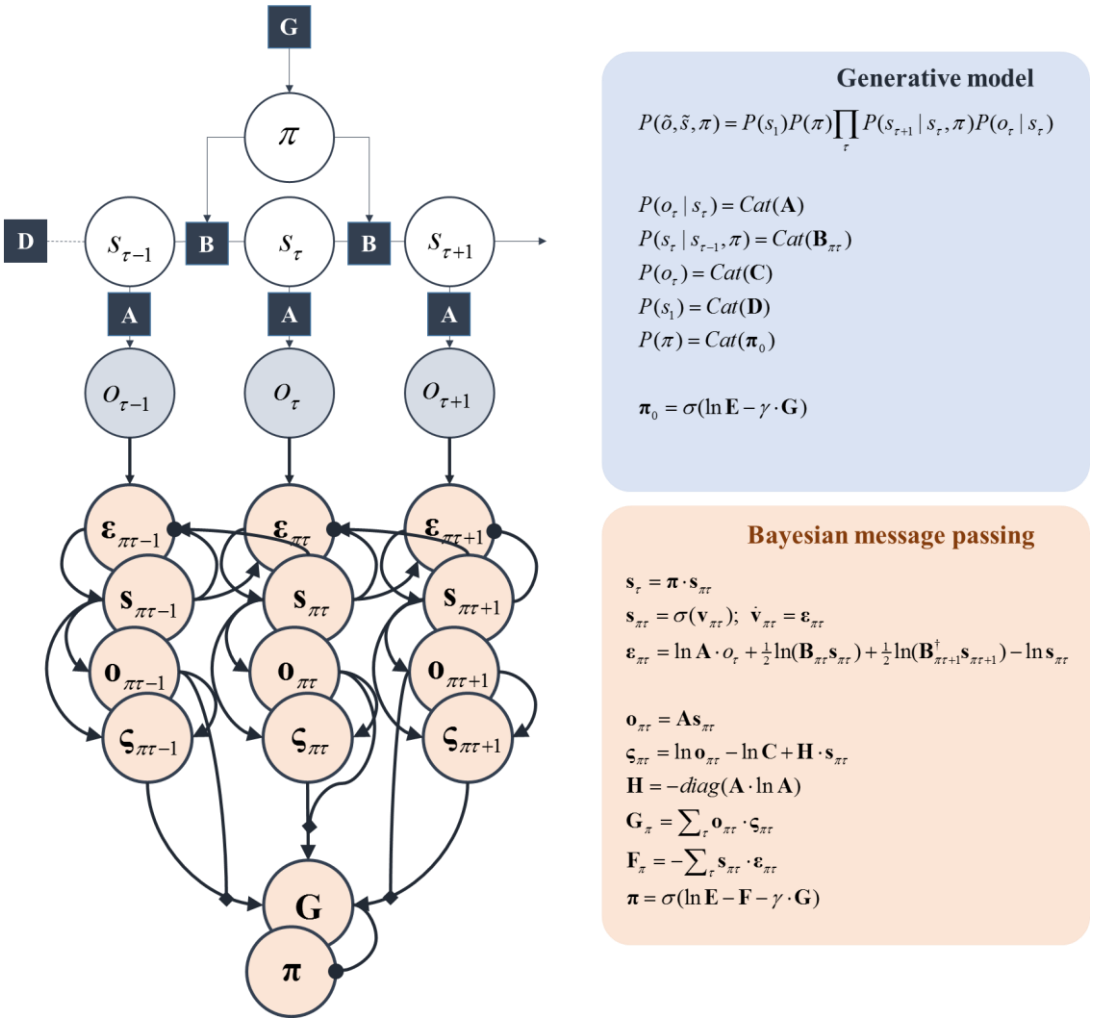

Figure 2.2 – Discrete state-space generative model. This figure adopts the same format as Figure 2.1, with a factor graph representation of a partially observed Markov decision process in blue, and the marginal message passing scheme used to draw inferences about this in pink. The softmax (σ) function is a normalised exponential that converts a log probability into a probability. The lower part of this figure makes use of expectations about states (**s**), predictive distributions over outcomes (**o**), errors (**ε**, **ς**), and an auxiliary variable (**v**) that plays the role of a membrane potential, or an un-normalised log probability. Several additional intermediate quantities are computed, including a conditional entropy (**H**), a bias in policy priors (**E**), a vector of free energies for each policy (**F**), and the expected free energy (**G**) for each policy, weighted by a precision, or inverse temperature, parameter (γ).

Variational inference

In this section, we take the generative models and variational principles outlined above, and illustrate how these may be used to derive Bayesian inferential schemes. This amounts to being able to write down the appropriate form of the free energy for a given generative model, and then performing a gradient descent. The sparsity of the generative models outlined above[6], and the factorisation this entails, is practically very useful. Because the factors of the model define local relationships across the graph, they can be used to define local update rules that ensure the free energy of the whole graph may be minimised without ever needing explicit computation at a global level. This fact has been exploited to derive a range of inferential message passing schemes (Minka 2005; Winn and Bishop 2005; Yedidia et al. 2005), two of which are used throughout this thesis. Biologically, this is crucial, given the local coupling through synaptic connections that underwrites computation in neural systems. In the following sections, we offer an overview of Bayesian filtering, used in inference for models defined in continuous time, and a marginal message passing scheme used to draw inferences about discrete time models. We then illustrate the performance of the latter in relation to other established Bayesian message passing schemes.

Continuous time

In this section, we start by writing out the free energy associated with the continuous time generative model outlined above. We follow the approach of (Buckley et al. 2017; Friston et al. 2010a), and show that a gradient descent on this quantity gives rise to a message passing scheme that can be used to perform inferences on time-series data of the sort reported by sensory receptors in biological systems. The free energy for the model expressed in Equation 2.4 is:

[6] Sparsity here means that each variable is not directly dependent upon all other variables. This means the joint distribution may be factorised into a set of conditional distributions, each of which involves only a subset of the variables in the entire model.

$$
\begin{aligned}
F[\tilde{\mu},\tilde{y}] &= E_q[\ln q(\tilde{x},\tilde{v}) - \ln p(\tilde{y},\tilde{x},\tilde{v})] \\
&\approx \underbrace{-\tfrac{1}{2}\ln 2\pi|C|}_{E_q[\ln q(\tilde{x},\tilde{v})]} + \underbrace{\tfrac{1}{2}\tilde{\varepsilon}\cdot\tilde{\Pi}\tilde{\varepsilon} + \tfrac{1}{2}\ln 2\pi|\tilde{\Pi}|}_{-E_q[\ln p(\tilde{y},\tilde{\mu})]} \underbrace{-\tfrac{1}{2}tr(C\partial^2_{\tilde{\mu}}\ln p(\tilde{y},\tilde{\mu}))}_{-\frac{1}{2}E_q[\Delta\tilde{\mu}\cdot\partial^2_{\tilde{\mu}}\ln p(\tilde{y},\tilde{\mu})\Delta\tilde{\mu}]}
\end{aligned}
\qquad (2.6
$$

$$
\tilde{\varepsilon} \triangleq \begin{bmatrix} \tilde{\varepsilon}_y \\ \tilde{\varepsilon}_x \\ \tilde{\varepsilon}_v \end{bmatrix} = \begin{bmatrix} \tilde{y} - \tilde{g}(\tilde{\mu}_x,\tilde{\mu}_v) \\ D\tilde{\mu}_x - f(\tilde{\mu}_x,\tilde{\mu}_v) \\ \tilde{\mu}_v - \tilde{\eta} \end{bmatrix}
$$

$$
\tilde{\Pi} \triangleq \begin{bmatrix} \tilde{\Pi}_y & & \\ & \tilde{\Pi}_x & \\ & & \tilde{\Pi}_v \end{bmatrix}
$$

The second line assumes (the Laplace assumption) a Gaussian form for the variational posterior density[7] (see Appendix A.1) with mode μ and covariance C, such that the first term is the entropy of a Gaussian distribution. The second, third, and fourth terms come from (the expectation of) a second order (quadratic) Taylor series expansion of the log joint probability around the mode of the variational posterior[8]. The form of the free energy here has a useful consequence. Taking the derivative with respect to the posterior covariance, we find that the covariance that minimises the free energy can be expressed in terms of an analytic function of the posterior mean:

$$
\begin{aligned}
&\partial_C F = -\tfrac{1}{2}C^{-1} - \tfrac{1}{2}\partial_{\tilde{\mu}}\ln p(\tilde{y},\tilde{\mu}) = 0 \Leftrightarrow \\
&C^{-1} = -\partial^2_{\tilde{\mu}}\ln p(\tilde{y},\tilde{\mu})
\end{aligned}
\qquad (2.7
$$

Substituting this into Equation 2.6 turns the final term into a constant. In addition (due to the quadratic approximation) the first and third terms (representing curvatures) are constant with respect to the expectation. We can now express inference in terms of a gradient descent on the free energy with respect to the posterior mean, computing the covariance using Equation 2.7:

[7] $p(x,v \mid y) \approx q(x,v) = \mathcal{N}(\mu, C^{-1})$

[8] Note that the linear term of the expansion vanishes under expectation, as the expected difference between a random variable and its mode is zero under Gaussian assumptions.

$$
\begin{aligned}
&\dot{\tilde{\mu}} = D\tilde{\mu} - \nabla_{\tilde{\mu}}\tilde{\varepsilon}\cdot\tilde{\Pi}\tilde{\varepsilon} \\
&\Rightarrow \\
&\dot{\tilde{\mu}}_x - D\tilde{\mu}_x = \nabla_{\tilde{\mu}_x}\tilde{g}\cdot\tilde{\Pi}_y\tilde{\varepsilon}_y - D\cdot\tilde{\Pi}_x\tilde{\varepsilon}_x + \nabla_{\tilde{\mu}_x}\tilde{f}\cdot\tilde{\Pi}_x\tilde{\varepsilon}_x \\
&\dot{\tilde{\mu}}_v - D\tilde{\mu}_v = \nabla_{\tilde{\mu}_v}\tilde{g}\cdot\tilde{\Pi}_y\tilde{\varepsilon}_y + \nabla_{\tilde{\mu}_v}\tilde{f}\cdot\tilde{\Pi}_x\tilde{\varepsilon}_x - \tilde{\Pi}_v\tilde{\varepsilon}_v
\end{aligned}
\tag{2.8}
$$

Under active inference, we also need to minimise the free energy through action. As the only variable in the above that depends upon action is *y*, we can express action as:

$$
\dot{a} = -\nabla_a\tilde{y}(a)\cdot\tilde{\Pi}_y\tilde{\varepsilon}_y \tag{2.9}
$$

The involvement of action only at the level of sensory input is highly consistent with the notion of a reflex arc (Adams et al. 2013a), of the sort found in the brainstem and spinal cord, where efferent motor neurons project to muscles and correct any deviation between incoming proprioceptive data and descending predictions about these data. These equations have been used extensively to model a wide range of neurobiological phenomena, including attention (Feldman and Friston 2010), perceptual illusions (Brown and Friston 2012), action-observation (Friston et al. 2011), communication (Friston and Frith 2015), and motor control (Baltieri and Buckley 2019; Perrinet et al. 2014). They have also been used in numerical simulations of self-organisation (Friston 2013) and morphogenesis (Friston et al. 2015a). We will apply these in Chapter 3 in the context of oculomotor control (Parr and Friston 2018a).

Discrete time

The free energy defined for a model using categorical distributions has a much simpler form, given that the sufficient statistics of a categorical distribution are simply vectors of probabilities. In this context, expectations become dot-products, and the free energy (and its gradients) can be expressed in linear algebraic terms:

$$
\begin{aligned}
F &= \boldsymbol{\pi} \cdot (\ln \boldsymbol{\pi} - \ln \mathbf{E} + \mathbf{F} + \gamma \cdot \mathbf{G}) \\
\mathbf{F}_{\pi} &= \sum\nolimits_{\tau} \mathbf{F}_{\pi\tau} \\
\mathbf{F}_{\pi\tau} &= -\sum\nolimits_{n} \mathbf{s}^{n}_{\pi\tau} \cdot (\boldsymbol{\varepsilon}^{n}_{\pi\tau} - \tfrac{N-1}{N} \sum\nolimits_{m} \ln \mathbf{A}^{m} \mathbf{s}^{\backslash n}_{\pi\tau} \cdot o^{m}_{\tau}) \\
-\nabla_{\mathbf{s}^{n}_{\pi\tau}} \mathbf{F}_{\pi\tau} &= \boldsymbol{\varepsilon}^{n}_{\pi\tau} \\
&= \sum\nolimits_{m} \ln \mathbf{A}^{m} \mathbf{s}^{\backslash n}_{\pi\tau} \cdot o^{m}_{\tau} + \tfrac{1}{2}\left(\ln(\mathbf{B}^{n}_{\pi} \mathbf{s}^{n}_{\pi\tau-1}) + \ln(\mathbf{B}^{n\dagger}_{\pi} \mathbf{s}^{n}_{\pi\tau+1})\right) - \ln \mathbf{s}^{n}_{\pi\tau} \\
\mathbf{B}^{n\dagger}_{\pi} &\propto \mathbf{B}^{nT}_{\pi}
\end{aligned}
\tag{2.10}
$$

The key points to draw from this equation are that the overall free energy (F) may be decomposed into local 'marginal' free energies ($\mathbf{F}_{\pi\tau}$). These are marginal in the sense that they approximate the free energies we would get if we were to marginalise over all variables in the generative model except for the state and outcome at a given time (under a given policy). This is seen in the form of the marginal free energy gradients in the penultimate line, where a marginal prior and likelihood are offset against the marginal posterior. The marginal prior is constructed by averaging the (log) prior we would get by running the model forwards and that we would get from running it backwards (see (Parr et al. 2019b) for details). The backwards transition probabilities are obtained through transposing the original transition matrix and renormalizing to ensure the columns sum to one (consistent with a probability distribution). The dagger notation (†) indicates the normalised transpose operation. Having defined a series of local free energies, we can set up a gradient descent on free energy for posterior expectations about states at a given time, conditioned upon a given policy[9]:

$$
\begin{aligned}
\mathbf{s}^{n}_{\pi\tau} &= \sigma(\mathbf{v}^{n}_{\pi\tau}) \\
\dot{\mathbf{v}}^{n}_{\pi\tau} &= \boldsymbol{\varepsilon}^{n}_{\pi\tau}
\end{aligned}
\tag{2.11}
$$

This says that expectations about states under policies may be updated based upon a series of linear (matrix multiplications) and non-linear (softmax) transforms, much as combinations presynaptic firing rates across a neural population are non-linearly transformed (via membrane depolarisations) into post-synaptic firing rates. These expectations may be used to compute the posterior probability of each policy (treating each policy as if it were a model, and performing a Bayesian model comparison):

[9] $P(s_{\tau} \mid o, \pi) \approx Q(s_{\tau} \mid \pi) = Cat(\mathbf{s}_{\pi\tau})$

$$\boldsymbol{\pi} = \sigma(\ln \mathbf{E} - \mathbf{F} - \gamma \cdot \mathbf{G})$$

$$\begin{aligned} \mathbf{G}_{\pi} &= \sum_{\tau} \mathbf{G}_{\pi\tau} \\ \mathbf{G}_{\pi\tau} &= \sum_{m} \mathbf{o}_{\pi\tau}^{m} \cdot \boldsymbol{\varsigma}_{\pi\tau}^{m} \\ \boldsymbol{\varsigma}_{\pi\tau}^{m} &= \ln \mathbf{o}_{\pi\tau}^{m} - \ln \mathbf{C}_{\pi\tau}^{m} + \mathbf{H}^{m} \cdot \mathbf{s}_{\pi\tau} \\ \mathbf{H}^{m} &= -diag(\mathbf{A}^{m} \cdot \ln \mathbf{A}^{m}) \end{aligned} \qquad (2.12$$

We will unpack the terms involved in Equation 2.12 in greater detail in the *Active Perception* section of this chapter but include them here for completeness. Equations 2.10-12 express a generic update scheme to solve an MDP model. Different models depend upon the choices for **A**, **B**, **C**, **D**, and **E**.

Simulated message passing

While the above is fairly abstract and technical, it lays the foundations for numerical simulation under specific generative models. To illustrate this, we present a simple simulation using a Hidden Markov Model (HMM) generative model (Parr et al. 2019b). An HMM is essentially an MDP without alternative policies or actions. It therefore does not need Equation 2.12 and may be simulated using only 2.10-11. The purpose of presenting this simulation is three-fold. First, it provides some intuition about the behaviour of the belief-update equations outlined above. Second, it is useful in thinking about the sorts of neuronal architectures required for the message passing, and the association between the variables involved in belief-updating with plausible electrophysiological correlates (e.g. firing rates). Finally, it allows for a numerical comparison with other Bayesian update schemes, establishing the construct validity of these biologically plausible inferential dynamics.

Simulations

The generative model for these simulations is shown in the upper part of Figure 2.3. It comprises two hidden state factors (s^1 and s^2) represented as light blue and green shapes respectively. At each time-step, the shapes within each of these factors may change

(probabilistically) into any other shape within that factor (i.e. if the state in factor 1 is a light blue square at the start, it may change into a light blue circle, triangle, or stay as a square at the next time-step). One way to think of this is as a simple symbolic language, where any given letter (shape) is followed by another with a certain probability based upon the statistics of that language. To illustrate the relative influences of prior information and observed data, we set the outcomes (darker blue shapes) depend only upon the light blue shapes in the first hidden state factor. Intuitively, this is as if the series of visual inputs (o^1) we are presented with are generated probabilistically from the words on the current page (s^1) but carry no information about the words on another page (s^2), which we cannot see. The purpose of this is to illustrate the behaviour of each scheme in the presence of informative and uninformative sensory input. The form of the HMM that mediates the influences between states at one time, and those at the next, and between current states and their associated outcomes, is shown in the upper right of Figure 2.3, along with the Bayesian update scheme used for the simulation. Note that this is a special case of the generative model and update scheme shown in Figure 2.2.

Below the generative model, Figure 2.3 shows the results of solving the equations above during sequential presentation of the outcomes. This is shown for three different Bayesian message passing schemes. These are variational message passing, belief propagation, and marginal message passing. The last of these is the scheme outlined above. The first two are established methods for approximate Bayesian inference. Each relies upon a different free energy functional, as unpacked in detail in (Yedidia et al. 2005) and Appendix A.2. These are useful points of comparison as variational message passing represents a very simple and neuronally plausible architecture but rests upon a severe (mean-field) approximation that lends itself to overconfident inference, while the belief propagation architecture is harder to justify from a biological viewpoint but performs very well from an inferential standpoint. These schemes accumulate evidence for different hidden states by assimilating successive outcomes into posterior beliefs. Each hidden state starts with a defined shape (blue triangle, green square), but undergoes stochastic transitions. This means that the future should always be more uncertain than the past. In what follows, we use belief propagation as a gold standard for inferential performance, against which the other two (more anatomically plausible) schemes are compared.

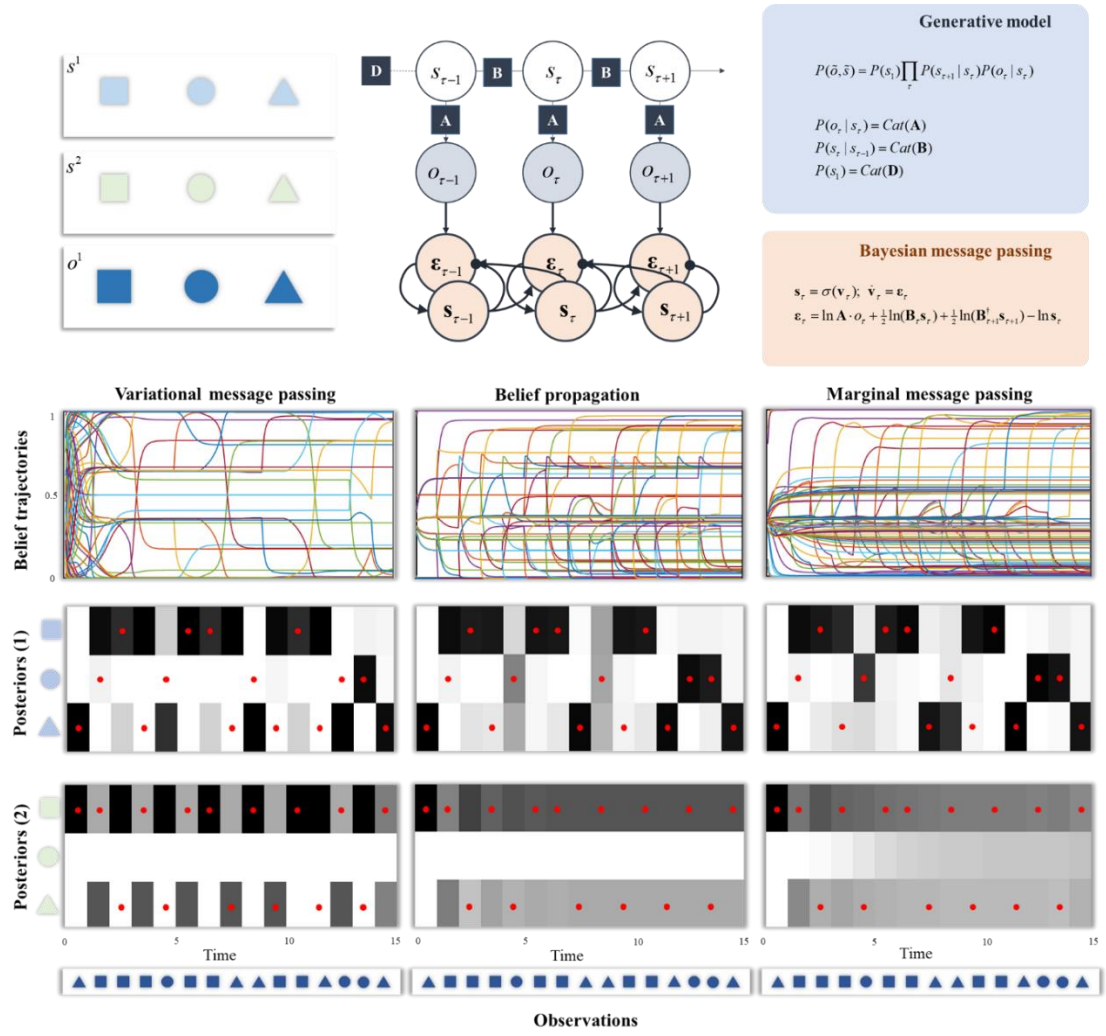

Figure 2.3 – Simulated neuronal message passing. *The upper part* of this figure shows the form of the generative model and Bayesian (marginal) message passing scheme used for the simulations. *The lower plots* illustrate inference using the same marginal message passing scheme (right), but also two other Bayesian message passing schemes for comparison. The first row of these ('belief trajectories') show the sufficient statistics (expectations, **s**) for each hidden state at each time step as they evolve over time. Given that there are three possible hidden states (circle, square, or triangle) in each factor, the probabilities of each all start at a third, and move closer to zero or one as new evidence is accumulated over time. The posterior probability plots below show the beliefs, after all the observations have been made, about the hidden states as they were at each time-point in the trial. Darker shades indicate a greater posterior probability, such that black indicates a posterior probability of one, and white of zero. The red dots superimposed upon these show the 'true' states that were used to probabilistically generate the observations. The sequence of observations presented over time are shown in the bottom row of the figure. These were presented sequentially, with a new observation at each time-step.

Relation to established inference schemes

Figure 2.3 shows the results of simulating inference via the three forms of neuronal message passing outlined above. This illustrates some cardinal features of the three schemes. The trajectories of beliefs following each outcome show that much of the belief updating occurs very early in variational message passing, before the presentation of most of the data. While a few revisions to these beliefs occur at later stages, it does not take long to arrive at highly confident beliefs about future states – this over-confidence of posterior beliefs is a well-recognised feature of variational inference under the mean-field approximation (Consonni and Marin 2007). In contrast, belief propagation and marginal message passing take a more restrained approach, with each new observation driving updating. This more tentative approach pays off, as they make fewer errors in estimating the true states that generated the data. This is consistent with the fact that belief propagation offers an exact estimate of marginal beliefs for these models, while the variational approach is only ever approximate.

The over-confidence of the variational approach manifests clearly in the posterior beliefs about the green shapes. Given the stochastic transitions, and the absence of any informative data about these states, posterior beliefs about the green shapes should become increasingly uncertain with distance from the (deterministic) initial state. The Belief Propagation scheme (based on Bethe free energy) clearly shows this, but the variational scheme does not, with highly confident beliefs about even the penultimate state. Marginal message passing compensates for this overconfidence issue, providing a much better approximation to an exact inference scheme than under the mean-field approach. In fact, it slightly overcompensates in the absence of precise data, leading to posteriors that are less confident than the belief propagation marginals. The temporal dynamics of belief updating (the upper plots) further illustrate the overconfidence of variational message passing relative to the other two schemes. Within the first time-step, the sufficient statistics of beliefs about the states over time (each represented as a line) approach extreme (zero or one) values. This means that, with only one observation, the mean-field variational approach exhibits an excessive confidence about present and future states, that is maintained as new observations are made. In contrast, the belief propagation and marginal message passing schemes afford more modest belief updates – following the first observation – that become more confident as new data are acquired.

Notably, the three schemes share some of the same errors (four errors in steps 2, 4, 9 and 10). By errors, we mean that the inferred state (darkest shade) at a given time-step does not match the state that actually generated the data (red dot). These errors happen when very unlikely

events occur, such as a dark blue square generated by a light blue triangle. Although incorrect, an inference that the light blue square caused the dark blue one is still Bayes optimal under the generative model we employed. In contrast, the additional four errors of variational message passing in steps 5, 8, 12 and 13 occur even though the data are highly consistent with the hidden states (e.g., a dark blue circle generated by a light blue circle). These errors reflect the excessive weight given to the empirical priors in variational message passing – it assumes the most probable *a priori* transition, ignoring the conflicting observation.

To quantify the performance of the mean-field and marginal approaches, we can exploit the fact that the belief-propagation approach is exact for the marginal posteriors for this inference problem. A simple way to do this is to compute the KL-Divergence between the marginal posteriors obtained through belief propagation and the solutions of the other two schemes. The smaller this divergence, the better the approximation to exact marginal beliefs. For the simulations of Figure 2.3, the divergences summed over marginal posteriors give the following:

$$\sum_{\tau} D_{KL}[Q_{BP}(s_\tau) \,||\, Q_{VMP}(s_\tau)] = 86.0563 \; nats$$
$$\sum_{\tau} D_{KL}[Q_{BP}(s_\tau) \,||\, Q_{MMP}(s_\tau)] = 3.7874 \; nats$$

This demonstrates quantitatively that, even for the relatively simple inference problem used here, there is a much greater divergence between the exact marginal posterior beliefs and those obtained using variational message passing, relative to marginal message passing.

Although we have presented this as a single simulation, the way in which the generative model is defined, and the sequential presentation of the data, actually induce several distinct inference problems that we have implicitly appealed to above in characterising these schemes. First, the factorisation of the hidden state-space into two different types of hidden state allows us to compare the extreme case in which data are uninformative about the hidden state (light green shapes) with the case in which there is only moderate uncertainty about the relationship between (light blue) states and the (dark blue) data. Figure 2.3 shows that, while marginal and belief propagation approaches attenuate their confidence – when data is uninformative compared to informative – the mean-field approach furnishes confident inferences in both cases. Note that these differences rely upon there being some uncertainty in the transitions from one state to the next. If we were to use deterministic transition probabilities, the

differences between these schemes would be largely abolished as all would make confident inferences.

The second comparison we have used relies upon sequential presentation of the outcomes. This means that each time-step represents a distinct inference problem, with more data available at later times than earlier. This is where the dynamics shown in the upper row of Figure 2.3 are revealing. At each successive time point, the inference problem becomes more constrained, as an additional observation is made. This allows us to compare the confidence using a small amount of data (at the start of the trial) with the confidence after more data have been seen (near the end of the trial). After making the first observation, variational message passing shows a consistent level of confidence until the end of the trial. This can be seen in the plot by noting that the distribution of lines in the vertical direction is relatively constant throughout the horizontal (temporal) axis. This contrasts with the other two schemes that show a greater proportion of lines reaching extreme values with each new observation.

Ultimately, all three schemes above are free energy minimising processes compatible with active inference (see (Schwöbel et al. 2018) for an example of the application of belief-propagation in this domain, and (van de Laar and de Vries 2019) for an example using variational message passing). However, we will appeal to marginal message passing throughout this thesis, due to its combination of the architectural simplicity of variational message passing with (nearly) the inferential performance of belief-propagation.

Neuronal process theories

Current circuit-level research shows a high degree of consistency between the form of the message passing shown in pink in the upper part of Figure 2.3 (and in Figure 2.2), and biological neuronal networks (Bastos et al. 2012; Haeusler and Maass 2007; Shipp 2016). Sensory input (via the thalamus) predominantly targets granular layers of cortex (Miller 2003; Shipp 2007), which excite more superficial cells and are disynaptically inhibited by them in turn (Thomson et al. 2002). This is consistent with the interactions between granular ($\boldsymbol{\varepsilon}$) and superficial ($\mathbf{s}$) cells in Figures 2.2 and 2.3. Associating these variables with messages passed between neural populations implies that these may be interpreted as normalised firing rates, with auxiliary variables ($\mathbf{v}$) playing the role of membrane potentials. This means we can treat the plots labelled ‘belief trajectories’ in Figure 2.3 as idealised firing rates. We can additionally associate the rates of change of their associated membrane potentials with the local field potentials induced by membrane depolarisation that might be measured using intracranial

electrodes, or whose combination manifests as measurable electromagnetic signals using sensors at the scalp (e.g. electroencephalography or magnetoencephalography). Committing to a process theory of this sort (Friston et al. 2017a) ensures that these simulations may be used to form empirical hypotheses about the anatomical structures required for and the electrophysiological correlates of belief updating processes, offering one way of disambiguating between alternative generative models that could be employed by an individual or group of individuals for a given experimental paradigm. To take this a step further, it is useful to be able to relate inference to the behaviours it causes, such that behavioural data may be used to compare hypothesised generative models. This requires that a generative model includes courses of action, and their sensory consequences. To incorporate alternative courses of action (i.e. plans or policies) into a generative model, we must be able to express a prior belief that characterises the plausibility of each policy that could be pursued. The final section of this chapter addresses this issue.

Active perception and planning

Clearly the processes simulated above are a poor account of biological perceptual inference. This is because biological systems are not simply passive recipients of data. Instead they move and seek out new data. While we save a more comprehensive numerical analysis of this active engagement to subsequent chapters, we conclude this chapter with an example that illustrates the importance of considering action in understanding perception, and then unpack the form of the prior over policies expressed in Figure 2.2. The example is that of Troxler fading, a phenomenon that occurs on suspension of exploratory eye-movements. Figure 2.4 shows the sort of stimulus that lends itself to this effect, where a central fixation cross is surrounded by static coloured shapes, of the sort poorly detected by the peripheral parts of the retina and magnocellular pathway (which are more sensitive to luminance or to high frequency temporal changes than to colour). Foveating the peripheral parts of the image ensure high quality data about the colour of these shapes, via the cone cells in the centre of the retina, and the associated parvocellular signalling pathway. This suggests that we can think of maintained central fixation as much like the situation illustrated by the second hidden state factor in Figure 2.3, where the absence of informative data leads to an accumulation of uncertainty over time. Consistent with this, on maintenance of central fixation, the peripheral colours appear to fade away into the background. They are restored as soon as exploratory eye-movements to peripheral locations are resumed.

This simple visual phenomenon provides a compelling demonstration of the role of action in perception, and the fallacy of studying the two in isolation of one another. In addition, it endorses the Helmholtzian view of perception as a process of unconscious inference (Von Helmholtz 1867) and the view that perceptions are hypotheses about the causes of sensory data (Gregory 1980). In the absence of the ability to precisely disambiguate between hypotheses, there is no precise percept. This suggests that the role of action in constructing a percept is in gathering high quality data to enable efficient (precise) inferences. This perspective on active perception may be formalised through an appeal to the expected free energy associated with alternative courses of action (Friston et al. 2015b). Simply put, this quantifies the potential for uncertainty resolution (or information gain) associated with a course of action (or policy). This can be used to score the prior probability of alternative policies. The next section motivates this more formally.

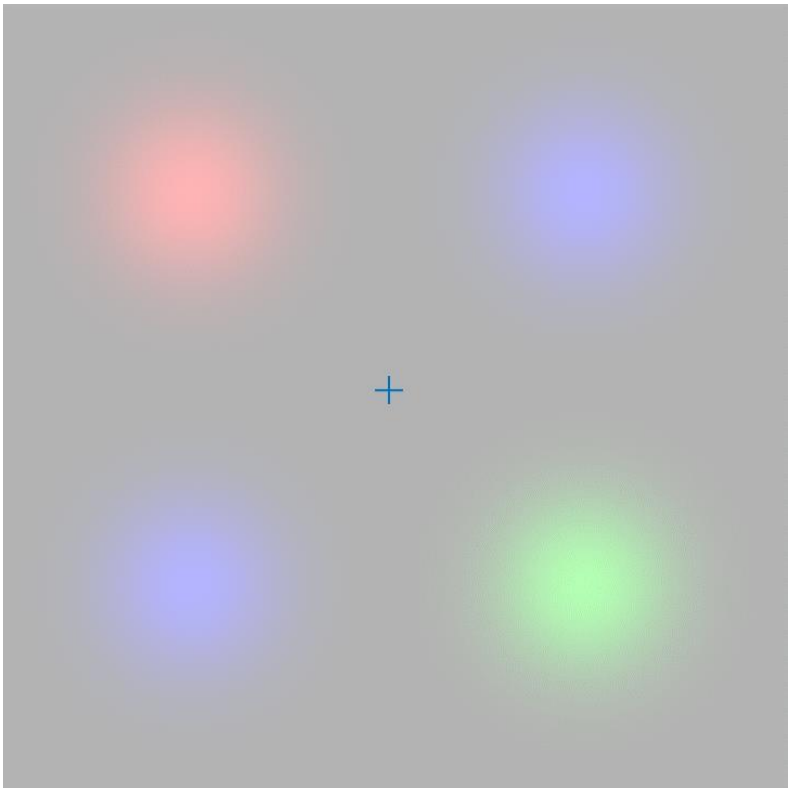

Figure 2.4 – Active vision and Troxler fading. The image shown here motivates the importance of action in perception in a simple way, through a phenomenon called Troxler fading. On maintenance of fixation on the central cross, the colours in the periphery appear to fade away. However, on resumption of normal exploratory eye movements, the colours reappear. This illustrates a crucial point that we will return to throughout this manuscript. Vision is not a passive process, with data presented to a static retina. Instead it is a process of exploration, in which the retina is moved around to best resolve uncertainty about a dynamically changing environment. This image is reproduced from (Parr et al. 2019a).

Expected free energy

To select a suitable prior over policies, we must take a step back, and think again from the perspective of homeostasis (or, more generally, from the perspective of non-equilibrium steady-state physics (Kwon et al. 2005)). This means thinking in terms of a distribution over states that a creature tends to occupy, and returns to following perturbation (Friston 2013; Friston and Ao 2012). We can then distinguish between this 'non-equilibrium steady-state' (NESS) distribution that is independent of the trajectory (or path) that achieves this[10] and the evolving distribution from beliefs about the current states into the future that depends upon the path (policy) pursued. Given that we have said that a creature will correct deviations from its NESS by returning to it, the divergence between the policy-dependent density over future outcomes and the NESS density will be small, at the endpoint of any plausible path. When the divergence is zero, the following relation holds (see Friston 2019, and Appendix A.3 for details):

$$
\begin{aligned}
&D_{KL}[P(o,s\mid\pi)\,\|\,P(o,s)]=0\Leftrightarrow\\
&\quad G(\pi)\approx H[P(o\mid\pi)]\\
&\quad G(\pi)\triangleq E_{P(o|s)Q(s|\pi)}[\ln Q(s\mid\pi)-\ln P(o,s)]\\
&\qquad =\underbrace{D_{KL}[Q(s\mid\pi)\,\|\,P(s)]}_{Risk}+\underbrace{E_{Q(s|\pi)}[H[P(o\mid s)]]}_{Ambiguity}
\end{aligned}
\tag{2.13}
$$

Here, we have defined the expected free energy (G) (Friston et al. 2015b) associated with a given policy (note the similarity in form to the free energy in Equation 2.2). The separation of the expected free energy into 'Risk' and 'Ambiguity' is useful in intuiting its properties. The closer the expected distribution of states under a given policy to the NESS states, the lower the expected free energy. Similarly, the greater the fidelity of the mapping from states to observations, the lower the ambiguity, and consequently expected free energy. For any plausible path or policy, where the divergence between policy-dependent and preferred (NESS) distributions is small, the approximate equality between the expected free energy and the entropy (i.e. negative expected log evidence) for that policy implies that, on average, those

[10] This is an important feature of a NESS in stochastic dynamics. Once a system has reached NESS, the density over its states conditioned upon the history (i.e. starting configuration and subsequent trajectory) of those states becomes equal to the density without conditioning upon the history.

policies with a smaller expected free energy will carry greater evidence, and will therefore be more probable. As such, the most appropriate choice of prior over policies (i.e. alternative paths to NESS) can be expressed:

$$\begin{aligned} P(\pi) &= Cat(\boldsymbol{\pi}_0) \\ \boldsymbol{\pi}_0 &= \sigma(\ln \mathbf{E} - \gamma \cdot \mathbf{G}) \end{aligned} \tag{2.14}$$

Here, **E** represents other potential influences over policies (that may be learned, or be computed through a hierarchical generative model) that may be thought of as habits (FitzGerald et al. 2014). The bold **G** is a vector of expected free energies for each policy, weighted by an inverse temperature parameter (γ) that determines the relative influences of **E** and **G**. These may be thought of in terms of a fixed prior and the expected (negative log) evidence for each policy. Given that minimising risk minimises expected free energy, we can interpret the NESS distribution over states as preferences; in the sense that a creature is most likely to pursue those courses of action that lead to those 'preferred' states. Often it is convenient to define a preference over observations, instead of states. To do this, we simply replace the risk term with an equivalent KL-Divergence expressed in terms of the distribution of observations that would have been generated by the associated states (intuitively, this is as if each state over which a preference is defined is deterministically associated with an outcome). This gives:

$$\begin{aligned} G(\pi,\tau) &\approx \underbrace{D_{KL}[Q(o_\tau \mid \pi) \,\|\, P(o_\tau)]}_{Risk} + \underbrace{E_{Q(s_\tau|\pi)}[H[P(o_\tau \mid s_\tau)]]}_{Ambiguity} \\ &= \mathbf{o}_{\pi\tau} \cdot (\ln \mathbf{o}_{\pi\tau} - \ln \mathbf{C}) + \mathbf{H} \cdot \mathbf{s}_{\pi\tau} \end{aligned} \tag{2.15}$$

This is the form[11] used for the prior defined in Figure 2.2 and that gives rise to the posterior expressed in Equation 2.12. Note that this expression only requires that we specify a final (anticipated or desired) distribution over outcomes (*Cat*(**C**)), which may be thought of as preferences. Equation 2.15, although very simple, has several very important properties that will be unpacked in Chapter 3. For now, we note that the first term of the expected free energy is the expected log probability of an observation, consequent on an action (e.g. 'what I would

[11] $Q(o_\tau \mid \pi) = Cat(\mathbf{o}_{\pi\tau})$

see if I looked there'). The expected log probability is the negative entropy (uncertainty) associated with this prediction. This says that the most probable policies, with the lowest expected free energy, will be those that involve seeking out those observations about which we are most uncertain. Returning to the example of Troxler fading, this says that the first thing we will do if we become uncertain about one of the coloured shapes is to look at it, resolving this uncertainty and precluding fading. It is only when we interrupt this policy that fading occurs. For detailed numerical simulations of Troxler fading using this approach, please see (Parr et al. 2019a).

Conclusion

In this chapter, we have outlined the generic aspects of active inference, its interpretation in terms of message passing between neuronal populations, and the importance of action in perception. We provided a simple example of passive perceptual inference, and the success of marginal message passing in approximating the inferences of belief propagation in this setting, while maintaining the architectural simplicity of variational message passing. We concluded with a discussion of planning, in the sense of inferring which actions to perform next. Under active inference, optimal plans are those associated with the lowest expected free energy, that best resolve our uncertainty about the causes of our sensory data. In subsequent chapters, we will apply the same principles outlined here, introducing specific generative models, based upon those used here, to solve specific tasks. In addition, we will introduce additional features as they are needed, including the role of precision in MDPs, learning of the parameters of the generative model, deep temporal modelling, and the role of Bayesian model reduction in combining MDP and continuous state-space models for tasks involving both decisions and movements.

3 - The computational anatomy of active vision

Introduction

In this chapter[12], we revisit the anatomy and physiology outlined in Chapter 1, but through the lens of the theoretical material in Chapter 2. Our aim is to understand the minimal generative model required for the emergence of the inferential architectures employed by the brain in implementing active vision. First, we address the problem of how to move the eyes. In doing so, we explore the relationship between the sort of model required to explain the proprioceptive and visual consequences of oculomotion and the anatomy of the brainstem (Parr and Friston 2018a). Second, we deal with the problem of deciding where to look (Parr and Friston 2017c). By varying the uncertainty associated with different variables in a generative model, we reproduce several behavioural phenomena (including 'the Streetlight effect' (Demirdjian et al. 2005) and 'inhibition of return' (Posner et al. 1985)). Finally, we combine these models, and hypothesise a role for the superior colliculus in translating beliefs about which location to foveate into beliefs about the Newtonian dynamics that realise this (Parr and Friston 2018c). On substituting the generative models for each of these processes into the belief-update equations outlined in Chapter 2, the resulting message passing bears a striking resemblance to the neuroanatomical systems engaged in oculomotor control and can be used to reproduce aspects of their physiology and pathology.

Movements

There are many neurological (Anderson and MacAskill 2013; Büttner et al. 1999; Perry and Zeki 2000; Sereno and Holzman 1995) and psychiatric (Holzman and Levy 1977; Lipton et al. 1983; Sereno and Holzman 1995) conditions that cause impairments of eye movement control. As such, assessment of oculomotion forms a crucial part of any neurological examination. In this section, we aim to characterise the functional anatomy of eye movement control by appealing to the continuous state-space formulation of active inference. Our agenda here is to try and understand the oculomotor system in terms of its computational anatomy, as

[12] The material in this chapter is adapted from (Parr and Friston 2017c; Parr and Friston 2018a; Parr and Friston 2018c)

a complement to similar attempts to understand the control of eye movements at higher levels of the visual system; e.g., (Bruce and Tsotsos 2009; Itti and Koch 2001). Previous active inference accounts of eye movements have focused on saccadic target selection (Friston et al. 2017f; Mirza et al. 2016) and ignored the mechanics of oculomotion, or have made use of the simplifying assumption that the position of the eyes can be altered directly through simple attractor dynamics (Friston et al. 2012a; Friston et al. 2017c). Here, we follow the example of models that have treated the eyes as physical objects, subject to Newton's laws (Adams et al. 2012; McSpadden 1998; Perrinet et al. 2014; Robinson 1964; Robinson 1968). We build upon these models by equipping each eye with separate kinetics, which are predicted by the brain using a model that is common to both eyes. We emphasise the anatomy and electrophysiology that emerge from this theoretical treatment and their striking resemblance to the properties of the brainstem (Büttner-Ennever and Büttner 1988; Büttner and Büttner-Ennever 2006).

The oculomotor system is an important interface between the inferential processes of the brain, and the Newtonian world that it inhabits. It forms a distributed network (Parr and Friston 2017a) that involves the cerebral cortex (Corbetta et al. 1998; Paus 1996), the cerebellum (Berretta et al. 1993), and the basal ganglia (Hikosaka et al. 2000; Hikosaka and Wurtz 1985b). Ultimately, neuronal messages from these regions combine to generate signals to the extraocular muscles to move the eyes. It is the brainstem that performs the translation of these instructions into motor nerve signals (Sparks 1986; Sparks 2002; Sparks and Mays 1990). In this section, we seek to understand the computations that must be performed to do this, and their neurobiological substrates. We begin by describing the mechanics of the eyes. We then describe a predictive (generative) model of eye movements. We demonstrate through simulation that this reproduces eye movements consistent with health and disease and show the emergence of established electrophysiological observations from these simulations.

A generative model for oculomotion

In Chapter 1, we outlined the connections between the brainstem and the oculomotor muscles. These are bidirectional connections, carrying motor commands to the muscles, and proprioceptive data to the brainstem. Chapter 2 highlighted the importance of a generative model to explain sensory data. Combining these, we begin our analysis of the oculomotor system by specifying the sort of generative model that would account for proprioceptive (and visual) data of the sort received by the midbrain and pons.

Saccadic eye movements implement the transition from one stationary fixation to another. While we may select a new target for fixation, the physical world does not allow us to alter position directly. Instead, changes in position must be brought about by applying forces that accelerate the eyes towards their target. For simplicity, we assume only two forces acting on each eye. These are resultant forces in the horizontal and vertical dimensions. Each force gives rise to a torque, made up of an active term (muscle contraction), an elastic term, and a viscous term. Using Newton's second law in its rotational form, we arrive at the equations of motion shown on the upper left of Figure 3.1 (labelled 'generative process'). These equations are relatively simple, but could in principle be replaced by a set of more realistic equations that take account of, among other things, the non-linear relationship between muscle elasticity and length (McSpadden 1998).

In addition to the equation describing the movement of the eyes themselves, it is necessary to specify how the angular position and velocity of each eye gives rise to sensory data. The information carried from the eye to the brainstem can be classified into two broad categories. Visual information is passed through the optic nerve (Cranial nerve II), while proprioceptive data from the extraocular muscles travels through afferent fibres in the oculomotor nerves (CN III, IV, VI). We have assumed a simple visual signal in this chapter: it is generated through an identity mapping, with added noise, from the position of the eyes (Faisal et al. 2008). In other words, what the eyes see depends only upon where they look.

The nature of proprioceptive signals from the extraocular muscles is a controversial topic (Donaldson 2000), but the presence of muscle spindles – the sensory organs of proprioception – in human extraocular muscles has been convincingly demonstrated (Cooper and Daniel 1949), as has the type of reflex associated with these spindles in other muscles (Sherrington 1893). It is worth acknowledging that the structure of these spindles is simpler than those found in other muscles (Ruskell 1989), but the density is comparable (Lukas et al. 1994). In most skeletal muscle, afferent nerve fibres from the muscle spindles carry data about the velocity (type Ia afferents) and instantaneous length of a muscle (type II afferents). Similar signals have been recorded from the oculomotor nerve (Cooper et al. 1951; Tomlinson and Schwarz 1977), when the extraocular muscles are stretched. We therefore assume that there are two proprioceptive modalities from each eye, carrying signals analogous to the II (position) and Ia (velocity) afferent fibres. Each of these has a horizontal and a vertical component. The equations determining these outputs are shown on the upper left of Figure 3.1. Having specified these primary afferents, we turn to the treatment of these sensory signals by the brain.

At this point, it is useful to make a distinction between a generative *model* (the brain's beliefs about how data are generated) and a generative *process* (how the world, or simulation, actually

generates data). So far, we have dealt with the process, but now turn to the model. The interface between the generative model and process is illustrated in the Bayesian network Figure 3.1, and this highlights the important differences between the two. The model is much simpler than the process. This is because the model does not allow for each eye to move independently, whereas the position of one eye offers no constraint over that of the other in the physical world. This is a reasonable simplification for a human brain to make but would not be fit for purpose for creatures (e.g. chameleons) that move both eyes independently. The other key differences are that action is part of the generative process, while hidden causes are only found in the model. The former causes changes in angular velocity, while the latter changes angular position. The hidden cause acts as a point attractor, drawing the eyes towards this position.

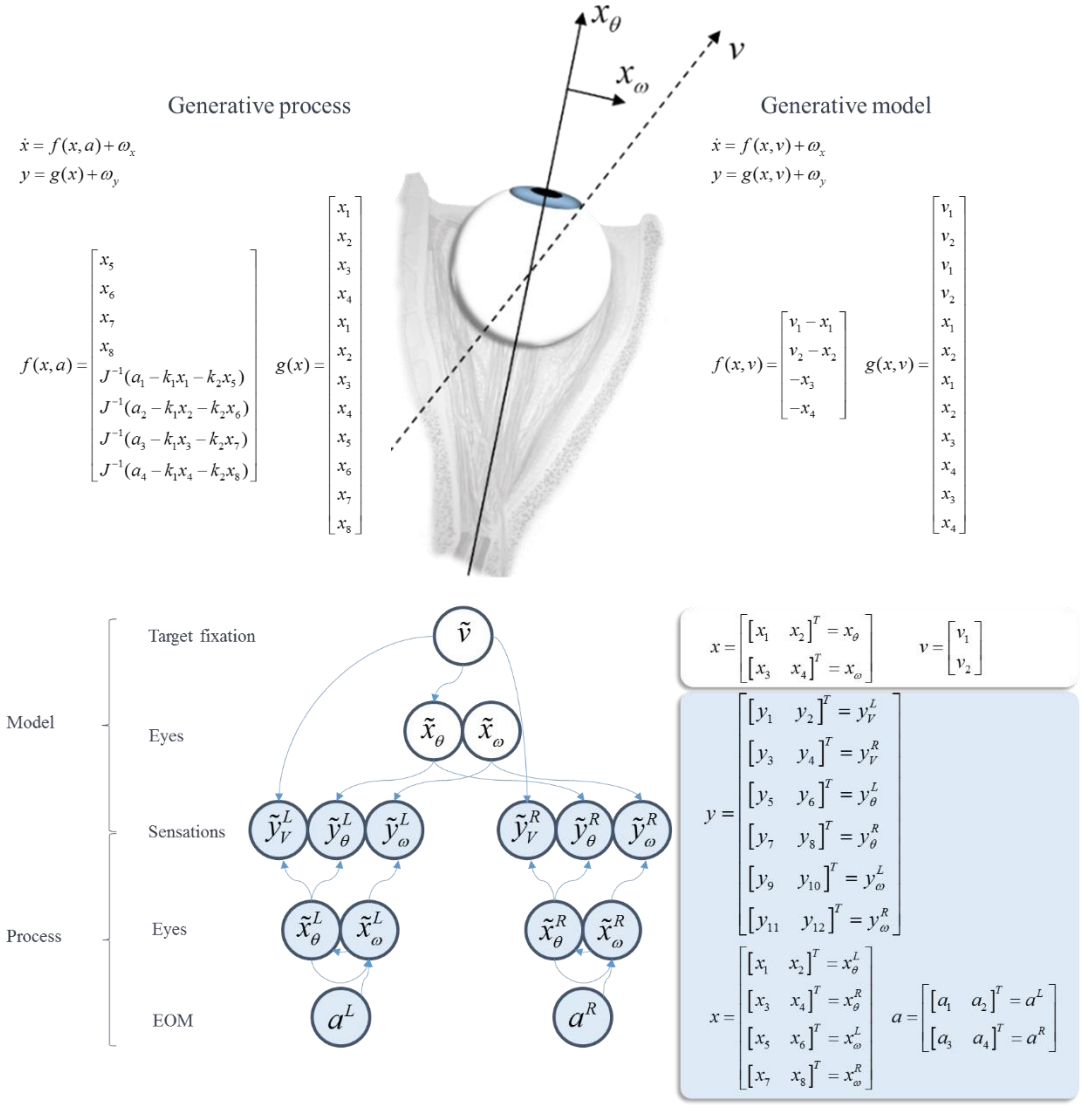

Figure 3.1 – Generative model and process. This Figure shows the generative process (*upper left*) used to generate the data presented to our simulation, and the generative model (*upper right*) used by our synthetic brainstem to explain and predict the input from the cranial nerves.

The blue panel on the *lower right* offers a key as to the interpretation of each of the elements of the hidden cause (v), hidden states (x), actions (a), and data (y). Each of these includes a horizontal and a vertical component. Subscripts indicate angular positions (θ) or velocities (ω) (and their associated proprioceptive inputs), and the visual modality (V). Superscripts index the left (L) or right (R) eye. Note that the variables in the generative model do not distinguish between the two eyes (see main text for details). The generative process uses several physical constants relating to the moment of inertia of the eyeballs (J) and the spring and viscous constants (k) of the oculomotor tendons and orbit respectively. The Bayesian network at the *lower left* shows the dependencies between these variables, and the point at which the external world (circles with light blue shading) interfaces with the brain's model of it (white circles). Note the asymmetry of this interface, with the force generated by the extraocular muscles (EOM) present only in the generative process, and the target fixation only in the model.

On the right-hand side of Figure 3.2, we illustrate how these equations could be implemented by passing messages between populations of neurons (Bastos et al. 2012; Friston and Kiebel 2009; Shipp 2016). Ascending messages here are (excitatory) prediction errors, while descending messages are (inhibitory) predictions. It is this pattern that characterises predictive coding (Friston and Kiebel 2009; Rao and Ballard 1999). This is exactly the same structure as in Figure 2.2, but where each we have summarised the generalised coordinates of motion for each variable with a single population but have unpacked each variable in terms of the generative model variables of Figure 3.1. This means that this network employs the generic aspects of the neuronal message passing of Chapter 2 but applies them to the specific problem of oculomotor control.

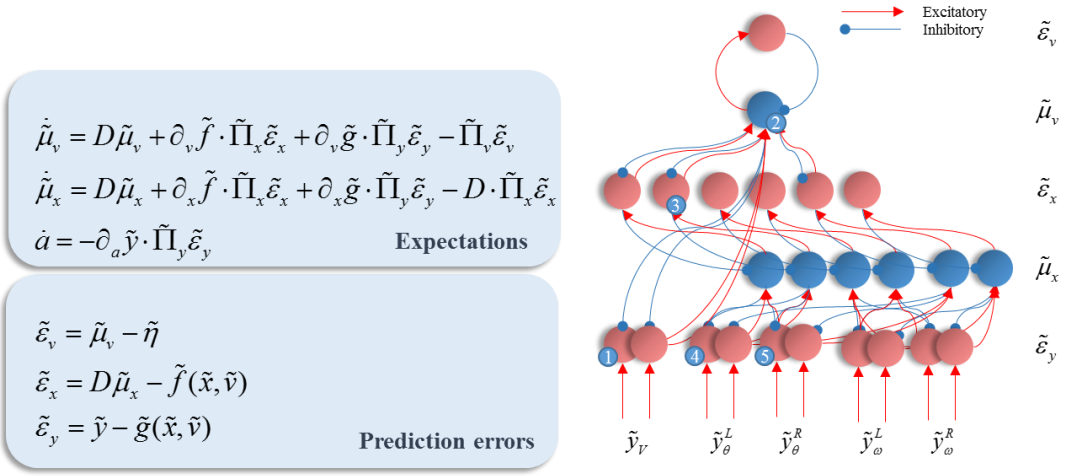

Figure 3.2 – Neuronal message passing for oculomotor inference. *On the left* are the equations describing a gradient descent on variational free energy. *On the right*, we show how these equations map to a neuronal message passing scheme for the generative model outlined above. To do so, we have simply assigned the terms on the left-hand side of each equation to a neuronal population and mapped the influences between each population with excitatory and inhibitory connections. We have separated the states representing positions and velocities into right and left components; for consistency with the representation of each hemifield on the contralateral side of the sagittal plane in the brain. The numbers in little blue circles refer to the anatomical designation of expectation and error units in Figure 3.4.

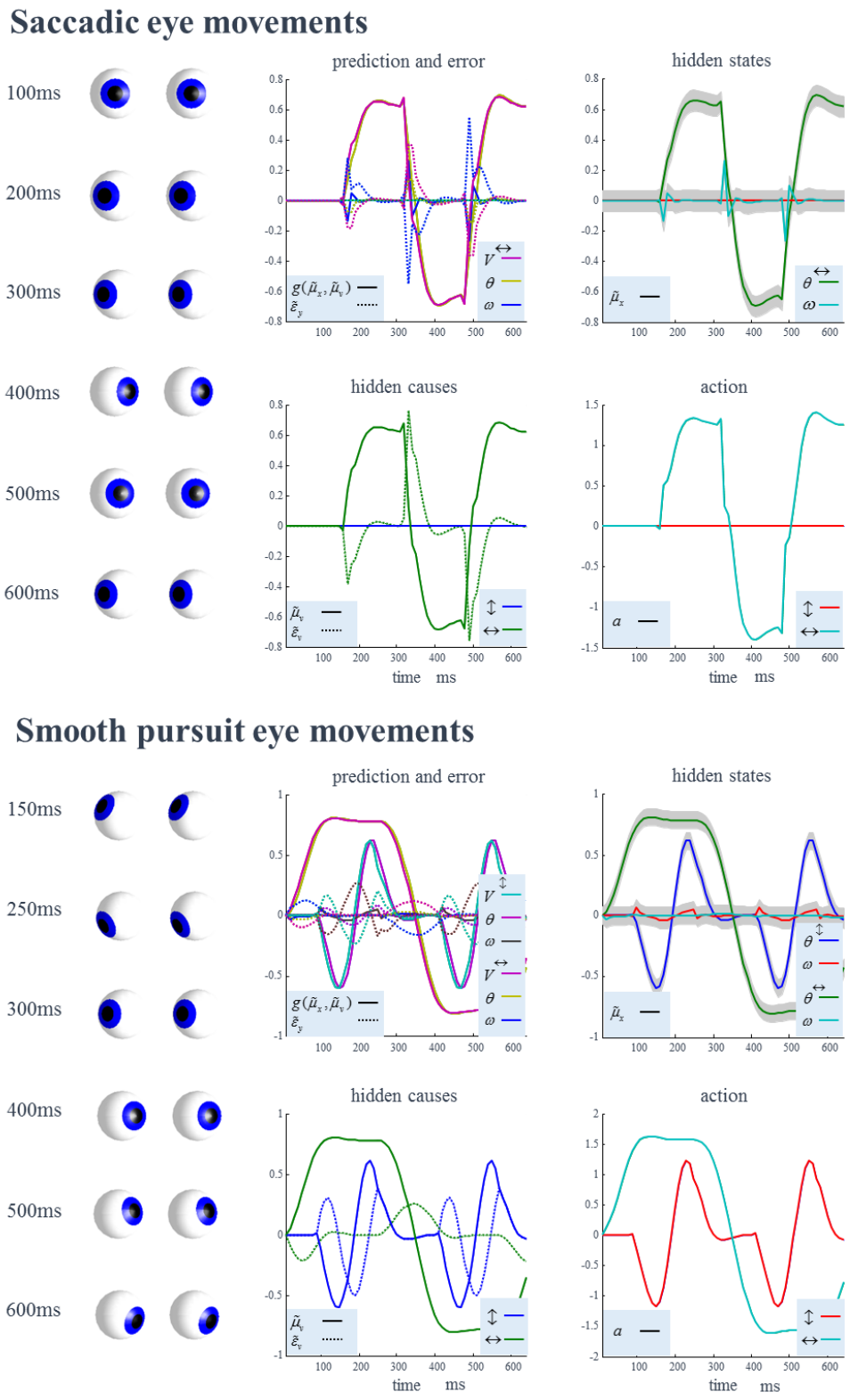

Figure 3.3 – Simulated eye movements. These plots show the changes in expectations (solid lines) and prediction errors (dotted lines) over time for the hidden causes and states during saccadic eye movements (*upper*), and smooth pursuit movements (*lower*). The eye positions at various times are shown on the left of each set of plots. The grey regions correspond to 90%

Bayesian confidence intervals around the inferred hidden states; namely the vertical and horizontal angular positions and velocities. The legend in the lower right of each plot indicates the modality represented by each line (visual = V, type II afferent/position = θ, type Ia afferent/velocity = ω). For example, a dotted line with a colour associated with V represents a prediction error in the visual domain. To see the key variables plotted individually, please refer to Figure 3.4, where these are represented in separate raster plots.

Oculomotor behaviour

Figure 3.3 shows the results of numerically solving these equations, with two different prior distributions over the trajectory of a fictive fixation location (v). The first is a discontinuous function that changes discretely to different values, inducing saccades. The second is a sinusoidal function that gives rise to smooth pursuit eye movements. For both priors, the active inference scheme successfully computes the forces required to fulfil these beliefs. The common generative model for both eyes ensures the eye movements are conjugate – i.e. the eyes move together. In summary, using a plausible generative model and standard (active inference or filtering) dynamics we can reproduce the control of eye movements. Notice that we have not appealed to any control theory: in active inference, motor control follows naturally from the suppression of prediction errors generated by prior expectations: see Figure 3.2. In other words, the active filter has prior beliefs about where it should be looking and action fulfils those beliefs in a Bayes optimal fashion. The plausibility of this sort of scheme has been addressed in the context of visual search (Friston et al. 2012a) and oculomotor delays (Perrinet et al. 2014).

We now turn to the question of the biological substrates of the active filtering equations used to generate oculomotor behaviour *per se*.

Brainstem anatomy and electrophysiology

The biological implementation of the equations in Figure 3.2 is anatomically constrained in several ways. First, sensory inputs must reach the brain by the cranial nerves that carry that information. The neuronal populations that receive these inputs directly must reside in regions of the brain that contain the terminals of the relevant sensory afferent fibres. Similarly, neurons

encoding actions should be lower motor neurons that contribute efferent fibres to the cranial nerves. The abducens nucleus mediates movements in the horizontal dimension only, and all movements in the vertical dimension are mediated by cranial nerves originating in the midbrain. The computational anatomy shown in Figure 3.4 satisfies these constraints, and is remarkably consistent with the patterns of excitatory and inhibitory connectivity of the brainstem (Chapter 1 and (Parr and Friston 2017a)).

To illustrate the neuronal plausibility of this computational anatomy, electrophysiological responses of cells in each region were simulated by taking the representations of each variable, as shown in the plots in Figure 3.3, and converting them into raster plots. The first two raster plots in Figure 3.4 show the firing rates of two of the three neuronal populations in the superior colliculus. As outlined in Chapter 1, the colliculus contains cells with three distinct electrophysiological phenotypes: 'burst', 'fixation', and 'build-up' cells (Munoz and Wurtz 1995a). Burst cells fire at the start of a saccade, as can be seen in the first raster plot. This cell type is known to di-synaptically inhibit cells in the Raphe nucleus interpositus (RIP) (Yoshida et al. 2001). This is consistent with the computational anatomy here, as there is an excitatory connection to a second collicular population that has inhibitory connections to the RIP. Both physiologically and anatomically, this cell type appears to be consistent with prediction error units signalling visual prediction (or 'retinal-slip') errors of the type implicated in models of eye movement (Krauzlis and Lisberger 1989). Fixation cells are active while a fixation is maintained. The second firing rate plot shows a cell that is active maximally only during fixations in one direction. These cells are known to project directly to cells in the RIP (Gandhi and Keller 1997), again showing consistency with our proposed anatomy. These cells appear to signal the expected hidden cause. Build-up cells have yet another distinct phenotype and must be assigned to the only remaining collicular cell type in Figure 3.4, which signals the error in the expected hidden cause. We discuss this cell type in more detail below, but first turn to a key target of projections from the superior colliculus.

The RIP contains a population of cells known as 'omnipause' cells (Büttner-Ennever et al. 1988). These cease firing at the start of a saccade, but are active during fixations. This corresponds well to the third raster plot that shows a decrease in activity locked to each saccade. This signal is the prediction error related to the hidden states encoding current eye position. Neurons in the RIP inhibit those in the rostral interstitial nucleus of the medial longitudinal fasciculus (thought to coordinate vertical saccades (Büttner-Ennever and Büttner 1978)) and in the parapontine reticular formation (that coordinates horizontal saccades (Cohen et al. 1968; Henn 1992)) (Strassman et al. 1986). The fourth and fifth rows of raster plots show neurons in the latter area. These neurons show bursting activity that triggers a saccade, here

related to the error in positional (proprioceptive) sensations. We have simulated such neurons representing saccades to either side of space.

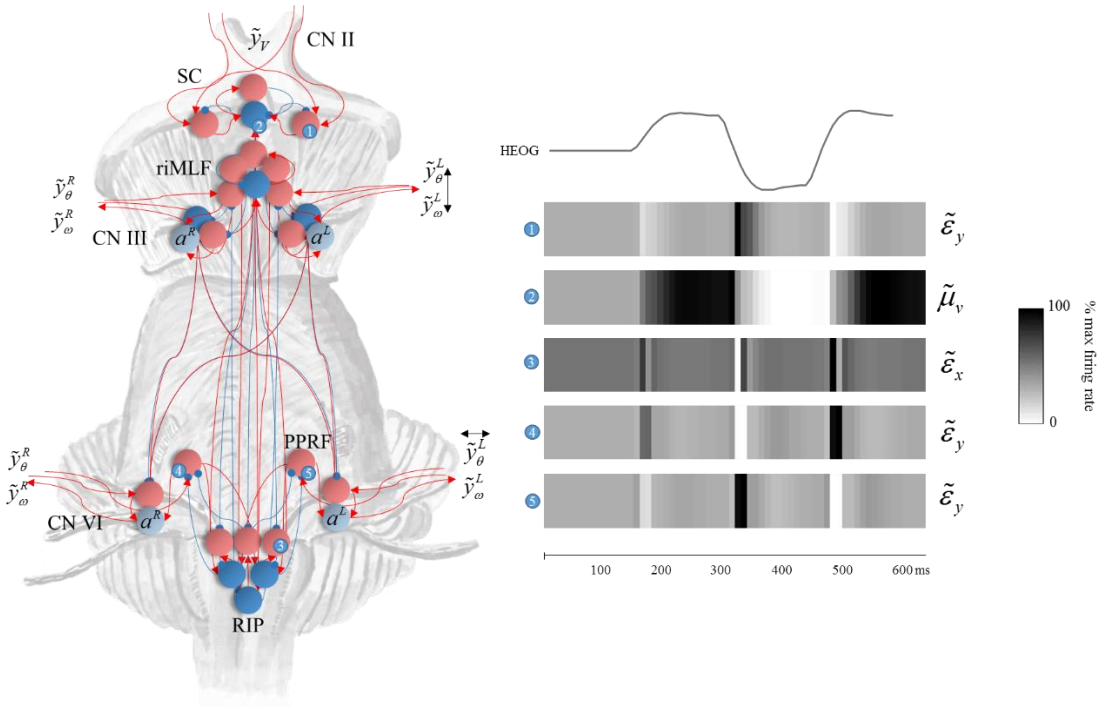

Figure 3.4 – The computational anatomy of oculomotion. *On the left* of this schematic, we show a plausible anatomical implementation of the Bayesian filtering equations in Figure 3.2. This satisfies the connectivity constraints described in the main text. Note that we have included motor neurons (grey) that represent action. As Figure 3.2 indicates, these only receive direct influences from the prediction error units at the sensory level. *On the right*, we show the simulated neuronal activities, along with a horizontal electro-oculographic (HEOG) trace indicating the eye position. Each of the numbered raster plots is associated with a particular neuronal population indicated by numbers in little blue circles. See the main text for a description of these units and Figure 3.2 for their equivalent location in the computational architecture. SC = superior colliculus; riMLF = rostral interstitial nucleus of the medial longitudinal fasciculus; PPRF = para-pontine reticular formation; RIP = raphe interpositus nucleus. Compare the anatomy here with that illustrated in Figure 1.1.

The pattern of activity of the build-up cells is very interesting, when viewed at a population level (Lee et al. 1988; Munoz and Wurtz 1995b). To simulate the spatiotemporal characteristics of electrophysiological responses in collicular build-up cells during saccades,

we treated the retinotopic location vectors (i.e. the horizontal and vertical components of the error) as encoding the peaks of activity in the superior colliculus. This enabled us to generate simulated responses of colliculus neurons in which (Gaussian) 'bumps' of activity moved over a retinotopic map, similar to those elicited in computational models of the superior colliculus (Bozis and Moschovakis 1998; Richert et al. 2013; Seung 1998; Seung et al. 2000; Trappenberg et al. 2001). In turn, this enabled us to simulate spatiotemporal responses that would have been observed (by assuming a fixed shape of bump); either by imaging perisaccadic population responses in the deep layers of the superior colliculus (see Figure 3.5A) – or unit responses at any particular location – over time – in terms of perisaccadic time histograms (see Figure 3.5B). The post stimulus (saccade) time histograms bear a remarkable similarity to empirical results of the sort shown in Figure 3.5C (Munoz and Wurtz 1995b).

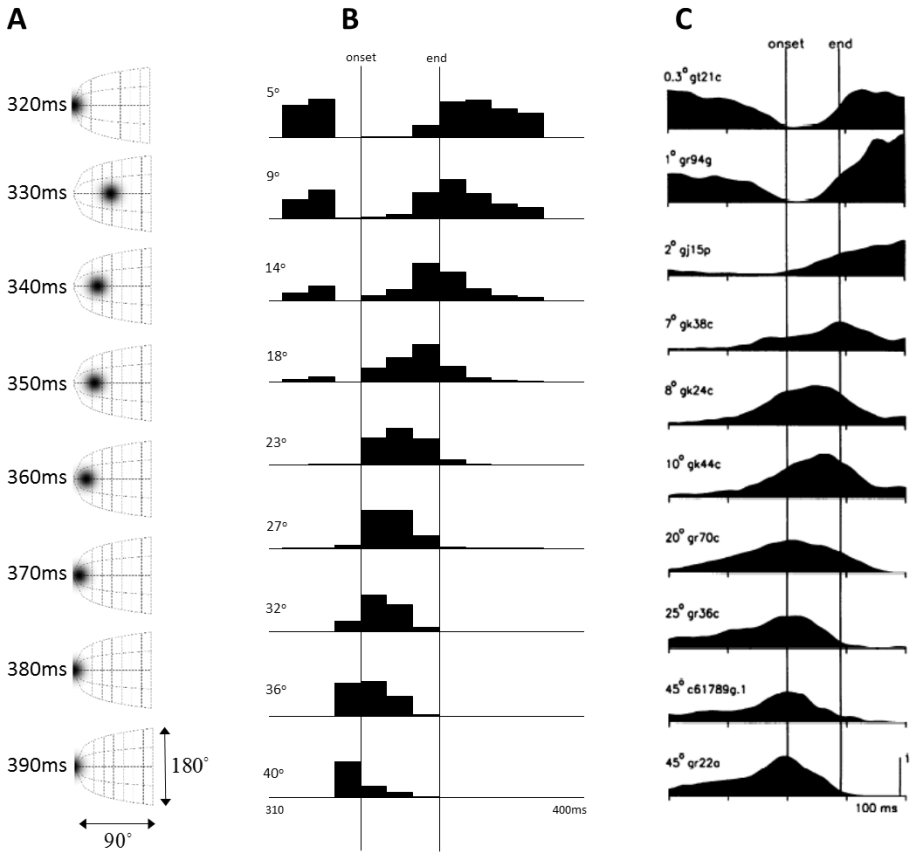

Figure 3.5– Collicular 'build-up' cells. This shows the population activity in collicular build-up cells during one of the saccades illustrated in Figure 3.3 (left). Our simulated build-up cells are those that signal the error in the hidden cause (target fixation location). A shows this as if we had imaged the right superior colliculus, which represents the left side of space. We have made use of the known retinotopy of the colliculus (Quaia et al. 1998) to plot this activity. B shows a set of simulated recordings of single cells from the onset to end of the saccade. Each

cell represents a different retinotopic location, indicated by the angles given for each plot. The the eccentricity increases with each row. C shows real data (adapted from (Munoz and Wurtz 1995b)) from single unit recordings of build-up cells in the superior colliculus.

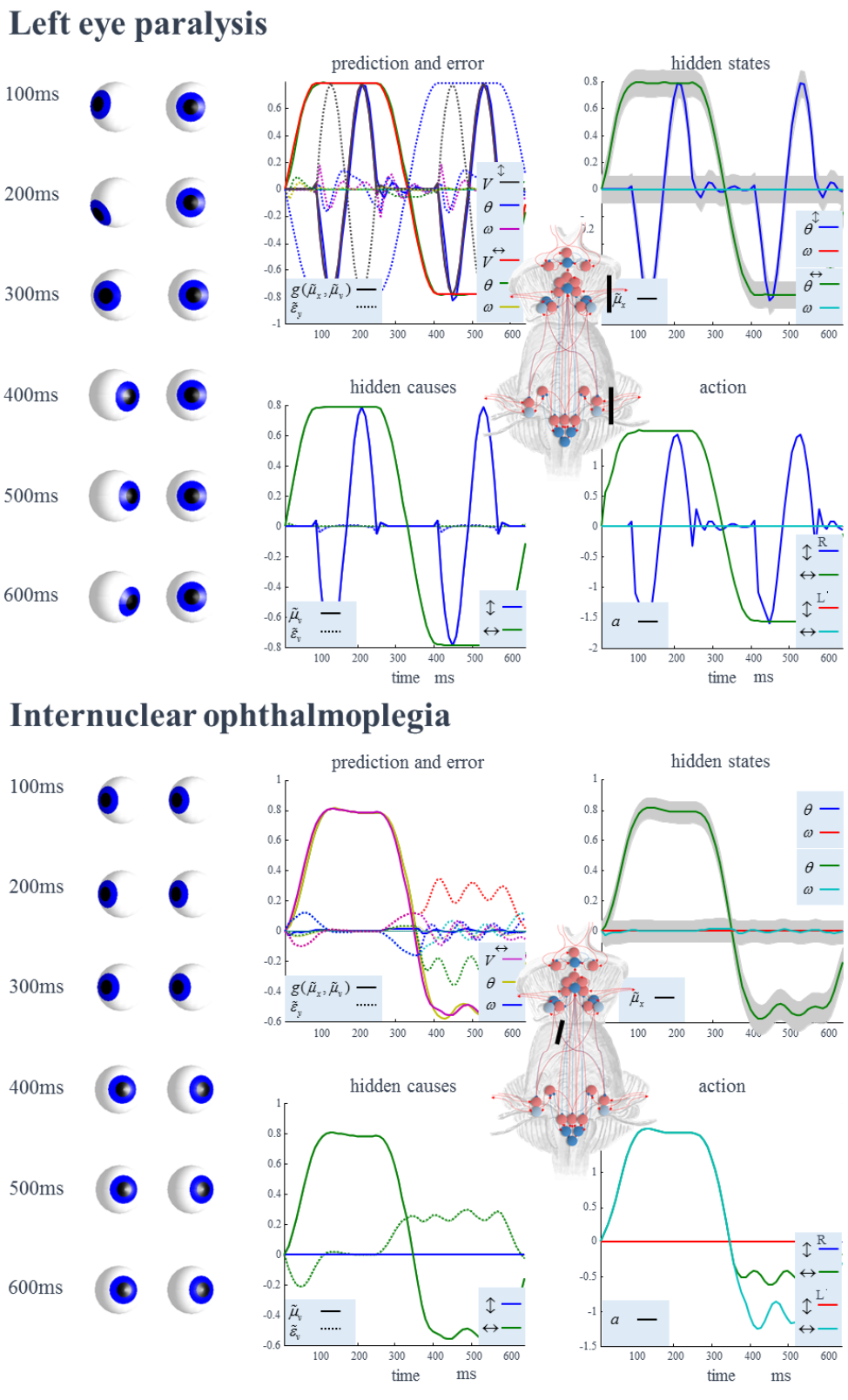

Figure 3.6 – Computational lesions. These plots demonstrate the consequences of simulated lesions. The first is a lesion of all the connections between the brainstem and the extraocular muscles of the left eye. As both the plots and the simulated eyes show, this causes a paralysis of the left eye, in keeping with what we would expect. Below this, we show the consequences of a lesion to the medial longitudinal fasciculus. The images and the plot of 'action' show that rightward gaze occurs normally in both eyes, but that leftward gaze reveals a deficit. The right eye fails to adduct to the same degree as the left abducts, and this induces nystagmus in both eyes – primarily the left. This is known clinically as an internuclear ophthalmoplegia. Please see refer to Figure 3.3 for an explanation of these plots.

Computational lesions

Having demonstrated the anatomical and physiological plausibility of an active inference formulation of oculomotor control, we used the anatomical constraints underwriting the computational anatomy in Figure 3.4 to motivate simulated lesions. Our first lesion removed all the connections that travel in the oculomotor cranial nerves on the left. This is to demonstrate that the simulation reproduces sensible results; i.e. the paralysis of the left eye (Figure 3.6). Computationally, this disconnection precludes the receipt of sensory data by proprioceptive prediction error units for the left eye and disconnects action units from the extraocular muscles.

The second simulation aims to model a subtler lesion: damage to the medial longitudinal fasciculus, that travels from the abducens (CN VI) nucleus in the pons to the contralateral oculomotor (CN III) nucleus in the midbrain, causes a clinical sign referred to as an 'internuclear ophthalmoplegia'. This is commonly seen in demyelinating conditions, such as multiple sclerosis, that induce white matter lesions. This pathology represents a disconnection syndrome (Catani and ffytche 2005) that manifests as a failure of conjugate control of eye movements. Figure 3.6 shows the results of performing this lesion *in silico*. Our lesion disrupts the signal from the left CN VI to the right CN III (see Figure 3.4). Computationally, this represents a disconnection between error and expectation units encoding horizontal positional error and angular velocity respectively. As in real patients, both eyes can look to the right normally. However, when looking to the left, the left eye can look laterally, but the right eye fails to keep up while moving medially. This violation of conjugacy induces nystagmus in the

(healthy) left eye. In our simulation, nystagmus is seen in both eyes, but more the left than the right. The deficit is most obvious in the plot labelled ‘action’.

Bayesian filtering in the brainstem

We have demonstrated in the above that, given a prior belief about anticipated fixation locations (η), Bayesian filtering can be used to generate movements that fulfil these beliefs. An important issue relates to the source of these priors. In predictive coding, there are typically higher hierarchical levels in play that send descending messages (predictions) to the lower level (Kiebel et al. 2008). These are used to derive the (empirical) prior beliefs at the lower level. In short, in this section we have focused on the lowest level of deep (hierarchical) active vision that translates predictions about "where I am going to look next" into oculomotion that realises these predictions. As the predictions (η) enter the Bayesian filtering equations to form prediction errors (ε), any descending connections would have to target units encoding these prediction errors. The anatomy of connections to the superior colliculus therefore hints at the anatomy of higher levels generating top-down predictions (Parr and Friston 2017a). This anatomy includes projections from the frontal eye fields (Fries 1984) and the substantia nigra pars reticulata (Hikosaka and Wurtz 1983). We will attempt to address the role of these connections in the third (*From decisions to movements*) section of this chapter, and to link them to the decision processes we have previously attributed to cortical and subcortical regions (Parr and Friston 2017b). This will be essential in order to account for more complex, oculomotor behaviour, including the spatial patterns of saccadic searches their resemblance to ‘Lèvy flights’ (Brockmann and Geisel 1999; Roberts et al. 2013).

There are some subtle differences in the neuronal responses we have simulated (Figure 3.4) compared to those measured in real neurons. For example, our simulated burst neurons show not only an increase in firing before a saccade in a given direction, but also a decrease in firing rate before a contralateral saccade. When these neurons have been interrogated *in vivo* (Munoz and Wurtz 1995a), a directional sensitivity of this type has been demonstrated. The firing rate of a burst neuron is higher when a saccade is performed in one direction compared to a saccade in the opposite direction. However, there is no clear decrease in activity, relative to baseline firing rate, in response to a saccade contralateral to the preferred direction of a burst neuron – as seen in our simulations. There are several possible explanations for this discrepancy. One is that, as firing rates cannot be negative, the positive and negative parts of the variables encoded by our synthetic neurons are really represented by different groups of burst neurons.

A second possibility is that the mapping between these variables and neuronal firing rates is a convex function. If this is the case, we would expect very small changes in firing rate for a change in a variable at the lower end of the scale compared to those induced by the same change at higher values. The low baseline firing rate of burst neurons (Munoz and Wurtz 1995a) supports this interpretation.

In addition to the oculomotor syndromes simulated here, an interesting next step would be to consider a broader range of pathologies. For example, schizophrenia is a psychiatric disorder associated with subtle oculomotor abnormalities, including changes in smooth pursuit eye movements (Thaker et al. 1998). Previous research using this form of modelling has been useful in characterising this kind of deficit in terms of abnormal estimates of precision in the generative model (Adams et al. 2012). In addition, eye movement signs are ubiquitous in neurology (Anderson and MacAskill 2013). To take this model forward – to address cardinal oculomotor deficits in psychiatry and neurology – we may need to develop a more complete model that, in addition to accounting for visual and proprioceptive data, accounts for vestibular inputs. This is likely to be important in the development of nystagmus due to cerebellar or brainstem damage (Troost 1989).

Summary

In this section, we have demonstrated that active inference provides a sufficient and principled account of oculomotor forces that fulfil prior beliefs about eye movements. By using a generative model that is common to both eyes, we enforce conjugate eye movements. When we map the ensuing Bayesian filtering equations to their associated process theory; namely, predictive coding, we find a connectivity structure that is remarkably consistent with the neuroanatomy of the oculomotor brainstem. Once this anatomical assignment is made, it is possible to simulate saccade-related responses we would expect to record from these regions with an electrode. These were formally very similar to recordings from the homologous anatomical regions in the electrophysiological literature. Finally, we showed that anatomically motivated computational lesions reproduced the eye movement deficits seen in neurological patients. Two important outstanding questions based upon this account of how eye movements to a desired target are achieved are ‘how do we decide between alternative saccadic targets?’ and ‘how do we map these decisions to desired fixation locations in a continuous space?’ These are the focus of the next two sections of this chapter.

Decisions

In this section, we take a step back from the mechanics of oculomotion and consider how saccadic targets may be selected in the first place. Decisions involve the selection of one from several alternatives – in our setting, saccadic targets. As such, (Markov) decision processes (see Chapter 2) are a natural form for the generative model because they are defined on discrete state spaces, where planning and decision-making may be thought of as disambiguating between competing hypotheses about how to act. This perspective on 'planning as inference' (Attias 2003; Botvinick and Toussaint 2012) implies we must be able to score alternative plans according to their prior probability. As outlined in Chapter 2, this may be done by computing the expected free energy for each policy (e.g. saccade) and pursuing those policies with the lowest expected free energy. In the following, we will unpack the properties of the expected free energy in terms of its role in driving exploration and exploitation. We then introduce a simple parameterisation of the likelihood and transition probabilities of a Markov decision process using a Gibb's distribution. This affords the opportunity to simply manipulate an inverse temperature parameter (or precision) for each of these distributions, augmenting or attenuating different sorts of uncertainty. Exploiting this parameterisation, we illustrate through simulation how the expected free energy drives exploratory behaviour under different sorts of uncertainty. We find that established psychological phenomena emerge from this treatment.

Planning as inference

In Chapter 2, we introduced the expected free energy as a means of scoring the prior probability of alternative paths (or policies) that we could pursue:

$$\begin{aligned} G(\pi) &= E_{P(\tilde{o}|\tilde{s})Q(\tilde{s}|\pi)}[\ln Q(\tilde{s}\mid\pi) - \ln P(\tilde{o},\tilde{s})] \\ &\approx E_{Q(\tilde{s}|\pi)}[H[P(\tilde{o}\mid\tilde{s})]] + D_{KL}[Q(\tilde{o}\mid\pi)\,\|\,P(\tilde{o})] \\ &= \underbrace{\underbrace{E_{Q(\tilde{s}|\pi)}[H[P(\tilde{o}\mid\tilde{s})]]}_{\textit{Ambiguity}} - \underbrace{H[Q(\tilde{o}\mid\pi)]}_{\substack{\textit{Predicitve}\\ \textit{Entropy}}}}_{(-)\ \textit{Information gain}} - \underbrace{E_{Q(\tilde{o}|\pi)}[\ln P(\tilde{o})]}_{\textit{'Preferred' outcomes}} \end{aligned} \tag{3.1}$$

The decomposition of the expected free energy in Equation 3.1 lends some intuition as to the sorts of policies that are favoured under active inference. The first term in the third line is the ambiguity, which quantifies the uncertainty in the mapping between a hidden state and the outcome it causes. Low ambiguity would imply a high fidelity in the relationship between the two. The second term is the uncertainty associated with predicted outcomes (Hwa 2004; Lewis and Gale 1994; Shewry and Wynn 1987). Heuristically, this may be thought of as how uncertain I am about what I would see if I looked there. Together, these terms quantify the information gain expected on performing a given policy. They appear in a range of fields under various names, including Bayesian surprise or salience (Itti and Baldi 2006), intrinsic motivation (Biehl et al. 2018; Oudeyer and Kaplan 2007), and as an objective function for experimental design (Lindley 1956). The last of these endorses the metaphor of the brain as a scientist, seeking to perform those experiments that elicit unambiguous data conditioned upon those variables it aims to infer, but avoiding those experiments for which it can already confidently predict the data it would be measure. In other words, the best experiments are those that address those things about which we are most uncertain (predictive entropy), but only if there is a potential to resolve this uncertainty (ambiguity). The final term in the expression provides a caveat to the scientific metaphor (Bruineberg et al. 2018). This biases policy selection towards those likely to result in preferred data (i.e. those data considered most probable *a priori*). This suggests a trade-off between choosing epistemically valuable (exploratory) policies and those policies that fulfil prior preferences (exploitative). In this section, we specify all outcomes to be equally preferred, such that the exploratory drive dominates. To illustrate how the potential for information gain influences policy selection, we manipulate two forms of uncertainty, and simulate the resulting behaviours.

Uncertainty and precision

Biological systems – like ourselves – are constantly faced with uncertainty. Despite noisy sensory data, and volatile environments, creatures appear to actively maintain their integrity. To account for this remarkable ability to make optimal decisions in the face of a capricious world, we propose a generative model that represents the beliefs an agent might possess about their own uncertainty. In this section, we address the computational basis for the representation of uncertainty by the brain, and its consequences for epistemic behaviour. We focus on two sources of uncertainty; uncertainty concerning the temporal evolution of environmental states, and uncertainty about the mapping from (hidden) states of the world to sensory observations. The first source of uncertainty corresponds to the ‘volatility’ of state transitions, while the

second corresponds to sensory noise and ambiguity. The latter has previously been addressed in the context of predictive coding, in which sensory precision (i.e., inverse variance) modulates the (possibly attentional) gain of ascending prediction errors (Feldman and Friston 2010). This modulatory effect is a direct consequence of (Bayes) optimal evidence accumulation (c.f., the Kalman gain of Bayesian filters in engineering). This formulation of attention appeals to the notion of the brain as a statistical organ: an organ that infers the causes of its sensations using internal models of how sensory impressions are generated by continuous states of the world. Here, we consider the role of precision in discrete state space models.

To equip our generative model (Figure 2.2) with beliefs about the uncertainty in both the transitions of hidden states (i.e., state precision), and the likelihood mapping from hidden states to outcomes (i.e., sensory precision), we introduce precision parameters. These are inverse temperature parameters, analogous to γ used for the policy prior (Equation 2.15). We first augment the likelihood distribution with a sensory precision, ζ:

$$P(o_\tau = i \mid s_\tau = j, \zeta_j) = \frac{\mathbf{A}_{ij}^{\zeta_j}}{\sum_i \mathbf{A}_{ij}^{\zeta_j}} \tag{3.2}$$

This is a Gibbs measure, commonly expressed as a softmax function, for which the denominator is a normalising constant (partition function). In this equation, ζ is the analogue of precision in predictive coding formulations of attentional gain (Feldman and Friston 2010). Note that each value *s* can take is associated with its own precision. The same approach can be followed to define the precision of state transitions (ω):

$$P(s_{\tau+1} = i \mid s_\tau = j, \pi, \omega) = \frac{\mathbf{B}_{\pi ij}^{\omega}}{\sum_i \mathbf{B}_{\pi ij}^{\omega}} \tag{3.3}$$

Having specified a simple means of parametrically changing these precisions, we show through simulation that epistemic foraging is heavily influenced by the beliefs an agent has about the dynamic and sensory precisions of their environment (Doya 2008). The temporal dynamics of visual search, including the phenomenon of 'inhibition of return', follow naturally from this formulation. The formal contribution of sensory precision to the information gain (via the ambiguity) in Equation 3.1 dissolves the 'dark room problem' (Friston et al. 2012c)

associated with active inference, without needing to invoke additional prior beliefs (Friston et al. 2012a).

Simulated visual foraging

To illustrate the influence of beliefs about uncertainty on behaviour, the generative model of Figure 3.7 (upper right) was used to simulate epistemic foraging. The generative model includes four stimuli, whose identity can change stochastically. These stimuli are mapped, noisily, to observable outcomes. Each stimulus is associated with a hidden state that defines its identity. An additional hidden state is the current eye position, that determines which of the stimuli is observed. This is associated with an identity mapping to a proprioceptive outcome indicating the current eye position, in a manner consistent with previous MDP models of saccadic eye movements (Mirza et al. 2016). In brief, this means that given the hidden states (namely, where the agent looks and the states of the stimulus at that position) one can generate probabilistic outcomes (proprioceptive information about where the agent is looking and exteroceptive outcomes reporting stimulus identity).

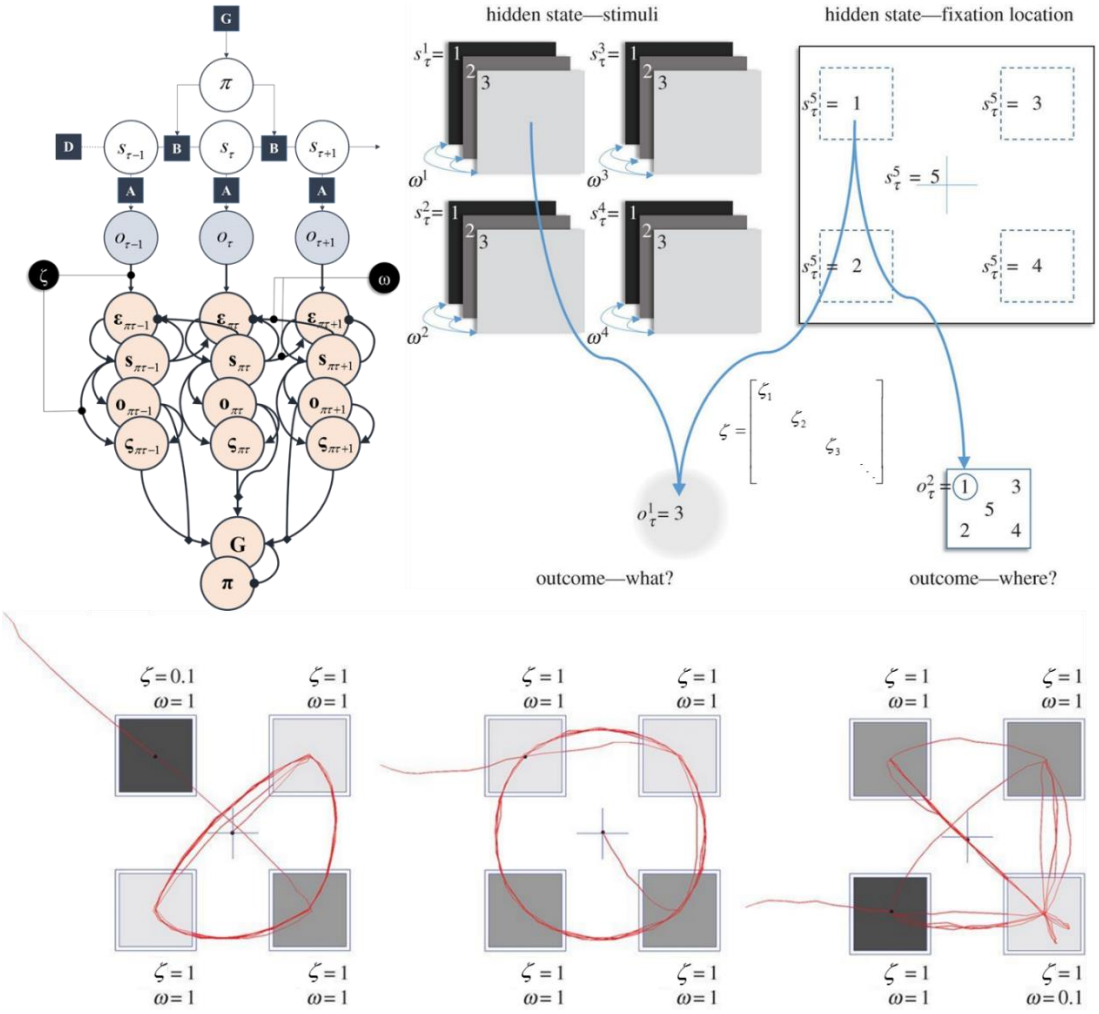

Figure 3.7 – Epistemic visual foraging. The *upper left* part of this figure shows the generic structure of the neuronal message passing from Figure 2.2 but highlights the points at which the parameters introduced in this section (ζ and ω) influence belief-updating. The likelihood precision (ζ) acts to control the gain of the influence of the outcomes (via the **A**-matrix) on prediction errors (**ε**), and the contribution of the expected ambiguity to the expected free energy (via **ς**). The precision of transitions (ω) determines the influence of beliefs about the present on beliefs about the future and vice versa. The *upper right* schematic shows the form of the specific generative model used for these simulations, with 5 hidden state factors (stimuli in four locations and eye position) giving rise to proprioceptive and visual data. Each eye position is associated with its own likelihood precision parameter, and each of the stimuli is associated with its own transition precision. The three plots along the *bottom row* of this figure show simulated eye-tracking traces constructed by solving the MDP for 8 saccades. These illustrate uniform saccadic sampling of the stimuli when all locations are associated with equal precisions, neglect of stimuli with low likelihood precision, and a bias towards volatile stimuli (with low ω).

The behaviour observed in the simulations can be explained by referring to the agent's beliefs about policies, and specifically the information gain (epistemic value, or salience) that contributes to the expected free energy (Equation 3.1). This says that the greater the expected change in beliefs, the lower the expected free energy. For a location associated with a low sensory precision (i.e. poor quality visual data), an observation is unlikely to elicit a substantial change in the posterior, so a saccade to such a location is less likely to be selected, as is shown in the lower left simulation in Figure 3.7. Heuristically, this is why a well-lit room, with precise sensory information, is preferable to a dark room; i.e., precise sensory cues that resolve ambiguity have greater epistemic affordance and are more likely to be sampled. More colloquially, this sort of behaviour recapitulates the joke about the drunkard looking for a lost key under a streetlamp (the "Streetlight effect" (Demirdjian et al. 2005)). Notably, the drunkards 'cognitive bias' is entirely Bayes optimal on an active inference view.

The greater frequency of saccades to stimuli with a higher volatility (lower right panel of Figure 3.7) can be similarly explained. On making an observation at a location and updating our beliefs about the hidden state causing it, we should be able to predict with greater confidence what we would observe by looking there again. More formally, the posterior predictive entropy (see Equation 3.1) will be smaller for this location relative to others. Recent observations are thus associated with a lower salience that then gradually increases over time, as the probability that the hidden state has transitioned to a new value increases, and is associated with a corresponding increase in the contribution of the predictive entropy to the expected free energy. In summary, knowing the state of a stimulus is a particular location means there is no further information to be gained by sampling that location, and it loses its salience. Note that salience is an attribute of both the world and the agent's beliefs about the world. However, if the stimulus can change, the salience of its location will increase slowly over time with uncertainty about its current status.

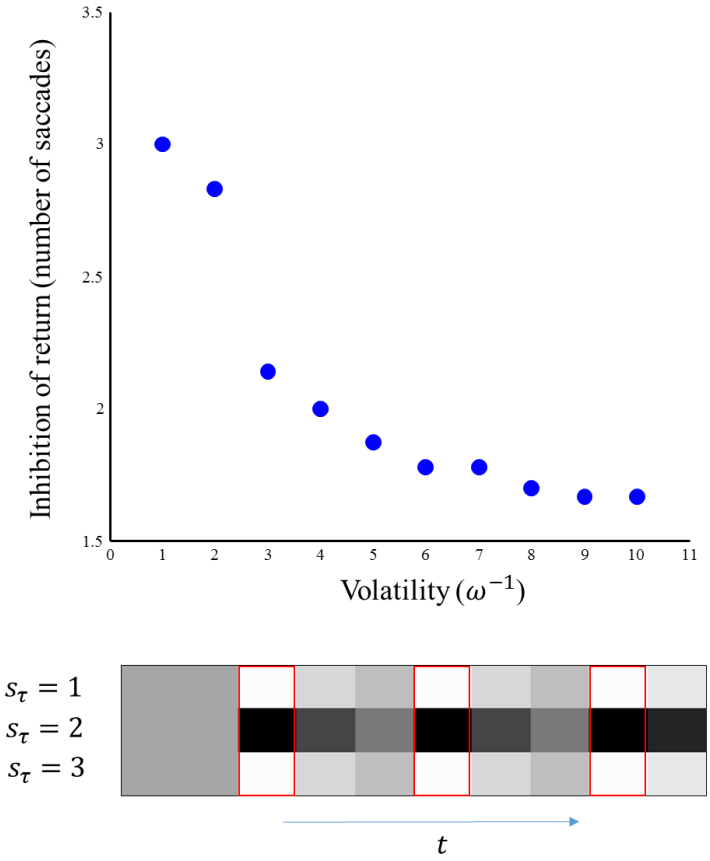

Figure 3.8 – Accumulating uncertainty and inhibition of return. This figure shows the inhibition of return, quantified by the average number of saccades before revisiting a location, under different levels of volatility. This illustrates the point that, as transitions become less deterministic, the salience of a location accumulates more rapidly. The plot below highlights one of the hidden state factors (one of the four stimuli), and represents the posterior probability that it is in one of three states (rows) at each time-step. As in Figure 2.3, black and white indicate probabilities of one and zero respectively, while intermediate probabilities are in shades of grey. Each column represents a single fixation. The location in question was fixated during the times outlined in red. Note the confident inference that the state was 2 during these fixations, but the gradual decrease in this confidence between fixations. This location is re-fixated only when the uncertainty in beliefs has dropped sufficiently that the posterior predictive entropy (and consequently salience) exceeds that of alternative fixation locations.

This phenomenon is consistent with the 'forgetting slopes' determined by calculating the error in reports about a stimulus at different times following presentation (Pertzov et al. 2013), and with theoretical analyses of the properties of the synthetic networks used to explain the maintenance of working memory signals (Burak and Fiete 2012). The concept of 'inhibition of return' (Klein 2000; Posner et al. 1985) naturally emerges from this formulation, as an agent

becomes less likely to return to the same location for a temporally limited period following a fixation. Figure 3.8 shows the accumulation of uncertainty over time since fixation, and illustrates how different levels of volatility influence the duration of inhibition of return. Formulating visual foraging in this way means that the ω parameter can be estimated from real subjects simply by measuring the length of the inhibition of return.

Summary

In summary, using a very simple but plausible formulation of active inference in the context of searching a simple visual scene, we find a natural explanation for two key phenomena in visual search; namely the attractiveness of salient, uncertainty reducing target locations and inhibition of return that depends upon the volatility of a visual scene. Crucially, both of these phenomena are rest on encoding the uncertainty or precision of state transitions and the generation of (visual) outcomes from hidden states. These simulations illustrate key aspects of the use of expected free energy to score alternative saccadic policies, and to select among competing fixation locations. In Chapter 4, we will consider the neuronal (i.e. neuromodulatory) encoding of uncertainty and precision. Before doing so, we will attempt to synthesise the (discrete-state space) modelling employed in this section with the continuous dynamics outlined at the start of this chapter. In other words, the final section of this chapter will deal with the reciprocal interaction between inferring where to look, and how to look there.

From decisions to movements – and back again

The nervous system faces a dual challenge in shaping behaviour. To induce changes in the external world, it is necessary to contract muscles or to secrete chemicals. Such processes necessarily involve the manipulation of continuous variables; muscle length or chemical concentration. In addition, animals must make decisions. To do so, they must entertain several different possible courses of action, or 'policies'. Ultimately, they must select one of these actions or policies that are necessarily discrete. We draw upon recent work that considers the interactions between the neuronal processing of discrete and continuous quantities (Friston et al. 2017c). To make this more concrete, and consistent with the discussion so far, we focus on the oculomotor system. As outlined in the preceding sections, sampling the visual world entails

making decisions about where to look, and implementing these decisions by contraction of the extraocular muscles.

We use perceptual inference performed by the networks supporting eye movements as a way of motivating and illustrating the theoretical challenge we want to address. However, the treatment offered below generalises to any system that involves the physical implementation of categorical decisions. The ideas presented in this section complement previous treatments of cognitive time (VanRullen and Koch 2003) including the notion of a 'perceptual moment' (Allport 1968; Shallice 1964; Stroud 1967), and the suggestion that brain oscillations act as discrete clocks to support this type of computation (Buschman and Miller 2009; Buschman and Miller 2010b). They also resonate with recent developments in machine learning (Linderman et al. 2017) and some of the problems faced in modern robotics (Cowan and Walker 2013; Schaal 2006). In short, the coupling of categorical decision making and dynamic perception (and motor control) raises some deep questions about the temporal scheduling of perception (and action).

Oscillatory rhythms in measured brain activity have been linked to cyclical perceptual processes (Buzsaki 2006), with theta and alpha cycles as the most popular hypothesised units of perceptual time (VanRullen 2016). In endorsement of this, the timing of processing relative to the phase of certain oscillations appears to be important (Buzsáki 2005). While there is some controversy concerning the frequency of the perceptual clock, an advantage to focusing on the oculomotor system is that we can evade this issue. The frequency of spontaneous saccadic sampling is around 4 Hz, allowing us to commit to a theta rhythm. Conveniently, this is the frequency often associated with attentional and 'central executive' (decision) functions (Chelazzi et al. 1993; Duncan et al. 1994; Hanslmayr et al. 2013; Landau and Fries 2012; VanRullen 2013), as opposed to sensory processes associated with faster frequencies (Drewes and VanRullen 2011; Dugué et al. 2011; Ergenoglu et al. 2004; van Dijk et al. 2008).

As reviewed in Chapter 1 (and (Parr and Friston 2017a)), the oculomotor system is a distributed network that includes brainstem, cortical, and subcortical regions. An important point of contact between the cortical oculomotor networks and those in the brainstem is the superior colliculus (Raybourn and Keller 1977), found in the midbrain. This structure receives a dual input from the cortex (Fries 1984) and the basal ganglia (Hikosaka and Wurtz 1983), and provides an important input to the brainstem oculomotor nuclei. In the following, we argue that the connectivity implied by active inference is consistent with a role for the superior colliculus as an interface between the discrete and continuous processing of the oculomotor system.

The inference problem addressed here is summarised in the Bayesian network of Figure 3.9, which extends the network of Figure 3.1 through the addition of a sequence of discrete (target) states that depends upon a (saccadic) policy. We first address the problem of translating between categorical and continuous random variables (Friston et al. 2017c), appealing to a technique known as 'Bayesian model reduction' (Friston et al. 2018; Friston et al. 2016c). We then consider the anatomical structures that could implement the requisite belief-updating and simulate a simple saccadic task that employs a mixed (MDP-Bayesian filter) generative model.

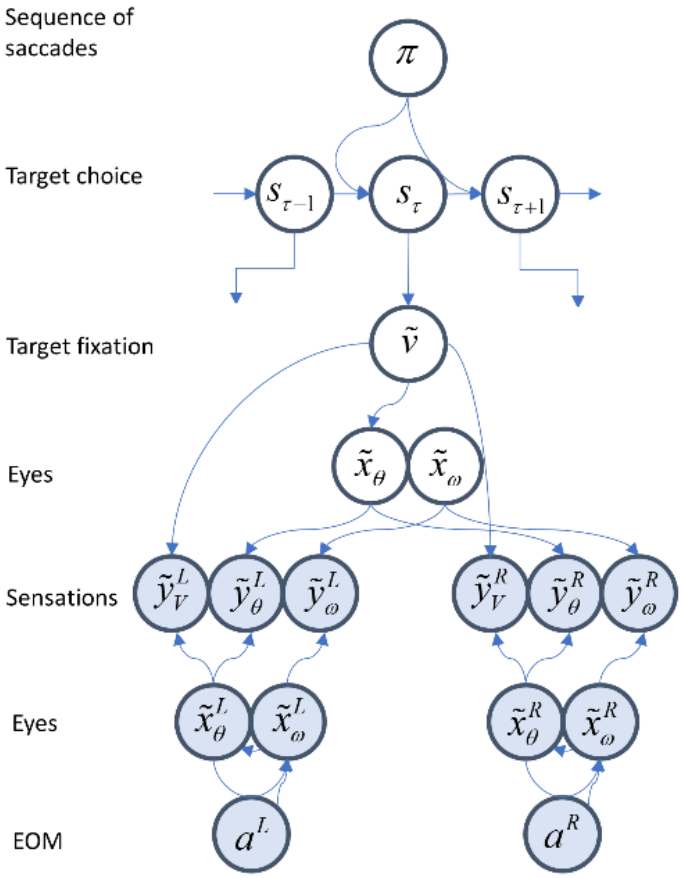

Figure 3.9 – From decisions to eye movements. This graphical model extends that shown in Figure 3.1 through the addition of an MDP structure describing a sequence of saccadic targets (*s*). At each time-step, these generate a target location in continuous space (*v*). The conditioning of continuous states upon discrete states is similar to the approach of a Gaussian mixture model, of the sort that underwrites clustering algorithms. This effectively treats each

discrete state as an alternative hypothesis, where each hypothesis is associated with a mean and precision (for simplicity, we assume that all of these hypotheses differ only in their means). The consequence of this is that the anticipated (prior) target fixation (v) is constructed through a weighted average of the means under each hypothesis, weighted by the probability of the discrete state associated with that hypothesis. This accounts for the influence of the discrete model over the continuous. To account for the reciprocal influence from the continuous to discrete, we can again appeal to the perspective of comparing hypotheses. This means we can compute the evidence the continuous model affords to each discrete hypothesis, and use this to form a posterior belief about the alternative targets that could have been chosen.

Bayesian model reduction

An MDP outcome – representing fixation location – corresponds to one of several discrete saccadic targets, defined in continuous coordinates (v). If we associate each target location with the attracting (fixed) points of some continuous oculomotor dynamics, the prediction from the MDP effectively defines an equilibrium point that will attract the subsequent eye movement; c.f., the equilibrium point hypothesis (Feldman 2009). If there is some uncertainty about the particular location of the target, we can specify the predicted target location through a Bayesian model average of each location associated with a discrete outcome hypothesis:

$$\begin{aligned} p(v \mid \mathbf{o}_\tau) &= \mathcal{N}(\mathbf{o}_\tau \cdot \eta, \Pi_v) \\ \mathbf{o}_\tau &= \boldsymbol{\pi} \cdot \mathbf{o}_{\pi\tau} \\ \mathbf{o}_{\pi\tau} &= \mathbf{A}\mathbf{s}_{\pi\tau} \end{aligned} \tag{3.4}$$

To recap, we have specified both discrete and continuous state space models. We have now supplemented these with Equation 3.4, which dictates how predictions of the former model can play the role of (empirical) prior beliefs in the latter. The next thing to specify is the process by which the continuous state space model informs the discrete model. In other words, what sort of evidence is passed from the continuous to the discrete part of the (active) inference scheme? In brief, the discrete part of the generative model provides prior constraints on the continuous part, while the continuous part reciprocates with Bayesian model evidence for the discrete hypotheses entertained by the discrete part to enable Bayesian belief updating. This updating entails the selection of alternative hypotheses (outcomes) that constitute empirical priors at the continuous level.

To adjudicate between these hypotheses, we need to compute the posterior probabilities over each outcome (e.g., target fixation). We give an abbreviated outline of this (Bayesian model reduction) procedure here: for more technical accounts, please see (Friston et al. 2018; Friston et al. 2016c). Given that the only difference between the models entailed by each discrete hypothesis is in their priors, we can write Bayes' rule for a full model, and for a 'reduced' model which assumes outcome *m* but has the same likelihood distribution. Dividing the terms on either side of the equality for the reduced model by the analogous terms for the full model (such that the likelihood cancels), we have:

$$\frac{p(v \mid y, \mathbf{o}_m)\, p(y \mid \mathbf{o}_m)}{p(v \mid y, \mathbf{o})\, p(y \mid \mathbf{o})} = \frac{p(v \mid \mathbf{o}_m)}{p(v \mid \mathbf{o})} \tag{3.5}$$

We can replace the posteriors here with those that we compute using the Bayesian filtering approach outlined above. Rearranging this, we get:

$$p(y \mid \mathbf{o}_m) q(v \mid \mathbf{o}_m) = \frac{p(v \mid \mathbf{o}_m)}{p(v \mid \mathbf{o})} q(v \mid \mathbf{o})\, p(y \mid \mathbf{o}) \tag{3.6}$$

Integrating both sides with respect to the hidden cause gives:

$$p(y \mid \mathbf{o}_m) = E_{q(v \mid \mathbf{o})}\left[\frac{p(v \mid \mathbf{o}_m)}{p(v \mid \mathbf{o})} \right] p(y \mid \mathbf{o}) \tag{3.7}$$

As the free energy approximates the negative log evidence, this can be rewritten as:

$$F[y, \mathbf{o}_m] = \ln E_{q(v \mid \mathbf{o})}\left[\frac{p(v \mid \mathbf{o})}{p(v \mid \mathbf{o}_m)} \right] - F[y, \mathbf{o}] \tag{3.8}$$

Notably, this means that the free energy for any hypothesis, *m*, can be calculated from the free energy of the full model without having to explicitly compute the posteriors associated with the latent variables in *m*. This is slightly technical point that, from a computational perspective, affords a very simple and efficient form of Bayesian model comparison. In other words, the evidence for different hypotheses or models at the discrete level can be computed directly and easily from the sufficient statistics of posterior beliefs encoded at the continuous level. In terms of neurobiology, Equation 3.8 speaks to the biological plausibility of message passing from the continuous to discrete domains.

To convert the evidence for each model back to discrete time, we integrate the model evidence (free energy) over the time period corresponding to one theta cycle. This is then combined with the prior over the model to give a vector (**L)** that can be passed through a softmax function to give the posterior over each outcome model, **r** (Friston et al. 2017c). This plays the role of a discrete observation or outcome from the point of view of the MDP (see Figure 3.10). This concludes our technical description of message passing between discrete and continuous parts of a generative model. A worked example of how this sort of could work in the brain is provided in the final section (using the update equations in Figure 3.10). To motivate the interpretation of these simulations we now consider the basic neurobiology of the oculomotor system and how its computational architecture could support inference of this sort.

The neuroanatomy of oculomotion

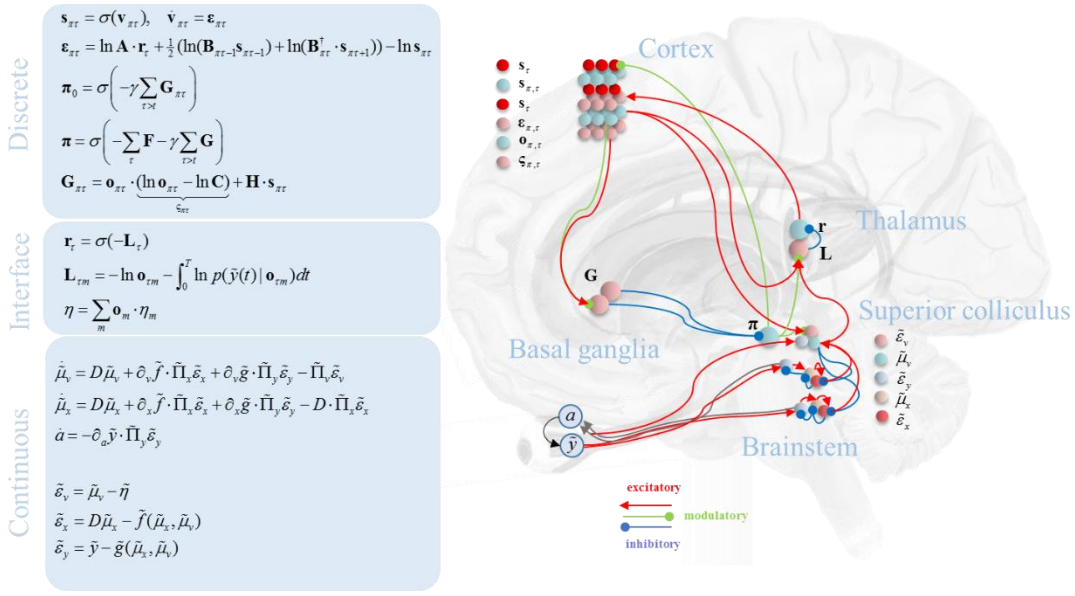

Figure 3.10 – The anatomy of oculomotion. This schematic illustrates the dependencies between the variables in the equations described in the main text, and summarised on the left.

It does so in the form of a neural network with populations of neurons assigned to plausible anatomical locations. There is a remarkable degree of neuroanatomical plausibility to these assignments; including a common laminar origin for cortical projections to the striatum, superior colliculus, and higher order thalamic nuclei. In addition, a dual cortico-subcortical input to the colliculus is necessitated by this scheme, as are the excitatory-inhibitory connections of the direct pathway through the basal ganglia. The equations in the box on the upper left describe marginal message passing in a Markov Decision Process. The lower box gives the Bayesian filtering equations of the sort usually associated with predictive coding. The middle box expresses the descending messages derived from Bayesian model averaging, and the ascending messages that result from model reduction. In previous papers, we show that this is likely to be represented in the mapping from higher cortical areas (Friston et al. 2017f), such as the dorsolateral prefrontal cortex – an area that houses representations that endure over a longer temporal scale (Parr and Friston 2017d) and connects to the frontal eye fields.

In the above, we described the problem the brain faces in making discrete decisions about where to look and the continuous inferences required to realise and update these decisions. We outlined the computations mandated by active inference in solving these problems, with a special focus on the message passing between discrete and dynamic domains. In this section, we associate these computations with their neurobiological substrates. While this assignment is speculative, it is constrained by both the anatomy of message passing and the presence (or absence) of connections in the brain. Figure 3.10 shows the consistency between the computational anatomy of oculomotion and the networks known to support oculomotor function. In the following, we describe the cortical, subcortical, and brainstem components of this network (Parr and Friston 2017a). This section concludes with an analysis of the superior colliculus, a structure uniquely placed to translate discrete decisions into target locations in a continuous state space.

The cerebral cortex is a laminar structure, with layer specific projections and terminations (Felleman and Van Essen 1991). The connectivity implied by inference using a Markov decision process closely resembles this pattern (Friston et al. 2017f). Specifically, the inference scheme we described above involves several distinct types of variable that receive messages from a subset of the other variables. This implies a stereotyped pattern of connectivity between these groups (or layers) of computational units. Consistent with cortical laminae, external input targets only one layer. Outputs of different types arise from defined

populations. In this section, we use known neuroanatomy to constrain the assignment of computational units to their appropriate laminae.

Layer IV of the cortex receives ascending connections from lower areas (Shipp 2007), or from first order thalamic nuclei. The computational units that receive this input are the error units (**ε**), suggesting that these occupy this layer. This also implies that **r**, the subcortical projection to layer IV, is likely to be represented by neurons in first order thalamic relay nuclei, such as the lateral geniculate nucleus (Herkenham 1980). Layer III gives rise to ascending connections. These are not shown here, but would arise from neurons encoding the state (**s**) at that hierarchical level (Friston et al. 2017f). For simplicity, we consider a single cortical area – the frontal eye field – omitting the parietal (Corbetta et al. 1998; Gaymard et al. 2003; Parr and Friston 2017b; Shipp 2004) and occipital (Bruce and Tsotsos 2009) contributions to this system.

Layer V of the cortex has several subcortical targets (Kasper et al. 1994; Ojima et al. 1996). It is the layer that houses the pyramidal cells of Betz in the motor cortex that project to lower motor neurons in the spinal cord. In addition, layer V gives rise to projections to the second order thalamic nuclei, such as the pulvinar, the superior colliculus, and the basal ganglia (Fries 1985; Hübener and Bolz 1988; Shipp 2007). As Figure 3 shows, units encoding predicted outcomes (**o**) send messages to all of these anatomical homologues. They participate in the evaluation of the expected free energy in the striatum, the model averaging of continuous time models by the superior colliculus, and Bayesian model reduction by the thalamus. That the **L** units of the thalamus receive cortical projections from layer V suggests that these neurons must be located in second order thalamic nuclei (Crick and Koch 1998; Rockland 1998; Sherman 2007).

As noted above, the basal ganglia receive input from cortical layer V, encoding predictions about discrete outcomes. This input, in addition to a signal from the error units (ς), is used to compute the expected free energy (**G**) for each policy. The basal ganglia are well recognised to be involved in policy evaluation (Gurney et al. 2001; Jahanshahi et al. 2015). Most of the cortical inputs to the basal ganglia target the striatum (Alexander and Crutcher 1990; Shipp 2017), implying the expected free energy is represented by medium spiny neurons in this structure. These give rise to inhibitory GABAergic projections to the substantia nigra pars reticulata, which itself projects to the superior colliculus (Hikosaka and Wurtz 1983). This 'direct pathway' connectivity is remarkably consistent with the influence of **G** on $\boldsymbol{\pi}$, and $\boldsymbol{\pi}$ on ε_v. The latter influence is in the Bayesian model averaging over expected outcomes to generate a prior mean (η) for the implementation of the policy in continuous time. The output nuclei of the basal ganglia participate in an additional Bayesian model averaging of hidden states. This

is mediated by modulatory projections (via thalamic relays) to superficial layers of the cortex (Haber and Calzavara 2009; McFarland and Haber 2002).

The brainstem is the source of the cranial nerves to the extraocular muscles. This suggests that brainstem structures engage in continuous message passing. We have previously demonstrated that the anatomy of this message passing is not only consistent with the connectivity of the brainstem, but also that it reproduces electrophysiological responses in these structures, and the same deficits as in neurological patients when lesioned ((Parr and Friston 2018a), and Figure 3.6). In addition, these nerves carry proprioceptive information from the muscles (Cooper and Daniel 1949; Cooper et al. 1951), while the midbrain receives optic nerve fibres from the retinotectal pathway (Linden and Perry 1983). Given beliefs about the current position and velocity of the eyes (μ_x), it is possible to make predictions about the resulting sensory input (Adams et al. 2012; Friston et al. 2012a; Perrinet et al. 2014). This induces a sensory prediction error (ε_y) that is minimised by action. This implies that the midbrain and pontine nuclei responsible for signals to the extraocular muscles must contain neurons that broadcast these errors. As the brainstem nuclei form the nodes of the network engaged in continuous inference, they must receive input from the region mapping decisions into this space. The obvious candidate for this region is the superior colliculus.

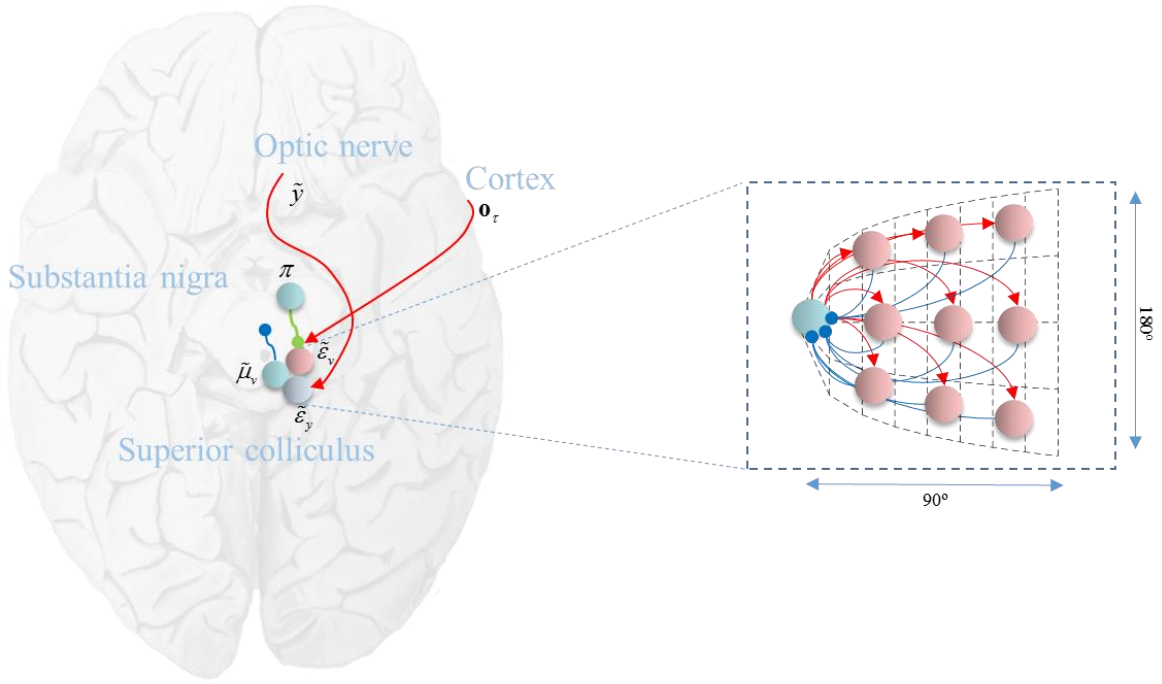

Figure 3.11 – The discrete-continuous interface. This schematic shows the connectivity between the neuronal populations in the superior colliculus in greater detail. The transverse section through the midbrain allows us to depict the terminations in the optic tectum from the optic nerve. We also illustrate the topographical arrangement of the fixation (rostral pole) and build-up (distributed throughout) cells, and the connectivity between these implied by our formulation. Note that burst neurons are the most dorsal, with build-up neurons found more ventrally. As the schematic shows, this would be consistent with the proposed extrinsic (between regions) connectivity, as each population is oriented towards the regions it is connected to. The intrinsic (within region) connectivity between build-up and fixation neurons is shown on the right, conforming to the known retinotopy of the colliculus (Paré et al. 1994; Quaia et al. 1998). The angles indicate the coordinates of the visual field represented at each point in the colliculus.

The superior colliculus is the interface between the forebrain and brainstem networks. It is the recipient of cortical (Hanes and Wurtz 2001) and basal ganglia projections (Hikosaka and Wurtz 1983) (Figure 3.11), and is intimately connected to the oculomotor system within the brainstem (Sparks 2002). As such, it sits at the anatomical boundary between the discrete and continuous networks. It is found in the dorsal midbrain, at the same level as the oculomotor

nucleus. Like the cortex, it is a laminar structure, with different electrophysiological responses in different subsets of cells. As outlined in the first part of this chapter (and in Chapter 1), there are three broad groups of these neurons, as illustrated in Figure 3.11. These are the 'burst', 'fixation', and 'build-up' cells (Ma et al. 1991; Munoz and Wurtz 1995a). To recap, we argued above (Parr and Friston 2018a) that these groups correspond to three different types of computational unit (Figure 3.4). Burst cells, which fire at the start of a saccade, have the properties we would expect from neurons signalling visual prediction error (ε_y). This is consistent with the fact that a subset of retinal ganglion cells synapse within the colliculus, and that some collicular cells respond to visual stimuli (Mays and Sparks 1980; Wurtz and Mohler 1976). Fixation neurons are active during fixations, and we have associated these with the expectation neurons encoding target fixation locations (μ_v). Consistent with the computational anatomy of Figures 3.10 and 3.11, it is this group that projects to the brainstem centres for saccade generation (Gandhi and Keller 1997).

Build-up neurons show a pattern of activation consistent with a population encoding (Anderson et al. 1998; Lee et al. 1988). A travelling 'hill' of excitation moves from a peripheral location towards the rostral pole of the colliculus during a saccade (Munoz and Wurtz 1995b). At a population level, these neurons can be thought of as expressing a prediction error between a target fixation and the current eye position (ε_v) (Sparks 1986). The movement towards the pole, representing the foveal location, can be thought of as encoding the reduction in prediction error as the eye moves closer to its target. That this occurs at the population level suggests that build-up neurons individually code for discrete spatial regions.

The discretised encoding of continuous variables by these units is consistent with their computational role, evaluating the difference between η, parameterising competing hypothetical models, and the estimated position in continuous coordinates, μ_v. Specifically, each neuron may encode the prediction error associated with a particular hypothesis, with activity weighted by the prior probability of that hypothesis (**o**). The conversion from discrete to continuous coordinates then simply requires that the connection strengths between these neurons and the fixation neurons at the pole vary with distance. The inhibitory connections (Munoz and Istvan 1998) from build-up to fixation neuron should be of a greater strength if the anatomical distance between the two is greater. In summary, electrophysiological properties corroborate the neuroanatomical evidence that the superior colliculus is the discrete-continuous interface of the oculomotor system, and the topography of build-up and fixation neurons hints at the computational mechanisms that map between them. In the final section, we use the neurobiological pointers established in this section to interpret simulated oculomotor control in terms of established electrophysiological responses in the oculomotor system.

Simulated electrophysiology

Our aim in performing these (minimal) simulations is to illustrate the interactions between the discrete and continuous domains of the oculomotor system. As such, we chose a simple behavioural paradigm: there are three possible fixation locations (left, right, and centre). We then set the prior preferences (through the **C** matrix); so that proprioceptive data is preferred that is consistent with central fixation initially, then with leftward fixation, rightward fixation, and finally central fixation again. This is consistent with the instruction to look at a sequence of targets at each of these locations. The structure of the model we used is as depicted in Figure 3.9. The continuous part employs the belief that the eyes are drawn towards an attracting location, and this is implemented by an action that has the effect of a Newtonian torque, as in the model used in the *Movement* section of this chapter above.

Figure 3.12 shows the results of simulating active vision in this model. Crucially, this type of simulation allows us to show what is happening at different neuroanatomical locations simultaneously, and gives a sense of the functional interaction of these areas. The results are presented in the form of raster plots, as if we had recorded from single neurons in each of the simulated brain areas. Representations in the cortex and basal ganglia update on a theta cycle (4Hz), while superior colliculus build-up cells translate this into continuous time for collicular fixation neurons. These induce (via brainstem circuits) changes in eye position, as shown in the simulated electro-oculography trace. Note that these neurons vary continuously with eye position, unlike the neurons in the discrete compartments – that encode the probability of an alternative fixation point.

Intuitively, we can see how the mapping from discrete to continuous occurs. At 250 ms, the cortex updates its representations. These cause the expected free energy under each policy to change, inducing updates in the striatum. Through the direct pathway, this causes inhibition in the substantia nigra pars reticulata, resulting in the selection of a new policy (‘look right’). A Bayesian model average over outcomes then results in the selection of superior collicular build-up neurons that represent the error between the current belief about eye position (given by the fixation neurons) and the anticipated rightward location. These then induce changes in the fixation neuron activity. During the next 100 ms, as the eyes move to fulfil this belief, we can see the resolution of the prediction error in the superior colliculus build-up layer through the movement of the ‘hill’ of activity towards the pole.

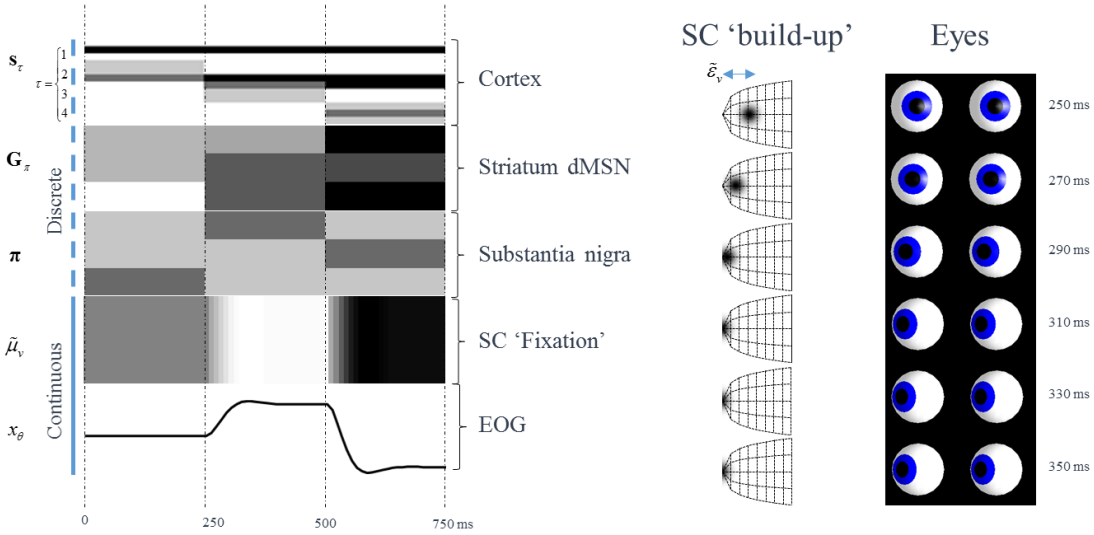

Figure 3.12 – Simulated electrophysiological responses. This figure shows the electrophysiological responses we would expect to observe under the process theory associated with active inference. On the left, we show neuronal firing rates (representing approximate posterior beliefs), depicted in the form of a raster plot. These are synchronised across all of the neurons shown. Darker colours indicate a greater firing rate. In the cortex (frontal eye field), we show neurons representing the possible fixation targets. The first three rows indicate the neurons representing left, centre, and right, at the first discrete time step (τ = 1). These three options are then replicated in the next three rows, but reporting the second time step. The third and fourth steps are similarly represented. In the striatum, direct pathway medium spiny neurons (dMSN) represent the expected free energy of each of the three policy options (saccade right, saccade centre, and saccade left). These then inhibit their corresponding neurons in the substantia nigra pars reticulata that represent the posterior beliefs about these policies. Note the inversion of striatal activity in the substantia nigra as a result of this inhibition. We show the activity of a superior collicular (SC) 'fixation' neuron to illustrate neuronal firing representing a continuous variable, with the horizontal electro-oculography trace below to depict the movement of the eyes. The right panel shows simulated activity across the superior collicular 'build-up' layer during a saccade, with the corresponding eye positions. Population activity is depicted in terms of a Gaussian intensity where the distance between the mean and the collicular pole is equal to the prediction error (as in (Parr and Friston 2018a)).

Summary

The approach we have described here is capable of reproducing a wide range of physiological and behavioural phenomena in the oculomotor system. Specifically, we have shown that the signals we have simulated bear a close resemblance to those measured in brainstem nuclei (Parr and Friston 2018a). Most strikingly, we found that simulated collicular 'build-up' neuron responses qualitatively reproduced single unit recordings published in the experimental literature (Munoz and Wurtz 1995b). In addition to these electrophysiological similarities, lesions to these models induce similar behavioural syndromes to those found in neurological patients with damage to the associated neuroanatomy. By disrupting neuronal message passing (i.e. inducing 'disconnection' syndromes (Geschwind 1965b)), we can simulate visual neglect (Chapter 5 and (Parr and Friston 2017b)) and internuclear ophthalmoplegia (Figure 3.6 and (Parr and Friston 2018a)). The white matter disconnections associated with these syndromes are the superior longitudinal fasciculus (Doricchi and Tomaiuolo 2003) and the medial longitudinal fasciculus (Virgo and Plant 2017) respectively. The locations of these synthetic lesions constrain the computational anatomy, and their nature endorses the notion that the brain engages in variational inference.

More generally, models based upon active inference have a high degree of face validity, in that they reproduce a wide range of neurobiological phenomena. These range from single cell responses, including place fields (Friston et al. 2017a) and midbrain dopamine activity (Friston et al. 2014), to evoked responses, including those associated with classic working memory tasks (Parr and Friston 2017d). They have been used to generate behaviours as diverse as exploration (Friston et al. 2015b; Mirza et al. 2016), handwriting (Friston et al. 2011), eye-blink conditioning (Friston and Herreros 2016), habit formation (FitzGerald et al. 2014), communication (Friston and Frith 2015), and insight (Friston et al. 2017b). In addition to these theoretical accounts, active inference has been used pragmatically to model behaviour and to characterise individuals according to the parameters of their prior beliefs (Adams et al. 2016; Mirza et al. 2018; Schwartenbeck and Friston 2016).

Given this broad applicability, the issues described in this chapter generalise beyond eye movements. Any neurobiological system that needs to make decisions and implement these via some physical effector must solve the problem we have described here. This is vital for (but not exclusive to) speech, locomotion, and autonomic regulation. Language is made up of discrete units (phonemes, words, sentences) that are expressed as continuous changes in auditory frequencies generated by contraction of the laryngeal (and pharyngeal) muscles (Simonyan and Horwitz 2011). Walking involves taking a series of discrete steps, each of

which requires a careful coordination of skeletal muscles (Ijspeert 2008; Winter 1984). Interoceptive states are frequently divided into discrete dichotomies including fed versus fasting (Kalsbeek et al. 2014; McLaughlin and McKie 2016; Roh et al. 2016), diastole versus systole, sympathetic versus parasympathetic (McDougall et al. 2014; Owens et al. 2018). Each of these induces continuous changes in enzyme activity, blood pressure, or smooth muscle contractions. The form of the message passing will be very similar for each of these processes, but the variables represented will differ. This suggests a similar pattern of cortico-subcortical connectivity, but differing regions of cortex, and different subcortical components.

In this chapter, we have chosen to focus on a fairly concrete problem – deciding where to look, and how to do this. For more abstract decisions, perhaps at higher hierarchical levels in the brain (Badre 2008; Badre and D'Esposito 2009; Christoff et al. 2009; Rasmussen 1985), it may be necessary to integrate beliefs across multiple modalities. A challenge for future work is to incorporate the set of beliefs that constitute an emotional state, as emotions are often thought to contribute to 'irrational' behaviours. It is not always easy to intuit how such behaviours might be Bayes optimal. One line of research into these issues frames them as questions about interoceptive inference (Ondobaka et al. 2017; Seth 2013; Seth and Friston 2016). Given beliefs about (abstract) variables that have both interoceptive and exteroceptive sensory consequences (Allen et al. 2019), it becomes clear that policies must minimise expected free energy in both domains. For example, a belief that a predatory animal is present implies that the sympathetic nervous system should be active, but also that visual data are consistent with the presence of said animal. Anatomically, these dependencies are consistent with the sensory and autonomic targets of the amygdala (LeDoux et al. 1988; Ressler 2010). A tachycardia then carries (weak) evidence for the presence of a scary animal, and could influence policy selection even in the absence of exteroceptive evidence. This suggests a framework in which an emotional state may influence decision making in an apparently irrational way that is entirely compatible with the formulation we have described here.

We hope to further the ideas in this work both theoretically and empirically. There are several important theoretical issues that will need to be addressed in greater depth than we have space for here. Among these is the need for a generalisation of the inferences required for oculomotor decisions (and their motoric implementation) to other systems. While the ideas we have presented are generally applicable, it will be necessary to specify the generative models required to solve locomotive, autonomic, and abstract decision making problems. Finally, although the computational anatomy we have proposed has a high degree of face validity, it will be necessary to establish its predictive validity. One way to do so would be to use computational fMRI (Schwartenbeck et al. 2015b), fitting this model to oculomotor behaviour, and looking for brain regions that show activity patterns consistent with the simulated neuronal

responses. One might hypothesise that these regions will match the computational anatomy illustrated in Figure 3.11. An alternative would be to use single unit responses from each brain area, recorded during an oculomotor task. One could then compare each simulated neuronal response to each recording, and construct a confusion matrix of the evidence for each synthetic signal in each region. We would expect a greater evidence for each signal that we have associated with each region above. We will see an example of this approach later.

Conclusion

In this chapter, we have described the discrete and continuous message passing that must be performed in an oculomotor system that realises a sequence of saccadic fixations. We have illustrated the remarkable consistency between the message passing implied by active inference and the anatomy of the oculomotor system. This accounts for several neuroanatomical observations, including the dual input from frontal eye fields and the substantia nigra to the superior colliculus, and the common laminar origin of axons that target the striatum, second order thalamus, and midbrain tectum. Finally, we simulated electrophysiological responses as saccadic targets are selected, and as the eyes move to implement that saccade. This shows, functionally, how the superior colliculus is uniquely positioned to act as the interface between the discrete and continuous oculomotor systems. Over the next few chapters, we exploit and extend aspects of the generative model used here. In Chapter 4, our focus is on the precision parameters introduced in the *Decisions* section of this chapter, and their biological (neuromodulatory) substrates. In Chapter 5, we discuss optimisation of the parameters of a generative model (i.e. learning), the novelty seeking this induces, and investigate the physiology of this using magnetoencephalography.

4 - Precision and neuromodulation

Introduction

In the previous chapter, our focus was on the behaviour that emerges from a given generative model architecture. However, the effective circuitry of the brain depends upon more than just its structural makeup. In this chapter[13], we build upon the notion of precision (introduced in Chapter 3) and consider how neuromodulatory influences can change the way in which a structurally similar generative model may behave very differently when the precision of different distributions is altered. This has important consequences for understanding synaptic pathologies and the neuromodulatory influences of pharmacological interventions. In the first (*Precision and pathology*) section of this chapter, we specify the message passing required for updating of precision parameters, hypothesise biological substrates for these, and illustrate this belief-updating in some simple scenarios. In doing so, we illustrate how pathological prior beliefs may render a model (or a brain) insensitive to informative sensory data. Through introducing a hierarchical model, we illustrate how pathological priors of this sort could occur. The phenomena that emerge from this could provide a useful way of looking at the distributed changes that occur in response to a pathological insult, and we illustrate this in relation to the visual hallucinations that sometimes result from synucleinopathies affecting the temporal lobes. The second section of this chapter (*Computational pharmacology*) takes a similar sort of approach, going from a biological (or therapeutic) perturbation to its associated computational perturbation and consequences for behaviour. This involves a series of synthetic pharmacological manipulations to the oculomotor model of Chapter 3 that provides additional face validity to the account of neuromodulation presented here and offers a computational mechanism for oculomotor biomarkers of therapeutic effects.

Precision and pathology

Accurate perceptual inference fundamentally depends upon accurate beliefs about the reliability of sensory data. In this section, we describe a Bayes optimal and biologically plausible scheme that refines these beliefs through a gradient descent on variational free

[13] This chapter uses material adapted from (Parr et al. 2018a; Parr and Friston 2017c; Parr and Friston 2019b)

energy. To illustrate this, we simulate belief updating during visual foraging and show that changes in estimated sensory precision (i.e. confidence in visual data) are highly sensitive to prior beliefs about the contents of a visual scene. In brief, confident prior beliefs induce an increase in estimated precision when consistent with sensory evidence, but a decrease when they conflict. Prior beliefs held with low confidence are rapidly updated to posterior beliefs, determined by sensory data. These induce much smaller changes in beliefs about sensory precision. We argue that pathologies of scene construction may be due to abnormal priors, and show that these can induce a reduction in estimated sensory precision. Having previously associated this precision with cholinergic signalling, we note that several neurodegenerative conditions are associated with visual disturbances and cholinergic deficits; notably, the synucleinopathies. On relating the message passing in our model to the functional anatomy of the ventral visual stream, we find that simulated neuronal loss in temporal lobe regions induces confident, inaccurate, empirical prior beliefs at lower levels in the visual hierarchy. This provides a plausible computational mechanism for the loss of cholinergic signalling and the visual disturbances associated with temporal lobe Lewy body pathology.

Inferring uncertainty

Just as we previously defined prior beliefs for hidden states and policies, we can also do the same for the parameters of the conditional probability distributions that make up the generative model. Appealing to the Gibbs parameterisation of Equations 2.15, 3.2, and 3.3, we can express the following prior beliefs about the precisions (inverse temperatures) associated with the likelihood (ζ), transition probability (ω), and prior over policies[14] (γ):

$$\begin{aligned} p(\zeta) &\propto \beta_\zeta \exp\left(-\beta_\zeta \zeta\right) \\ p(\omega) &\propto \beta_\omega \exp\left(-\beta_\omega \omega\right) \\ p(\gamma) &\propto \beta_\gamma \exp\left(-\beta_\gamma \gamma\right) \end{aligned} \tag{4.1}$$

[14] For the purposes of this section, we assume the **E** term in Equation 2.15 is uniform (i.e. the prior over policies is determined exclusively by the expected free energy), noting that the results presented here may be simply extended to incorporate this.

The posterior probabilities are assumed to have the same form, and are distinguished using a bold **β**. These gamma distributions have the useful property that there is a very simple form for the expectation:

$$\begin{aligned}\boldsymbol{\zeta} &= E_Q[\zeta] = \boldsymbol{\beta}_\zeta^{-1} \\ \boldsymbol{\omega} &= E_Q[\omega] = \boldsymbol{\beta}_\omega^{-1} \\ \boldsymbol{\gamma} &= E_Q[\gamma] = \boldsymbol{\beta}_\gamma^{-1}\end{aligned} \tag{4.2}$$

Appendix A.4 derives the variational updates required for estimating the posterior probabilities of these parameters. The result takes a relatively simple form:

$$\begin{Bmatrix}\dot{\boldsymbol{\beta}}_\zeta \\ \dot{\boldsymbol{\beta}}_\omega \\ \dot{\boldsymbol{\beta}}_\gamma\end{Bmatrix} = \begin{Bmatrix}\sum_\tau (\mathbf{o}_\tau^\zeta - o_\tau) \cdot \ln \mathbf{A} + \beta_\zeta - \boldsymbol{\beta}_\zeta \\ \sum_\tau \boldsymbol{\pi} \cdot (\mathbf{s}_{\pi\tau}^{\boldsymbol{\omega}} - \mathbf{s}_{\pi\tau}) \cdot \ln \mathbf{B}_\pi \mathbf{s}_{\pi\tau-1} + \beta_\omega - \boldsymbol{\beta}_\omega \\ (\boldsymbol{\pi} - \boldsymbol{\pi}_0) \cdot \mathbf{G} + \beta_\gamma - \boldsymbol{\beta}_\gamma\end{Bmatrix}$$

$$\begin{aligned}\mathbf{o}_\tau^\zeta &\triangleq \boldsymbol{\pi} \cdot \left(\tfrac{\mathbf{A}^\zeta}{\mathbf{Z}(\zeta)} \mathbf{s}_{\pi\tau}\right) \\ \mathbf{s}_{\pi\tau}^{\boldsymbol{\omega}} &\triangleq \tfrac{\mathbf{B}_\pi^{\boldsymbol{\omega}}}{\mathbf{Z}(\boldsymbol{\omega})} \mathbf{s}_{\pi\tau-1} \\ \mathbf{Z}(\boldsymbol{\zeta})_j &= \sum_i \mathbf{A}_{ij}^\zeta \\ \mathbf{Z}(\boldsymbol{\omega})_j &= \sum_i \mathbf{B}_{\pi ij}^{\boldsymbol{\omega}}\end{aligned} \tag{4.3}$$

Equation 4.3 prescribes intuitively sensible updates to the precision parameters. These depend upon a dot product between an error (the difference between the prediction based upon current beliefs about precision and the inferred or observed value) and the log probability assuming a precision of one. If this dot product is positive, this suggests overconfidence, and leads to an increase in **β**, with the corresponding decrease in its reciprocal (the expected precision). The relative simplicity of these equations speaks to their neurobiological plausibility.

Neuromodulatory systems

Having derived Bayes optimal updates for these parameters, we can now simulate creatures who infer the precision of its environment, in terms of both likelihood mappings and state transitions. To interpret these simulations, it is worth considering the likely biological substrates of these dynamics. Table 4.1 summarises the evidence implicating various neuromodulatory systems in these inferences. This suggests a role for noradrenaline in signalling ω, and for acetylcholine in signalling ζ. This implicates connectivity between the cortex and the noradrenergic and cholinergic systems in mediating the updating of Equation 4.3. These systems are related to cortical areas via the cingulum, and the dorsal noradrenergic bundle. Damage to the latter has been linked to deficits in epistemic behaviour (Mason and Fibiger 1977; Wendlandt and File 1979) and attentional set-shifting (Tait et al. 2007). Disruption of the dorsal noradrenergic bundle has also been associated with impaired extinction of a conditioned stimulus (Fibiger and Mason 1978), perhaps reflecting a representation of very low volatility.

If volatility is signalled by noradrenaline, the networks computing this quantity should interact with the locus coeruleus, a noradrenergic brainstem nucleus that projects to much of the cortex (Berridge and Waterhouse 2003). Anterograde tracing has demonstrated that the prefrontal cortex is a source of projections to the locus coeruleus (Arnsten and Goldman-Rakic 1984). Pharmacological manipulations (Sara and Hervé-Minvielle 1995) show that these projections influence the activity of brainstem noradrenergic neurons. Specifically, inactivation of frontal regions causes a sustained increase in locus coeruleus firing. This makes these regions good candidate sites for the computation of volatility. Given the close association between central noradrenaline and pupillary diameter (Koss 1986), the dynamics of the Bayesian updates given here can be incorporated into an MDP based generative model of pupillary data, first to establish the validity of the updates as a description of noradrenergic signalling, and then as part of an generative model of empirical responses that can be elicited experimentally (Schwartenbeck and Friston 2016). While outside the scope of this thesis, we pursued this pupillometric approach (Vincent et al. In press), using a convolutional linear model to demonstrate evidence in favour of models that incorporate updates in ω in explaining pupillary responses. This complements other computational approaches in understanding the role of noradrenaline (Nassar et al. 2012).

Prefrontal regions also project to the basal forebrain (Zaborszky et al. 1997), the primary source of cholinergic projections to the cortex. To reach the cortex, fibres from the basal forebrain pass the corpus callosum rostrally, before joining the cingulum (Eckenstein et al. 1988). Cholinergic axons leave this white matter bundle to diffusely innervate the cerebrum. The resemblance between the update equations for each precision parameter suggests that sensory precision can be calculated in a manner analogous to volatility, with outcome

prediction errors in sensory areas propagated to frontal regions that calculate the required update in this precision. This must then be communicated to the nucleus basalis of Meynert; a forebrain nucleus that provides a cholinergic signal to sensory cortices. This view of the cholinergic system is endorsed by several empirical observations. First, nicotinic acetylcholine receptors are found on the presynaptic terminals of cells in layers 3 and 4 of the cortex (Lavine et al. 1997; Sahin et al. 1992). These laminae are the targets of sensory relays from the thalamus (Shipp 2007). Secondly, cholinergic manipulations modulate the gain of visually evoked responses (Disney et al. 2007; Gil et al. 1997). This renders it almost tautologically true that one of the roles of acetylcholine is to modulate the precision of (some types of) sensory input, as precision and gain are mathematically identical. Finally, in both behavioural (Marshall et al. 2016) and neuroimaging (Moran et al. 2013b) studies in humans that explicitly test the association between precision and acetylcholine, differences following cholinergic manipulations are best accounted for by altered precision.

The association between computational parameters and neuromodulatory systems is important in understanding the sorts of pathology that arise from failures of neuromodulatory precision control. The disruption of the dopaminergic modulation of policy precision can result in disease states, including Parkinson's disease (Frank 2005; Friston et al. 2014; Galea et al. 2012), and schizophrenia (Adams et al. 2013b; Goldman-Rakic et al. 2004; Howes and Kapur 2009). Similarly, the neurotransmitter systems associated here with sensory precision and volatility are disrupted in a range of neuropsychiatric disorders. Depletion of acetylcholine is associated with Alzheimer's disease (Lombardo and Maskos 2015; Whitehouse et al. 1981), while disruptions of noradrenaline signalling are thought to contribute to anxiety (Blier and El Mansari 2007), post-traumatic stress disorder (Gören and Cabadak 2014), depression (Moret and Briley 2011), and Wernicke-Korsakoff encephalopathy (Halliday et al. 1993; Mair et al. 1985). Additionally, the lateral asymmetry of noradrenergic projections in the forebrain (Oke et al. 1978), reflected in pupillary responses (Kim et al. 1998), hints at a role in visual neglect (Malhotra et al. 2006). A formal description of the computational processes that are disrupted in these disorders allows for the development of a computational phenotyping (Schwartenbeck and Friston 2016) of patients. This may aid in the characterisation of defective neurophysiology, making use of the process theory (Friston et al. 2017a) associated with active inference.

The story on offer here provides a coherent and formal account of neuromodulation in the brain that is broadly consistent with previous neurobiological accounts of perception and decision-making (Daw and Doya 2006; Doya 2002; Doya 2008). In brief, there are three fundamental sorts of beliefs that determine behaviour: (i) beliefs about outcomes given hidden or latent states of the world, (ii) beliefs about states of the world and (ii) beliefs about policies

given states of the world. Each of these sets of beliefs is equipped with an uncertainty or precision that may be encoded by specific modulatory neurotransmitter systems. The evidence reviewed above – and in (Doya 2008; Yu and Dayan 2005) – speaks to the following (summarised in Table 4.1): (i) cholinergic systems encode the precision of beliefs about outcomes given states of the world (c.f., attention and expected uncertainty); (ii) noradrenergic systems encode the precision of state transitions (c.f., volatility and unexpected uncertainty) and (iii) dopaminergic systems encode the precision of beliefs about policies (c.f., action selection). The coherent aspect of this account rests on the fact that all three systems play the same computational role; namely, an encoding of precision. Furthermore, all three neurotransmitter systems have the same basic effects on synaptic transmission; namely, a neuromodulatory gain control. For the remainder of this section, we focus upon the cholinergic system and its disorders.

Simulated scene-construction

The brain's visual system must overcome formidable inferential challenges. Despite receiving sequentially sampled, spatially limited sensory information from a two-dimensional array of photoreceptors, we perceive spatially and temporally continuous three-dimensional visual scenes, populated with complex objects. However, despite this remarkable capacity for scene construction (Mirza et al. 2016; Parr and Friston 2017a), the visual system is not infallible. It depends upon a delicate balance between prior beliefs about perceptual hypotheses (Gregory 1980), and the sensory evidence that supports or refutes them (Brown and Friston 2012; Geisler and Kersten 2002). The simulations in this section address the computational mechanisms that could maintain this balance, and the consequences of their failure; e.g. (Collerton et al. 2005).

We will present two sets of simulations. The first uses a simple (single hierarchical level) generative model to illustrate the basics of perceptual inference – and how this depends upon the precision afforded sensory evidence, relative to (empirical) prior beliefs about state transitions. This is formally very similar to the model used in Figure 3.7. In the second simulation, we equip the model with a second (hierarchical) level that embodies the belief that outcomes are generated by a scene (i.e., a combination of visual objects at four spatial locations) that remains constant over successive (five saccade) visual searches. While this, deliberately simple, form of visual scene does not capture the rich phenomenology associated with real scene construction (Hassabis et al. 2007), it enables us to simulate visual processing under lesions that are hierarchically remote from the (neuromodulatory) effects of expected

precision. We offer this as a formal explanation for the sort of (functional) diaschisis that characterises synucleinopathies; particularly those associated with visual hallucinosis.

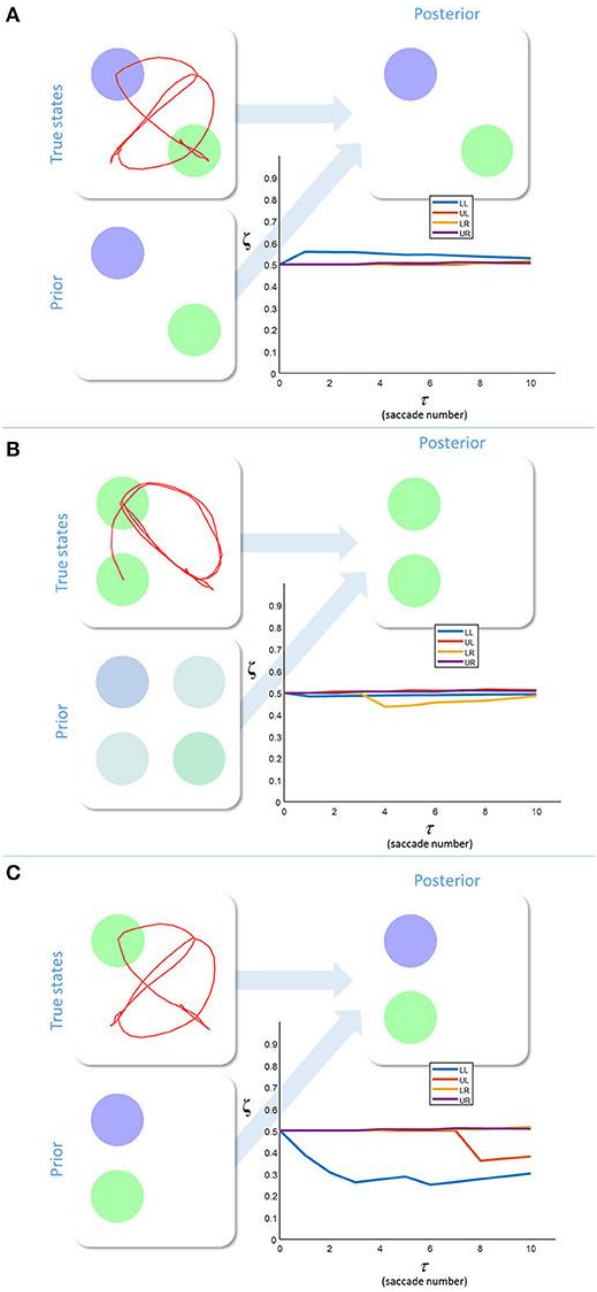

Figure 4.1 – Inferring uncertainty. The simulations illustrated here use Equation 4.3 to make inferences about the likelihood precision during a simple visual foraging set-up. With the exception of explicit prior beliefs about the precision, the generative model is identical to that used in Figure 3.7. As before, each location is associated with a hidden state factor (with an additional factor for eye-position). Here, the hidden state at each location takes one of three values: absent (white), green, or blue. The prior beliefs about the visual stimuli are depicted by setting the intensity of each colour equal to the probability of that colour. The posterior beliefs are represented similarly. The true states are presented along with the saccadic

trajectory (red line) that determines the sequence in which the stimuli were sampled. The (posterior) sensory precision is shown in the line plots. There is one precision term associated with each location (LL = lower left, LR = lower right, UL = upper left, UR = upper right). Figure 4.1A shows inference with a prior belief that is consistent with the true states. Figure 4.1B shows a relatively imprecise prior that is inconsistent with sensory states. Here, the sensory evidence dominates the inference. Figure 4.1C shows the result of setting a precise prior belief against contradictory sensory data. In this case, the prior dominates, but must induce a decrease in sensory precision in order to do so.

Figure 4.1 shows the results of simulating a visual search for 10 saccades under different prior beliefs and stimuli. First, we chose a set of prior beliefs that matched the true states of the world (Figure 4.1A). There is little change in the estimated sensory precision over time, and the posterior belief matches both the prior and the true states. We then tested the case for which the prior and the true states are different. Figure 4.1B shows a prior belief with the same content as 4.1A, but held with a lower degree of confidence (i.e. the prior belief is less precise). Again, there is little change in sensory precision, but now the posterior reflects the true states and not the prior. In other words, the sensory likelihood dominates perceptual inference. In Figure 4.1C, we simulate another mismatch between the prior and sensory evidence. This time, the prior belief is held with a high degree of confidence (i.e. a very precise prior), and this dominates inference. The posterior belief matches the prior, and is inconsistent with the sensory data sampled. The conflict between the prior and the sensory evidence is resolved in this case by a decrease in the precision associated with the contradictory locations. Heuristically, if a prior belief is held very confidently, evidence to the contrary is disregarded or ignored.

We have demonstrated that excessively precise prior beliefs lead to a compensatory decrease in the precision of the likelihood distribution. Given the association between sensory precision and cholinergic signals (Dayan and Yu 2001; Marshall et al. 2016; Vossel et al. 2014; Yu and Dayan 2002), this provides a possible mechanism for the decrease in activity in the nucleus basalis of Meynert in several neurodegenerative disorders (Candy et al. 1983). Of these, the synucleinopathies, including Lewy body dementia, show an especially dramatic decrease in cholinergic signalling (Perry et al. 1994), and are associated with false visual inferences (i.e. hallucinations). These are, by definition, the imposition of prior beliefs on perception in the absence of supportive sensory evidence. Furthermore, decreased sensory gain means a smaller response to visual stimulation, consistent with the combination of reduced occipital

cholinergic activity (Kuhl et al. 1996) and occipital hypometabolism (Heitz et al. 2015; Lobotesis et al. 2001) in synuclein disorders.

The above raises an important question: What is the source of the abnormal prior beliefs in conditions such as Lewy body dementia? We have previously argued that pathological prior beliefs might arise through anatomically defined vascular lesions (Parr and Friston 2017b). Here, too, we can appeal to the anatomical distribution of the lesions to try to understand the relationship between tissue pathology and computational (network level) dysfunction. Lewy body pathology occurs in many brain regions, but it is their presence in parts of the temporal lobe that is associated with visual hallucinations (Harding et al. 2002). This leads us to consider the ventral visual hierarchies and their computational homologues.

Hierarchical inference

The visual system, like other sensory systems, is known to be hierarchically organised (Desimone et al. 1985; Felleman and Van Essen 1991; Markov et al. 2013; Zeki and Shipp 1988). We have previously appealed to this hierarchical structure to model reading (Friston et al. 2017f) and visual working memory tasks (Parr and Friston 2017d). We now draw upon the same idea to account for the source of the prior beliefs above, and to show how inaccurate but highly precise beliefs can develop. The visual system is organised into two broad hierarchical streams. These are the ventral (what) and the dorsal (where) pathways (Ungerleider and Haxby 1994). It is the former that is of relevance here, as it leads from the occipital cortex to the temporal cortex, and represents stimulus identity at increasing levels of abstraction. While regions earlier in this pathway tend to respond to simple visual features (Hubel and Wiesel 1959), later regions are selective for more complex visual objects (Valdez et al. 2015) or scenes (Epstein et al. 1999), constructed from lower level features. This is very important in accounting for the phenomenology of visual hallucinations in neurodegenerative conditions, as hallucinatory components of the percept appear in a consistent and plausible way in the context of the scene. This implies there is no impairment in scene construction per se. Instead, it is the wrong scene that is constructed. Crucially, this suggests hallucinated scenes are constructed based upon hierarchical principles, leading to the integration of a false percept in a way that is contextualised by the rest of the scene. This does not imply any impairment in the posterior precision of the overall percept.

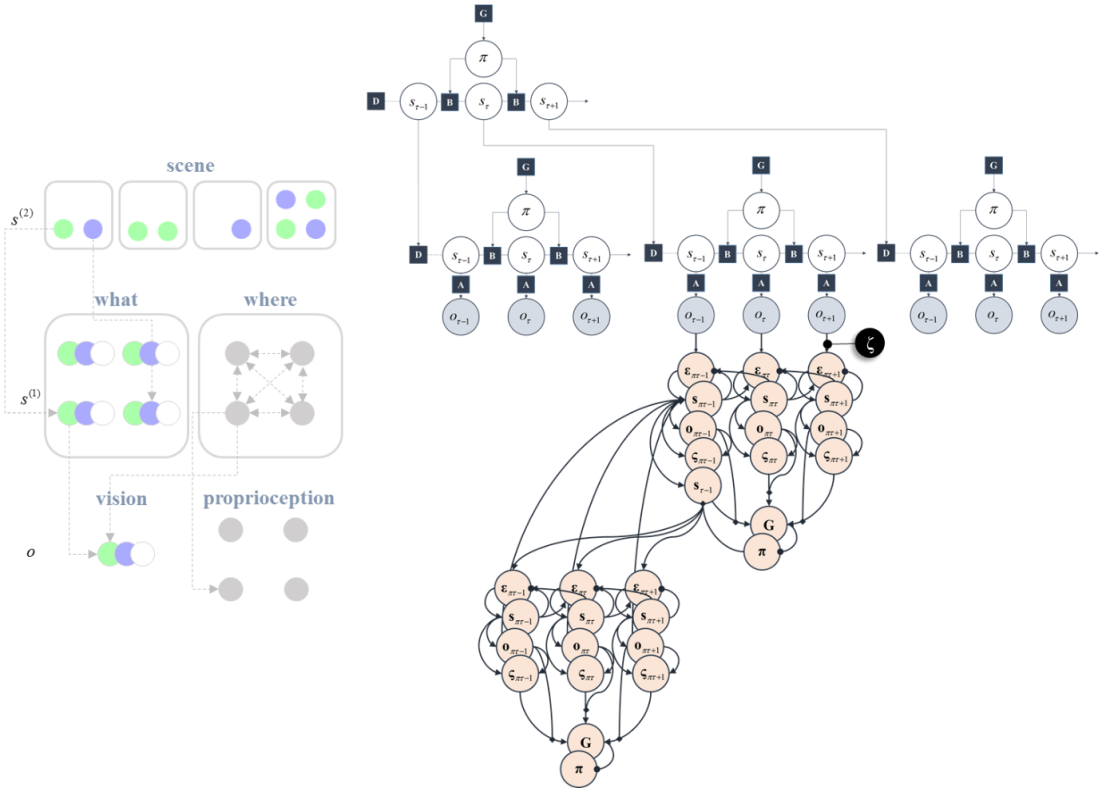

Figure 4.2 – Deep temporal models. *The right part* of this figure illustrates the extension of the MDP model from Figure 2.2 to include a second hierarchical level. The outcome of the MDP at the higher level is the initial hidden state in the MDP at the lower level. Crucially, this means each time-step at the higher level transcends multiple time-steps at the lower level. In other words, a state at a high level corresponds to a short trajectory of states at the lower level, just as a word corresponds to a short trajectory (or sequence) of letters. The message passing used to solve this sort of generative model mirrors this structure, with slower states contextualising faster states and faster states playing the role of observations from the perspective of the slower states. For the simulations that follow, we optimised beliefs about the precision (black circle) associated with the likelihood at the lower level. *The left part* of this figure shows the specific structure of the hierarchical generative model used here. The lower level ($s^{(1)}$) is exactly the same as the model used for the simulations in Figure 4.1, with four hidden state factors accounting for the colour of a stimulus at each point in the visual scene ('what') and one for the current fixation location ('where'). These give rise to the proprioceptive and visual (foveal) outcomes (o). This model has been supplemented with a second level ($s^{(2)}$) that represents four alternative hypothesised scenes (configurations of the coloured circles) and generates the 'what' states at the lower level.

To account for this hierarchical structure, we can augment our generative model so that visual stimuli (the 'what' panel in Figure 4.2) are themselves generated by 'scene' states. Both the

'what' and the 'scene' variables are types of hidden state. We refer to the former as a 'first level' and the latter as a 'second level' state. The second level is much simpler in this case (Figure 4.2), as there is only one type of state with no policies. Furthermore, all transitions at the second level are taken to be identity matrices, expressing the belief that the scene remains constant over time. This type of generative model allows the first level (empirical) priors to be generated by the second level. While this generative model is too abstract to map directly to the real visual system, this type of hierarchy does express cardinal features of the organisation of the ventral visual stream.

Importantly, although inference about a visual feature can be performed within a given fixation, it takes multiple saccades to make inferences about the scenes at the second level. This implies that beliefs at this level should be updated more slowly (Friston et al. 2017f), consistent with the slower response properties higher in sensory hierarchies (Hasson et al. 2008; Kiebel et al. 2008; Murray et al. 2014).

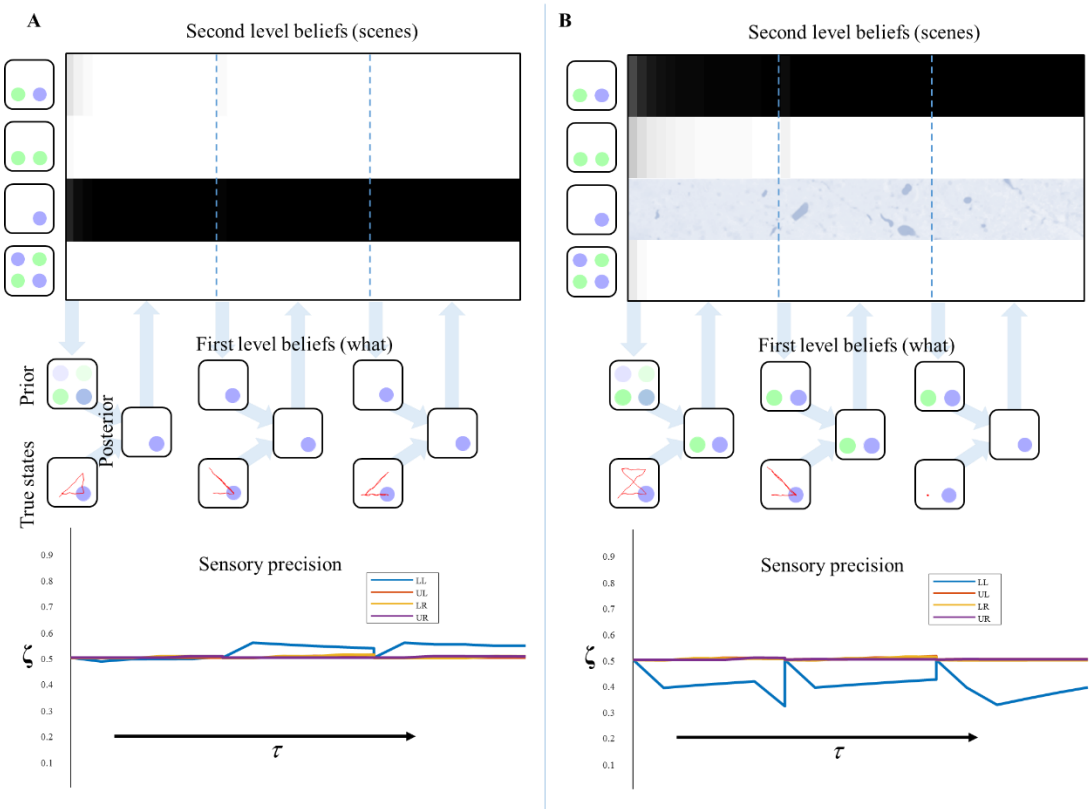

Figure 4.3 – Empirical priors and pathology. These plots illustrate the evolution of beliefs about second level states, first level states, and sensory precision. The upper plots show the beliefs about scenes over time. Each row of these represents a given scene (indicated by the images on the left). The shading indicates the belief that this is the scene responsible for the sensory input. Black indicates a belief that the probability is 1, white indicates 0. The

descending arrows represent the computation of a first level empirical prior from the second level beliefs. A new empirical prior is generated after every 5 saccades (demarcated by dashed blue lines). The empirical prior and the sensory consequences of saccadic exploration combine to form first level posterior beliefs (exactly as in Figure 4.1). The beliefs from each set of 5 fixations are used to update the second level beliefs (ascending arrows). The lower plots show the beliefs about the sensory precision, aligned to the beliefs at the higher level. The precision is reset at the vertical dotted lines. 4.3A shows 'healthy' second level priors that associate an equal probability to each scene at $\tau = 0$. Under these priors, the correct scene is inferred and the consistency between priors and sensory data leads to an increase in sensory precision. 4.3B shows the same model but with the prior probability of the third scene set to zero to simulate the loss of this neuron (or neuronal population). Here, the conflict between priors and sensory evidence leads to a decrease in precision. This also demonstrates the importance of action in perception as, at the final time-step, consistently fixating on the lower left location leads to a correct percept. This illustrates the point that collecting more data can compensate for the diminished precision of those data.

Figure 4.3 shows the inferences made when we simulate responses with this hierarchical model. The upper part of the figure shows the beliefs about each of the second level states through time. At the start, the second level beliefs are combined (weighted by their probabilities) to generate an empirical prior at the first level. In both Figure 4.3A and 4.3B, this prior is relatively imprecise. A sequence of 5 saccades is performed, and the observations made are used to refine the first level posteriors in exactly the same way as in Figure 4.1. These posterior beliefs are used to update the second level beliefs. These then generate a new empirical prior and this sequence repeats. The sensory precision is reset to its prior value whenever a new empirical prior is set (at the start of each sequence of 5 saccades).

The reason the precision re-sets to its prior value periodically is due to the separation of temporal scales inherent in the generative model. This is analogous to processes like reading, for which a sentence provides a high-level context linking sequential words. Letters in one word only inform inferences about the next word via sentence-level representations. This means all lower level representations are set to their (empirical) prior values every time we move from one word to the next. In our setting, the same is true of all lower level representations, including the precision. In a more complex model, it would be possible to condition the prior belief about the precision upon slowly changing variables at the higher level. While beyond the material presented in this chapter, this would allow inferences about the precision to transcend the time-scale of the lower level.

Figure 4.3A illustrates this process in a model with 'healthy' second level priors. There is a very rapid inference that the third scene is the most likely cause at the second level, and the first level beliefs, following saccadic interrogation, invariably match the true states. A moderate increase in estimated precision occurs under the second empirical prior, because confident prior beliefs match the sensory inputs. Note that the start location is in the lower left, so there is a larger effect for the precision at this location. This illustrates the fact that the prior has a greater influence at the start of the trial, where fewer observations have been made. Figure 4.3B shows a simulated synucleinopathy. Neuronal loss (or disconnection) high in the ventral stream has been simulated by setting the second level prior belief for one of the scenes (the correct one) to zero. This is as if we had removed the neuron that represents this second level hypothesis. Interestingly, this does not impede the formation of confident (false) first level empirical priors. As we saw earlier, these induce a decrease in the estimated sensory precision, and false perceptual inference. This demonstrates that pathology high in ventral visual hierarchies can, in principle, induce changes in distant brain areas – something that has been characterised in terms of a functional or dynamic diaschisis (Carrera and Tononi 2014; Price et al. 2001). The idea that damage to a neuronal population preserves the confidence in the beliefs they represent may seem counterintuitive. However, even with the loss of neurons representing the correct inference, there is still a clear 'best' explanation at the level of scenes. This leads to confident posterior beliefs about the scene, giving rise to confident (but incorrect) empirical prior beliefs about the contents of that scene. This is further facilitated by the permissive decrease in sensory precision.

This result recapitulates the idea that, for hallucinations to occur, prior beliefs must be held with a high degree of confidence (precision) relative to that associated with contradictory sensory evidence. This has previously been demonstrated in the context of auditory hallucinations in schizophrenia (Adams et al. 2013b). Our account does, however, provide a different perspective on the initial computational insult. While this has previously been formulated as a false prior belief that something is present, we have demonstrated that hallucinations may be induced by a false prior that a given scene is not a good explanation for sensory data. This forces the brain to resort to an alternative explanation, associated with other, spurious, perceptual content.

Computational neuropathology

In the above, we have presented a model that relates temporal lobe pathology to the development of complex visual hallucinations and reduced cholinergic signalling to the

occipital cortex. Crucially, although the primary pathology only affects temporal components of the simulated network, its computational consequences are felt throughout the brain via a dynamic diaschisis (from Greek διάσχισις meaning 'shocked throughout'). This type of account is necessary in explaining the patterns of diaschisis observed in neuropathological processes.

The synucleinopathies (including Lewy body dementia, Parkinson's disease, and Multiple system atrophy (McCann et al. 2014; Tsuboi and Dickson 2005)) provide important examples that illustrate the need to connect tissue pathology to computational dysfunction. Despite the presence of physiological changes in the occipital cortices (Kuhl et al. 1996), and visual symptoms (McKeith et al. 2004; Weil et al. 2016), the histopathological processes in these disorders tend not to affect occipital cortex directly (Khundakar et al. 2016). While impaired dopamine signalling to the cortex in these disorders might contribute, occipital regions tend to receive relatively few dopaminergic projections (Javoy-Agid et al. 1989). The absence of these processes in primary visual areas, and the association between visual symptoms and temporal lobe Lewy bodies (Harding et al. 2002), calls for an explanation of physiological changes in the former in terms of their computational relationship to the latter.

We note that, for prior beliefs to dominate inference, the sensory precision must be low relative to the precision of prior beliefs. This means that hallucinations could occur with intact prior beliefs and a primary lesion to systems encoding sensory precision, or an increase in prior confidence without any change in sensory precision. Computationally, these are equivalent as they each change the balance of precisions in the same way. However, they are not necessarily biologically equivalent. The former implies a primary lesion to neuromodulatory systems that modulate synaptic gain in sensory cortices, while the latter implies damage to higher regions of cortex that provide empirical priors to sensory areas. In the context of visual hallucinations in Lewy body disease, both of these are present. While these may be two independent primary lesions, a simpler explanation would be that one is a downstream effect of the other. In this chapter, we have suggested a mechanism by which damage to higher cortical areas could lead to disruption of synaptic gain in early visual cortex.

In short, the formal account of active inference or vision on offer here also provides an explanation in terms of a functional diaschisis – a dysfunction of one region as a consequence of a distant lesion (Carrera and Tononi 2014; Price et al. 2001). Figure 4.4 illustrates a plausible computational anatomy that could underwrite this account. While this anatomy is speculative, it serves to illustrate the importance of the functional interactions between brain regions to the understanding of neurological disease. Damage to temporal regions, representing second level beliefs, induces changes in the first level beliefs. This leads to

inconsistencies between perceptual beliefs and sensory data, which down regulates cholinergic projections to the occipital cortex. Decreased cholinergic signalling uncouples beliefs about states from sensations they cause, facilitating hallucinations (O'Callaghan et al. 2017; Perry et al. 1991). As this model would predict, treatment with cholinesterase inhibitors increases occipital blood flow, while attenuating visual hallucinations (Mori et al. 2006). A complementary, and more biophysically detailed, perspective on this is formulated in terms of impaired conductance (Tsukada et al. 2015) in the synapses between visual cortical neurons, and those in the ventral visual stream.

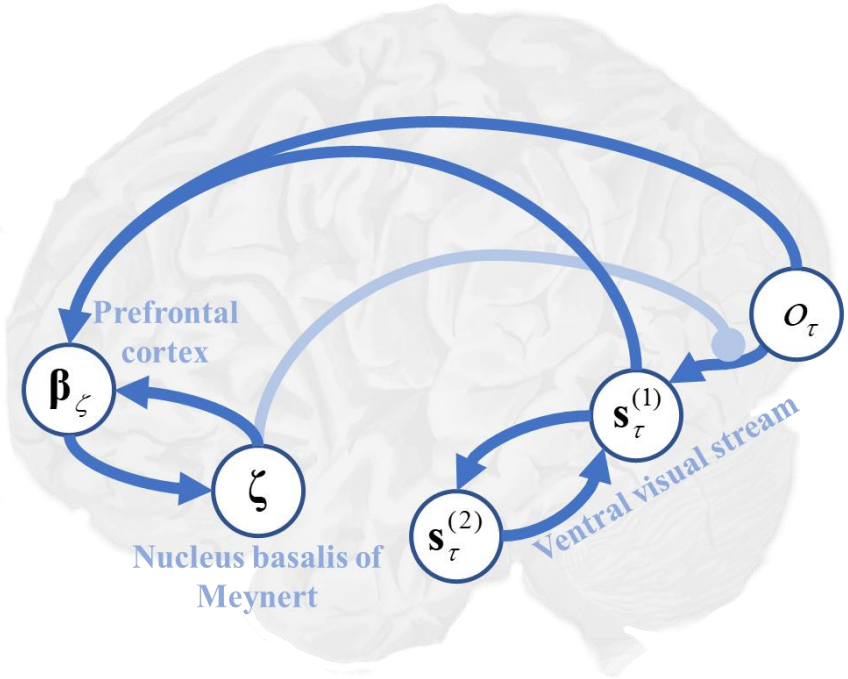

Figure 4.4 – Precision, hierarchy, and the ventral visual stream. This schematic illustrates the hypothetical computational anatomy of the ventral visual stream and its cholinergic modulation. Visual outcomes (*o*) are shown in the primary visual cortex. These inform first level beliefs ($\mathbf{s}^{(1)}$) early in the ventral stream, and the connection between these is modulated by cholinergic projections from the basal nucleus in the forebrain. First level beliefs are reciprocally influenced by second level beliefs ($\mathbf{s}^{(2)}$) in the temporal lobe. We have (speculatively) suggested that the prefrontal cortex may be engaged in computing the expected precision, utilising its inputs from those regions representing first level beliefs, and its connections to the basal forebrain (Zaborszky et al. 1997).

An influential model of recurrent complex visual hallucinations (Collerton et al. 2005) implicates these same regions, but makes the point that many other disorders involve similar

changes. For example, cholinergic deficits are also associated with Alzheimer's disease (Minoshima et al. 2004), although to a lesser extent (Perry et al. 1994; Tiraboschi et al. 2000). Like those for Lewy body disease, pharmacological therapeutics have focused on correcting this neurochemical deficit (Lam et al. 2009). There is evidence to implicate changes in the temporal lobes in this disorder, as it tends to impact these structures early. However, this is typically more medial than in Lewy body dementia (Minoshima et al. 2002). Furthermore, it is unlikely that the cholinergic deficits in Alzheimer's disease are consequences of temporal lobe changes, as there is good evidence for a primary pathological insult to the nucleus basalis (Etienne et al. 1986; Liu et al. 2015; Samuel et al. 1994). This renders it improbable that this condition exhibits a similar set of computational deficits to those described above. The lower prevalence of visual hallucinations in Alzheimer's disease, despite overlapping pathological features with Lewy body disease, illustrates an important point. It is not sufficient to have temporal lobe damage and cholinergic dysfunction to give rise to hallucinations. The interplay between the two is crucial in characterising this type of diaschisis.

In this chapter, we have focused upon false positive inferences (i.e. hallucinations). However, brain damage often leads to false negative inferences (i.e. agnosia) (Warrington and James 1967). These manifest as a failure to perceive a stimulus, despite it being present. The approach we have described could be used to account for these phenomena in several ways. We outline these here, but emphasise that determining which of these best accounts for agnosia remains an open question that requires further investigation. The first way in which we could account for this is by setting a prior belief that a given object is present to zero. If the most probable alternative explanation is the absence of any object, this inference will result. It is important to distinguish this inference of absence from uncertain inferences, in which the presence or absence of an object cannot be inferred with any certainty. These could result from disconnections that render this object conditionally independent from sensory data in the generative model. This would ensure beliefs about the presence or absence of a given object would not depend upon these data. A third way in which certain stimuli may fail to enter into perceptual awareness is the failure to attend to certain kinds of stimuli, as in visual neglect (Halligan and Marshall 1998). We have previously argued that this syndrome, in which stimuli on the left of space are ignored, depends upon a failure to actively engage with stimuli on the left (Parr and Friston 2017b). We unpack this idea in more detail in Chapter 6.

A number of outstanding questions are raised by the approach we have taken, which require empirical resolution. The first concerns our use of the term 'visual features'. We have illustrated a feature as the colour of a circle in a given location, but this is not mandated by the mathematics used in our generative model. In principle, relevant features could be shape, luminance, contrast, or any other experienced attribute. We would need to present patients

with a task like that illustrated above, but with different sorts of stimuli, to elucidate which of these afford the right level of description – and whether the ensuing responses are conserved over patients. The second question concerns the fixed parameters of the generative model – such as the prior belief about sensory precision. These are likely to be subject specific, but could be estimated from eye-tracking data collected during the above task (Mirza et al. 2018).

Summary

In the first part of this chapter, we illustrated the computational mechanisms that could act to maintain the perceptual balance between prior beliefs and sensory evidence. We simulated inferences about the precision associated with the likelihood, and demonstrated that confident, but incorrect, prior beliefs cause a decrease in the expected sensory precision, and false perceptual inferences. In the second part, we asked what the computational mechanisms might be that give rise to pathological empirical priors, and motivated this through an appeal to the neurobiology of synuclein disorders. We described a plausible mechanism by which tissue pathology in higher visual areas could cause in occipital hypometabolism, cholinergic deficits, and visual hallucinations. Crucially, this calls upon the computational (network level) pathologies induced by regional synucleinopathies. This accounts for several empirical findings, including the association of temporal lobe changes with hallucinations in Lewy body disease and the improvement in hallucinations and occipital metabolism when these patients are treated with cholinesterase inhibitors. The ideas and simulations presented here emphasise the importance of relating neuropathological processes to computational dysfunction to understand neurological disease.

Computational Pharmacology

As reviewed in Chapter 1, oculomotor behaviour relies upon the coordination of a distributed network of regions throughout the brain (Parr and Friston 2017a; Robinson 1968). Assessment of oculomotion therefore offers a simple (non-invasive) way to measure brain function. While disruption of normal neurological (Anderson and MacAskill 2013) or psychiatric (Lipton et al. 1983) function can induce a range of characteristic eye-movement deficits; subtler modulations of neuronal function may also be detected in oculomotion. In this section, we focus upon the neurochemical aspects of oculomotor control, and the sorts of oculomotor syndromes that may be induced by therapeutic agents (Naicker et al. 2017; Reilly et al. 2008).

In doing so, we draw from recent theoretical work (outlined in Chapter 3) addressing the computational anatomy of oculomotion (Parr and Friston 2018a; Parr and Friston 2018c), and emerging themes in computational accounts of neuromodulation [(Friston et al. 2014; Marshall et al. 2016; Parr et al. 2018a; Parr and Friston 2017c; Sales et al. 2018; Schwartenbeck et al. 2015b), and the *Precision and Pathology* section above]. These accounts are based upon the idea (formalised in Chapter 2) that the brain uses a generative model to infer the causes of its sensations, and that this model is equipped with beliefs about the precision (inverse variance) of the relationships between different kinds of latent (i.e., unobserved) variables generating sensory (i.e., observed) samples. The precision of a belief can be thought of as the confidence in that belief (as opposed to its content). As such, precisions are generally associated with *neuromodulatory* influences over synaptic gain (Feldman and Friston 2010; Marder and Thirumalai 2002; Nadim and Bucher 2014), as opposed to *driving* postsynaptic responses (i.e., modulating transmembrane conductance as opposed to depolarisation).

We have previously argued for an association (i) between acetylcholine and beliefs about how precisely hidden variables in the world give rise to sensory data, (ii) between noradrenaline and beliefs about how hidden variables in the present cause those in the future, and (iii) between dopamine and beliefs about how we will act upon the world (Friston et al. 2014; Parr and Friston 2017c). In what follows, we first provide an overview of oculomotion in terms of these three kinds of precision, their associated neurotransmitter systems, and active inference. We then introduce a simple delay-period oculomotor task – of the sort used extensively in primate electrophysiological studies (Funahashi 2015). Through manipulating various precision terms, we will see that the resulting oculomotor syndromes reproduce those induced by pharmacological agents acting upon their associated neurochemical systems. The implication here is that if one can generate pathological eye movements from selective deficits in neuromodulatory systems *in silico*, it is possible to estimate these deficits using empirical observations, such as eye tracking [see Adams et al (2016) for a proof of principle using slow pursuit eye movements].

The neurochemical anatomy of oculomotor control

In this section, we briefly review the neuroanatomical networks involved in ocular control, with a special focus on the synapses on which different neurotransmitters are thought to act. In describing this functional anatomy, one can associate these neurotransmitters with putative computational roles. The direct cortical control of eye movements involves predominantly dorsal brain areas, including the frontal eye fields (Künzle and Akert 1977), which

communicate with the nearby dorsolateral prefrontal cortex (Buschman and Miller 2007). The former area is thought to represent the position of the eyes (Moore and Fallah 2001), while the latter is associated with the maintenance of beliefs about cued targets (Goldman-Rakic 1987). Like most of the cortex, these areas receive distributed projections from the locus coeruleus and the basal forebrain via the cingulum (Avery and Krichmar 2017; Doya 2008). As such, these cortical regions are modulated by noradrenaline and acetylcholine. Under active inference, noradrenaline is thought to represent the precision of transitions (i.e., confidence in probabilistic beliefs about the dynamics of the world – such as motion and occlusion). This encoding of precision is crucial in prefrontal cortical regions involved in the maintenance of a remembered stimulus, as the persistence of a belief over time rests upon a precise belief that the target does not change between viewing the stimulus and enacting the appropriate response (Parr and Friston 2017d).

Acetylcholine has been linked to the precision of beliefs about how latent or hidden states of the world – that cannot be directly observed (e.g., eye position) – give rise to sensory (visual or proprioceptive) data (Marshall et al. 2016; Moran et al. 2013b; Vossel et al. 2014). Loss of acetylcholine, as observed in conditions such as Lewy body dementia, can lead to a failure of sensory data to constrain perceptual inference in the right sort of way. Complex visual hallucinations – namely, false positive perceptual inference (Collerton et al. 2005; Parr et al. 2018a) – represent a dramatic example of this failure. In the context of motor control, imprecise predictions about desired movements may lead to a failure of descending predictions (i.e., motor commands) to elicit the predicted proprioceptive signals via motor reflexes.

The cortical regions described above communicate with brainstem oculomotor regions via two main pathways. The first is a direct cortico-collicular projection (Künzle and Akert 1977). The second is via the basal ganglia (Hikosaka et al. 2000). The output nuclei of the basal ganglia include the substantia nigra pars reticulata, which monosynaptically inhibits the superior colliculus via GABAergic projections (Hikosaka and Wurtz 1983). The other part of the substantia nigra – the pars compacta – provides dopaminergic innervation to the striatum (Moss and Bolam 2008). In terms of active vision, the cortico-collicular pathways may be thought of as predicting the proprioceptive and visual consequences of alternative saccades *that could be performed*. The nigro-collicular pathway then weights each alternative, depending upon striatal evaluations of the 'goodness' (technically, expected free energy) of each possible saccade. This goodness is simply the capacity of that saccade to fulfil prior beliefs about the sensory outcomes of a visual sampling (e.g., to comply with experimental instructions or to resolve uncertainty by accumulating evidence during visual scene construction). This sets up a biased competition in the superior colliculus, resulting in the selection of a saccadic target (Veale et al. 2017; Zelinsky and Bisley 2015). The superior

colliculus then propagates this signal to other oculomotor brainstem areas (Parr and Friston 2018a; Robinson 1968), resulting in a saccade towards this target. The GABAergic signal here is vital in setting up the competition between alternative saccades (Hall 1999); leading to a precise representation of the chosen saccadic target. Finally, in active inference formulations, the nigro-striatal pathway is responsible for maintaining precise beliefs about which saccadic policy to pursue. Please see Figure 4.5 for a description of this computational anatomy in terms of Bayesian belief updating and neuronal message passing. This may be thought of as a simplification of Figure 3.10 that emphasises the role of neurochemical modulation.

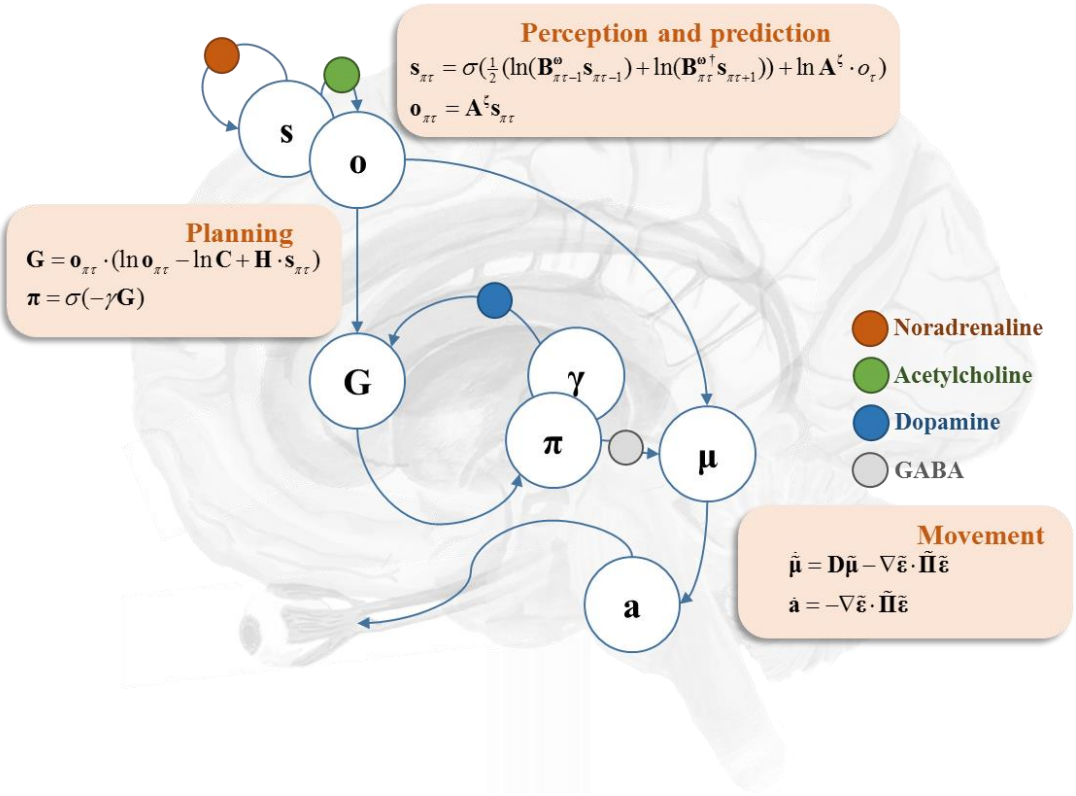

Figure 4.5 – Computational neuropharmacology and oculomotion. This schematic illustrates a simplified (computational) anatomy of oculomotor control, highlighting some of the key synapses at which neuromodulatory transmitters act. The cortical components of this network include the frontal eye fields, and the dorsolateral prefrontal cortex. We have associated these regions with beliefs about hidden states (**s**), and predictions about the (categorical) outcomes (**o**) that these states entail. The 'Perception and prediction' panel specifies how these are computed from beliefs about the way in which states give rise to observations (**A**), and beliefs about how states at a given time evolve (**B**). These likelihood and prior transition terms are equipped with precisions – superscripts **ζ** and **ω**, respectively – that quantify the confidence (inverse variance) of associated conditional beliefs. The

likelihood and prior precisions have been associated with cholinergic and catecholaminergic modulation respectively. These cortical regions project to both the basal ganglia (i.e., the striatum) and the superior colliculus. The direct pathway through the basal ganglia itself targets the superior colliculus, via the substantia nigra pars reticulata. Striatal neurons are modulated by dopaminergic projections from the substantia nigra pars compacta ($\boldsymbol{\gamma}$), while the projections from the pars reticulata to the superior colliculus provide a GABAergic modulation of the cortico-collicular pathway ($\boldsymbol{\Pi}$). The 'Planning' panel shows how the basal ganglia may evaluate alternative saccades by computing the expected free energy ($\mathbf{G}$) associated with each eye movement and subsequent sensory samples. Dopamine modulates confidence in beliefs about the best saccade to select ($\boldsymbol{\pi}$), given this evaluation. The 'Movement' panel provides the (Bayesian filtering) equations that may be used to implement the next saccade. These rely upon prior beliefs about where the eyes should be that are obtained from the predictions from the cortex ($\mathbf{o}$) modulated by plans evaluated in the basal ganglia ($\boldsymbol{\pi}$). Together, these are used to compute the average belief about where to look next, which is then equipped with a precision ($\boldsymbol{\Pi}$). The error ($\boldsymbol{\varepsilon}$) between current beliefs about the position of the eyes ($\boldsymbol{\mu}$) and the target is then used to drive brainstem reflexes that act ($\mathbf{a}$) to minimise this error – implementing the motor command from the cerebrum. From our perspective, the key feature of these equations (see Figures 2.1 and 2.2), is that they suggest a modulatory role for the precisions described above. For a more detailed technical account of these equations, please see (Friston et al. 2017c) and Chapter 2, and for a conceptual overview of their relationship to anatomy, please see (Parr and Friston 2018b).

Delayed oculomotor task

A range of oculomotor changes have been observed following different therapeutic interventions. These include changes in the characteristics of a saccade (e.g. hypo or hypermetric), and in the decision processes leading to a saccade (e.g. inappropriate saccade targets). For the purposes of this chapter, we adopt a single oculomotor task (Funahashi et al. 1989) that showcases a simple decision process but also allows us to inspect the trajectory of the saccade itself (Figure 4.6). This involves presentation of a cue at the saccadic target location, followed by maintenance of fixation after the disappearance of the target. When cued, the task is to saccade to the remembered target. This oculomotor delay-period task has been used extensively in primate research, notably in the study of working memory, so has well-described neurophysiological correlates.

To simulate oculomotor performance under this paradigm, we have to specifying a model of how sensory outcomes are generated by latent or hidden states of the world – and how those states can be changed by selecting particular actions or movements. This generative model then specifies belief updating in a synthetic brain under ideal Bayesian assumptions (see the equations in Figure 1). The beliefs in question here are expectations about states of the world generating sensations – and the plausible actions that can change those states. Crucially, the synthetic subject believes she will select those actions that maximise the evidence for her model of the world. It transpires that this is the same as selectively sampling in sensory outcomes (e.g., directing saccades to particular parts of the visual field) that resolve uncertainty. This resolution of uncertainty comes in two flavours. First, the information gained by sampling new observations and, second, ensuring that these observations are consistent with the generative model (i.e., conform to prior preferences). In the following generative model, we have to deal with two sorts of states: namely, discrete and continuous states. Discrete states correspond to different locations, different stages of each trial *etc.*, while continuous states refer to things like eye position and velocity.

The generative model we use to simulate this task uses a Markov Decision Process (MDP) model of *discrete* states (Friston et al. 2015b) that generates a set of hypothetical saccadic targets. Each of these hypotheses represents the equilibrium (attracting) point (Feldman and Levin 2009) in a *continuous* state-space model (described in detail in (Parr and Friston 2018a)) of the eyes themselves (Figure 3.1). The MDP part of the model comprises three types of hidden (unobservable) state – that jointly generate discrete predictions for the continuous part of model dealing with continuous oculomotor trajectories (Friston et al. 2017c)). The discrete states generating predictions include the current fixation location, the target location, and the current stage of the trial. The last of these states includes the target presentation, delay period, and saccade-to-target stages. Fixation location is a controllable state, meaning that any of the five (fixation cross, up, down, left, right) locations may change to any other location depending upon the saccade selected. The target location is static over time, ensuring it is the same at the end of the trial as it was at the beginning. It is this enduring context that gives rise to the delay-period activity – thought to be supported by recurrent glutamatergic connections in Layer III of the cortex (Kritzer and Goldman-Rakic 1995) – characteristic of the prefrontal cortex. The third hidden state models transitions to the next stage of the trial, at each time-step.

During the fixation stage of the trial, the visual outcome indicates the target location. During the delay-period, the cross turns red and no cues are shown. An outcome ‘incorrect’ occurs if a saccade is performed during this step. A priori, this outcome is not preferred and is therefore avoided. At the final stage of the trial, the cross changes from back to blue, and a saccade is permitted. ‘Correct’ outcomes ensue if the fixation location hidden state at this time matches

the target location, and ‘incorrect’ pursues otherwise. Prior preferences dictate that ‘correct’ outcomes mark a particular saccade as more likely (in terms of maximising the model evidence or minimising free energy expected following a saccade). At all stages, the proprioceptive outcome is generated through an identity mapping from the fixation location (i.e., the subject has precise sensory evidence about where she is looking). We do not explicitly model free free-viewing during inter-trial intervals, and assume that feedback is given immediately following the response. Equipped with this model, we can now examine the effects of changing the precisions (i.e., simulated neuromodulators) on task performance.

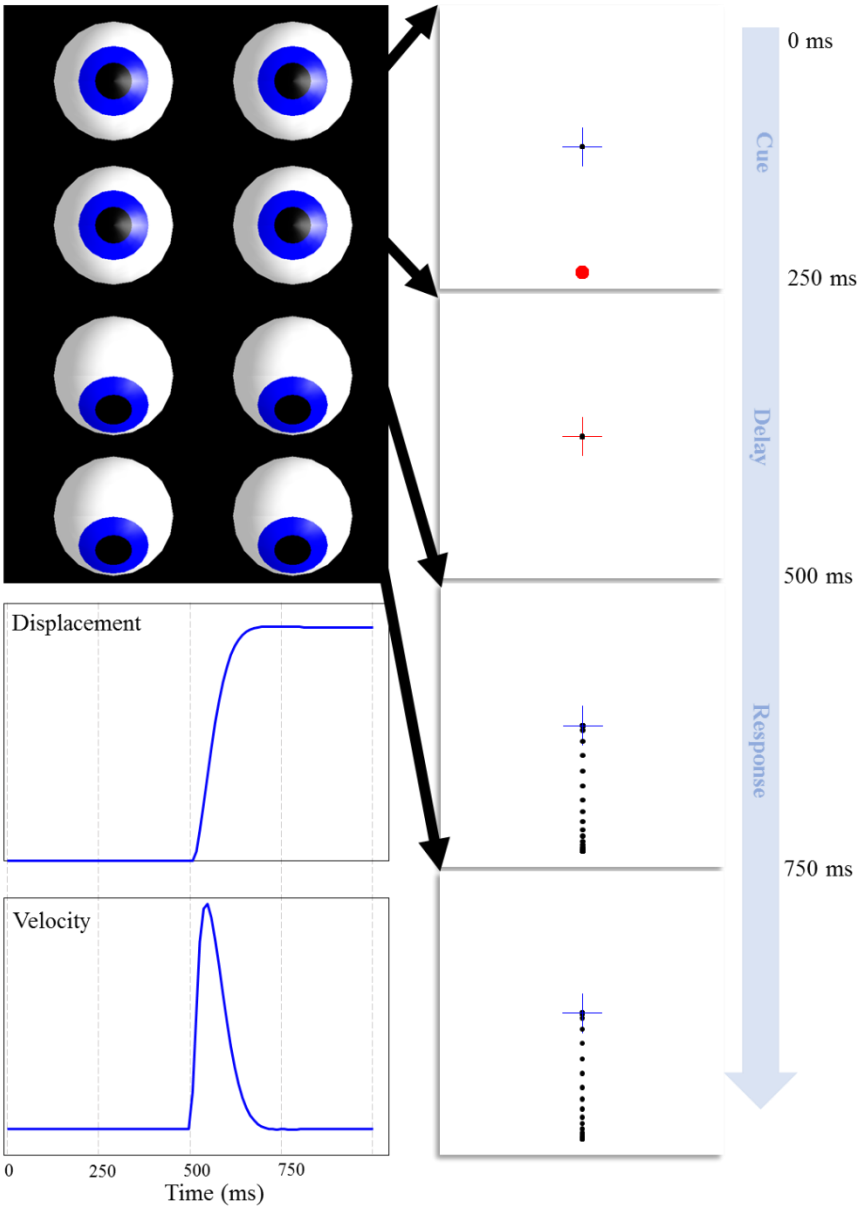

Figure 4.6 – Delayed oculomotor task. This illustrates the sequence of a single trial of our simulated task. The task begins with fixation in the centre, while a peripheral target is presented. The target then disappears, but the cross changes to red, indicating that fixation must be maintained. When the cross changes back to blue, this indicates that a saccade should be made to the target location. The sequence shown represents correct task performance, where the saccade is withheld until the appropriate time, and is then directed to the correct location. The lower left panels show the displacement from the fixation cross and the velocity of the eyes as a function of time, while the upper left images show the position of the eyes at the end of each discrete time-step (i.e. the dotted vertical lines in the lower left plots). Note that the trial sequence takes place over four time-steps that each represent a 250ms continuous trajectory. This was chosen for consistency with the frequency of saccadic sampling. While much longer delay periods are normally employed in practice, the model could be extended to deal with these simply by adding in additional delay periods (each of 250ms). The dashed vertical lines in the plots on the lower left indicate the phases of the trial, as outlined here. These will be used in all subsequent figures for to aid comparison.

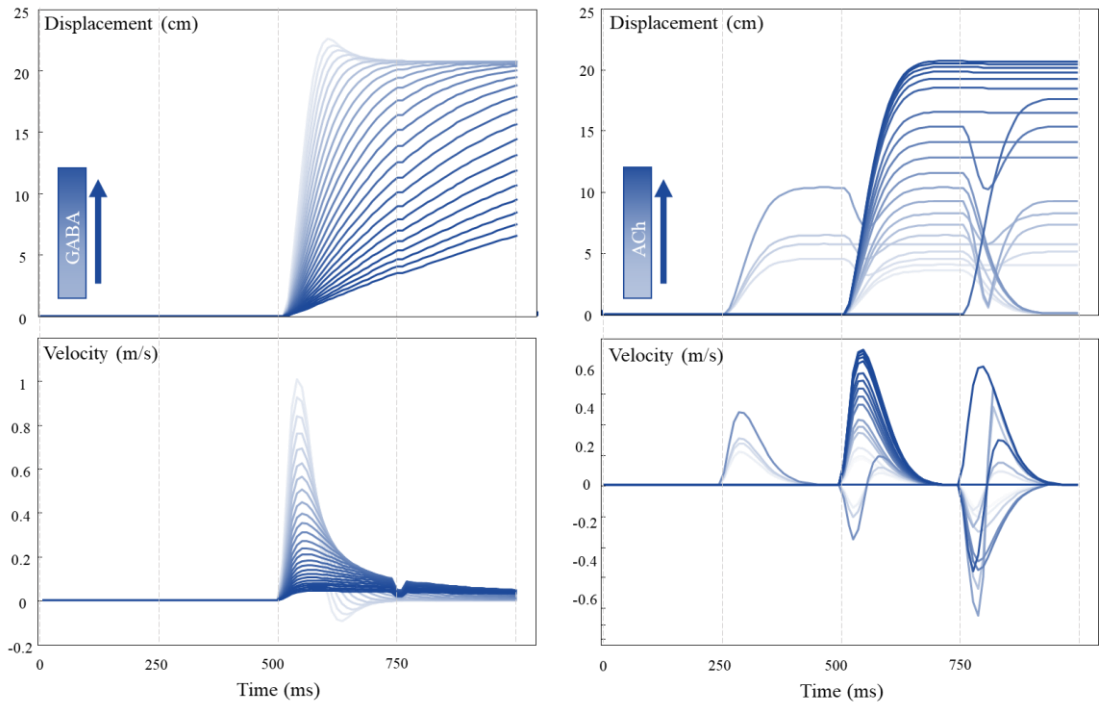

Figure 4.7 – Saccade characteristics. The plots on the left show a set of trials with varying levels of GABA (**II** from Figure 4.5). The upper plot shows the displacement over time through the trial, while the lower plot shows the associated velocity. Note that very low levels of GABA result in an overshoot, that is subsequently corrected, and a higher velocity. In contrast, high levels of GABA lead to slow, hypo-metric saccades. These become broken when the

velocity is sufficiently slow that the saccade takes more than one discrete time-step (vertical dashed lines) to complete. The plots on the right show the same characterisation of saccades with varying levels of cholinergic modulation (ζ from Figure 4.5). These show a similar, but inverted, phenomenology; with increasing levels of acetylcholine leading to faster saccades. Unlike with the GABAergic changes, there is no hypermetric overshoot. The saccades instead converge to the optimal distance. At low levels of acetylcholine, saccades start to occur too early or late, and in some cases, more than one saccade occurs during a given trial. It is useful to try to infer where the normal physiological range of these parameters may lie – to understand the difference between overdoses or depletions. While this is really an empirical question, best answered by fitting these models to data, we can try to address this issue heuristically. Given that healthy eye-movements tend not to overshoot, and that they reach their target displacement quickly, this suggests normal physiological ranges are at the lower end of the GABA scale, and that most of the traces shown above represent excesses above this (with the exception of those that overshoot, which may be depleted). Similarly, the relatively low frequency of inappropriate saccades in healthy people suggest that physiological ranges of acetylcholine are at the higher end of the scale shown here. The improvement elicited by some cholinergic drugs (see main text) suggests that the normal range is not quite at the higher limit shown here.

Gamma-Aminobutyric acid (GABA)

Benzodiazepines are a class of pharmacological agents that act through modulation of GABAergic activity (Griffin et al. 2013). Specifically, they bind to the $GABA_A$ receptor, and facilitate action of the endogenous neurotransmitter. They are commonly used in clinical practice to treat a range of conditions including, but not limited to, anxiety disorders, insomnia, and (in an acute setting) epilepsy. Oculomotor changes during use of these agents are sufficiently robust that they have been proposed as biomarkers for the pharmacological effects (de Visser et al. 2003). These effects include a clear (inverse) dose-response relationship (Bittencourt et al. 1981) with saccadic peak velocity. Although the actions of systemic benzodiazepine administration are difficult to localise, this has also been demonstrated using anatomically precise injections of muscimol (a GABA agonist) directly into the superior colliculus (Hikosaka and Wurtz 1985). This induced a similar attenuation of saccade peak velocity. This is consistent with Figure 4.5, which associates the oculomotor effects of GABA with inhibition of the superior colliculus, and with Figure 4.7, which shows the effect of increasing the precision of beliefs about the anticipated eye position (empirical prior) on the

displacement and velocity of a saccade over time. Notably, the peak velocity decreases with increasing precision, consistent with the effect of increasing the dose of a benzodiazepine. Intuitively, the greater the prior precision is over the dynamics represented in the brainstem, the harder it is to update these beliefs such that the eyes can move to a new location.

Acetylcholine

Cholinergic or anticholinergic effects are common to many drug classes (Campbell et al. 2009; Ness et al. 2006) and also represent an important mode of action of several toxins (e.g. organophosphate pesticides (Minton and Murray 1988)). As with the benzodiazepines, cholinergic effects have been associated with the velocity of a saccadic eye-movement (Naicker et al. 2017). This is interesting from the perspective of the scheme in Figure 4.5, as cholinergic modulation is hypothesised to occur at the level of inference about categorical variables (which saccade to perform and which location is the target). It is not immediately obvious how such inferences could influence continuous variables such as velocity. In addition to its cortical site of action, acetylcholine is vital in the normal function of the striatum and has actions on brainstem nuclei directly (Dautan et al. 2014; Kobayashi and Isa 2002; Maurice et al. 2015). The cholinergic (precision) manipulations illustrated in Figure 4.7 suggest that these categorical inferences can influence velocity in a consistent way. As cholinergic transmission increases so does the peak saccade velocity. This is an example of a functional diaschisis (Carrera and Tononi 2014; Fornito et al. 2015; Price et al. 2001), in which altering one part of a network has implications for all other parts. The effect here is due to the fact that the precision of the state-outcome mapping determines the precision of the predictive distribution over alternative saccadic locations. If this distribution becomes less precise, the expected (average) anticipated location in continuous coordinates becomes a mixture of all of the possible locations, weighted by their relative probability. When the precision is low, this means saccades towards more central locations become more probable and that, as precision increases, the anticipated location will move further towards a specific target.

Hyoscine (a.k.a. scopolamine), an antimuscarinic (anticholinergic) drug (Corallo et al. 2009) used to treat motion sickness, slows the velocity of saccades (Oliva et al. 1993), consistent with Figure 3. Interestingly, it additionally causes saccades to become hypo-metric, and impairs the stability of fixation. These effects are shown clearly in Figure 4.7, with lower levels of cholinergic signalling leading to shorter saccades, and an increase in the number of inappropriate saccades. The latter are due to the fact that precision sharpens or flattens the distribution of plausible saccadic targets. As this distribution becomes flatter, saccades

previously deemed inappropriate acquire plausibility. Agonists of the cholinergic system, notably nicotine, improve the performance of saccade (specifically 'anti-saccade') tasks if performance is suboptimal (e.g. on first exposure to a task) (Rycroft et al. 2006). These effects tend to saturate fairly quickly, implying that nicotine adds no additional benefit (or deficit) when task performance has already been optimised – and this may be why some studies report no improvement or changes in velocity with nicotine administration (Sherr et al. 2002). This is consistent with the saturation of responses we see in our simulations, with increasing levels of precision asymptotically approaching optimal saccadic trajectories. It is encouraging that these studies show similar results to the computationally focal manipulations performed in our simulations, despite the fact that these drugs do not act in an anatomically specific way.

Dopamine

Pharmacological manipulation of the dopamine system can be highly effective in treating both neurological and psychiatric disorders. Parkinson's disease, in which the substantia nigra pars compacta degenerates, responds to L-Dopa (a dopamine precursor) (Smith et al. 2012), in addition to dopamine agonists (Jenner 1995). In contrast, suppression of dopaminergic activity is a key part of the pharmacological strategy adopted in antipsychotic medications (Kapur et al. 2000). In oculomotor tasks, impairments in dopamine signalling cause deficits in saccades, most pronounced in memory-guided saccades (Kato et al. 1995; Kori et al. 1995). Similar deficits have been described in Parkinson's disease (Chan et al. 2005), in which there is a degeneration of dopaminergic nuclei. This is highly consistent with the active inference account of dopamine as representing the precision of beliefs about temporally deep policies, or plans about how to act (Friston et al. 2017a). While a visually guided saccade requires a planning depth of one-step-ahead, a memory guided task required the inference of the appropriate plan over multiple time-steps (from presentation of the cue to the execution of the action). We have previously demonstrated the influence of dopamine on simulated saccades in a memory guided paradigm (Parr and Friston 2017d), and here replicate this influence in the oculomotor delay period task described above. Figure 4.8 illustrates the effect of changing the prior precision on simulated dopamine firing (updates in the precision over time) and its behavioural consequences. Memory-guided saccades are disrupted once dopamine levels drop. Not only are saccades performed to incorrect locations, they also occur at inappropriate times, consistent with the impairment in sequential planning induced here.

These simulations provide further face validity to the idea that dopamine is involved in signalling the precision of beliefs about deep policies. Previous theoretical accounts have

reproduced aspects of the phenomenology of dopamine signalling based upon this (Friston et al. 2013), and have inspired empirical studies, including the use of simulated precision updates as regressors in functional imaging studies (implicating the dopaminergic midbrain) (Schwartenbeck et al. 2015b), and modelling of behavioural responses under pharmacological manipulations (Marshall et al. 2016). A simple experiment that could be performed to further test these ideas would be to fit the model described here to the (saccadic) decisions made in this task (Mirza et al. 2018), and to see whether the prior precision over policies estimated from real participants correlates with their spontaneous blink rate – a peripheral manifestation of central dopamine function (Karson 1983).

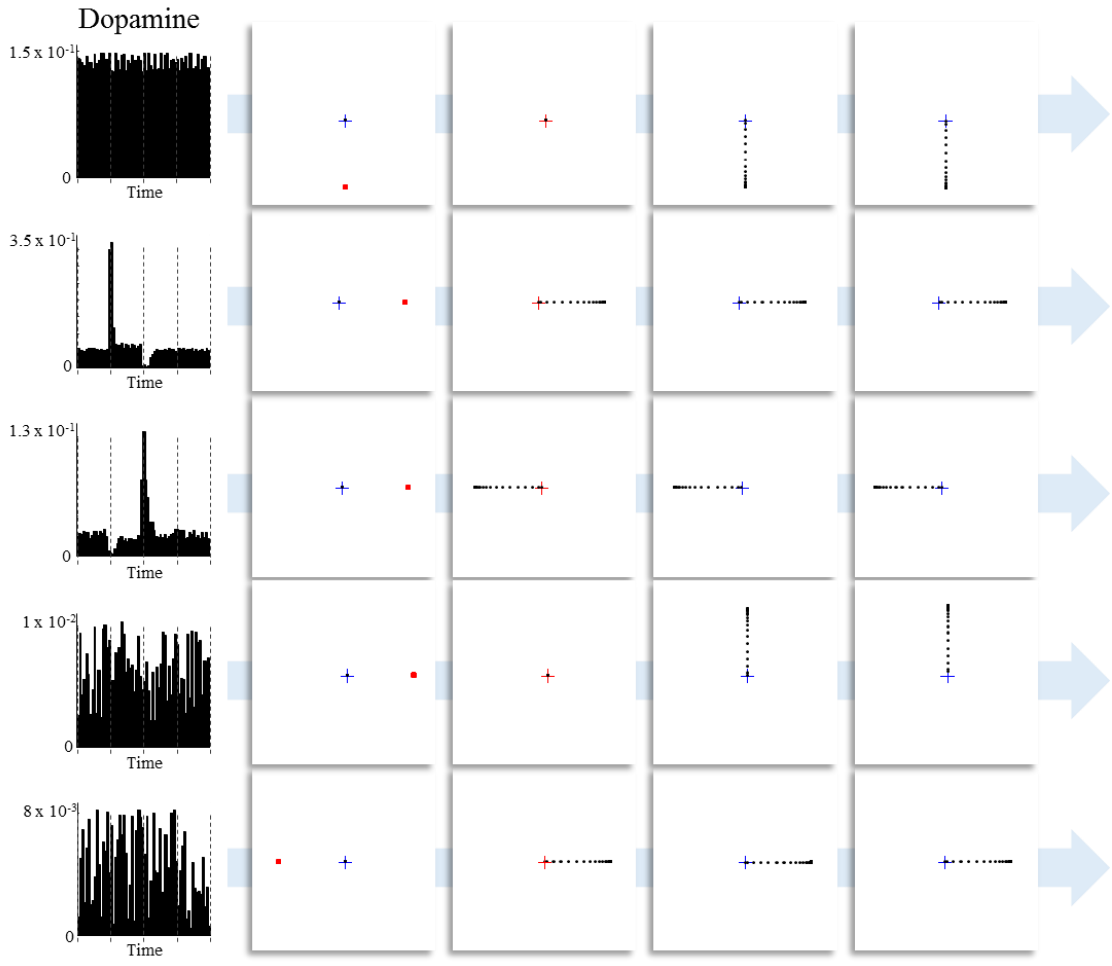

Figure 4.8 – Dopaminergic modulation of saccadic choices This Figure illustrates the effect of dopaminergic modulation of decision making during the delay period task. Each row shows a different level of prior precision over policies (highest for the first row, and lowest for the last). Simulated dopaminergic firing rates are shown on the left (note the differences in axis ranges). When the precision is very low, the selected saccadic target is random, as all possible saccadic policies become (nearly) equally probable. As the prior precision is increased, the

first notable change occurs in the dopamine plots (third row). Here, there is a decrease in dopamine firing during the first saccade, as uncertainty about the policy pursued increases. This is because this saccade is inconsistent with the policy consistent with reaching the target. Having committed to the incorrect policy, there is a dopamine spike coinciding with a confident inference that this policy is being pursued. As dopamine levels increase further, this spike moves earlier, as the saccade performed (although still premature) is still consistent with reaching the target. When sufficiently high, dopamine levels show little change throughout the trial, with the correct policy inferred quickly and confidently from the first time-step. The key message to take away from this Figure is that, in the absence of dopamine, oculomotor decisions become increasingly random. This is because the distribution over action sequences becomes less precise. As the precision tends towards zero, all plans of action become equally probable.

Noradrenaline

The evidence for a modulation of oculomotor responses by noradrenaline is less clear (Reilly et al. 2008). Although some saccadic tasks are reported to vary with noradrenergic modulation (e.g. using methylphenidate (Klein et al. 2002; O'Driscoll et al. 2005)), including upon the timing of saccades (Suzuki and Tanaka 2017), there is little evidence for a systematic influence over behaviour in the delayed oculomotor task. Despite this, there is evidence for neurophysiological changes in circuits implicated in task performance when prefrontal α2-adrenoreceptors are modulated (Arnsten 2011; Arnsten and Li 2005; Sawaguchi et al. 1990). Specifically, administration of clonidine, an α2-agonist, facilitates the delay period activity associated with maintenance of working memory (Li et al. 1999; Suzuki and Tanaka 2017). This is highly consistent with the simulations shown in Figure 4.9, where increasing noradrenergic signalling improves the propagation of information about the past to the future (columns of the raster plots).

Intuitively, an inability to project precise beliefs into the future, or from the past to the present, should undermine the performance of this task. Under active inference, a simple explanation for the preservation of performance even in the absence of precise beliefs about transitions rests upon the use of deep (sequential) policies. If we are able to infer, based upon early observations, the course of action we will pursue, performance becomes robust to the degradation of memories about those observations. In other words, if I know I have to perform a saccade to the left location, whether I believe that this is the target location or not has no influence over my task performance. This illustrates the dissociation between beliefs about

states of the world and beliefs about 'how I will act'. This affords an opportunity to investigate the interaction between neuromodulators and the influence of, for example, noradrenergic modulations on dopamine firing.

Interestingly, selective noradrenaline uptake inhibitors have been shown to be efficacious in treating anxiety disorders (Montoya et al. 2016). These drugs increase signalling at α2-receptors (Grandoso et al. 2004), implying that anxiety may be partly mitigated through an induction of the belief that environmental dynamics are more precise. Under the view that stress is a manifestation of uncertainty (Peters et al. 2017), this conclusion makes a great deal of sense. The reduction in uncertainty about what will happen next, by setting up a belief that the world is actually quite predictable, may be an important part of the computational mechanism of action of these pharmacological agents.

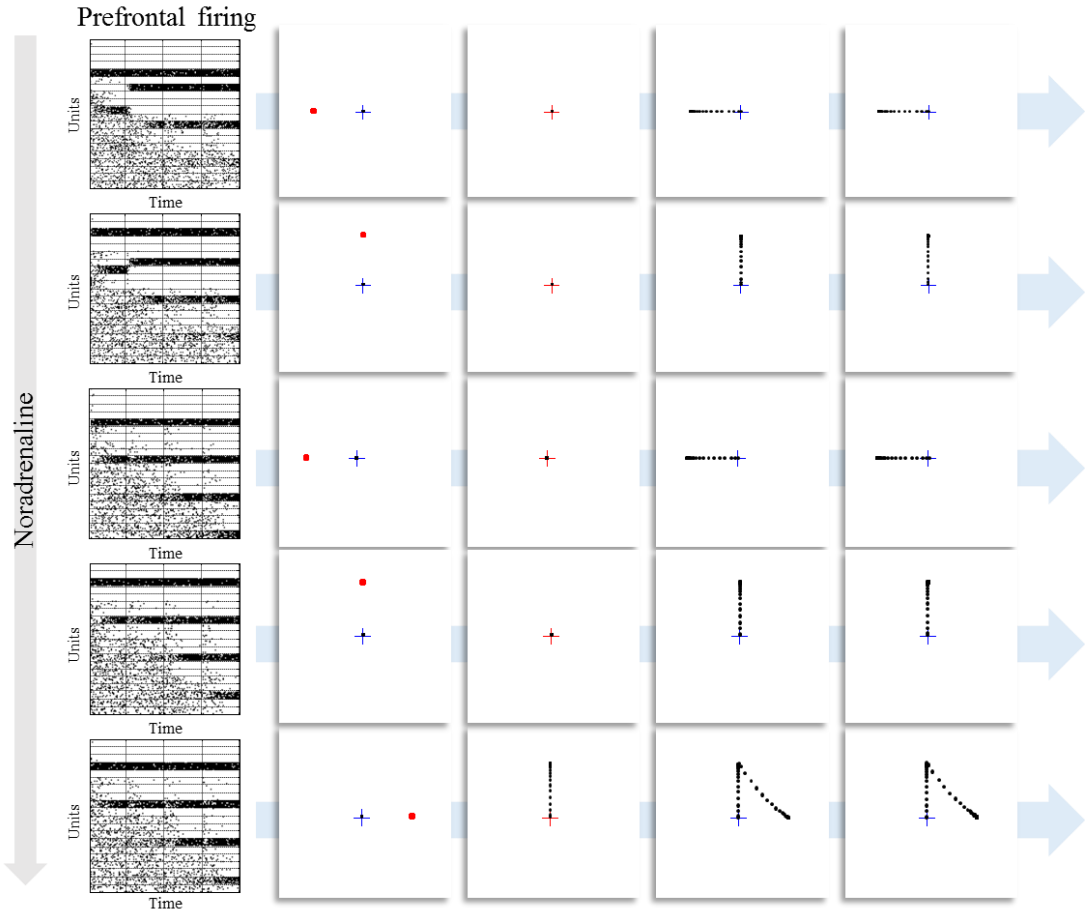

Figure 4.9 – Noradrenergic modulation of prefrontal firing Each row illustrates a single trial of the delay period oculomotor task, but with different levels of noradrenaline. This has little effect on the performance of the task. Even in the lowest row, where a premature saccade takes place, this mistake is corrected at the next time-step. Note that this error is a consequence of the random sampling of actions from beliefs about policies, and does not occur on the majority of trials. It has been retained here to illustrate the change in strategy that leads to the

successful completion of the task. While there are no clear behavioural consequences of this manipulation, the physiological implications are much more striking. These are shown as raster plots of prefrontal cortical neurons representing the remembered target location. Each row represents the firing of a population of neurons representing the probability of one of the target locations at specific times throughout the trial. The lower rows within these plots indicate later times. This means that at the first time step (first column), the last row represents beliefs about the future. By the final time step (last column), the last row represents beliefs about the present. As the concentration of noradrenaline increases, more structure becomes apparent in the lower parts of the plots, indicating a more successful propagation of the inferences drawn from observing the initial cue to the later points in the trial. Note that the increase in persistent activity is accompanied by a decrease in the firing of other neurons, representing the probability of alternative states.

Summary

In the above, we have demonstrated the face validity of the use of active inference to simulate the effects of pharmacological therapies on oculomotor behaviour. Previous accounts of these behaviours have proposed their utility as biomarkers for the action of therapeutic (or toxic) agents in individual patients (de Visser et al. 2001; de Visser et al. 2003; Reilly et al. 2008). Complementing this approach, we offer a mechanistic (computational) account that bridges the gap between chemical and behavioural changes. The advantage of casting this in computational terms is that the model used here can be fit to empirical (eye-tracking) data to estimate the changes in precision brought about by specific drugs (Adams et al. 2016; Mirza et al. 2018; Schwartenbeck and Friston 2016). This offers a mechanistically informed method for non-invasive evaluation of synaptic function in individual patients. In doing so, it may be possible to titrate drug doses to achieve an optimal change in central nervous system function, or to avoid adverse psychopharmacological effects.

Part of the strength of this method is the appeal to behaviours that depend upon inferences in two different, but connected domains (Parr and Friston 2018c). These are categorical decisions between alternative saccadic targets that depend upon working memory and delay period activity, and the continuous implementation of these decisions through oculomotion. This means that it is possible to draw inferences, based upon behaviour, about the function of anatomically disparate brain regions using a single model. An important caveat here is that the model we have used is overly simple from a pharmacological perspective. Notably, we have neglected the fact that neuromodulatory compounds act at many different anatomical sites, and

have different effects on different receptor subtypes. These omissions undoubtedly have important computational consequences. This may be why, although we have captured some aspects of the oculomotor changes induced by different drugs, there are others that are not reproduced. For example, the changes in speed of saccades associated with dopaminergic changes (Lynch et al. 1997) were not seen here. Despite these limitations, the correspondence between drug effects and the behaviours resulting from changes to precision parameters adds further weight to computational accounts of neuromodulatory systems, and offers a tool to evaluate these theoretical accounts empirically.

Conclusion

This chapter comprised two sections dealing with key aspects of neuromodulatory systems in a clinical context. The *Precision and pathology* section addresses the ways in which faulty neuromodulation may permit false inference, and a possible mechanism whereby anatomically distant lesions may cause a diaschisis that impairs (for example) cholinergic function. The *Computational pharmacology* section offers a computational perspective on the influence of commonly used drugs on behaviour. We considered oculomotor behaviour as a specific example that is known to vary with drug administration, and that is easy to measure with non-invasive techniques. The simulations presented above illustrate that some of the key features of oculomotor responses to pharmacological interventions can be replicated *in silico* through an appeal to active inference. Both sections rest upon the idea that planning is inference about how to act, and that these inferences entail predictions about the sensory consequences of action. Each stage of this process is sensitive to the precision associated with the relationship between different kinds of variable, and these precisions are thought to manifest biologically as synaptic gain – subject to neuromodulatory chemicals. Ultimately, we hope that this approach will be useful in a clinical setting, enabling quantitative characterisations of disease processes and pharmacologically induced synaptic modulations using non-invasive measures.

Table 4.1 – Putative roles of neurotransmitters in active inference (Parr and Friston 2018b)

Neurotransmitter	Precision	Evidence

Acetylcholine	Likelihood	• Presence of presynaptic receptors on thalamocortical afferents (Lavine et al. 1997; Sahin et al. 1992) • Modulation of gain of visually evoked responses (Disney et al. 2007; Gil et al. 1997) • Changes in effective connectivity with pharmacological manipulations (Moran et al. 2013b) • Modelling of behavioral responses under pharmacological manipulation (Marshall et al. 2016; Vossel et al. 2014)
Noradrenaline	Transitions	• Maintenance of persistent prefrontal (delay-period) activity (requiring precise transition probabilities) depends upon noradrenaline (Arnsten and Li 2005; Zhang et al. 2013) • Pupillary responses to surprising (i.e. imprecise) sequences (Krishnamurthy et al. 2017; Lavín et al. 2013; Liao et al. 2016; Nassar et al. 2012; Vincent et al. In press) • Modelling of behavioral responses under pharmacological manipulation (Marshall et al. 2016)
Dopamine	Policies	• Expressed post-synaptically on striatal medium spiny neurons (Freund et al. 1984; Yager et al. 2015) • Computational fMRI reveals midbrain activity with changes in precision (Schwartenbeck et al. 2015b) • Modelling of behavioral responses under pharmacological manipulation (Marshall et al. 2016)
Serotonin	Preferences or interoceptive likelihood	• Receptors expressed on layer V pyramidal cells (Aghajanian and Marek 1999; Elliott et al. 2018; Lambe et al. 2000) in medial prefrontal cortex

	• Medial prefrontal cortical regions heavily implicated in interoceptive processing and autonomic regulation (Marek et al. 2013; Mukherjee et al. 2016)

5 – Novelty, neglect, and dynamic causal modelling

Introduction

Visual neglect is a debilitating neuropsychological phenomenon that has many clinical implications and – in cognitive neuroscience – offers an important lesion deficit model. In this chapter[15], we describe a computational model of visual neglect based upon active inference. Our objective is to establish a computational and neurophysiological process theory that can be used to disambiguate among the various causes of this important syndrome; namely, a computational neuropsychology of visual neglect. In the *Novelty and neglect* section of this chapter, we introduce a Bayes optimal model based upon Markov decision processes that reproduces the visual searches induced by the *line cancellation task* (used to characterise visual neglect at the bedside). We then consider three distinct ways in which the model could be lesioned to reproduce neuropsychological (visual search) deficits. Crucially, these three levels of pathology map nicely onto the neuroanatomy of saccadic eye movements and the systems implicated in visual neglect.

A key aspect of this model comes back to the idea, expressed in Figure 1.4, that high dimensional beliefs may be more efficiently represented through short-term (synaptic) plastic changes than through delay-period activity (of the sort exploited in Chapter 4, Figure 4.9). This implies we need to move beyond inferences about states, or the relatively course precision modulations in the last chapter, and to optimise beliefs about the parameters of the model. When we consider the contribution of this additional set of beliefs to the expected free energy, we find that these supplement the expected free energy with an information gain associated with parameters. In analogy with the *salience* we associated with information gain about states of the world, we now associated a *novelty* value with the corresponding information gain about the probabilistic structure of the world.

Given that performance of the dot cancellation task, in our model, depends upon parameter learning of the sort that could be mediated by changes in synaptic efficacy, this predicts changes in effective connectivity between different brain regions over the time-course required to perform this task. The neuropsychology of visual neglect, in combination with the computation lesions required to induce neglect *in silico*, motivates a specific hypothesis about the brain regions whose coupling we would expect to change over time. As detailed in the *Dynamic causal modelling* section below, we collected magnetoencephalography data from

[15] This chapter is adapted from (Parr and Friston 2017b; Parr et al. 2019c)

healthy participants while they performed the same cancellation task as our simulation. We use these data to test the hypotheses arising from the *Novelty and neglect* section, finding evidence in favour of time-dependent changes in the coupling between dorsal frontal and ventral parietal regions during the performance of this task.

Novelty and neglect

This section has three key aims. The first is to provide an example of the role of novelty and learning under active inference in the context of active vision. The second is to illustrate a computational approach to neuropsychology, developing a belief-based differential diagnosis (i.e. a set of alternative computational lesions that could explain pathological behaviour), and illustrating how these may be disambiguated using non-invasive eye-tracking and Bayesian model comparison. The third aim of this section is to appeal to the neuropsychology of visual neglect to form specific, anatomically informed, hypotheses about the neurophysiological changes that underwrite healthy active vision, amenable to testing using non-invasive neuroimaging techniques.

Visual neglect

Visual neglect is a common syndrome in which patients neglect one side (typically the left) of space (Halligan and Marshall 1998). It is often caused by right middle cerebral artery strokes, but has also been reported as a consequence of inflammatory (Gilad et al. 2006), metabolic (Auclair et al. 2008), and degenerative (Andrade et al. 2010; Ho et al. 2003) diseases. It has also been observed as a feature of seizure activity (Heilman and Howell 1980; Schomer and Drislane 2015; Turtzo et al. 2008), and as part of a migraine aura (Di Stefano et al. 2013). In addition to the wide range of pathological processes which can cause the syndrome, visual neglect can be caused by a range of anatomical lesions. These include both cortical (Corbetta and Shulman 2002) and subcortical (Karnath et al. 2002) insults. There is evidence that the heterogeneity of the causes of visual neglect map on to distinct behavioural phenotypes (Grimsen et al. 2008; Hillis et al. 2005; Medina et al. 2009; Verdon et al. 2009), and this has the potential to be exploited clinically and scientifically.

Eye tracking provides one way in which to characterise behavioural deficits in visual neglect. These measurements have demonstrated that patients with visual neglect perform saccades to

the right side of space with a disproportionately high frequency, compared to leftward saccades. This occurs both spontaneously (Fruhmann Berger et al. 2008; Karnath and Rorden 2012) and during search tasks (Husain et al. 2001). While these biases will form the main subject of this chapter, it is important to note that it may be possible to elicit signs of neglect in patients with no deficit in ocular exploration. For example, in tasks requiring a manual response, it is possible that patients may exhibit a normal pattern of saccadic eye movements, but that they may be impaired in executing a response (Bourgeois et al. 2015; Ladavas et al. 1997). In this section, we consider the control of eye movements, and the conditions that would have to be fulfilled in order to explain the saccadic patterns observed in visual neglect. We aim to show that there is a well-defined and distinct set of conditions that can reproduce the neglect syndrome.

Active inference provides a principled framework in which to define these conditions – in terms of the prior beliefs that a patient would have to possess for their behaviour to be Bayes optimal. The notion of optimal pathology might seem a strange one, but the existence of a set of prior beliefs that renders any behaviour optimal is mandated by the complete class theorems (Daunizeau et al. 2010; Wald 1947). This means that we can characterise pathology in terms of optimal inference, but in a system or subject that operates under a poor model of its environment (Conant and Ashby 1970). In the following, we briefly review active inference and show how this normative approach can be used to identify the functional lesions that could cause visual neglect. We then propose a neuroanatomical network that is consistent with the neuronal message passing implied by active inference. This allows us to equate functional lesions to anatomical lesions, and to simulate saccadic eye movements for each lesion *in silico*. We explore the influence of subcortical structures (Karnath et al. 2002) in visual neglect, and the notion that visual neglect is a type of disconnection syndrome (Bartolomeo et al. 2007; He et al. 2007). This section concludes by asking the question whether the different sorts of (saccadic) behaviour induced by distinct sorts of lesions is sufficient to identify the locus of the lesion. We address this question using *in silico* neuropsychology and Bayesian model selection.

The purpose of this section is to describe the active inference scheme and establish its predictive validity in (simulated) visual neglect. In the following section, we will validate the underlying functional anatomy using eye tracking and MEG in real (normal) subjects. Our ultimate objective is to translate this model into clinical studies – to provide a functionally and biologically grounded characterisation of neuronal computations in patients with visual neglect.

Learning, novelty, and expected free energy

Chapter 2 detailed the importance of free energy minimisation, and of the selection of policies that minimise expected free energy. Specifically, policies which are associated with a smaller expected free energy should be considered more likely than those associated with a larger expected free energy. Chapter 3 illustrated the consequences of this for active visual sampling in selection of salient (uncertainty resolving) saccades. Here, we generalise this to the resolution of uncertainty about the parameters of a generative model. To do so, we must specify a generative model that includes prior beliefs about these parameters, and the Bayesian belief-updating that realises the learning of the generative model. Given that the probability distributions are specified as categorical distributions, the appropriate conjugate prior for the likelihood (**A**) matrix is a Dirichlet distribution. This means that the probability can be represented simply in terms of Dirichlet concentration parameters. For each state (*s*) there are a set of Dirichlet parameters (*a*) one for each outcome, which could be associated with this state. These are initially 'pseudo-observations', as no observation has yet been made. The belief about the probability of an outcome (*o*) given a state is:

$$P(o \mid s, \mathbf{A}) = Cat(\mathbf{A})$$
$$P(\mathbf{A} \mid a) = Dir(a) \Rightarrow \begin{cases} E_{P(\mathbf{A}|a)}[\mathbf{A}_{ij}] = \dfrac{a_{ij}}{\sum_k a_{kj}} \\ E_{P(\mathbf{A}|a)}[\ln \mathbf{A}_{ij}] = \psi(a_{ij}) - \psi(\sum_k a_{kj}) \end{cases} \quad (5.1$$

In Equation 5.1, ψ is a digamma (derivative of a gamma) function. As observations are made, the agent is able to learn – that is, accumulate its Dirichlet parameters – to better fit its observations. This process of learning simply involves increasing the Dirichlet parameter representing a particular outcome when that is observed (Beal 2003; Blei et al. 2003). The amount it is increased by the (approximate) posterior probability that each hidden state was occupied when the observation was made. This allows a creature to remember the observations they sampled when they believed they were in a particular state (Friston et al. 2016b). The notion that the mapping between representations of two variables should be increased when the two are simultaneously active is strikingly similar to Hebbian plasticity (Brown et al. 2009; Hebb 1949). This analysis suggests that this form of memory could be implemented by short-term changes in synaptic efficacy. More formally, the free energy, incorporating these priors, is:

$$F = \boldsymbol{\pi} \cdot (\ln \boldsymbol{\pi} - \ln \mathbf{E} + \mathbf{F} + \gamma \cdot \mathbf{G}) + \underbrace{(\mathbf{a} - a) \cdot E[\ln \mathbf{A}] + \ldots}_{D_{KL}[Q(\mathbf{A}) \| P(\mathbf{A})]}$$

$$\mathbf{F}_{\pi} = \sum_{\tau} \mathbf{F}_{\pi\tau}$$

$$\mathbf{F}_{\pi\tau} = \ldots - (E[\ln \mathbf{A}]\mathbf{s}_{\pi\tau}) \cdot o_{\tau} + \ldots \quad (5.2$$

$$Q(\tilde{s}, \pi, \mathbf{A}) = Q(\tilde{s}, \pi) Q(\mathbf{A})$$

$$Q(\mathbf{A}) = Cat(\mathbf{a})$$

Taking the gradient of this expression with respect to the expected sufficient statistics of the likelihood:

$$\nabla_{E[\ln \mathbf{A}]} F = \mathbf{a} - a - \sum_{\tau} \mathbf{s}_{\tau} \otimes o_{\tau} = 0 \Leftrightarrow \mathbf{a} = a + \sum_{\tau} \mathbf{s}_{\tau} \otimes o_{\tau} \quad (5.3$$

The final expression here expresses the activity dependent plastic changes described above (i.e., the accumulation of Dirichlet parameters). If we interpret the beliefs about states and their corresponding observations as represented in firing rates, and the **A-**matrix as a connectivity matrix, Equation 5.3 simply says that the strength of a given synapse is incremented whenever the neurons representing the state and observation connected by that synapse are simultaneously active. Note that this Hebbian perspective on learning emerges from the generative model and the minimisation of free energy. An important consequence of the Dirichlet parameterisation concerns the scaling of parameters. The scaling of the Dirichlet parameters does not influence the values in the likelihood matrix. However, it does influence the degree to which these change following an observation. If all the concentration parameters are very large (as would be the case if many past observations had been made), a single observation will make a very small difference to the likelihood. If the parameters are very small, an observation can trigger one-shot learning, suggesting a rapid short term plasticity effect[16]. Such effects have been proposed as one mechanism underlying working memory (Mongillo et al. 2008). This behaviour is of particular interest in the current context, as will

[16] Intuitively, imagine flipping a coin 5 times, and getting 5 heads in a row. This might lead us to update our beliefs to favour the hypothesis that this is an unfair coin. However, if this had been preceded by 100 flips with 50 heads and 50 tails, the final 5 heads would do little to influence our beliefs about whether or not this is a fair coin.

become apparent in the next section, where the form of the MDP used to model visual neglect is described.

Having specified the way in which beliefs about parameters are updated, we now need to supplement the expected free energy (Equation 2.16) with priors and posteriors about the parameters.

$$\begin{aligned} G(\pi) &= E_{Q(o,s,\mathbf{A}|\pi)}[\ln Q(s,\mathbf{A}\mid\pi) - \ln P(o,s,\mathbf{A})] \\ &= E_{Q(o,s,\mathbf{A}|\pi)}[\ln Q(s\mid\pi) - \ln P(o,s)] \\ &\quad + E_{Q(\mathbf{A}|s,o)Q(o,s|\pi)}[\ln Q(\mathbf{A}) - \ln P(\mathbf{A}\mid s,o)] \\ &\approx \underbrace{D_{KL}[Q(o\mid\pi)\,\|\,P(o)]}_{Risk} + \underbrace{E_{Q(s|\pi)}[\ln P(o\mid s)]}_{Ambiguity} \\ &\quad - \underbrace{E_{Q(o,s|\pi)}[D_{KL}[Q(\mathbf{A}\mid s,o)\,\|\,Q(\mathbf{A})]]}_{Novelty} \end{aligned} \tag{5.4}$$

Remembering that lower expected free energies are associated with more probable policies, this expression implies that the greater the expected update from prior to posterior beliefs about the parameters (the *novelty* term), the more probable the policy. This implies that the salience discussed in Chapter 3 in relation to inference is supplemented with the novelty afforded by the potential to learn. Appendix A.5 derives the form of the expression for the novelty, which may be expressed in terms of a matrix (**W**):

$$\begin{aligned} \mathbf{G}_{\pi\tau} &= \mathbf{o}_{\pi\tau}\cdot(\ln\mathbf{o}_{\pi\tau} - \ln\mathbf{C}) + \mathbf{H}\cdot\mathbf{s}_{\pi\tau} - \mathbf{o}_{\pi\tau}\cdot\mathbf{W}\mathbf{s}_{\pi\tau} \\ \mathbf{W}_{ij} &\triangleq \tfrac{1}{2}\left(\frac{1}{a_{ij}} - \frac{1}{\sum_k a_{kj}}\right) \end{aligned} \tag{5.5}$$

Connecting this back to the accumulation of Dirichlet parameters, we see that very large prior Dirichlet parameters yield very little novelty. This is consistent with the fact that an observation is unlikely to cause a substantial change in the expected likelihood. Equation 5.5 makes an important connection between the capacity to increment a connection strength (which depends upon the size of the initial Dirichlet parameters) and the novelty value associated with the observations expected under a given policy.

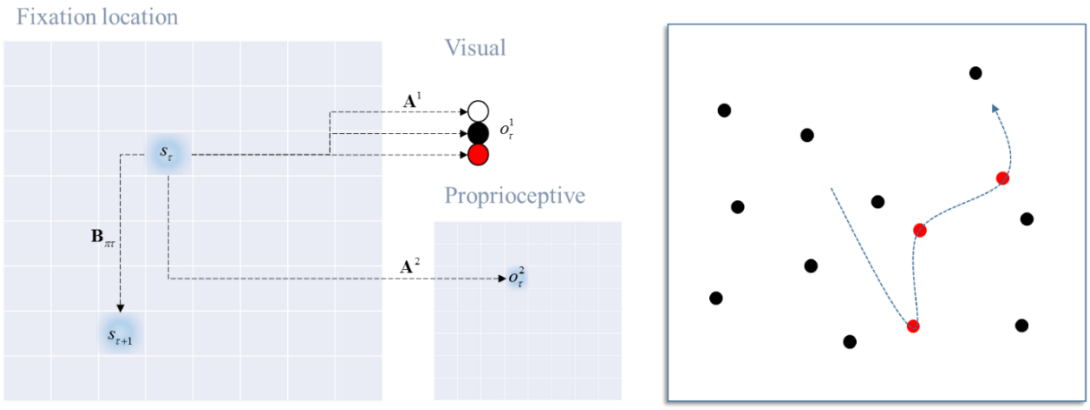

Figure 5.1 - Generative model for the saccadic cancellation task. The structure of the particular generative model used for the saccadic cancellation task is shown on the left. The hidden states in this model are the locations in the visual field that are fixated. These are indicated by the 8x8 grid on the left of this figure. The agent may saccade to any location on the grid, and the particular saccade is defined by the policy (π) which selects the appropriate **B**-matrix. Each component of this matrix defines the probability of a saccade to a given location, given a current location. There are two A matrices which provide a probabilistic mapping from the hidden states to the visual ($\mathbf{A}^1$) or proprioceptive ($\mathbf{A}^2$) outcome modalities. Prior preferences are defined by the **C**-vectors for each modality. On the right is a depiction of the structure of the task resulting from the generative model. The dotted line is the saccade path, and this demonstrates the change from black to red of targets as they are cancelled.

Saccadic cancellation task

The task performed by the particular MDP model used in this work is based on the pen-and-paper line cancellation task (Albert 1973; Fullerton et al. 1986). This task is used to assess visual neglect clinically, and is very sensitive (Ferber and Karnath 2001). Despite its popularity, it is worth noting that there are many possible reasons that performance of this task might be impaired. We will demonstrate this for a few of these reasons below. We will use a saccadic version, which involves presenting the subject with an array of targets that can be placed at various locations on an 8x8 grid. The task is to look at each of the targets until all targets have been sampled (i.e., cancelled). When a target has been fixated, it changes colour from black to red (see the right panel in Figure 5.1), indicating that it has been seen. The model used to emulate this behaviour is shown in Figure 5.1, in terms of the variables in the MDP. The only hidden states in this model correspond to the location currently foveated. An identity

matrix maps these deterministically to proprioceptive observations, ensuring there is no uncertainty about the hidden state (i.e., where the subject is currently looking). The uncertainty in the model is contained in the (likelihood) mapping from hidden states to visual outcomes. There are three possible outcomes: no target (white), target (black), and cancelled target (red). The prior preferences of the simulated agent are that it has equal preferences over all proprioceptive outcomes, prefers to see targets that have not been cancelled, and does not expect to see targets that have already been cancelled. Our synthetic subject begins with (almost) uniform beliefs about the **A**-matrix (i.e., what will happen if she looks at a particular location). However, these incorporate very weak, but accurate, beliefs concerning the locations of the targets. On foveating a target, the first visual outcome is a black target. This observation allows the appropriate Dirichlet parameters to be accumulated. During fixation, the target changes from black to red, and this causes further changes in the Dirichlet parameters, so that the subject remembers she has already cancelled that location. This implements a synaptic form of spatial working memory (Mongillo et al. 2008), of exactly the sort shown in Figure 1.4. The subject may saccade to any location at any time, meaning there are 64 possible actions, each of which is associated with a corresponding transition (**B**) matrix. Having established the basic form of the generative model, sufficient to simulate visual search, we now turn to the finer details of the implicit epistemic foraging, how novel targets are selected – and what can go wrong under pathological priors.

Computational neuropsychology

In principle (under the complete class theorem), all neuropsychological syndromes can be formulated in terms of active inference. The challenge is to find the prior beliefs a subject would have to possess to render their behaviour Bayes optimal. For visual neglect, we consider the abnormal patterns of saccadic eye movements in patients (Bays et al. 2010; Fruhmann Berger et al. 2008; Husain et al. 2001; Karnath and Rorden 2012), and the beliefs which would engender these patterns. For each saccadic policy, the generative model specifies the prior probability that the policy will be pursued. By analysing the form of this prior belief, one can develop a differential diagnosis for the computational lesions in visual neglect. As noted above, the prior belief about policy should depend on the expected free energy. The smaller the expected free energy under a policy, the more likely it will be pursued. Rearranging the expected free energy from 5.4, we can dissect out the terms related to different sorts of information gain:

$$G(\pi) = -\underbrace{E_{Q(o|\pi)}[\ln P(o)]}_{Preferences} - \underbrace{E_{Q(o|\pi)}[D_{KL}[Q(s \mid o,\pi) \,||\, Q(s \mid \pi)]]}_{Salience} - \underbrace{E_{Q(o,s|\pi)}[D_{KL}[Q(A \mid o,s) \,||\, Q(A)]]}_{Novelty} \quad (5.6$$

$$Q(s \mid o,\pi) \triangleq \frac{Q(s \mid \pi)P(o \mid s)}{Q(o \mid \pi)}$$

The second (salience) term in this equation, in the context of the generative model used here, is identical for all policies. This is because the identity mapping from the hidden states representing locations to the proprioceptive outcomes allows the subject to infer location in visual space with certainty. This means there is no information gain or epistemic value that would otherwise resolve uncertainty about the hidden states. The key terms that determine policy selection are the first and third. The former implies that a policy which is expected to fulfil the agent's prior beliefs (preferences) about outcomes has a lower expected free energy than one which does not. The latter suggests that a policy which affords the greatest change in the beliefs about the likelihood mapping, from beliefs prior to seeing counterfactual outcomes has the lowest expected free energy. Heuristically, policies that elicit observations that enable large Bayesian belief updates become more attractive. In other words, the subject will be attracted to novel contingencies that resolve uncertainty about the consequences of being in a particular state; i.e., the likelihood mapping.

In short, prior preferences and novelty are both important factors in determining the selection of a location to saccade to. This implies two possible computational mechanisms for visual neglect. A subject may have a prior belief that she will experience the proprioceptive outcomes corresponding to the right side of space with a greater probability than those corresponding to the left. Alternatively, the subject may be more confident in her beliefs about the mapping from states to outcomes on the left, and therefore consider the right side of visual space novel. This is equivalent to starting with very large Dirichlet parameters (corresponding to a large number of pseudo-observations) for locations on the left. This follows because an observation resulting from a saccade to the left will induce a small change in beliefs about the likelihood mapping.

A third possibility relates to (baseline) prior beliefs about policies that may not depend upon expected free energy. Although active inference mandates that an agent believes it will pursue policies which minimise its expected free energy, it does not preclude fixed prior beliefs over policies which, in visual neglect, might identify saccades to the right to be *a priori* more likely

than those to the left. To express this formally, we can augment the expression for priors over policies as in Equation 2.15 (omitting the salience terms, given the argument above):

$$
\begin{aligned}
P(\pi) &= Cat(\boldsymbol{\pi}_0) \\
\boldsymbol{\pi}_0 &= \sigma(\underset{(2)}{\ln \mathbf{E}} - \boldsymbol{\gamma} \cdot \mathbf{G}) \\
\mathbf{G}_\pi &= -\sum\nolimits_\tau \mathbf{o}_{\pi\tau} \cdot (\underset{(3)}{\ln \mathbf{C}} - \underset{(1)}{\mathbf{W}\mathbf{s}_{\pi\tau}})
\end{aligned}
\tag{5.7}
$$

Here, **E** expresses the prior beliefs about policies that do not depend on the expected free energy. In this form, the (log) priors over policies are expressed as a linear function of expected free energy, where **E** corresponds to the y-intercept and precision is the sensitivity or slope. The numbers labelling various terms in Equation 5.7 index the terms that we lesioned for the simulations, and correspond to the numbers in black circles shown in Figure 5.2.

In summary, the above formal considerations have led us to identify three possible synthetic lesions which could give rise to visual neglect. These are changes in the priors over policy **E**, the Dirichlet parameters of the beliefs about **A**, and the priors concerning proprioceptive outcomes, contained in the matrix **C**. In the following section, we review plausible neurobiological substrates for each of these computational pathologies.

The neuroanatomy of visual neglect

As we detailed in Chapters 1 and 3, the superior colliculus, in the midbrain, is a key site for the control of saccadic eye movements (Raybourn and Keller 1977). It is also a point of convergence for the cortical and subcortical structures involved in oculomotor control (Berson and McIlwain 1983; Fries 1984; Fries 1985; Gaymard et al. 2003; Künzle and Akert 1977; Shook et al. 1990). The substantia nigra pars reticulata, a GABAergic output nucleus of the basal ganglia, projects directly to the colliculus (Hikosaka and Wurtz 1983), as do cortical areas including the frontal eye fields (Künzle and Akert 1977) and the lateral intraparietal cortex (Gaymard et al. 2003) (sometimes called the parietal eye fields (Shipp 2004)). These dorsal frontal and parietal areas constitute the dorsal attention network (Corbetta and Shulman 2002), and communicate via the first branch of the superior longitudinal fasciculus (Bartolomeo et al. 2012; Makris et al. 2004). The frontal eye fields are well placed to house the hidden states representing eye position, while dorsal parietal areas are suited to the

representation of proprioceptive information. The former are known to contain spatial maps in egocentric space, as evidenced by demonstrations that stimulation of neurons in this region induce saccades that end in specific egocentric eye positions (Bruce et al. 1985; Sajad et al. 2015). The latter contain neurons that are modulated by multiple spatial reference frames (Andersen et al. 1985; Pouget and Sejnowski 2001).

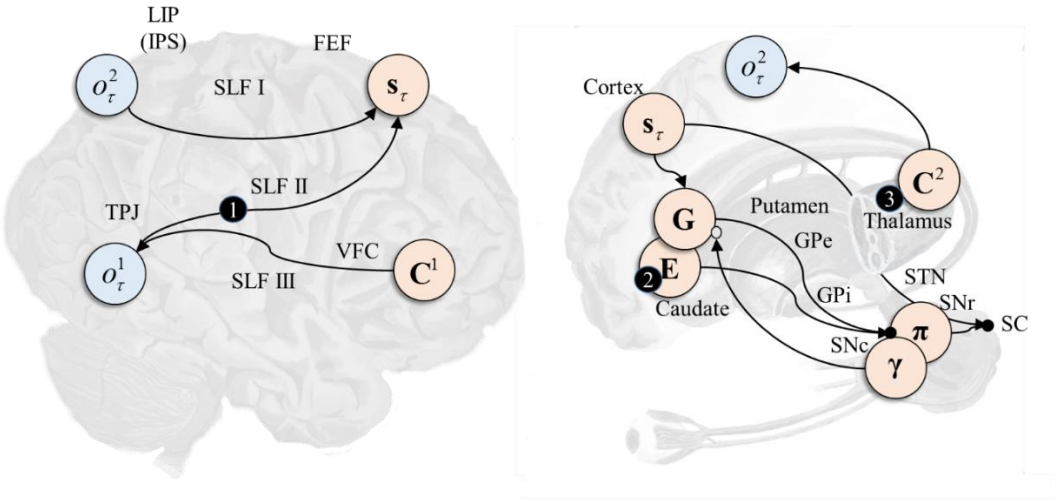

Figure 5.2 - Computational anatomy and lesion sites. This schematic illustrates the proposed mapping from the computational entities implicated by the model (see Figure 5.1 and 2.2) and their neuroanatomical substrates. On *the left* the dorsal and ventral attention networks are shown. The former involves the frontal eye fields (FEF) and posterior parietal areas in the region of the lateral intraparietal area (LIP) and intraparietal sulcus (IPS). The frontal areas of this network are assumed to represent the hidden states, corresponding to the current fixation location. The parietal component represents proprioceptive outcomes (eye position). The connection between these frontoparietal areas is the first branch of the superior longitudinal fasciculus (SLF I), mediating the likelihood mapping between the hidden states and proprioceptive outcomes ($\mathbf{A}^2$). The ventral attention network includes the ventral frontal cortex (VFC) and the temporoparietal junction (TPJ). These are connected by SLF III, which could carry prior preferences about visual outcomes ($\mathbf{C}^1$). Visual outcomes are assumed to be represented in the TPJ, which suggests the SLF II is the mapping from hidden states to visual outcomes ($\mathbf{A}^1$), and it is in these connections that the beliefs about the target locations are encoded. Prior preferences for proprioceptive outcomes are assigned to the pulvinar, a nucleus of the thalamus. On *the right* the connections from the pulvinar to the dorsal parietal cortex (LIP) are shown. These are portrayed as conveying expectations about (proprioceptive) outcomes in $\mathbf{C}^2$. In addition, the pathways through the basal ganglia are also shown. The policy

evaluation processes are depicted as stages in the direct pathway. In this scheme, the putamen evaluates the expected free energy, and baseline policy priors, **E**. These are modulated by dopaminergic inputs from the substantia nigra pars compacta (SNc), in proportion to their precision γ, and the output of the putamen is transformed by the substantia nigra pars reticulata (SNr) into a distribution over policies. The simulated lesions we considered are numbered: 1 – SLF II; 2 – Putamen; 3 – Pulvinar.

The parietal cortex is part of the dorsal visual stream, thought to carry information about the location of a stimulus (Goodale and Milner 1992; Ungerleider and Haxby 1994). In the present context, the first branch of the superior longitudinal fasciculus would perform a coordinate transformation, bringing spatial information about a stimulus into egocentric coordinates; suitable for planning eye movements. This suggests that the superior longitudinal fasciculus corresponds to the connectivity or mapping encoding the likelihood matrix **A** (see Figure 5.2). In our model, this is an identity mapping, but this is only the case when the head is assumed to be in a fixed position. A model which allowed for head movements would require this matrix to represent a more complex coordinate transform. Given proprioceptive outcomes are represented in the dorsal parietal regions; inputs to this region must represent prior beliefs concerning proprioception. A candidate structure providing this information is the pulvinar, which is involved in visual search behaviours (Ungerleider and Christensen 1979). The connections from this region would then encode the **C** matrix.

Despite the important role of dorsal frontoparietal areas in the generation of saccadic movements (Corbetta et al. 1998), it is more ventral frontoparietal lesions which are associated with the visual neglect syndrome (Corbetta et al. 2000; Corbetta and Shulman 2002; Corbetta and Shulman 2011). These regions are the constituents of the ventral attention network, and are connected by the third branch of the superior longitudinal fasciculus (Bartolomeo et al. 2012; Rushworth et al. 2005). The parietal part of this network includes areas in the region of the temporoparietal junction, closer to the temporal regions associated with the ventral visual stream. This component of the visual system has been described as the 'what' pathway (Ungerleider and Haxby 1994), propagating information concerning stimulus identity to complement the 'where' information of the dorsal stream. The ventral temporoparietal regions are then good candidates for the representation of the visual outcome modality of the model, allowing them to influence eye movements in a stimulus-driven manner (Shomstein et al. 2010). Connections from the ventral frontal cortex could then carry information concerning prior beliefs (equivalent here to the instructions a subject would be given), consistent with the proposed role of areas in this region in representing task demands (Dosenbach et al. 2006;

Duncan 2001) and in target detection (Stevens et al. 2005). This suggests that the third branch of the superior longitudinal fasciculus is the anatomical substrate of **C**.

Notably, the ventral attention network is lateralised to the right cerebral cortex, while the dorsal network is much more symmetrical (Corbetta et al. 2002; Thiebaut de Schotten et al. 2011; Vossel et al. 2012). This is consistent with the notion that temporal regions could represent the 'what' modality, as identity is largely independent of location, and therefore does not require a bilateral representation (Parr and Friston 2017a). There is evidence to suggest that this unilateral representation of identity is right lateralised (Warrington and James 1967; Warrington and James 1988; Warrington and Taylor 1973), while left sided homologues relate to object naming (Kirshner 2003). We note that, although temporoparietal regions are thought to play a role in target detection (Corbetta et al. 2000), they do not appear to be necessary for object recognition. The involvement of the ventral network is consistent with the fact that visual neglect is frequently associated with right hemispheric lesions.

This leaves the question of how lesions in ventral regions produce the saccadic deficits that might be expected from dysfunction of areas which are directly involved in saccadic control. One answer to this question is that visual neglect involves dysfunction of the dorsal network as a consequence of the failure of the ventral network, or of the interaction of the two networks (He et al. 2007). The two networks are joined by the second branch of the superior longitudinal fasciculus (Thiebaut de Schotten et al. 2011), and it has been proposed that visual neglect represents a functional disconnection syndrome involving this pathway. Given that this branch connects the parietal part of the ventral system to the frontal part of the dorsal system, this corresponds exactly to the mapping described by **A**. It is interesting that this tract, heavily implicated in visual neglect (Doricchi and Tomaiuolo 2003; Thiebaut de Schotten et al. 2005), appears to be the anatomical homologue of the mathematical entity identified above as a candidate for pathological priors – on purely theoretical grounds.

As stated above, an important input to the superior colliculus is the substantia nigra pars reticulata. This structure is a point of convergence for the direct and indirect pathways through the basal ganglia. Both of these originate from the striatum, which comprises the caudate nucleus and putamen. In visual neglect patients with subcortical lesions, there is substantial lesion overlap found in the putamen, and to a lesser degree in the caudate (Karnath et al. 2002). As indicated in Figure 5.2, the putamen is involved in the evaluation of policies. This fits with the proposed role of the basal ganglia. Additionally, as policies that are independent of the expected free energy are equivalent to habitual behaviour, it makes intuitive sense that pathological biasing of policies would take place within a structure which is involved in habit formation; i.e., the striatum (Yin and Knowlton 2006). The consistency of the anatomy of the

basal ganglia with the policy update equations is further enhanced when the hierarchical extension of these equations is considered (Friston et al. 2017f). These imply multiple parallel loops, originating and ending in the cortex, closely resembling those described in subcortical structures (Haber 2003).

The pulvinar is another subcortical region that is strongly implicated in visual search and neglect – and, as mentioned above, connects to dorsal parietal areas (Behrens et al. 2003; Weller et al. 2002). This makes it a plausible anatomical substrate for the representation of prior beliefs about proprioceptive outcomes. This is consistent with accounts of the pulvinar in directing attention (Kanai et al. 2015; Shipp 2003) and eye movements (Petersen et al. 1985), and as a 'salience map' (Robinson and Petersen ; Veale et al. 2017).

There are other possible lesions which could be accommodated by this model. For example, unilateral disruptions of the connections from the substantia nigra pars reticulata to the superior colliculus (Hikosaka and Wurtz 1985; Schiller et al. 1987; Schiller et al. 1980), or of the dopaminergic modulation of the striatum (Kato et al. 1995; Kori et al. 1995), have been shown to cause visual neglect-like syndromes. However, these lesions are rarely reported as causes of neglect in human patients. We have prioritised the lesions corresponding to the white matter tract that connects the dorsal and ventral attention networks, in addition to two common subcortical lesions; the putamen and pulvinar. These closely resemble the theoretically motivated lesions of **A**, **E**, and **C**.

In the above, our focus has been on disruption of the communication between posterior and frontal cortices, and on subcortical disconnections within the right hemisphere. Importantly, there is good evidence (Dietz et al. 2014; Rushmore et al. 2006; Vuilleumier et al. 1996) that neglect involves inter-hemispheric imbalances in addition to intra-hemispheric disruptions (Bartolomeo 2014; Bartolomeo et al. 2007). This is a key feature of an existing model of neglect (Kinsbourne 1970). Fortunately for our framework, the two are inherently linked. Examination of the equations in Figure 2.2 reveals two key features in the belief updates for hidden states. The first feature is that beliefs about states are conditionally dependent upon policies. This means that any bias towards policies favouring saccades to the right will increase the probability, on taking a Bayesian model average over policies, of a fixation location on the right. Given the contralateral cortical control of eye movements, this corresponds to increased left hemispheric activity. The second important computational feature is the softmax function, which ensures posterior beliefs over allowable fixation locations sum to one (i.e., ensures a proper probability distribution). Such a constraint could be biologically implemented by inhibitory interactions within and between the two frontal eye fields. In other words, if fixations on the right side of space are considered more probable, it must be the case that

leftward fixations are less probable. This necessarily implements a form of inter-hemispheric competition – a competition that is won by the left hemisphere if any of the lesions described in the previous section bias policies towards rightward saccades.

Simulating visual neglect

Figure 5.3 shows the results of running the simulation for 20 saccades, under different prior beliefs (i.e., lesions). Strikingly, all three lesioned models produce very similar behavioural patterns. This heterogeneity of functional lesions is consistent with the diverse set of anatomical lesions known to cause visual neglect. While the non-lesioned model samples both sides of space, all three lesions cause a bias towards sampling the right side of space. This biased sampling is very similar to that observed in visual neglect patients (Bays et al. 2010). It is worth noting that people may have additional priors over their policies (possibly contained in **E**), that result in a slightly different pattern of saccadic search than that depicted in Figure 5.3. For example, people might have a prior bias towards performing a saccade to a nearby target. We have omitted this additional prior, as our aim is to present a minimal model that reproduces the important features of neglect.

The functional disconnection induced by altering the Dirichlet parameters of **A** effectively increases the novelty associated with saccades to the right hemifield. This corresponds to the functional disconnection of the dorsal and ventral attention networks, and can be thought of as impairing the 'capture' of attention by salient stimuli, consistent with existing theories of visual neglect (Ptak and Schnider 2010) and attention (Shulman et al. 2009). The simulated pulvinar lesion causes the agent to fulfil their prior beliefs that they are more likely to be looking at the right side of space, and the lesioned putamen biases policy selection in favour of saccades in this direction.

While mechanistically distinct, the behavioural profiles of each of these lesions do not appear to lend themselves to precise diagnoses in terms of observable behaviour. In the next subsection, we consider a more realistic approach to spatial representations. We follow this with an attempt to determine whether the syndrome generated by these lesions is really as homogenous as it appears, or whether it is possible to identify the lesion from saccadic behaviour.

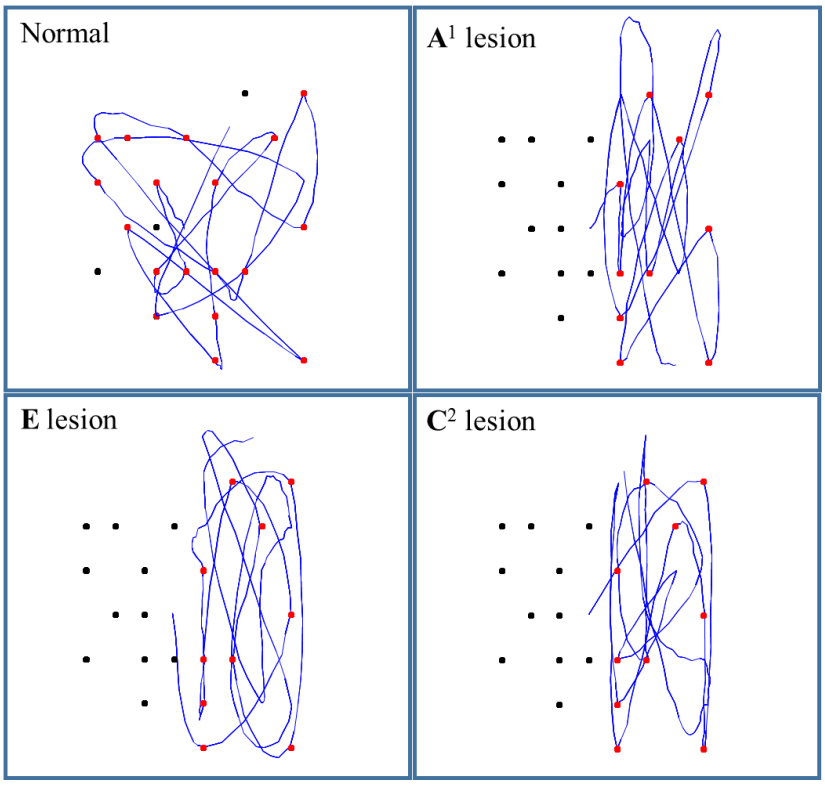

Figure 5.3 - Simulated saccadic cancellation task. Each of the panels shows the simulated eye tracking data (blue) during 20 saccades. In all cases, the target array was the same. The *upper left panel* shows the performance of the model with no simulated lesions. The *upper right panel* shows the results when the $\mathbf{A}^1$ Dirichlet parameters were increased for the left hemifield, corresponding to a functional disconnection of the second branch of the right superior longitudinal fasciculus. The *lower left panel* shows performance when there is a biasing of policy selection, simulating a lesion of the putamen. The *lower right panel* represents a lesion of the prior beliefs about proprioceptive outcomes, which relates to a deficit in the inputs to the dorsal parietal cortex, likely from the pulvinar.

Multiscale representations of space

Spatial representations in the brain involve multiple segregated spatial scales, and resolutions. The magnocellular system, for example, carries information with a relatively low spatial resolution, while the parvocellular system provides higher resolution information (Livingstone and Hubel 1988; Nealey and Maunsell 1994; Zeki and Shipp 1988; Zeki and Shipp 1989). Visual neglect provides further evidence for the brain's use of multiscale spatial representations. One example of this is the Ota search task (Ota et al. 2001) in which participants are asked to identify, from an array of shapes, which shapes are complete. While

some visual neglect patients fail to address any of the left hand side of the array ('egocentric visual neglect'), others address all shapes, but are impaired in determining which shapes are complete ('allocentric visual neglect'). Specifically, those shapes which have a deficiency on their right side are correctly identified as incomplete, while those deficient on their left side are incorrectly identified as complete. While many accounts have described the two perceptual deficits in terms of different spatial reference frames (Medina et al. 2009), it has been argued that both forms are actually different manifestations of an egocentric visual neglect (Corbetta and Shulman 2011; Driver and Pouget 2000). If both are considered to take place in the same reference frame, the two behavioural patterns would be consistent with visual neglect operating at a coarse spatial scale in the first case, and a finer scale in the second.

Equipping the model with a multiscale representation is simple to do in our generative model: instead of representing each of the 64 locations at a high resolution, we can encode each location using three levels (i.e. factors) of resolution, each level divided into four quadrants that, collectively, specify $4^3 = 64$ locations (see Figure 1.4 for a conceptual overview of this sort of factorisation). Technically, this means the **A** matrix now becomes three matrices encoding the likelihood mappings at low, intermediate, and high levels of resolution. Functionally, this means that the subject perceives visual input at three levels of resolution – and can entertain uncertainty (and novelty) at any level. This also means we have the opportunity to model pathological (prior) biases at the level of quadrants of the visual field, quadrants within each quadrant and quadrants within those quadrants.

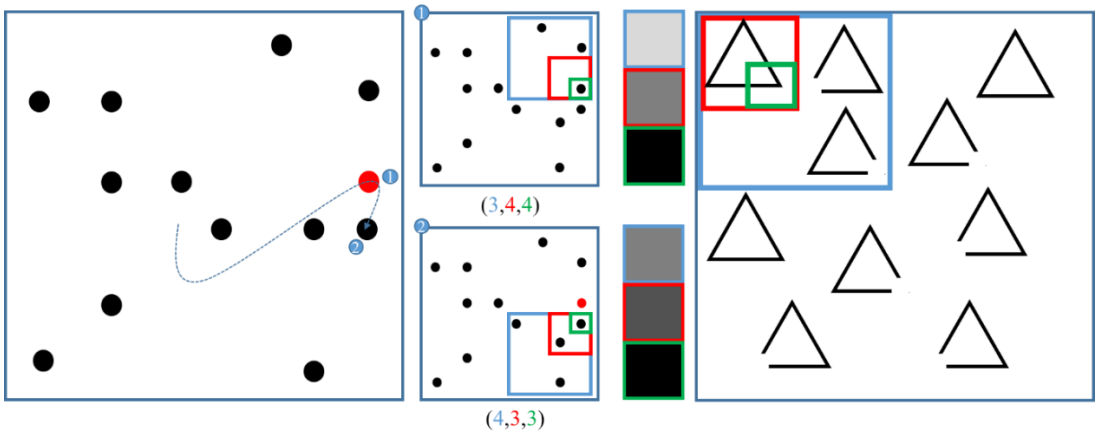

Figure 5.4 - Multiscale representations of space. In the illustration on *the left*, two fixation points in a sequence of saccades are highlighted. This is to demonstrate their representation in terms of a multiscale spatial state space. In the *centre left*, this state space is shown for each fixation point. This specifies a location in an 8x8 space, as before. However, the location is specified in terms of which quadrant (blue), which subquadrant (red) and which

subsubquadrant (green) the location is found. These three specifications constitute the hidden states of the multiscale model. An advantage of this model is that it allows visual outcomes to be defined at different resolutions. This is shown in the *centre right*. Each outcome corresponds to the density of targets in the quadrant, subquadrant, and subsubquadrant currently fixated. Darker shades indicate a greater density. Note that the finest resolution is at the level of individual locations, so density is equivalent to the presence or absence of a target. Cancelled targets appear red at this level only – lower resolutions are considered to be colour-blind; consistent with the properties of the magnocellular system (Hubel and Livingstone 1987). As a saccade is made from a quadrant containing three targets to one containing four, the lowest resolution (blue frame) outcome becomes denser. Similarly, the subquadrant representation (red frame) becomes darker, as a subquadrant containing only one target is followed by a subquadrant containing two. The finest resolution (green frame) represents the maximum density (one target) for both fixation locations. The illustration on *the right* motivates the multiscale representation in terms of the Ota task. This shows one quadrant of an array of shapes. If the blue frame was biased towards occupying the right side of the array, this would resemble an egocentric hemineglect. If the green frame were biased towards the right, this would be closer to an allocentric hemineglect.

Figure 5.4 shows a multiscale representation in our model, and its application to the saccadic cancellation task. The right panel shows how the Ota task was used to motivate this approach. If visual neglect is induced at a coarse scale – i.e., quadrant enclosed by the blue frame – an egocentric behavioural pattern of saccadic sampling would be expected. However, if induced at a finer scale (green frame), neglect would cause an allocentric pattern. Figure 5.5 shows the simulated eye tracking data generated under this multiscale representation. Lesions are shown at each spatial scale and are induced by scaling the corresponding Dirichlet concentration parameters. The other two types of functional lesion produce similar results. Crucially, different spatial scales of visual neglect could reflect different lesion topologies, as more ventral lesions have been associated more with neglect at the object scale (Grimsen et al. 2008; Medina et al. 2009; Verdon et al. 2009).

A simplification we have made in the generative model we have used is that we have assumed the head position is stationary. This allows us to treat the coordinate transform, performed by the first branch of the superior longitudinal fasciculus, as an identity transformation. If we did not make this assumption, the transformation would have to be modulated by a set of hidden states representing the head position, as in established models of parietal contributions to attention (Pouget and Sejnowski 1997; Pouget and Sejnowski 2001). The influence of head

position over the reference frame – in which neglect is induced – allows for the possibility of different egocentric coordinate systems. However, it may be that a set of egocentric reference frames are insufficient, on their own, to explain some neglect phenomena. There is evidence that the orientation of the axes of reference frames can be influenced by the spatial configuration of visual stimuli (Driver et al. 1994; Li et al. 2014), but the inferences involved in these processes lie outside the scope of this chapter. Importantly, deficits that are classically described as 'object-centred' are rarely seen in the absence of 'egocentric' deficits (Rorden et al. 2012; Yue et al. 2012). This suggests that such deficits are not an essential part of the neglect syndrome, but may occur in larger lesions that compromise additional connections.

In summary, we have seen that a normative (active inference) model of visual searches and biased (visual) sampling can provide a sufficient, if minimal, account of the functional deficits observed in patients during line cancellation tasks. The computational architecture and message passing implied by the active inference scheme is remarkably consistent with the known functional anatomy of visual search and saccadic eye movements – and the deficits in epistemic foraging seen in patients with neglect. In the final section, we turn to the practical issues of using this sort of model to make inferences about lesions on the basis of saccadic eye movements.

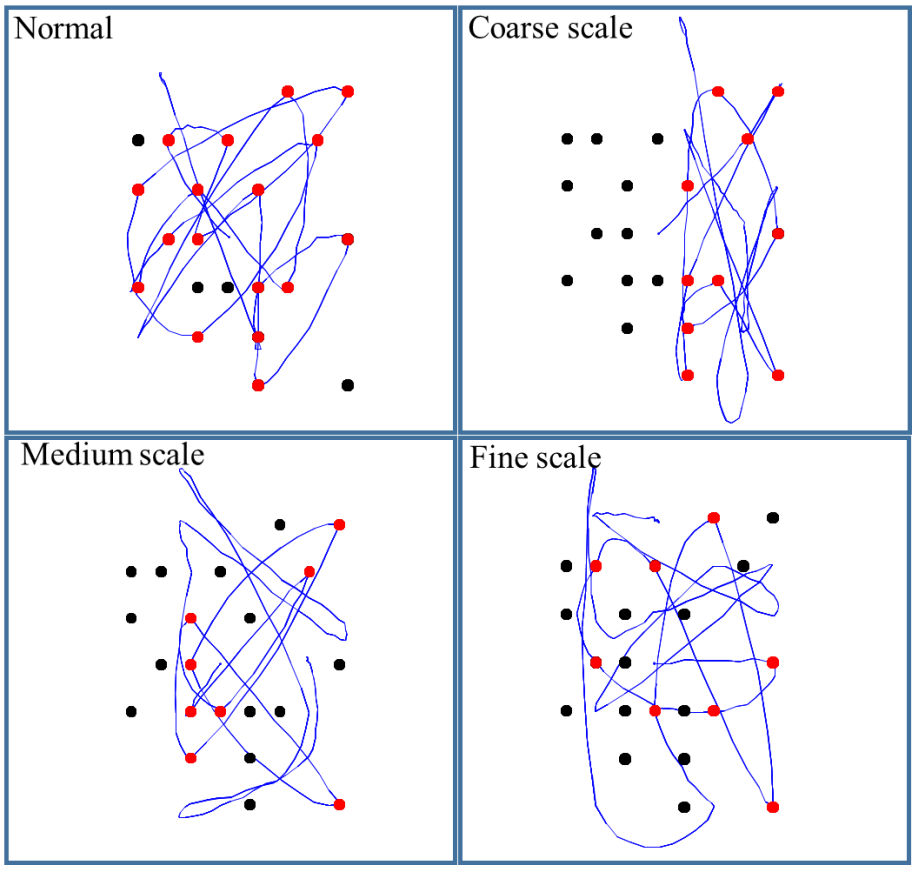

Figure 5.5 – **Lesions at different spatial scales.** By changing the number of the initial Dirichlet parameters, we have simulated hemineglect at three resolutions. As can be seen in the above, the course scale representation biases saccades to the right side of the array, similar to the patterns seen in Figure 5.3. The medium scale representation biases saccades to the right side within each of the four quadrants of visual space. Neglect at the finest scale biases saccades to the right of each subquadrant (comprising four possible locations). For larger targets, but the same spatial scales, each of these biased sampling policies would produce results very similar to those observed in patients performing the Ota task.

Previous models have addressed attentional processes in general (Bundesen 1998; Heinke and Humphreys 2005), and neglect specifically (Heinke and Humphreys 2003; Kinsbourne 1970). Our approach complements many of these models, while making use of more recent theoretical developments. The belief update scheme we have employed has been used to reproduce a range of other behaviours (FitzGerald et al. 2014; Friston et al. 2015b; Moutoussis et al. 2014), physiological responses (Friston et al. 2014; Schwartenbeck et al. 2015b), and pathologies (Schwartenbeck et al. 2015a), emphasising its plausibility as a description of brain function. Additionally, our use of active inference allows us to appeal to a physiologically plausible process theory (Friston et al. 2017a), that facilitates the formation of empirical hypotheses about electrophysiological data. For example, we would expect that there would be an increase in the effective connectivity (in healthy subjects) between the regions connected by the second branch of the superior longitudinal fasciculus as Dirichlet parameters are accumulated. We anticipate that this should be reflected in the activities of neurons in these brain regions, and that dynamic causal modelling for evoked responses (David et al. 2006) provides a means to test this hypothesis experimentally. We pursue this in the next section.

Computational lesion deficit analysis

We have established that, just as with anatomical lesions, there are several functional lesions that can induce very similar behaviour. This raises an important question. Is the mapping from lesion to behaviour truly a many-to-one mapping? In other words, is it possible, given the (simulated) behavioural data, to determine which lesion model generated it? If so, this could have important implications for clinical diagnosis, as it would allow the separation of distinct functional categories of visual neglect.

To answer this question, we used synthetic eye tracking data from each of the lesion models. To assess the ability of the paradigm to disambiguate among lesions, we computed the log likelihood of simulated behaviour for every combination of lesion and model. This log likelihood or evidence was computed by summing the likelihood of each saccade under the posterior probability of saccade, under each MDP model. Clearly, in a practical application, one would need to estimate (subject specific) parameters that best accounted for the observed behaviour (Schwartenbeck and Friston 2016). However, in this instance, there are no unknown parameters and the log evidence for any given model reduces to the expected log likelihood, under that model. Given that we know each set of lesion data was generated by one of the models, we can calculate the posterior probabilities of each model using a softmax function of the likelihoods for each synthetic dataset.

The results of this Bayesian model comparison are shown in Figure 5.6. It is clear from the confusion matrix shown in the figure that one can reliably disambiguate between health and pathology. Furthermore, the lesions in **A**, (i.e., a synthetic disconnection between dorsal and ventral attentional systems) although visually very similar to those of the other two lesion models, produce a characteristic behavioural pattern, allowing the lesion identity to be recovered. This disambiguation speaks to an empirical test of our anatomical model: if one were to take patients with visual neglect, with known anatomical lesions as determined by imaging, one might expect to find (for example) that a lesion deficit analysis using eye movements for patients with disconnection of the superior longitudinal fasciculus will find greater evidence for an **A** lesion than for any of the other lesion models. Lesions of **E** and **C** could not be disambiguated from one another using only synthetic saccades. That the latter model has a greater posterior probability for both sets of simulated data suggests that this is a simpler explanation for the data, and has incurred a lower complexity penalty during Bayesian model comparison. However, they were clearly identified as being abnormal, and not due to simulated lesions of the superior longitudinal fasciculus; i.e. **A**. This suggests that distinguishing between the two may require an additional data modality, such as reaction time, pupillometry, or electrophysiology.

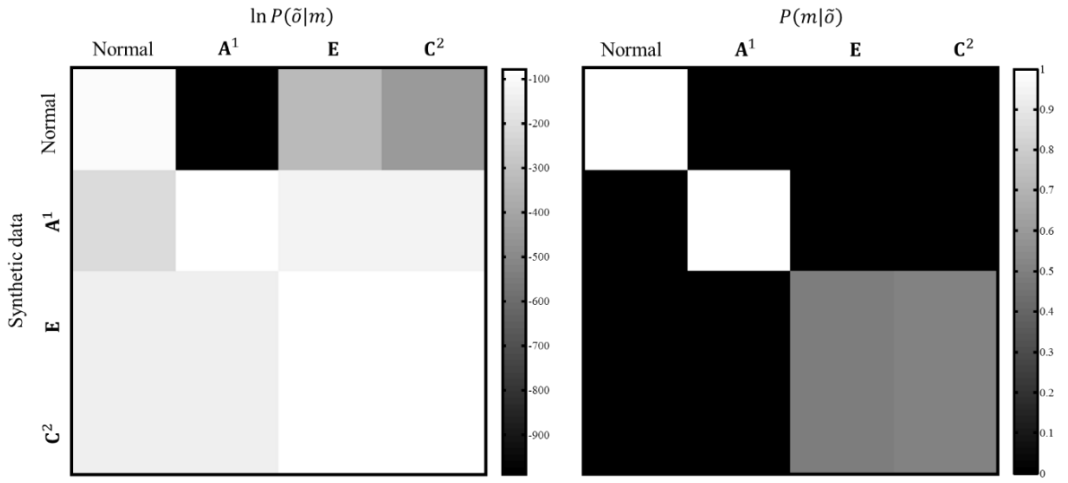

Figure 5.6 - Confusion matrices constructed from 40 saccades. The matrix *on the left* shows the (log) model evidence for each model, *m* (columns), given synthetic eye tracking data, *o*, generated from each model (rows). This is equivalent to the (log) likelihood or model evidence, as there were no unknown parameters. These results were generated using multiscale representations with lesions at the coarsest resolution in all cases. *On the right* is the matrix of posterior probabilities. This is obtained from the matrix on the left, using a softmax function applied to the log evidence is in each row (i.e., for different models of each synthetic dataset).

Summary

Visual neglect can be formulated as a computational bias in an active inference scheme that can be quantified in terms of abnormal prior beliefs. In the above, we identified three, theoretically motivated, functional lesions. On defining a generative MDP model that performed a cancellation task, we found that the connectivity implied by the model structure corresponded well to the anatomy of the dorsal and ventral attention networks, in addition to their subcortical influences. The functional lesions in this anatomical assignment matched lesions associated with visual neglect; namely, in the second branch of the superior longitudinal fasciculus, the putamen, and the pulvinar. The saccadic behaviour generated under these lesion models closely resembles that of patients with visual neglect. To provide a more realistic spatial representation, we used a multiscale encoding of visual state space, which implements a multiscale resolution. This allowed us to demonstrate visual neglect at different scales. Encouragingly, although the saccadic behaviours appeared homogenous across each lesion model, we found that we could recover distinct groups of lesions by

comparing the evidence for each lesion in synthetic data. In principle, this demonstrates that computational phenotyping of visual neglect patients is possible.

Dynamic causal modelling

As we have seen, perception is a fundamentally active process. While this is true across modalities, it is especially obvious in the visual system, where what we see depends upon where we look (Andreopoulos and Tsotsos 2013; Ognibene and Baldassarre 2014; Parr and Friston 2017a; Wurtz et al. 2011). In this section, we consider the anatomy that supports decisions about where to look, and the fast plastic changes that underwrite effective saccadic interrogation of a visual scene. We appeal to the metaphor of perception as hypothesis testing (Gregory 1980), treating each fixation as an experiment to garner new information about states of affairs in the world (Mirza et al. 2018; Mirza et al. 2016; Parr and Friston 2017c). Building upon recent theoretical work (Parr and Friston 2017b) outlined in the first part of this chapter, which includes a formal model of the task used here, we hypothesised that the configuration of a visual scene is best represented in terms of expected visual sensations contingent upon a given saccade ('what I would see if I looked there') (Zimmermann and Lappe 2016). This implies a form of short-term plasticity following each fixation, as the mapping from fixation to observation is optimised.

The purpose of this study is not to evaluate whether we engage in active vision, as there is already substantial evidence in favour of this (Mirza et al. 2018; Yang et al. 2016a), but to try to understand how the underlying computations manifest in terms of changes in effective connectivity. Our aim is to establish whether there is neurobiological evidence in favour of optimisation of a generative model (Yuille and Kersten 2006) that represents visual consequences of fixations as a series of eye-movements are performed.

In the following, we describe our experimental set-up, including our gaze-contingent cancellation task. Through source reconstruction, we demonstrate the engagement of frontal, temporal, and parietal sources, and note the right-lateralisation of the temporal component. We then detail the hypothesis in terms of network models or architectures and use dynamic causal modelling (DCM) to adjudicate between models that do and do not allow for plastic changes in key connections. This model comparison revealed a decrease, from early to late fixations, in the inhibition of neuronal populations in the ventral network by those in the dorsal network.

Network architecture

Paraphrasing the description of the active visual system given above, functional neuroimaging, neuropsychological, and structural connectivity studies converge upon a system that can be separated into a bilateral dorsal frontoparietal network, and a right lateralised ventral network. In brief, functional imaging experiments (Corbetta and Shulman 2002; Vossel et al. 2012) during visuospatial tasks reveal activation of the frontal eye fields (FEF) and the intraparietal sulcus (IPS) in both hemispheres, but greater involvement of the right temporoparietal junction (TPJ) than its contralateral homologue. The volumes of the white matter tracts connecting the components of the dorsal attention network are comparable, while those connecting the ventral network sources are of a significantly greater volume in the right hemisphere (Thiebaut de Schotten et al. 2011). Neuropsychological asymmetries reinforce this network structure, with right hemispheric lesions much more likely than left to give rise to visual neglect (Halligan and Marshall 1998).

As we highlighted above, neglect (often) appears to be a consequence of a disconnection between the ventral and (right) dorsal networks (Bartolomeo et al. 2007; He et al. 2007). Given the dorsal frontoparietal origins of cortico-collicular axons (Fries 1984; Fries 1985; Gaymard et al. 2003; Künzle and Akert 1977), frontal control of eye position (Bruce et al. 1985; Sajad et al. 2015), and the representation of visual stimulus identity in the ventral visual ('what') stream (Goodale and Milner 1992; Ungerleider and Haxby 1994), this is consistent with the idea that the connection between these networks is the neural substrate of an embodied (oculomotor) map of visual space. It is worth noting that the temporo-parietal component of the ventral attention network is not within the ventral visual stream. However, it has been associated with target-detection operations (Chica et al. 2011; Corbetta and Shulman 2002; Serences et al. 2005) that rely upon a simple form of visually derived stimulus identity. Although our focus is in the visual domain, we note that similar networks appear to be involved in auditory attention and neglect (Dietz et al. 2014).

Synthesising these theoretical and neuroanatomical constructs, we hypothesised that the coupling between the dorsal and ventral attention networks changes with successive fixations in a saccadic task. This hypothesis is based upon the idea that, as an internal model of the task is optimised, the relationship between fixation locations and their visual consequences should become more precise (see (Parr and Friston 2017b) for a simulation that demonstrates this). If this is the case, this could manifest in one of two ways. The effective connectivity from the temporoparietal cortex to the frontal eye-fields could increase over time. Alternatively, plastic changes in connections in the opposite (dorsal-to-ventral) direction could decrease their

effective connectivity to relieve descending inhibition of the ventral-to-dorsal projections arising from superficial pyramidal cells. Ultimately, both of these would enhance the influence of ventral parietal over dorsal frontal regions. We used an oculomotor cancellation paradigm, based upon the classic pen-and-paper line cancellation task used to assess visual neglect (Albert 1973; Fullerton et al. 1986). In this task, patients with neglect tend to cancel (by crossing out) lines on the right side of a piece of paper but miss those on the left. Using magnetoencephalography (MEG) and dynamic causal modelling (DCM) for evoked responses (David et al. 2006) we assessed changes in effective connectivity between dorsal frontal and ventral temporoparietal sources during early and late cancellations (fixations) in healthy participants. Our task involved performing saccades to targets on a screen that, once fixated, changed colour and were considered cancelled.

Methods – Experimental design and imaging

We recruited 14 healthy right-handed participants (8 females and 6 males) between the ages of 18 and 35 from the UCL ICN subject pool under minimum risk ethics. Participants were seated in the MEG scanner (whole-head 275-channel axial gradiometer system, 600 samples per second, CTF Omega, VSM MedTech, Coquitlam, Canada), with a screen about 64cm in front of them, showing the stimulus display (size 40 x 29.5cm). This was presented using Cogent 2000 (developed by the Cogent team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience).

The sequence of stimuli is illustrated in Figure 5.7. Following a fixation cross, a set of 16 black dots appeared on the screen, simultaneously, in pseudo-random (using the Matlab random number generator) locations. When a dot was fixated, it changed from black to red (i.e. was ‘cancelled’). Participants were asked to look at the black dots, but to avoid looking at the red dots. We tracked the eyes of the participants while the dots were on screen using an SR Research eye-tracker (Eyelink 1000 – operated using Psychtoolbox) sampling at a frequency of 1kHz. We divided the cancellation events into two categories: early (first 8) and late (last 8).

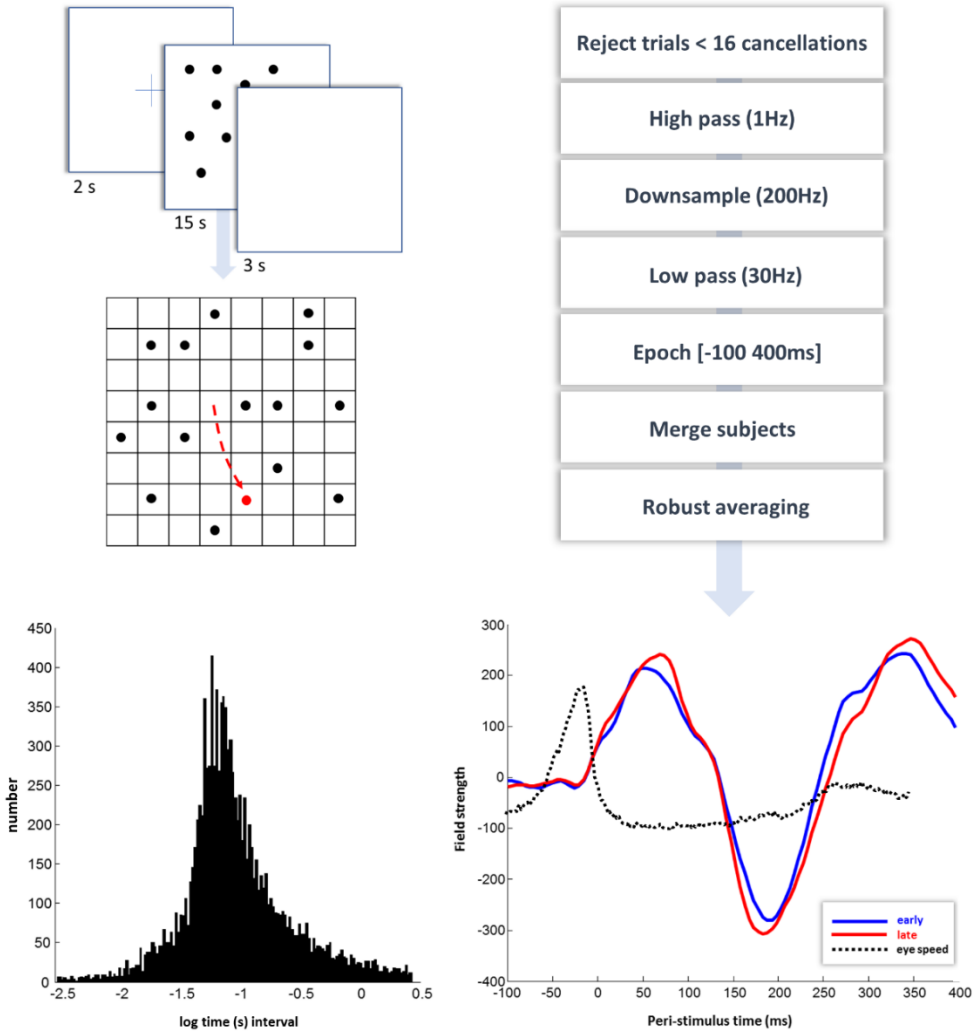

Figure 5.7 - Oculomotor cancellation task and pre-processing. The graphic on the upper left illustrates the sequence of events for a given trial. First, a fixation cross is presented for 2 seconds. After this, a display with 16 black dots is randomly generated and presented for 15 seconds. This is followed by a blank screen for 3 seconds. The dots were placed within an 8x8 grid (not visible to the participants), as shown in the lower left schematic. When the dots were visible on screen, we tracked the eyes of the participant. Whenever their gaze entered a square containing a black dot, this changed from black to red and remained red for the rest of the trial. Participants were instructed to look at the black dots, and to avoid looking at red dots. Events were defined as the time at which the eye crossed into the square, causing a change in colour (i.e. a cancellation). There were 15 of these trials per block, with 6 blocks per participant. The lower left plot shows a histogram of the time intervals between saccadic dot cancellations, to give a sense of the latency between saccades. These latencies are reported using a (natural) logarithmic time scale (with time in seconds) over the first 2.5 standard deviations above and below the mean. The mean here is -1.0597, corresponding to roughly 3 cancellations per

second (consistent with the 3-4Hz frequency of saccadic sampling (Hoffman et al. 2013)). On the right, we show the sequence of pre-processing steps used and the first principal component of the ensuing evoked response. The evoked response to early cancellations is averaged from 6738 events, and the response to late cancellations from 6571. Superimposed upon this is a trace of the eye speed in peristimulus time in arbitrary units. This is aligned so that zero corresponds to the average speed during the time in which the fixation cross was present.

Although almost all perceptual tasks call upon some sort of engagement with the sensorium, this task emphasises the active nature of visual processing through making the visual element of the task as simple as possible. This still calls upon optimisation of beliefs under an internal model, as formalised in (Parr and Friston 2017b). As outlined above, this has some validity in relation to disorders in which active vision is impaired. However, it is worth noting that other approaches to studying these processes, particularly those that focus on behavioural (as opposed to neurophysiological) measures (Mirza et al. 2018; Yang et al. 2016a), make use of more complicated visual stimuli – so that different saccades afford different levels of information gain about a particular scene category.

Our pre-processing steps (using SPM 12 - http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) are specified in Figure 5.7. As participants generally had no trouble in cancelling all 16 dots, we rejected all trials for which they were unable to do so (assuming these were due to eye-tracker calibration errors). We merged the epoched data from all participants, and averaged the epochs corresponding to the first 8, and the last 8, cancellations over all participants to create a grand average. This meant we averaged over fixations preceded by saccades from all possible directions, ensuring any directional eye movement induced artefacts following cancellation were averaged away. Using robust averaging provides an additional protection against artefactual signals, as this iterative procedure rejects those trials that deviate markedly from the mean response. The average eye-speed is shown in Figure 5.7 (black dotted line) to illustrate that it falls to its minimum at about the same time as the target is cancelled. The first principal component, across spatial channels, of the averaged evoked response (to a cancellation) in each condition is shown on the same plot. To further interrogate the changes in effective connectivity, we additionally constructed grand averaged responses to each of the 16 cancellations in a trial. These were used for the more detailed model of (parametric) time-dependent responses described below.

In Figure 5.8, we show the reconstructed source activity obtained using multiple sparse priors (Friston et al. 2008). This scheme tries to infer the sources in the brain that generated the data

measured at the sensors. There are an infinite number of possible solutions to this problem, but Bayesian methods attempt to find the simplest of these. Our results – using standard settings (Litvak et al. 2011) – show a relatively symmetrical distribution of frontal and posterior cortical sources, and a right lateralised (asymmetrical) temporal component. While the inferred locations are more ventromedial than we might expect (likely due to the ill posed nature of the MEG inverse problem), it is encouraging that we can recover sources that are broadly consistent with the known anatomy, and lateralisation, of the attention networks (Corbetta and Shulman 2002) from these data.

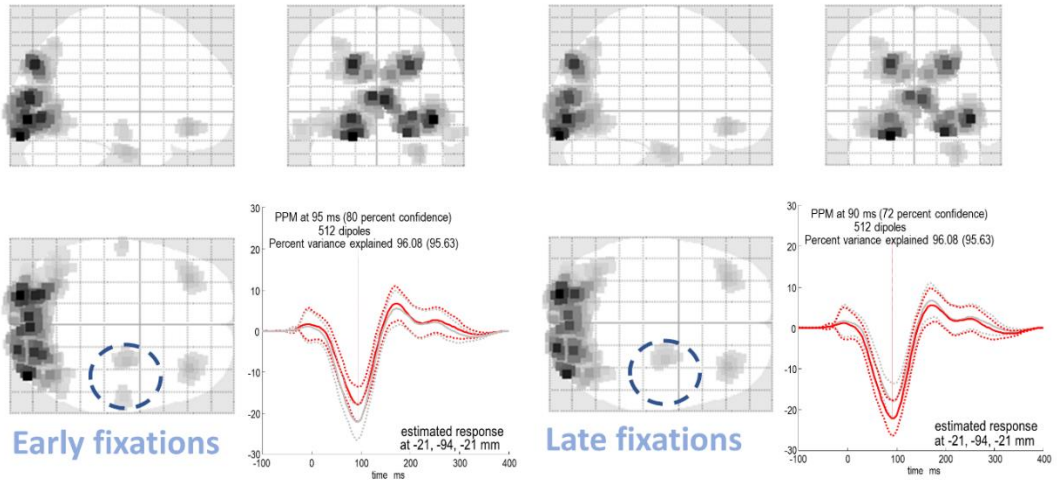

Figure 5.8 - Source reconstruction with multiple sparse priors. These images show the Bayes optimal source reconstruction under multiple sparse priors (and following application of a temporal Hanning window) for the first 8 cancellations (left) and the second 8 cancellations (right) in a trial. This reveals a set of symmetrical sources in both the frontal and posterior cortical sources, with a right lateralised temporal component. The striking asymmetry of these temporal sources (dashed circles) is encouraging, considering the known rightward lateralisation of the ventral attention network. While we might expect the frontal sources to be more dorsal, this may reflect the ill-posed nature of MEG source localisation – there are many possible combinations of sources in 3D space that could give rise to the same pattern of activation over the 2D sensory array. The estimated responses show the greatest amplitude at around 100ms. In the left plot (showing the maximal response for the first condition), the red lines indicate the reconstructed activity from the early cancellations and grey from the late cancellations. In the right plot (maximal response for the second condition), red is late and grey is early. Bayesian credible intervals are shown as dotted lines for each response. The confidence associated with the posterior probability maps (PPM) (Friston and

Penny 2003), in addition to the variance explained, are included in the upper left of each plot, and the location at which the response is estimated is given in the lower right.

Dynamic causal modelling

Dynamic causal modelling (DCM) tries to explain measured electrophysiological data in terms of underlying neuronal (i.e., source) activity (Friston et al. 2003). This rests upon optimising the model evidence (or free energy) for a biophysically plausible neural mass model. The (log) evidence that data (**y**) affords a model (*m*) is:

$$\begin{aligned} \ln p(\mathbf{y}\mid m) &\geq E_q[\ln \underbrace{p(\mathbf{y},\mathbf{x},\boldsymbol{\theta}\mid m)}_{\textit{Generative model}} - \ln \underbrace{q(\mathbf{x},\boldsymbol{\theta}\mid m)}_{\textit{Approx. posterior}}] \\ &= \underbrace{E_q[\ln p(\mathbf{y}\mid \mathbf{x},\boldsymbol{\theta})]}_{\textit{Accuracy}} - \underbrace{D_{KL}[q(\mathbf{x},\boldsymbol{\theta}\mid m)\,||\,p(\mathbf{x},\boldsymbol{\theta}\mid m)]}_{\textit{Complexity}} \end{aligned} \tag{5.8}$$

DCM makes use of a Variational Laplace procedure (Friston et al. 2007), a close relative of the scheme expressed in Figure 2.1, to optimise beliefs (*q*) about neuronal activity (**x**) and the parameters (**θ**) that determine this activity (e.g., connection strengths) and the (likelihood) mapping (e.g., lead field) from **x** to **y**. The lead field matrix maps source activity to the measured sensor data on the scalp (Kiebel et al. 2006). In maximising model evidence, DCM finds the most accurate explanation for the data that complies with Occam's principle; i.e., is minimally complex (as measured by the KL-Divergence between posteriors and priors). By comparing different generative models, we can test hypotheses about biologically grounded model parameters; here, condition-specific changes in connectivity under a particular network architecture.

The generative model we used is the canonical microcircuit model (Bastos et al. 2012; Moran et al. 2013a), which incorporates four distinct neuronal populations per source of a distributed (hierarchical) network (Figure 5.9). These are spiny stellate cells, superficial and deep pyramidal cells, and inhibitory interneurons. The connections associated with each of these populations conforms to known patterns of laminar specific connectivity in the cerebral cortex (Felleman and Van Essen 1991; Shipp 2007; Zeki and Shipp 1988), allowing us to distinguish between ascending and descending extrinsic (i.e. between source) connections. This accounts

for the prior probability density ($p(\mathbf{x},\boldsymbol{\theta}|m)$) that, supplemented with a lead-field provides a likelihood ($p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta})$) and completes the forward or generative model.

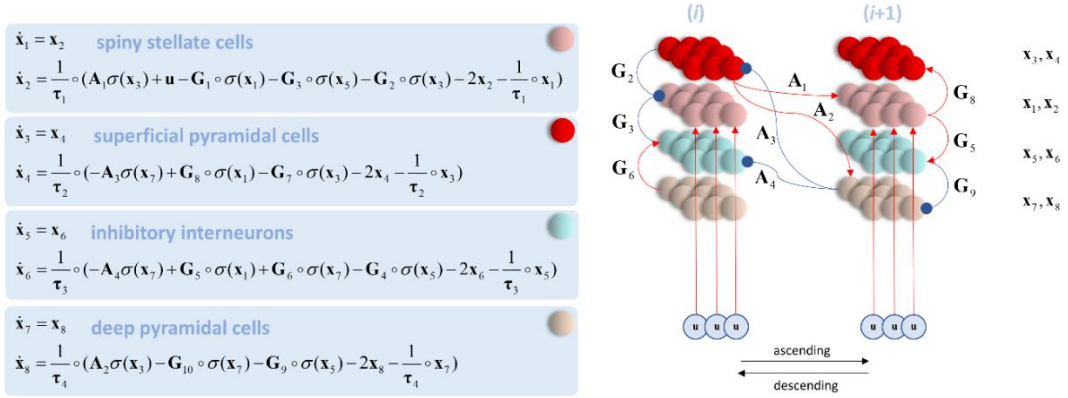

Figure 5.9 - The canonical microcircuit. The equations on the left of this schematic describe the dynamics of the generative model that underwrites the dynamic causal modelling in this chapter. The **x** vectors represent population specific voltage (odd subscripts) and conductance (even subscripts). Each element of the **x** vectors represents a distinct cortical source. The notation **a**◦**b** means the element-wise product of **a** and **b**. The matrix **A** determines extrinsic (between-source) connectivity (here illustrated as connections between a lower source *i* and a higher source *i*+1), while **G** determines the intrinsic (within-source) connectivity. Subscripts for these matrices indicate mappings between specific cell populations. For example, $\mathbf{A}_1$ describes ascending connections from superficial pyramidal cells (source *i*) to spiny stellate cells (source *i*+1), while $\mathbf{A}_3$ describes descending connections from deep pyramidal cells (source *i*+1) to superficial pyramidal cells (source *i*). Experimental inputs – in our case, the cancellation of the target on fixation – are specified by **u**. On the right, we illustrate the neuronal message passing implied by these equations. Red arrows indicate excitatory connections and blue inhibitory. Superficial pyramidal cells give rise to ascending connections that target spiny stellate and deep pyramidal cells in a higher cortical source. Descending connections arise from deep pyramidal cells that target superficial pyramidal cells and inhibitory interneurons.

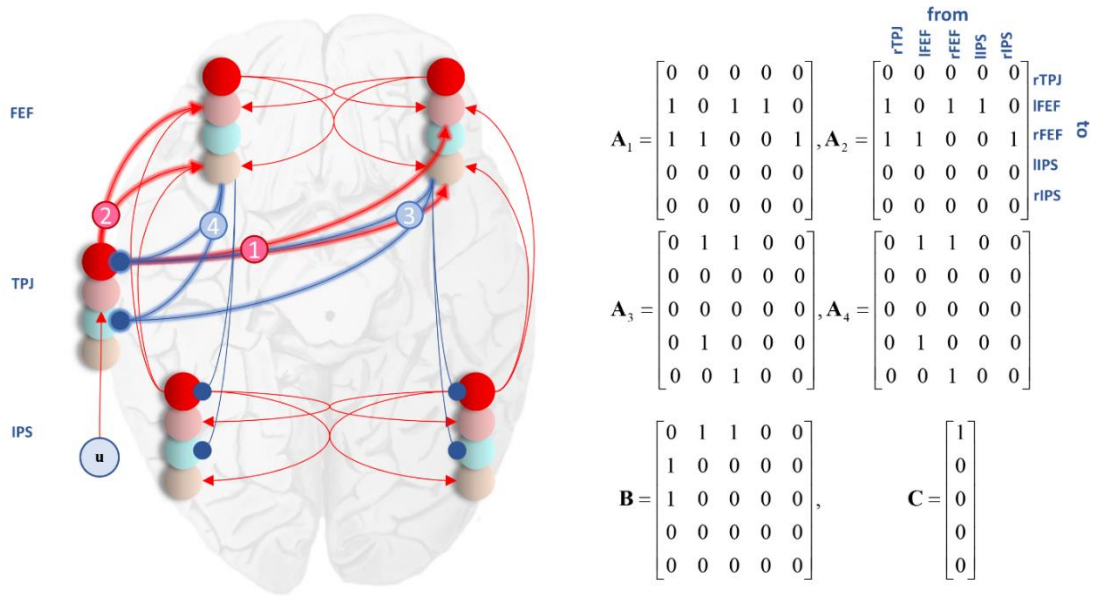

Figure 5.10 – Network architecture. This schematic illustrates the form of the network model we used to test our hypothesis. The dorsal network is present bilaterally (FEF and IPS) and is connected to the ventral network – represented by the temporoparietal junction (TPJ) – on the right. The TPJ receives input as it sits lower in the visual hierarchy than the FEF (Felleman and Van 1991). Our hypothesis concerns the (highlighted) connections between the two networks. We compared models that allowed for changes or visual search-dependent plasticity in connections from the TPJ to left FEF (1), from the TPJ to right FEF (2), from the left FEF to TPJ (3), from the right FEF to TPJ (4), and every combination of the above. The matrices on the right illustrate the specification of these connections. The **A** matrices are the same as those in Figure 5.9 and represent extrinsic connections between sources (with subscripts indicating which specific cell populations in those sources). **B** specifies the connections that can change between the early and late cancellations and **C** specifies which sources receive visual (i.e., geniculate) input. To ensure that the signs of the **A** (and **C**) connections do not change during estimation, their logarithms are treated as normally distributed random variables. This ensures an excitatory connection cannot become an inhibitory connection and *vice versa*.

As we were interested in changes in the coupling of the dorsal and ventral attention networks, we specified our generative model as in Figure 5.10; incorporating the bilateral dorsal network and the right lateralised temporoparietal contribution to the ventral network (consistent with the source reconstruction above). The connections between the right temporoparietal junction (rTPJ) and the left frontal eye field (FEF) probably involve an intermediate thalamic relay (Guillery and Sherman 2002; Halassa and Kastner 2017), but this was omitted for simplicity.

Our hypothesis was that the connections between the rTPJ and each FEF would change between early and late target cancellations – as evidence is accumulated and the precision (c.f., efficacy) of likelihood mappings increases (Parr and Friston 2017b). Figure 5.10 highlights these ascending and descending connections. After fitting the full model (with modulation of all four connections) to our empirical data, we used Bayesian model reduction (Friston et al. 2016c) to evaluate the evidence for models with every combination of these condition-specific effects (early versus late) enabled or set, *a priori*, to zero.

Results

Figure 5.11 reports the results of a model comparison between 16 (2^4) models that allowed for different patterns of search-dependent changes in the forward and backward connections between each FEF and the rTPJ. Given our grand average data, model 8 has a posterior probability of 0.827. This model allows for changes in backwards connections, and the forward connection from rTPJ to the right FEF, but not to the left FEF. This provides evidence in favour of changes in the efficacy of dorsal-ventral connections. Acknowledging that other models, although improbable, were found to be plausible, we averaged our results across models, weighting each model by its posterior probability. Following this Bayesian model averaging, we still found striking changes in the backward connections, which show a decrease in effective connectivity for late compared to early cancellations. As backwards connections are (net) inhibitory, this corresponds to a disinhibition of the superficial pyramidal cells – the origin of ascending connections – in the TPJ. In other words, the effective connectivity during the later stages of the trial changed, compared to that during the first few cancellations, to relieve the inhibitory effect of the dorsal attention network on the source of its input from the ventral network.

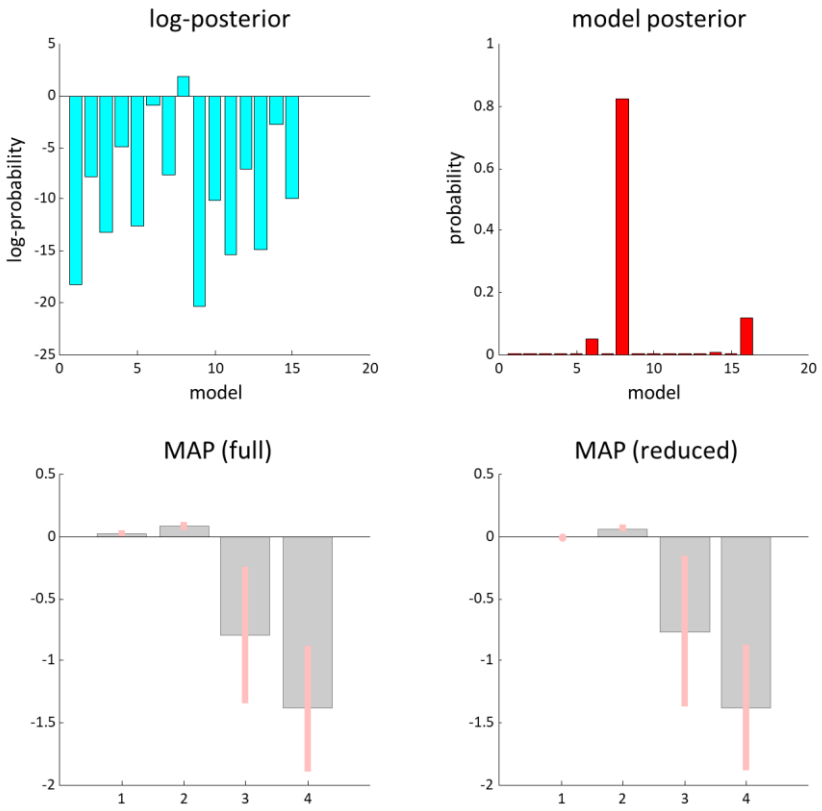

Figure 5.11 - Model comparison and Bayesian model averaging. This figure shows the results of comparing models with different combinations of condition-specific effects on the forward and backward connections between the right TPJ and the frontal eye fields. We performed this comparison using Bayesian model reduction (Friston et al. 2016c), which involves fitting a full model that allows all four connections to change and analytically evaluating the evidence for models with combinations of these changes switched off. The upper plots show the log posterior probabilities associated with each model, and the posterior probabilities. The winning model (number 8) allows for modulation in connections 2, 3, and 4 (see Figure 5.10). The lower plots show that, for the later fixations, there is a modest increase in the effective connectivity in connection 2, but a decrease in 3 and 4. These values correspond to log scaling parameters, such that a value of zero means no change. The lower left plot shows these parameter (maximum a posteriori) estimates for the full model (that allows for all connections). The lower right plot shows the Bayesian model average of these estimates (weighted by the probability of each reduced model to account for uncertainty over models). Bayesian 90% credible intervals are shown as pink bars.

The effects of this disinhibition can be seen in the reconstructed neuronal activities shown in Figure 5.12. During later cancellations, the activity of the superficial pyramidal cells in the rTPJ has a greater amplitude than evoked during earlier fixations. Figure 5.10 shows that this is the population inhibited by the descending connections (labelled 3 and 4). These are the connections that show the greatest change (both relative and absolute, despite being slightly weaker at baseline than the forward connections). Although the change in ascending synapses is small or absent, the increase in activity in these forward projecting TPJ cells has driven an increase in the amplitude of responses in all populations in each FEF. The most dramatic effect is in the deep pyramidal layer, which receives direct input from the superficial TPJ cells. Figure 5.12 additionally shows the resemblance between the activity in deep pyramidal cells in FEF and the simulated rate of belief updating obtained under a Markov Decision Process model of the same behavioural task: for details, please see (Parr and Friston 2017b). This model represents a formalisation of the ideas raised in the introduction; namely, that representations of visual space depend upon beliefs about the sensory consequences of actions. In brief, the differences in the rate of belief updating from early to late fixations are due to the optimisation of the mapping from fixation location to the presence or absence of a target. More precise beliefs later in the task enable faster and more confident belief updates (that translate physiologically into increased effective connectivity; i.e., the rate constants of neuronal belief updating).

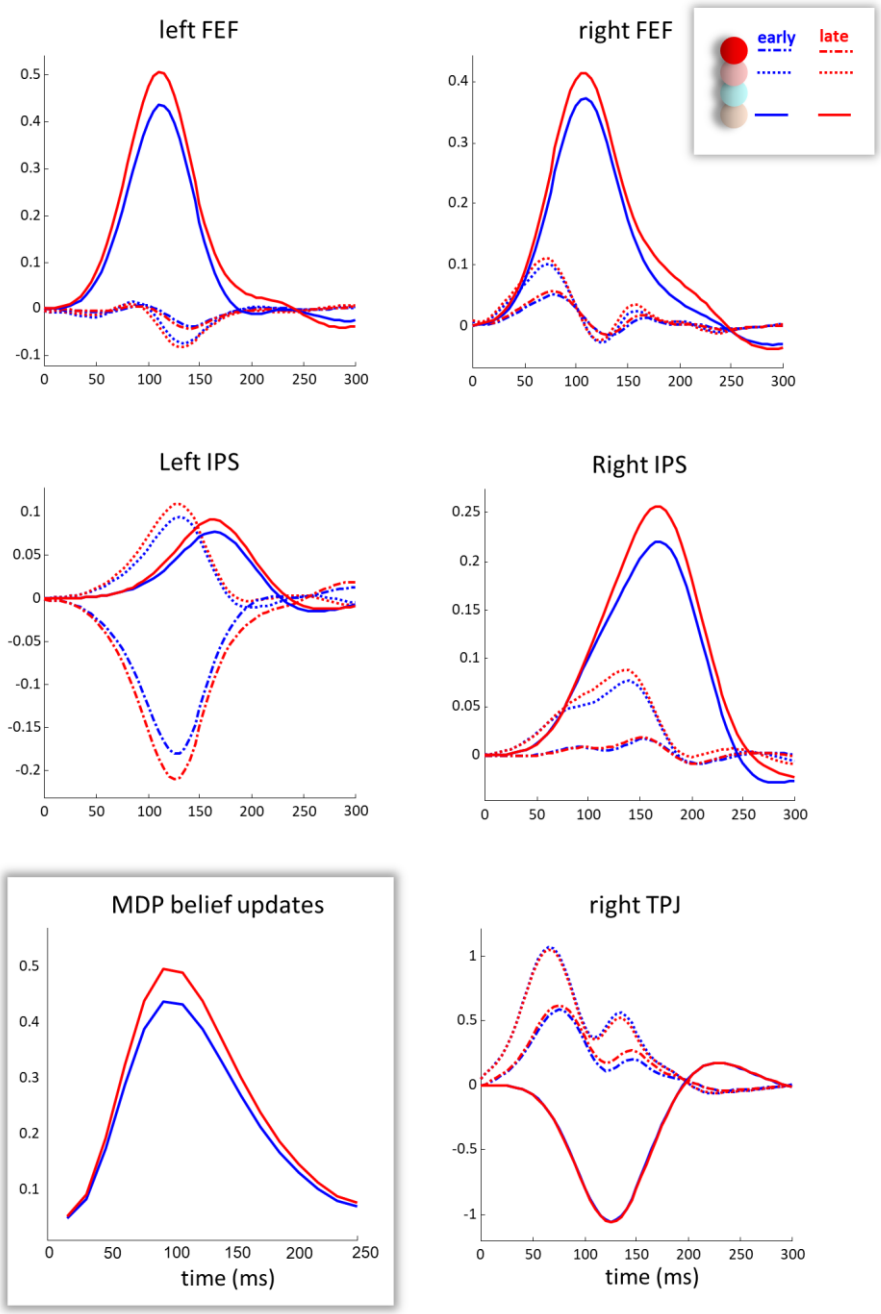

Figure 5.12 - Estimated neuronal activity. These plots show the estimated activity in each excitatory cell population. Dashed lines indicate the superficial pyramidal cells that give rise to ascending connections and are inhibited by higher cortical sources. Ascending connections target the spiny stellate cells (dotted lines), and the deep pyramidal cells (unbroken lines). The latter give rise to descending connections. The activity here is shown for early (blue) and late (red) cancellations, for each of the cortical areas shown in Figure 5.10. The lower left plot (highlighted) shows the simulated evoked responses obtained from the Markov Decision Process model described in (Parr and Friston 2017b) and earlier in this chapter, drawing from the process theory associated with active inference (Friston et al. 2017a). It is computed by

taking the absolute rate of change of the sufficient statistics of posterior beliefs about the current fixation location, summed over spatial scales (please see the discussion for details). While the y-axis here is arbitrary, the x-axis extends to 250ms, consistent with the theta frequency of saccadic eye movements. There is a striking resemblance between the simulated rate of belief updating and the frontal eye field neuronal activity estimated from our empirical data.

To explore the changes in coupling demonstrated above in a more parametric way, we inverted a DCM that was identical to that described above, but treated each cancellation as a separate event. This meant that, in place of the relatively coarse division into 'early' and 'late', we could test hypotheses about parametric changes in connection strength over 16 sequential cancellations. Figure 5.13 illustrates a model comparison that tests these hypotheses, endorsing the pattern of changes found in Figure 5.11. Due to the implicit model of time-dependent effects, this enables us to plot the estimated changes in coupling throughout the trial, as shown in Figure 5.13. These show a progressive decrease in the strength of inhibitory backwards connections, with a modest increase over time in excitatory forward connections.

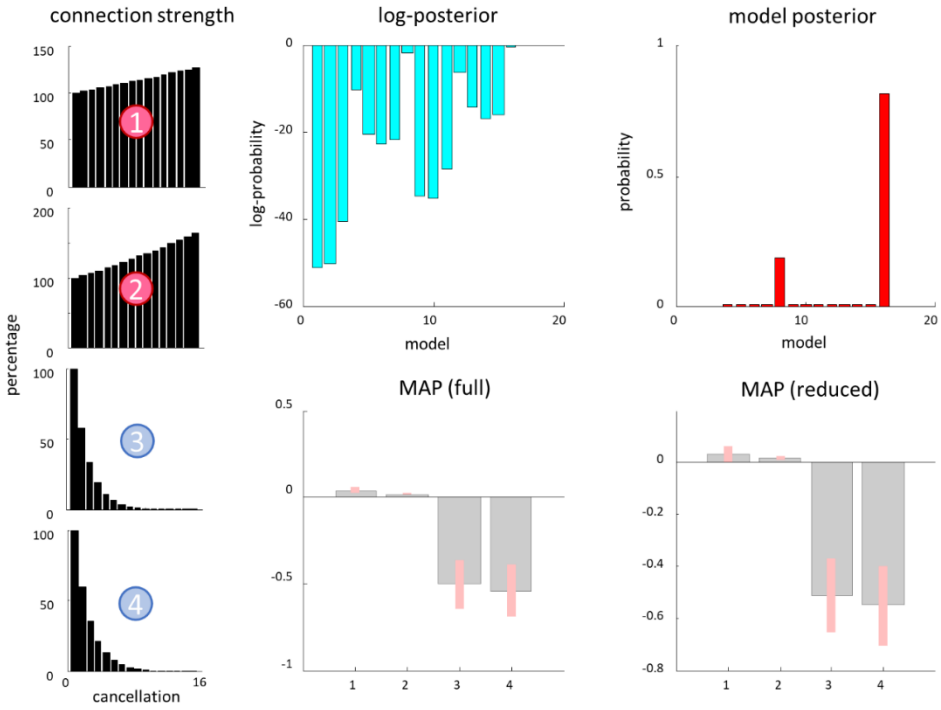

Figure 5.13 – Time-dependency of modulatory changes. The plots on the right of this Figure are the same as those in Figure 5.11, but modelling a parametric effect of number of previous cancellations. For this model, in place of the 'early' and 'late' conditions, we treated each

sequential cancellation as a separate event. Because the model is parameterised in terms of log-scaling parameters, linear (i.e., [0,1,…,15]) parametric effects of time (number of previous cancellations) correspond to a monoexponential change in coupling (starting from a strength of exp(0), corresponding to 100%). The two most probable models are the same as in Figure 5.11, and the overall pattern of changes shown in the MAP estimates is the same (but with some evidence in favour of a small change in connection 1). The plots on the left show the estimated changes in each connection with successive cancellation events, as a percentage of their initial values. These indicate an increase in the strength of forward excitatory connections over time, and a decrease in backward inhibitory connections.

Discussion

The results presented here provide evidence in favour of short term plastic changes in the connections between the dorsal and ventral attention networks during the active interrogation of a simple visual scene. This supports an enactive perspective on visual cognition (Bruineberg 2017; Hohwy 2007; Vernon 2008), as it is consistent with the idea that we represent visual sensations as the consequences of action – and that these contingencies may be learned over a short time period. While these results have interesting implications for active vision, they also constrain the way in which cortical neuronal circuits might implement inferential computations. That the descending connections appear to change the most is consistent with the idea that ascending signals in the brain carry evidence for or against hypotheses represented in higher areas. While this appears counterintuitive, the evidence afforded to a hypothesis about one variable (e.g. location on a horizontal axis) depends upon beliefs about other variables (e.g. location in the vertical axis). In other words, dorsally represented beliefs about eye position, if represented in any factorised coordinate system, must act to contextualise the ascending signals from the ventral to dorsal network. As this is learned over successive fixations, this contextualisation (i.e., interaction between factors) leads to increasingly precise mappings between eye position and its visual consequences – consistent with the disinhibition we observed here. This is analogous to the increase in amplitude of evoked responses following cueing in working memory paradigms (Lenartowicz et al. 2010) that can be reproduced *in silico* by appealing to beliefs about the context of ascending signals (Parr and Friston 2017d).

An interesting question that arises from this is what type of coordinate system the frontal eye fields might employ. The argument given above applies regardless of the choice of coordinate system but depends upon there being some factorisation (Parr and Friston 2017a). This

factorisation could be representation of a horizontal and a vertical axis (McCloskey and Rapp 2000) or could be closer to a wavelet decomposition – used in computational visual processing (Antonini et al. 1992). The latter separates an image into different spatial scales and resolutions. For example, we might represent which quadrant of space we are looking at, and which sub-quadrant within that quadrant. Either of these systems requires far fewer neurons than we would need if we were to independently represent each location in visual space. This is an important aspect of the normative (active inference) theory on which the simulations in Figure 7 were based. In brief, the sorts of generative models used by the brain to infer the causes of its sensory input are subject to exactly the same imperatives used in Bayesian model comparison; namely, the brain's generative or forward models must provide an accurate account of sensations with the minimum complexity. Reducing the number of parameters via factorisation is, in theory, an important aspect of minimising complexity or redundancy (Barlow 1961; Barlow 1974; Friston and Buzsaki 2016; Tenenbaum et al. 2011). We used a decomposition of location into quadrants to simulate the belief updating shown in the lower left of Figure 5.12, which enables us to reproduce visual neglect at different spatial scales (Parr and Friston 2017b), consistent with neuropsychological observations (Grimsen et al. 2008; Medina et al. 2009; Ota et al. 2001; Verdon et al. 2009).

Visual neglect is increasingly recognised as a disconnection syndrome (He et al. 2007). Specifically, it can arise through damage to the white matter tracts that link right dorsal frontal sources to ventral temporoparietal areas (Doricchi and Tomaiuolo 2003; Thiebaut de Schotten et al. 2005). A disconnection of this sort would preclude the changes we have observed in these connections. From the perspective of active inference, this means that saccades to the left side of space represent poor perceptual experiments, as the capacity to learn from them is diminished (Denzler and Brown 2002; Lindley 1956; MacKay 1992; Yang et al. 2016a). We have previously argued that syndromes in which active scene construction is impaired – visual neglect being an important example – may result from pathological prior beliefs about these action-sensation mappings (Parr et al. 2018b). An inability to change this mapping following observation, perhaps due to white matter disconnection (Catani and ffytche 2005; Geschwind 1965a), means that actions that would otherwise engage (and modify) a given connection afford a smaller opportunity for novelty resolution (Parr and Friston 2017b). The failure to update this mapping is consistent with the impairments in spatial working memory that have been elicited in saccadic tasks in neglect patients (Husain et al. 2001). This suggests that one could follow up this idea by temporarily disrupting changes in these (dorsal-ventral) connections using transcranial magnetic stimulation. We hypothesise that this will induce saccadic scan paths consistent with those observed in visual neglect (Fruhmann Berger et al. 2008; Karnath and Rorden 2012). Encouragingly, this approach has previously been used to

induce other features of visual neglect (Ellison et al. 2004; Platz et al. 2016), including changes in line bisection and visual search performance following stimulation of the right TPJ.

An additional direction for future research concerns the use of more complex visual environments. In this study we kept the visual stimuli as simple as possible. However, many interesting phenomena in active vision can be elicited using more sophisticated, and often dynamic, manipulations. An advantage to using stationary targets is that they induce scanning saccades as opposed to reactive saccades – of the sort associated with a suddenly appearing target. The former are accompanied by greater involvement of the frontal part of the dorsal network, while the latter implicates the parietal part (Pierrot-Deseilligny et al. 1995). Given that our hypothesis concerned the frontal regions of the dorsal network, the use of static targets facilitated the involvement of these regions. However, the inclusion of a second condition in which targets suddenly appeared would help us to further interrogate the respective contributions of the frontal and parietal cortices to these processes.

Specifically, it would be interesting to probe the computational mechanisms that underwrite differences between scanning and reactive saccades for both perception and neurobiological measurements (Zimmermann and Lappe 2016). This may relate to the time required for belief-updating, which itself is likely to depend upon the sorts of beliefs that are updated. Typically, cortical areas that sit higher in the anatomical hierarchy (Felleman and Van Essen 1991; Shipp 2007; Zeki and Shipp 1988) are thought to represent stimuli that evolve over longer time-periods (Hasson et al. 2015; Hasson et al. 2008; Kiebel et al. 2008; Murray et al. 2014), in relation to early sensory cortices. Given that the frontal eye fields are engaged in control of scanning saccades, which occur at about 3-4 Hz, it is plausible that the time-scale for updating beliefs about 'where I am looking' corresponds to this frequency. Speculatively, short-latency reactive saccades may be driven by lower cortical regions (e.g. parietal cortex) that represent the locations of fast-changing stimuli and may not leave enough time for completion of belief updating in frontal areas. This might account for the changes in spatial perception of stationary stimuli that follow adaptive changes in saccadic amplitude, but the absence of this phenomenon when dynamic stimuli induce reactive saccades. This is because, under the view that we represent visual space in terms of the visual consequences of saccades, a failure to complete belief updating – in brain regions representing alternative saccades – may preclude the sort of changes in coupling between frontal and temporoparietal areas observed here. Intuitively, this is sensible when constructing a motor map of visual space: there is little point in including transient stimuli, as they are unlikely to be there on looking back. This idea predicts that there should be a diminished inhibition of return following a reactive, as opposed to a scanning, saccade.

Summary

In this section, we tested the hypothesis advanced in the previous section that the coupling between dorsal and ventral frontoparietal networks is altered during visual exploration. To do so, we used dynamic causal modelling based upon a network motivated by pre-existing structural, functional, and neuropsychological data. We found greatest evidence for a model that allowed for modulation in connections from the dorsal to the ventral network. Bayesian modelling averaging revealed a decrease in the effective connectivity of these connections, resulting in a disinhibition of ventral sources by the dorsal attention network. These results are consistent with the idea that the visual data obtained following a saccade drive plastic changes, optimising beliefs about the sensory consequences of a given saccadic fixation. This has potentially important implications for syndromes in which visual exploration is disrupted – notably, visual neglect. We hope that understanding (and measuring) these changes in effective connectivity in health will yield insights into the pathophysiology of disconnection syndromes.

Conclusion

In this chapter, we started by developing a generative model capable of performing a saccadic variant line-cancellation task of the sort used to assess visual neglect. To efficiently represent memories of whether a given target had been cancelled or not, we framed this in terms of a model that represented visual input as the sensory consequence of eye-movements. This meant the status of the targets was 'stored' in the conditional probabilities of the former given the latter. This construction induced a *novelty* term in the expected free energy, which implemented a form of inhibition-of-return, preventing revisiting of previously cancelled targets. Using this generative model, we illustrated how a range of computational lesions could elicit a neglect syndrome, consistent with the range of anatomical lesions that can cause this. Based upon the architecture of the message passing implied by the generative model, and previous neuropsychological studies of visual neglect, we associated the computational lesions with hypothetical neuroanatomical substrates. The implication of this assignment was that the conditional probability of the target status (cancelled or not) given the fixation location could be represented in the connections between dorsal frontal and right temporoparietal regions. Given the optimisation of this probability distribution during performance of the task in question, we hypothesised that we would be able to detect a change in coupling between these

regions using magnetoencephalography. Using dynamic causal modelling in healthy human subjects, we found evidence in favour of this hypothesis.

6 – Computational neurology and Bayesian inference

Introduction

Computational theories of brain function have become very influential in neuroscience. They have facilitated the growth of formal approaches to disease, particularly in psychiatric research. In this chapter[17], we provide a narrative review of the body of computational research addressing neurological and neuropsychological syndromes, and focus on those that employ Bayesian frameworks. In doing so, we seek to place the investigations into active vision (and its disorders) presented in this thesis into the broader context of inferential pathology in neurology. Bayesian approaches to understanding brain function formulate perception and action as inferential processes. These inferences combine 'prior' beliefs with a generative (predictive) model to explain the causes of sensations. Under this view, neuropsychological deficits can be thought of as false inferences that arise due to aberrant prior beliefs (that are poor fits to the real world). This draws upon the notion of a Bayes optimal pathology – optimal inference with suboptimal priors – and provides a means for computational phenotyping. In principle, any given neuropsychological disorder could be characterised by the set of prior beliefs that would make a patient's behaviour appear Bayes optimal. We start with an overview of some key theoretical constructs and use these to motivate a form of computational neuropsychology that relates anatomical structures in the brain to the computations they perform. Throughout, we draw upon computational accounts of neuropsychological syndromes. These are selected to emphasise the key features of a Bayesian approach, and the possible types of pathological prior that may be present. They range from visual neglect through hallucinations to autism. Through these illustrative examples, we review the use of Bayesian approaches to understand the link between biology and computation that is at the heart of neuropsychology.

The process of relating brain dysfunction to cognitive and behavioural deficits is complex. Traditional lesion-deficit mapping has been vital in the development of modern neuropsychology but is confounded by several problems (Bates et al. 2003). The first is that there are statistical dependencies between lesions in different regions (Mah et al. 2014). These arise from, for example, the vascular anatomy of the brain. Such dependencies mean that regions commonly involved in stroke may be spuriously associated with a behavioural deficit (Husain and Nachev 2007). The problem is further complicated by the distributed nature of

[17] This chapter is adapted from (Parr et al. 2018b)

brain networks (Valdez et al. 2015). Damage to one part of the brain may give rise to abnormal cognition indirectly – through its influence over a distant region (Carrera and Tononi 2014; Price et al. 2001). An understanding of the contribution of a brain region to the network it participates in is crucial in forming an account of functional diaschisis of this form (Boes et al. 2015; Fornito et al. 2015). Solutions that have been proposed to the above problems include the use of multivariate methods (Karnath and Smith 2014; Nachev 2015) to account for dependencies, and the use of models of effective connectivity to assess network-level changes (Abutalebi et al. 2009; Grefkes et al. 2008; Mintzopoulos et al. 2009; Rocca et al. 2007) in response to lesions.

In this chapter, we consider a complementary approach that has started to gain traction in psychiatric research (Adams et al. 2015; Adams et al. 2013b; Corlett and Fletcher 2014; Friston et al. 2017e; Huys et al. 2016; Schwartenbeck and Friston 2016). This is the use of models that relate the computations performed by the brain to measurable behaviours (Iglesias et al. 2017; Krakauer and Shadmehr 2007; Mirza et al. 2016; Testolin and Zorzi 2016). Such models can be associated with process theories (Friston et al. 2017a) that map to neuroanatomy and physiology. This complements the approaches outlined above, as it allows focal neuroanatomical lesions to be interpreted in terms of their contribution to a network. Crucially, this approach ensures that the relationship between brain structure and function is addressed within a conceptually rigorous framework – this is essential for the construction of well-formed hypotheses for neuropsychological research (Nachev and Hacker 2014). We focus here upon models that employ a conceptual framework based on Bayesian inference.

Our motivation for pursuing a Bayesian framework is that it captures many different types of behaviour, including apparently suboptimal behaviours. According to an important result known as *the complete class theorem* (Daunizeau et al. 2010; Wald 1947), there is always a set of a prior beliefs that renders an observed behaviour Bayes optimal. This is fundamental for computational neurology as it means we can cast even pathological behaviours as the result of processes that implement Bayesian inference (Schwartenbeck et al. 2015c). In other words, we can assume that the brain makes use of a probabilistic model of its environment to make inferences about the causes of sensory data (Doya 2007; Knill and Pouget 2004), and to act upon them (Friston et al. 2012b). Another consequence of the theorem is that computational models that are not (explicitly) motivated by Bayesian inference (Frank et al. 2004; O'Reilly 2006) may be written down in terms of Bayesian decision processes. Working within this framework facilitates communication among models, and ensures they could be used to phenotype patients using a common currency (i.e., their prior beliefs). It follows that the key challenges for computational neuropsychology can be phrased in terms of two questions: ‘what

are the prior beliefs that would have to be held to make this behaviour optimal?' and 'what are the biological substrates of these priors?'

The notion of optimal pathology may seem counter-intuitive, but we can draw upon another theorem, *the good regulator theorem* (Conant and Ashby 1970), to highlight the difference between healthy and pathological behaviour. This states that a brain (or any other system) is only able to effectively regulate its environment if it is a good model of that environment. A brain that embodies a model with priors that diverge substantially from the world (i.e., body, ecological niche, culture, *etc*.) it is trying to regulate will fail at this task (Schwartenbeck et al. 2015c). As such, pathological regulators may be thought of creatures who model their world inaccurately. In other words, they are good regulators of a hypothetical alternative world that is more coherent with their model. If pathological priors relate to the properties of the musculoskeletal system, we might expect motor disorders such as tremors or paralysis (Adams et al. 2013a; Friston et al. 2010b). If abnormal priors relate to perceptual systems, the results may include sensory hallucinations (Adams et al. 2013b; Fletcher and Frith 2009) or anaesthesia. In the following, we review some important concepts in Bayesian accounts of brain function. These include the notion of a generative model, the hierarchical structure of such models, the representation of uncertainty in the brain, and the active nature of sensory perception. In doing so we will develop a taxonomy of pathological priors. While this taxonomy concerns types of inferential deficit (and is not a comprehensive review of neuropsychological syndromes), we draw upon examples of syndromes to illustrate these pathologies. We relate these to failures of neuromodulation and to the notion of a 'disconnection' syndrome (Catani and ffytche 2005; Geschwind 1965a).

The generative model

Bayesian inference

Much work in theoretical neurobiology rests on the notion that the brain performs Bayesian inference (Doya 2007; Friston 2010; Knill and Pouget 2004; O'Reilly et al. 2012). In other words, the brain makes inferences about the (hidden or latent) causes of sensory data. 'Hidden' variables are those that are not directly observable and must be inferred. For example, the position (hidden variable) of a lamp causes a pattern of photoreceptor activation (sensory data) in the retina. Bayesian inference can be used to infer the probable position of the lamp from the retinal data. To do this, two probability distributions must be defined (these are illustrated

graphically in Figure 6.1a). These are the prior probability of the causes, and a likelihood distribution that determines how the causes give rise to sensory data. Together, these are referred to as a 'generative model', as they describe the processes by which data is (believed to be) generated. Bayesian inference uses a generative model to compute the probable causes of sensory data (Beal 2003; Doya 2007; Ghahramani 2015). Many of the inferences that must be made by the brain relate to causes that evolve through time. This means that the prior over the *trajectory* of causes through time can be decomposed into a prior for the initial state, and a series of transition probabilities that account for sequences or dynamics (Figure 6.1b). These dynamics can be subdivided into those that a subject has control over (Figure 6.1c), such as muscle length, and environmental causes that they cannot directly influence.

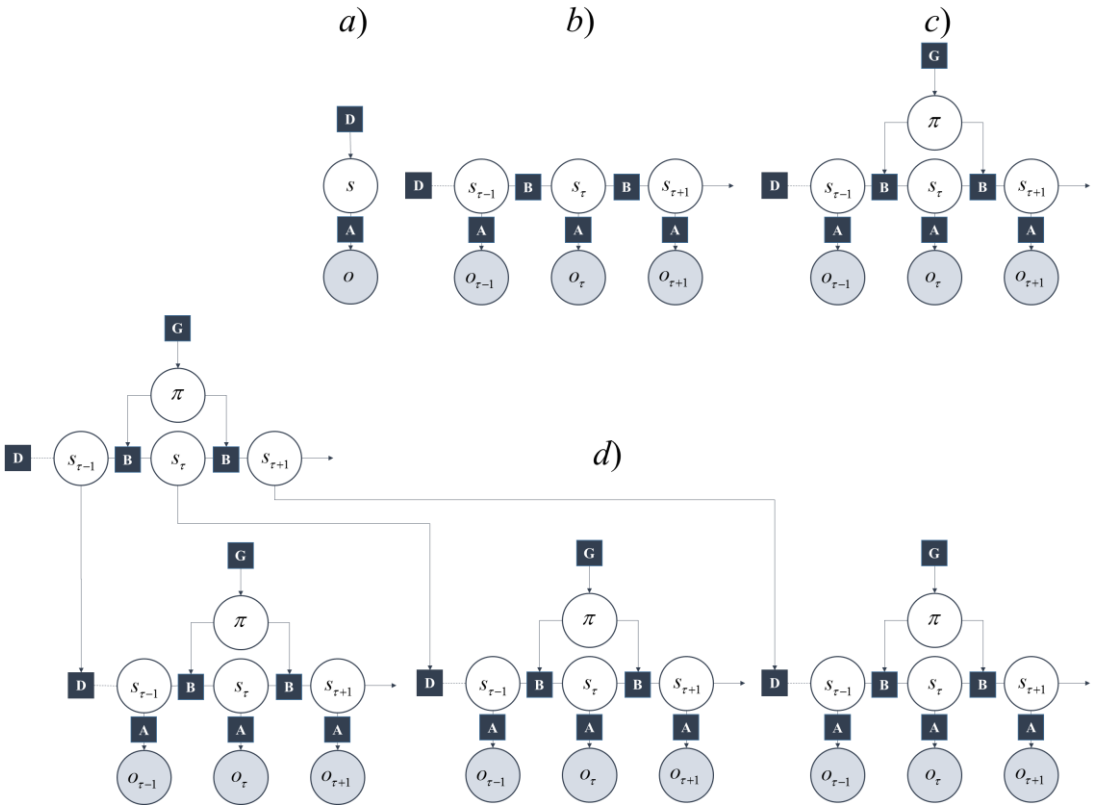

Figure 6.1 – Generative models. These networks graphically illustrate the structure of generative models, using the same notation as in Figure 2.2. (*a*) The simplest model that permits Bayesian inference involves a hidden state (s) that is equipped with a prior $P(s)$, labelled **D**. This hidden state generates observable data (o) through a process defined by the likelihood $P(o|s)$ (vertical arrow, labelled **A**). (*b*) It is possible to equip such a model with dynamically changing hidden states. To do so, we must specify the probabilities of transitioning between states $P(s_{\tau+1}|s_\tau)$ (horizontal arrows, labelled **B**). (*c*) Transitions between states may be influenced by the course of action (π) that is pursued. (*d*) Hierarchical levels can

be added to the generative model (Friston et al. 2017f). This means that the processes that generate the hidden states can themselves be accommodated in the inferences performed using the model.

Predictive coding

Predictive coding is a prominent theory describing how the brain could perform Bayesian inference (Bastos et al. 2012; Friston and Kiebel 2009; Rao and Ballard 1999). This relies upon the idea that the brain uses its generative model to form perceptual hypotheses (Gregory 1980) and make predictions about sensory data. The difference between this prediction and the incoming data is computed, and the ensuing prediction error is used to refine hypotheses about the cause of the data (see Figure 2.1). Under this theory, the messages passed through neuronal signalling are either predictions, or prediction errors. There are other local message passing schemes that can implement Bayesian inference (Dauwels 2007; Friston et al. 2017c; Winn and Bishop 2005; Yedidia et al. 2005), particularly for categorical (as opposed to continuous) inferences, and these are described in Chapter 2 and Appendix A.2. Although we use the language of predictive coding in the following, we note that our discussion generalises to other Bayesian belief propagation schemes.

The notion that hypotheses are corrected by prediction errors makes sense of the kinds of neuropsychological pathologies that result from the loss of sensory signals. For example, patients with eye disease can experience complex visual hallucinations (ffytche and Howard 1999). This phenomenon, known as Charles Bonnet syndrome (Menon et al. 2003; Teunisse et al. 1996), can be interpreted as a failure to constrain perceptual hypotheses with sensations (Reichert et al. 2013). In other words, there are no prediction errors to correct predictions. A similar line of argument can be applied to phantom limbs (De Ridder et al. 2014; Frith et al. 2000). Following amputation, patients may continue to experience ‘phantom’ sensory percepts from their missing limb. The absence of corrective signals from amputated body parts means that any hypothesis held about the limb is unfalsifiable. In the next sections, we consider some of the important features of generative models, and their relationship to brain function.

Hierarchical models

Cortical architecture

An important feature of many generative models is hierarchy. Hierarchical models assume that the hidden causes that generate sensory data are themselves generated from hidden causes at a higher level in the hierarchy (Figure 6.1d). As a hierarchy is ascended, causes tend to become more abstract, and have dynamics that play out over a longer time course (Kiebel et al. 2008; Kiebel et al. 2009). An intuitive example is the kind of generative model required for reading (Friston et al. 2017f). While lower levels may represent letters, higher levels represent words, then sentences, then paragraphs.

There are several converging lines of evidence pointing to the importance of hierarchy as a feature of brain organisation. One of these is the patterns of receptive fields in the cortex (Gallant et al. 1993). In primary sensory cortices, cells tend to respond to simple features such as oriented lines (Hubel and Wiesel 1959). As we move further from sensory cortices, the complexity of the stimulus required to elicit a response increases. Higher areas become selective for contours (Desimone et al. 1985; von der Heydt and Peterhans 1989), shapes, and eventually objects (Valdez et al. 2015). The sizes of receptive fields also increase (Gross et al. 1972; Smith et al. 2001).

A second line of evidence is the change in temporal response properties. Higher areas appear to respond to stimuli that change over longer time courses than lower areas (Hasson et al. 2015; Hasson et al. 2008; Kiebel et al. 2008; Murray et al. 2014). This is consistent with the structure of deep temporal generative models (Friston et al. 2017f) (a sentence takes longer to read than a word). A third line of evidence is the laminar specificity of inter-areal connections that corroborates the pattern implied by electrophysiological responses (Felleman and Van Essen 1991; Markov et al. 2013; Shipp 2007). As illustrated in Figure 6.2, cortical regions lower in the hierarchy project to layer IV of the cortex in higher areas. These ‘ascending’ connections arise from layer III of the lower hierarchical region. ‘Descending’ connections typically arise from deep layers of the cortex, and target both deep and superficial layers of the cortical area lower in the hierarchy.

In addition to the anatomical and physiological evidence for hierarchical organisation summarised above, neuropsychological data emphasises the functional importance of this organisation. For example, it has been argued that deficits in semantic knowledge can only be interpreted with reference to a hierarchically organised set of representations in the brain. This argument rests on observations that patients with agnosia, a failure to recognise objects, can present with semantic deficits at different levels of abstraction. For example, some neurological patients are able to distinguish between broad categories (fruits or vegetables)

but are unable to identify particular objects within a category (Warrington 1975). The preservation of more abstract knowledge, with impairment of within-category semantics, is taken as evidence for distinct hierarchical levels that can be differentially impaired. This is endorsed by findings that some patients have a category-specific agnosia (for example, a failure to identify living but not inanimate stimuli) (Warrington and Shallice 1984). A model that simulates these deficits relies upon a hierarchical structure that allows for specific categorical processing at higher levels to be lesioned while maintaining lower level processes (Humphreys and Forde 2001). Notably, lesions to this model were performed by modulating the connections between hierarchical levels. This resonates well with the type of computational 'disconnection' that predictive coding implicates in some psychiatric disorders (Friston et al. 2016a). We now turn to the probabilistic interpretation of such disconnections.

Ascending and descending messages

The parallel between the hierarchical structure of generative models and that of cortical organisation has an interesting consequence. It suggests that connections between cortical regions at different hierarchical levels are the neurobiological substrate of the likelihoods that map hidden causes to the sensory data, or lower level causes, that they generate (Friston et al. 2017f; Kiebel et al. 2008). This is very important in understanding the computational nature of a 'disconnection' syndrome. It implies that the disruption of a white matter pathway corresponds to an abnormal prior belief about the form of the likelihood distribution. This immediately allows us to think of neurological disconnection syndromes – such as visual agnosia, pure alexia, apraxia, and conduction aphasia (Catani and ffytche 2005) – in probabilistic terms. We will address specific examples of these in the next section. Under predictive coding, the signals carried by inter-areal connections have a clear interpretation (Shipp 2016; Shipp et al. 2013). Descending connections carry the predictions derived from the generative model about the causes or data at the lower level. Ascending connections carry prediction error signals.

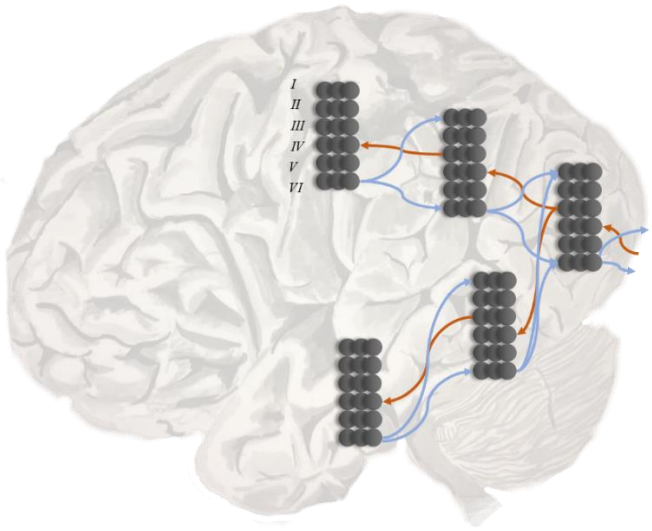

Figure 6.2 – Hierarchy in the cortex. This schematic illustrates two key features of cortical organisation. The first is hierarchy, as defined by laminar specific projections. Projections from primary sensory areas, such as area V1, to higher cortical areas typically arise from layer III of a cortical column, and target layer IV. These ascending connections are shown in red. In contrast, descending connections (in blue) originate in deep layers of the cortex and project to both superficial and deep laminae. The second feature illustrated here is the separation of visual processing into two, dorsal and ventral, streams. In terms of the functional anatomy implied by generative models in the brain, this segregation implies a factorisation of beliefs about the location and identity of a visual object (i.e., knowing what an object is does not tell you where it is – and *vice versa*).

Sensory streams and disconnection syndromes

What and where?

Figure 6.2 illustrates an additional feature common to cortical architectures and inference methods. This is the factorisation of beliefs about hidden causes into multiple streams. Bayesian inference often employs this device, known as a 'mean-field' assumption, which 'carves' posterior beliefs into the product of statistically independent factors (Beal 2003; Friston and Buzsáki 2016). The factorisation of visual hierarchies into ventral and dorsal 'what' and 'where' streams (Ungerleider and Haxby 1994; Ungerleider and Mishkin 1982) appears to be an example of this. A closely related factorisation separates the dorsal and ventral attention networks (Corbetta and Shulman 2002). This factorisation has important

consequences for the representation of objects in space. Location is represented bilaterally in the brain, with each side of space represented in the contralateral hemisphere. As it is not necessary to know the location of an object to know its identity, it is possible to represent this information independently, and therefore unilaterally (Parr and Friston 2017a). It is notable that object recognition deficits tend to occur when patients experience damage to areas in the right hemisphere (Warrington and James 1967; Warrington and James 1988; Warrington and Taylor 1973). Lesions to contralateral (left hemispheric) homologues are more likely to give rise to difficulties in naming objects (Kirshner 2003).

The bilateral representation of space has an important consequence when we frame neuronal processing as probabilistic inference. Following an inference that a stimulus is likely to be on one side of space, it must be the case that it is less likely to be on the contralateral side. If neuronal activities in each hemisphere represent these probabilities, this induces a form of interhemispheric competition (Dietz et al. 2014; Rushmore et al. 2006; Vuilleumier et al. 1996). An important role of commissural fibre pathways may be to enforce the normalisation of probabilities across space (although some of these axons must represent likelihood mappings instead (Glickstein and Berlucchi 2008)). This neatly unifies theories that relate disorders of spatial processing to interhemispheric (Kinsbourne 1970) or intrahemispheric disruptions (Bartolomeo 2014; Bartolomeo et al. 2007). Any intrahemispheric lesion that induces a bias towards one side of space necessarily alters the interhemispheric balance of activity (Parr and Friston 2017b).

Disconnections and likelihoods

The factorisation of beliefs into distinct processing streams is not limited to the visual system. Notably, theories of the neurobiology of speech propose a similar division into dorsal and ventral streams (Hickok and Poeppel 2007; Saur et al. 2008). The former is thought to support articulatory components of speech, while the latter is involved in language comprehension. This mean-field factorisation accommodates the classical subdivision of aphasias into fluent (e.g. Wernicke's aphasia) and non-fluent (e.g. Broca's aphasia) categories. The anatomy of these networks has been interpreted in terms of predictive coding (Hickok 2012a; Hickok 2012b), and this interpretation allows us to illustrate the point that disconnection syndromes are generally due to disruption of the likelihood mapping between two regions. We draw upon examples of aphasic and apraxic syndromes to make this point.

Conduction aphasia is the prototypical disconnection syndrome (Wernicke 1969), disconnecting Wernicke's area from Broca's area. The former is found near the temporoparietal junction, and is thought to contribute to language comprehension. The latter is in the inferior frontal lobe, and is a key part of the dorsal language stream. Disconnection of the two areas results in an inability to repeat spoken language. This connection between these two areas, the arcuate fasciculus (Catani and Mesulam 2008), could represent the likelihood mapping from speech representations in Wernicke's area to the articulatory proprioceptive data processed in Broca's area as in Figure 6.3 (left). While auditory data from the ventral pathway may inform inferences about language, the failure to translate these into proprioceptive predictions means that such predictions cannot be fulfilled by the brainstem motor system (Adams et al. 2013a).

The idea that a common generative model could generate both auditory and proprioceptive predictions, associated with speech, harmonises well with theories of about the 'mirror-neuron' system (Di Pellegrino et al. 1992; Rizzolatti et al. 2001). These neurons respond both to the performance of an action by an individual, and when that individual observes the same action being performed by another. Similarly, Wernicke's area appears to be necessary for both language comprehension and generation (Dronkers and Baldo 2009) (but see (Binder 2015)). Anatomically, there is consistency between the mirror neuron system and the connectivity between the frontal and temporal regions involved in speech. The former is often considered to include Broca's area and the superior temporal sulcus – adjacent to Wernicke's area (Frith and Frith 1999; Keysers and Perrett 2004).

A common generative model for action observation and generation (Kilner et al. 2007) generalises to include the notion of 'conduction apraxia' (Ochipa et al. 1994). As with conduction aphasia, this disorder involves a failure to repeat what another is doing. Instead of repeating spoken language, conduction apraxia represents a deficit in mimicking motor behaviours. This implies a disconnection between visual and motor regions (Catani and ffytche 2005; Goldenberg 2003). This must spare the route from language areas to motor areas. Other forms of apraxia have been considered to be disconnection syndromes in which language areas are disconnected from motor regions, preventing patients from obeying a verbal motor command (Geschwind 1965b). Under this theory, deficits in imitation that accompany this are due to disruption of axons that connect visual and motor areas. These also travel in tracts from posterior to frontal cortices.

Other disconnection syndromes include (Catani and ffytche 2005; Geschwind 1965a) visual agnosia, caused by disruption of connections in the ventral visual stream, and visual neglect (Bartolomeo et al. 2007; Ciaraffa et al. 2013; Doricchi and Tomaiuolo 2003; He et al. 2007).

Neglect can be a consequence of frontoparietal disconnections (Figure 6.3, right), leading to an impaired awareness of stimuli on the left despite intact early visual processing (Rees et al. 2000). We consider the behavioural manifestations of visual neglect in a later section. Before we do so, we turn from disconnections to a subtler form of computational pathology.

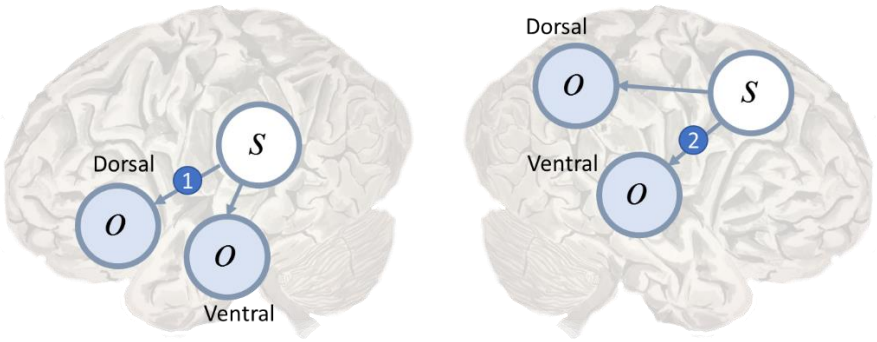

Figure 6.3 – Dorsal and ventral streams. Here we depict a plausible mapping of simple generative models to the dual streams of the language (left) and attention (right) networks. We highlight the likelihood mappings that correspond to white matter tracts implicated in disconnection syndromes. The number 1 in the blue circle on the left highlights the mapping from the left temporoparietal region, which responds to spoken words (Howard et al. 1992), to the inferior frontal gyrus, involved in the dorsal articulatory stream (Hickok 2012b). This region is well placed to deal with proprioceptive data from the laryngeal and pharyngeal muscles (Simonyan and Horwitz 2011). The connection corresponds to the arcuate fasciculus and lesions give rise to conduction aphasia. The number 2 indicates the mapping from dorsal frontal regions that represent eye fixation locations to ventral regions associated with target detection and identity. This corresponds to the second branch of the superior longitudinal fasciculus. Lesions to this structure are implicated in visual neglect (Doricchi and Tomaiuolo 2003; Thiebaut de Schotten et al. 2005).

Uncertainty, precision, and autism

Types of uncertainty

In predictive coding, the significance ascribed to a given prediction error is determined by the precision of the mapping from hidden causes to the data. If this mapping is very stochastic, the gain of the prediction error signal is turned down. A precise relationship between causes

and data leads to an increase in this gain – it is this phenomenon that has been associated with attention (Feldman and Friston 2010). In other words, attention is the process of affording a greater weight to reliable information (Parr and Friston 2019a).

The generative models depicted in Figure 6.1 indicate that there are multiple probability distributions that may be excessively precise or imprecise (Parr and Friston 2017c). One of these is the sensory precision that relates to the likelihood. It is this that weights sensory prediction errors in predictive coding (Feldman and Friston 2010; Friston and Kiebel 2009). Another source of uncertainty relates to the dynamics of hidden causes. It may be that the mapping from the current hidden state to the next is very noisy, or volatile. Alternatively, these transitions may be very deterministic. A third source of uncertainty relates to those states that a person has control over. It is possible for a person to hold beliefs about the course of action, or policy, that they will pursue with differing levels of confidence.

Beliefs about the degree of uncertainty in each of these three distributions have been related to the transmission of acetylcholine (Dayan and Yu 2001; Moran et al. 2013b; Yu and Dayan 2002), noradrenaline (Dayan and Yu 2006), and dopamine (Friston et al. 2014) respectively (Marshall et al. 2016). The ascending neuromodulatory systems associated with these transmitters are depicted in Figure 6.4. The relationship between dopamine and the precision of prior beliefs about policies suggests that the difficulty initiating movements in Parkinson's disease may be due to a high estimated uncertainty about the course of action to pursue (Friston et al. 2013). A complementary perspective suggests that the role of dopamine is to optimise sequences of actions into the future (O'Reilly and Frank 2006). Deficient cholinergic signalling has been implicated in the complex visual hallucinations associated with some neurodegenerative conditions (Collerton et al. 2005) (see Chapter 4 for a discussion of this).

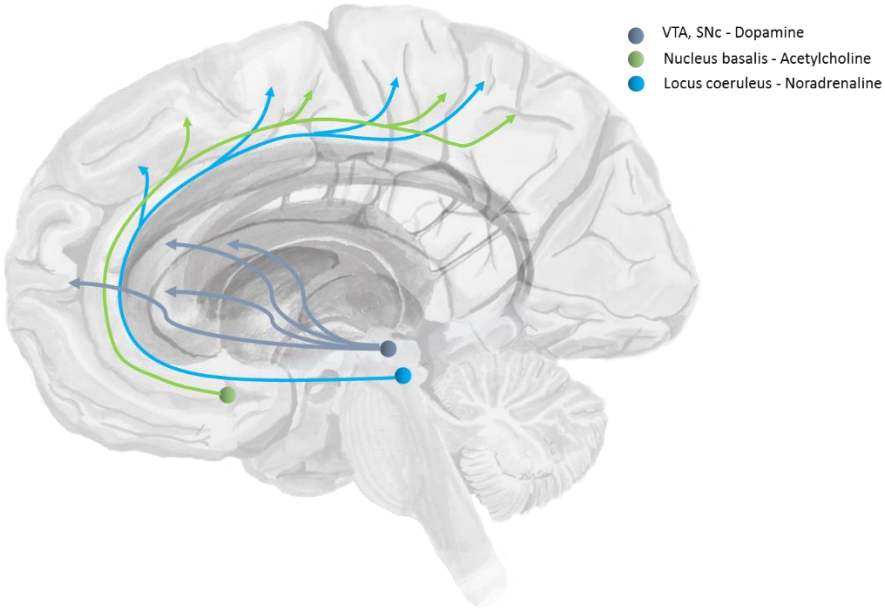

Figure 6.4 – The anatomy of precision. The ascending neuromodulatory systems carrying dopaminergic, cholinergic, and noradrenergic signals are shown (in a simplified form). Dopaminergic neurons have their cell-bodies in the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc) – two nuclei in the midbrain. The medial forebrain bundle contains the axons of these cells, and allows them to target neurons in the prefrontal cortex and the medium spiny neurons of the striatum. The nucleus basalis of Meynert is found in the basal forebrain. This is the source of cholinergic projections to the cortex (Eckenstein et al. 1988). Axons originating here join the cingulum. Neurons in the locus coeruleus project from the brainstem, through the dorsal noradrenergic bundle, and also join the cingulum to supply the cortex with noradrenaline (Berridge and Waterhouse 2003).

Precision and autism

One condition that has received considerable attention using Bayesian formulations is autism (Lawson et al. 2014; Pellicano and Burr 2012). This condition usefully illustrates how aberrant prior beliefs about uncertainty can produce abnormal percepts. An influential treatment of the inferential deficits in autism argues that the condition can be understood in terms of weak prior beliefs (Pellicano and Burr 2012). The consequence of this is that autistic individuals rely to a greater extent upon current sensory data to make inferences about hidden causes. This hypothesis is motivated by several empirical observations, including the resistance of people

with autism to sensory illusions (Happé 1996; Simmons et al. 2009), and their superior performance on tasks requiring the location of low-level features in a complex image (Shah and Frith 1983). The susceptibility of the general population to sensory illusions is thought to be due to the exploitation of artificial scenarios that violate prior beliefs (Brown and Friston 2012; Geisler and Kersten 2002). For example, the perception of the concave surface of a mask as a convex face is due to the, normally accurate, prior (or 'top-down') belief that faces are convex (Gregory 1970). Under this prior, the Bayes optimal inference is a false inference (Weiss et al. 2002). If this prior belief is weakened, the optimal inference becomes the true inference.

The excessive dependence on sensory evidence has been described in terms of an aberrant belief about the precision of the likelihood distribution (Lawson et al. 2014). This account additionally considers the source of this belief (Lawson et al. 2017). It suggests that this may be understood in terms of an aberrant prior belief about the volatility of the environment. Volatility here means the stochasticity of the transition probabilities that describe the dynamics of hidden causes in the world. Highly volatile transitions prevent the precise estimation of current states from the past, and result in imprecise beliefs about hidden causes. In other words, past beliefs become less informative when making inferences about the present. Sensory prediction errors then elicit a greater change in beliefs than they would do if a strong prior were in play. This theory of autism has been tested empirically (Lawson et al. 2017), providing a convincing demonstration of computational neuropsychology in practice. Using a Bayesian observer model (Mathys 2012), it was shown that participants with autism overestimate the volatility of their environment. Complementing this computational finding, pupillary responses, associated with central noradrenergic activity (Koss 1986), were found to be of a smaller magnitude when participants encountered surprising stimuli compared to neurotypical individuals.

A failure to properly balance the precision of sensory evidence, in relation to prior beliefs, may be a ubiquitous theme in many neuropsychiatric disorders. A potentially important aspect of this imbalance is a failure to attenuate sensory precision during self-made acts. The attenuation of sensory precision is an important aspect of movement and active sensing, because it allows us to temporarily suspend attention to sensory evidence that we are not moving (e.g., in the bradykinesia of Parkinson's disease). In brief, a failure of sensory attenuation would have profound consequences for self-generated movement, a sense of agency and selfhood. We now consider the implications of Bayesian pathologies for the active interrogation of the sensorium and its neuropsychology.

Active inference and visual neglect

Active sensing

In the above, we have considered how hypotheses are evaluated as if sensory data is passively presented to the brain. In reality, perception is a much more active process of hypothesis testing (Krause 2008; Yang et al. 2016a; Yang et al. 2016b). Not only are hypotheses formed and refined, but experiments can be performed to confirm or refute them. Saccadic eye movements offer a good example of this, as they turn vision from a passive into an active process (Gibson 1966; Ognibene and Baldassarre 2014; Parr and Friston 2017a). Each saccade can be thought of as an experiment to adjudicate between plausible hypotheses about the hidden causes that give rise to visual data (Friston et al. 2012a; Mirza et al. 2016). As in science, the best experiments are those that will bring about the greatest change in beliefs (Clark 2017; Friston et al. 2016b; Lindley 1956). A mathematical formulation of this imperative (Friston et al. 2015b) suggests that the form of the neuronal message passing required to evaluate different (saccadic) policies maps well to the anatomy of cortico-subcortical loops involving the basal ganglia (Friston et al. 2017f). This is consistent with the known role of this set of subcortical structures in action selection (Gurney et al. 2001; Jahanshahi et al. 2015), and their anatomical projections to oculomotor areas in the midbrain (Hikosaka et al. 2000). To illustrate the importance of these points, we consider visual neglect (paraphrasing Chapter 5), a disorder in which active vision is impaired.

Visual neglect

A common neuropsychological syndrome, resulting from damage to the right cerebral hemisphere, is visual neglect (Halligan and Marshall 1998). This is characterised by a failure to attend to the left side of space. This rightward lateralisation may be a consequence of the mean-field factorisation discussed earlier. Although space is represented bilaterally in the brain, there is no need for representations of identity to be bilateral. This means that the relationships between location and identity should be asymmetrical, complementing the observation that visual neglect is very rarely the consequence of a left hemispheric lesion.

A behavioural manifestation of this disorder is a bias in saccadic sampling (Fruhmann Berger et al. 2008; Husain et al. 2001; Karnath and Rorden 2012). Patients with neglect tend to perform saccades to locations on the right more frequently than to those on the left. There are several different sets of prior beliefs that would make this behaviour optimal. We will discuss three possibilities (Parr and Friston 2017b), and consider their biological bases (Figure 6.5). One is a prior belief that proprioceptive data will be consistent with fixations on the right of space. The dorsal parietal lobe is known to contain the 'parietal eye fields' (Shipp 2004), and it is plausible that an input to this region may specify such prior beliefs. A candidate structure is the dorsal pulvinar (Shipp 2003). This is a thalamic nucleus implicated in attentional processing (Kanai et al. 2015; Robinson and Petersen ; Ungerleider and Christensen 1979). Crucially, lesions to this structure have been implicated in neglect (Karnath et al. 2002).

A second possibility relates more directly to the question of good experimental design. If a saccade is unlikely to induce a change in current beliefs, then there is little value in performing it. One form that current beliefs take is the likelihood distribution mapping 'where I am looking' to 'what I see' (Mirza et al. 2016). As illustrated in Figure 6.3 (right) this likelihood distribution takes the form of a connection between dorsal frontal and ventral parietal regions (Parr and Friston 2017a). To adjust beliefs about this mapping, observations could induce a plastic change in synaptic strength following each saccade (Friston et al. 2016b). If the white matter tract connecting these areas is lesioned, it becomes impossible to update these beliefs. As such, if we were to cut the second branch of the superior longitudinal fasciculus (SLF II) on the right, disconnecting dorsal frontal from ventral parietal regions, we would expect there to be no change in beliefs following a saccade to the left. These would make for very poor 'visual experiments' (Lindley 1956). A very similar argument has been put forward for neglect of personal space that emphasises proprioceptive (rather than visual) consequences of action (Committeri et al. 2007). In these circumstances, optimal behaviour would require a greater frequency of rightward saccades. Lesions to SLF II (Doricchi and Tomaiuolo 2003; Lunven et al. 2015; Thiebaut de Schotten et al. 2005), and the regions it connects (Corbetta et al. 2000; Corbetta and Shulman 2002; Corbetta and Shulman 2011) are associated with neglect.

A third possibility is that the process of policy selection may be inherently biased. Above, we suggested that these computations may involve subcortical structures. The striatum, an input nucleus to the basal ganglia, is well known to be involved in habit formation (Graybiel and Grafton 2015; Yin and Knowlton 2006). Habits may be formalised as a bias in prior beliefs about policy selection (FitzGerald et al. 2014). It is plausible that a lesion in the striatum might induce a similar behavioural bias towards saccades to rightward targets. One of the subcortical structures most frequently implicated in visual neglect is the putamen (Karnath et al. 2002),

one of the constituent nuclei of the striatum. Such lesions may be interpretable as disrupting the prior belief about policies.

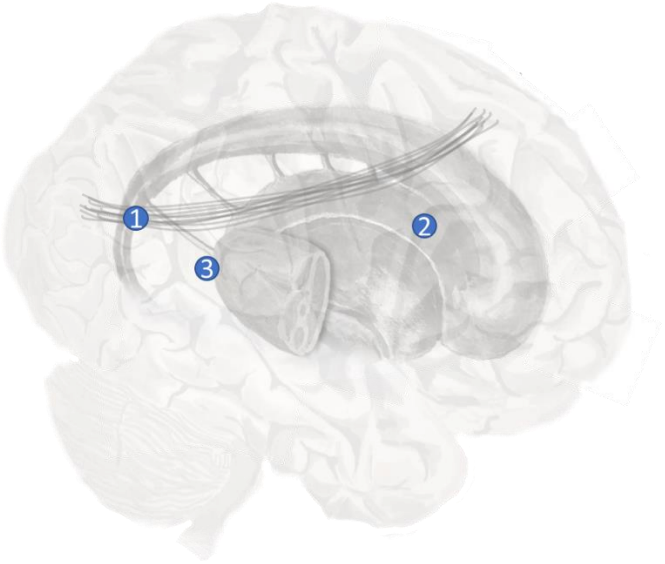

Figure 6.5 – The anatomy of visual neglect. Three lesions implicated in visual neglect are highlighted here. 1 – Disconnection of the second branch of the right superior longitudinal fasciculus (a white matter tract that connects dorsal frontal with ventral parietal regions); 2 – Unilateral lesion to the right putamen; 3 – Unilateral lesion to the right pulvinar (a thalamic nucleus). Note that lesion 1 here is the same as lesion 2 in Figure 6.3.

Anosognosia

The ideas outlined above, that movements can be thought of as sensory experiments, are not limited to eye movements and visual data. Plausibly, limb movements could be used to test hypotheses about proprioceptive (and visual) sensations. This has interesting consequences for a neuropsychological deficit known as anosognosia (Fotopoulou 2012). This syndrome can accompany hemiplegia, which prevents the performance of perceptual experiments using the paralysed limb (Fotopoulou 2014). In addition to the failure to perform such an experiment, patients must be able to ignore any discrepancy between predicted movements and the contradictory sensory data suggesting the absence of a movement (Frith et al. 2000). This distinction mirrors that between salience and attention in the visual domain (Parr and Friston 2019a). As this failure of monitoring movement trajectories can be induced in healthy subjects

(Fourneret and Jeannerod 1998), it seems plausible that this could be exaggerated in the context of hemiplegia, through a dampening of exteroceptive sensory precision.

This explanation is not sufficient on its own, as anosognosia does not occur in all cases of hemiplegia. Lesion mapping has implicated the insula in the deficits observed in these patients (Fotopoulou et al. 2010; Karnath et al. 2005a). This is a region often associated with interoceptive inference (Barrett and Simmons 2015) that has substantial efferent connectivity to somatosensory cortex (Mesulam and Mufson 1982; Showers and Lauer 1961). Damage to the insula and surrounding regions might reflect a disconnection of the mapping from motor hypotheses to the interoceptive data that accounts for what it 'feels like' to move a limb. This is consistent with evidence that the insula mediates inferences about these kinds of sensations (Allen et al. 2016). A plausible hypothesis for the computational pathology in anosognosia is then that a failure of *active* inference is combined with a disconnection of the likelihood mapping between motor control and its interoceptive (and exteroceptive) consequences (Fotopoulou et al. 2008).

A (provisional) taxonomy of computational pathology

In the above, we have described the components of a generative model required to perform Bayesian inference. We have reviewed some of the syndromes that may illustrate deficits of one or more of these components. Broadly, the generative model constitutes beliefs about the hidden states, their dynamics, and the mechanisms by which sensory data are believed to be generated. Each of these beliefs can be disrupted through an increase or decrease in precision, or through disconnections. Modulation of precision implicates the ascending neuromodulatory systems. This modulation may be important for a range of neuropsychiatric and functional neurological disorders (Edwards et al. 2012).

In addition to modulation of connectivity, disconnections can completely disrupt beliefs about the conditional probability of one variable given another. The hierarchical architecture of the cortex suggests that inter-areal white matter tracts, the most vulnerable to vascular or inflammatory lesions, represent likelihood distributions (i.e. the probability of data, or a low-level cause, given a high-level cause). Drawing upon the notion of a mean-field factorisation, we noted that such disconnections are likely to have a hemispheric asymmetry in the behaviours they elicit. It is also plausible that functional disconnections might occur within a cortical region. This would allow for disruption of transition probabilities. While not as

vulnerable to vascular insult, other pathologies can cause changes in intrinsic cortical connectivity (Cooray et al. 2015).

Epistemic, foraging, behaviour is vital for the evaluation of beliefs about the world. Unusual patterns of sensorimotor sampling can be induced by abnormal beliefs about the motor experiments that best disambiguate between perceptual hypotheses. These computations implicate subcortical structures, such as the basal ganglia. There are two ways that disruption of these computations may result in abnormal behaviour. The first is that prior beliefs about policies may be biased. This can be an indirect effect, through other beliefs, or a direct effect due to dysfunction in basal ganglia networks. The second is that an impairment in performing these experiments, due to paralysis, might impair the refutation of incorrect perceptual hypotheses. This may be compounded by a disconnection or a neuromodulatory failure, as has been proposed in anosognosia.

One further source of an aberrant priors, exploited in Chapter 4, is neuronal loss. In neurodegenerative disorders, there may be a reduction in the number of neurons in a given brain area. This results in a smaller number of possible activity patterns across these neurons and limits the number of hypotheses they can represent. This means that disorders in which neurons are lost may cause a shrinkage of the brain's hypothesis space. In other words, the failure to form accurate perceptual hypotheses in such conditions may be due to an attrition of the number of hypotheses that can be entertained by the brain. An important future step in Bayesian neuropsychology will be linking tissue pathology with computation more directly. This may be one route towards achieving this.

Conclusion

While Bayesian approaches are not in conflict with other methods in computational neuroscience, they do offer a different (complementary) perspective that is often very useful. For example, many traditional modelling approaches would not predict that disconnections in early sensory streams, such as the retino-geniculate system, could result in complex sensory hallucinations. Calling upon a hierarchical generative model that makes 'top-down' predictions about sensory data, clarifies and provides insight into such issues. In the above we have discussed the features of the generative models that underwrite perception and behaviour. We have illustrated the importance of these features through examples of their failures. These computational pathologies can be described in terms of abnormal prior beliefs, or in terms of their biological substrates. We noted that aberrant priors about the structure of a likelihood

mapping relate to disconnection syndromes, ubiquitous in neurology. Pathological beliefs about uncertainty may manifest as neuromodulatory disorders. The process of identifying the pathological priors that give rise to Bayes optimal behaviour in patients is promising both scientifically and clinically. If individual patients can be uniquely characterised by subject-specific priors, this facilitates a precision medicine approach grounded in computational phenotyping (Adams et al. 2016; Mirza et al. 2018; Schwartenbeck and Friston 2016). This also allows for empirical evaluation of hypotheses about abnormal priors, by comparing quantitative, computational phenotypes between clinical and healthy populations. Relating these priors to their biological substrates offers the further possibility of treatments that target aberrant neurobiology in a patient specific manner.

Summary of key publications contributing to this thesis

Title	Comment	Citation
The active construction of the visual world	Review of the anatomy of active vision and visual neglect as a pathology of active vision	(Parr and Friston 2017a)
The computational anatomy of visual neglect	Computational model of visual neglect, and demonstration that simulated lesions can be recovered from synthetic eye-tracking data	(Parr and Friston 2017b)
Uncertainty, epistemics, and active inference	Introduction of precision (attention) to discrete state space models for active inference, their role in driving exploratory behaviour, and their adrenergic and cholinergic substrates	(Parr and Friston 2017c)
Working memory, attention, and salience in active inference	Deep temporal modelling of classic delay-period working memory tasks, and simulations of electrophysiological correlates of working memory	(Parr and Friston 2017d)

Active inference and the anatomy of oculomotion	Modelling of brainstem oculomotor processes, using generalised Bayesian filtering to solve Newtonian equations of motion. Simulation of brainstem physiology and pathology, e.g. internuclear ophthalmoplegia	(Parr and Friston 2018a)
The discrete and continuous brain: from decisions to movement – and back again	Combining continuous models of oculomotion with discrete decision processes to simulate cortical, basal ganglia, superior collicular, and brainstem responses simultaneously	(Parr and Friston 2018c)
Generalised free energy and active inference	Technical account of inferences about the future, treating yet-to-be collected data as latent variables	(Parr and Friston 2019c)
Computational neuropsychology and Bayesian inference	Review of Bayesian accounts of brain dysfunction, calling upon the complete class theorems	(Parr et al. 2018b)
Precision and false perceptual inference	Simulation of a computational diaschisis, inspired by visual hallucinations in synucleinopathies	(Parr et al. 2018a)
Dynamic causal modelling of active vision	Demonstration, using DCM for MEG, that a form of visual working memory is represented through changes in synaptic efficacy	(Parr et al. 2019c)
Neuronal message passing: Mean field, Bethe, and Marginal approximations	Comparison of local Bayesian message passing schemes as descriptions of neuronal computation	(Parr et al. 2019b)
Attention or Salience?	Review of the distinctions between attentional gain processes and salience attribution	(Parr and Friston 2019a)
The anatomy of inference: Generative models and brain structure	Review of the anatomical process theories associated with active inference, with a focus upon the connectivity implied by the conditional independencies (Markov blankets) of a generative model	(Parr and Friston 2018b)

The computational pharmacology of oculomotion	Simulation of a delay-period oculomotor task under various synthetic neuromodulatory perturbations, investigating the oculomotor consequences of common pharmacological agents	(Parr and Friston 2019b)

Appendices

A.1 – The Laplace approximation

The Laplace approximation for a variational distribution is the assumption that it takes a Gaussian form. This may be motivated though a Taylor series expansion of the log probability around the mode of the distribution:

$$q(x) \approx p(x \mid y) \propto p(x, y)$$
$$\mu = \arg\max_x p(x \mid y)$$

$$\ln q(x) = \ln p(\mu, y) + (x-\mu)\cdot \underbrace{\partial_\mu \ln p(\mu, y)}_{0} + \tfrac{1}{2}(x-\mu)\cdot \partial_{\mu\mu} \ln p(\mu, y)(x-\mu) \quad \text{(A1}$$
$$\Rightarrow q(x) \propto \exp\left(\tfrac{1}{2}(x-\mu)\cdot \partial_{\mu\mu} \ln p(\mu, y)(x-\mu)\right)$$
$$q(x) = \mathcal{N}(\mu, C^{-1})$$
$$C^{-1} = -\partial_{\mu\mu} \ln p(\mu, y)$$

The second term of the expansion disappears, as the gradient of the log probability at the mode is zero (by definition). The first term is constant with respect to *x*, so can be absorbed into the partition function. The result is a quadratic (i.e. Gaussian) form for the log probability, justifying the Laplace approximation in the region around the posterior mode. Note that the precision of this distribution is the (negative) curvature of the log joint probability, evaluated at the mode.

A.2 – Bethe and mean-field approximations

This appendix, based upon that in (Parr et al. 2019b), outlines the two alternative message passing schemes used to compare the marginal message passing approach in Chapter 2.

Mean-field approximation

In place of the free energy expressed in Equation 2.10, variational message passing uses a free energy defined by choosing the family of posterior distributions to be those that fully factorise (in this case, over time):

$$Q(\tilde{s}^n \mid \pi) \triangleq \prod_\tau Q(s_\tau^n \mid \pi) \tag{A2}$$

This implies the following free energy functional:

$$\begin{aligned} \mathbf{F}_\pi &= \sum_\tau \mathbf{F}_{\pi\tau} \\ \mathbf{F}_{\pi\tau} &= -\sum_n \mathbf{s}_{\pi\tau}^n \cdot \left(\sum_m \ln \mathbf{A}^m \mathbf{s}_{\pi\tau}^{\setminus n} \cdot o_\tau^m + \ln \mathbf{B}_\pi^n \mathbf{s}_{\pi\tau-1}^n - \ln \mathbf{s}_{\pi\tau}^n \right) \end{aligned} \tag{A3}$$

Variational message passing is obtained simply by taking the gradients of this free energy with respect to the factors of the approximate posterior (Beal 2003).

$$\begin{aligned} \mathbf{s}_{\pi\tau}^n &= \sigma(\mathbf{v}_{\pi\tau}^n) \\ \dot{\mathbf{v}}_{\pi\tau}^n &= -\nabla_{\mathbf{s}_{\pi\tau}^n} \mathbf{F}_\pi \\ &= \sum_m \ln \mathbf{A}^m \mathbf{s}_{\pi\tau}^{\setminus n} \cdot o_\tau^m + \ln \mathbf{B}_\pi^n \mathbf{s}_{\pi\tau-1}^n + \ln \mathbf{B}_\pi^n \cdot \mathbf{s}_{\pi\tau+1}^n - \ln \mathbf{s}_{\pi\tau}^n \end{aligned} \tag{A4}$$

Note that, unlike in Equation 2.10, we take the gradient of the free energy summed over all time-points ($\mathbf{F}_\pi$). In contrast, Equation 2.10 treats each time-step as a separate generative model, so takes the gradients with respect to the (temporally local) free energies ($\mathbf{F}_{\pi\tau}$).

Bethe approximation

A less severe choice for the family of posterior distributions is that afforded by the Bethe approximation:

$$Q(\tilde{s}^n \mid \pi) = \left(\prod_\tau Q(s_\tau^n \mid \pi)\right)\left(\prod_\tau \frac{Q(s_\tau^n, s_{\tau-1}^n \mid \pi)}{Q(s_\tau^n \mid \pi)Q(s_{\tau-1}^n \mid \pi)}\right) \tag{A5}$$

This lets us write down the Bethe free energy[18]:

$$\begin{aligned}
\mathbf{F}_\pi &= \sum_\tau \mathbf{F}_{\pi\tau} \\
\mathbf{F}_{\pi\tau} &= -\sum_n \Big(\mathbf{s}_{\pi\tau}^n \cdot \Big(\sum_m \ln \mathbf{A}\mathbf{s}_{\pi\tau}^{\setminus n} \cdot o_\tau^m - \ln \mathbf{s}_{\pi\tau}^n\Big) \\
&\quad + tr(\mathbf{S}_{\pi\tau}^n \cdot (\ln \mathbf{B}_\pi - \ln \mathbf{S}_{\pi\tau}^n + \ln \mathbf{s}_{\pi\tau}^n \otimes \mathbf{s}_{\pi\tau-1}^n))\Big)
\end{aligned} \tag{A6}$$

We can then take the appropriate gradients and solve for their fixed points. To enforce the constraint that the singleton distributions are marginals of the pairwise distributions, we use Lagrange multipliers, which contribute the following gradients:

$$\begin{aligned}
\nabla_{\mathbf{S}_{\pi\tau}^n}(\boldsymbol{\lambda}_{\tau-1}^\tau \cdot (\mathbf{S}_{\pi\tau}^n \mathbf{1} - \mathbf{s}_{\pi\tau}^n)) &= \boldsymbol{\lambda}_{\tau-1}^\tau \\
\nabla_{\mathbf{S}_{\pi\tau}^n}(\boldsymbol{\lambda}_\tau^{\tau-1} \cdot (\mathbf{S}_{\pi\tau}^n \cdot \mathbf{1} - \mathbf{s}_{\pi\tau-1}^n)) &= \boldsymbol{\lambda}_\tau^{\tau-1} \\
\nabla_{\mathbf{s}_{\pi\tau}^n}(\boldsymbol{\lambda}_{\tau-1}^\tau \cdot (\mathbf{S}_{\pi\tau}^n \mathbf{1} - \mathbf{s}_{\pi\tau}^n)) &= -\boldsymbol{\lambda}_{\tau-1}^\tau \\
\nabla_{\mathbf{s}_{\pi\tau}^n}(\boldsymbol{\lambda}_{\tau+1}^\tau \cdot (\mathbf{S}_{\pi\tau+1}^n \cdot \mathbf{1} - \mathbf{s}_{\pi\tau}^n)) &= -\boldsymbol{\lambda}_{\tau+1}^\tau
\end{aligned} \tag{A7}$$

[18] $Q(s_\tau^n, s_{\tau-1}^n \mid \pi) = Cat(\mathbf{S}_{\pi\tau}^n)$

These gradients are added to those of the free energy, such that the sum of the gradients must be equal to zero to satisfy these constraints. In the following, we use '$\circ$' to indicate a Hadamard (elementwise) product, in addition to employing Kronecker tensor products and sums.

$$\nabla_{\mathbf{S}^n_{\pi\tau}} \mathbf{F}_\pi = \ln \mathbf{B}_\pi - \ln \mathbf{S}^n_{\pi\tau} + \ln \mathbf{s}^n_{\pi\tau} \otimes \mathbf{s}^n_{\pi\tau-1}$$

$$\left.\begin{aligned} \mathbf{S}^n_{\pi\tau} &= \arg\min_{\mathbf{S}^n_{\pi\tau}} \mathbf{F}_\pi \\ &\textit{subject to} \\ \mathbf{S}^n_{\pi\tau}\mathbf{1} &= \mathbf{s}^n_{\pi\tau} \\ \mathbf{S}^n_{\pi\tau}\cdot\mathbf{1} &= \mathbf{s}^n_{\pi\tau-1} \end{aligned}\right\} \Leftrightarrow \mathbf{S}^n_{\pi\tau} = \mathbf{B}_\pi \circ (\mathbf{s}^n_{\pi\tau} \otimes \mathbf{s}^n_{\pi\tau-1}) \circ e^{\boldsymbol{\lambda}^{\tau}_{\tau-1} \oplus \boldsymbol{\lambda}^{\tau-1}_{\tau}} \tag{A8}$$

$$\nabla_{\mathbf{s}^n_{\pi\tau}} \mathbf{F}_\pi = \sum\nolimits_m \ln \mathbf{A}\mathbf{s}^{\backslash n}_{\pi\tau} \cdot o^m_\tau - \ln \mathbf{s}^n_{\pi\tau}$$

$$\left.\begin{aligned} \mathbf{s}^n_{\pi\tau} &= \arg\min_{\mathbf{s}^n_{\pi\tau}} \mathbf{F}_\pi \\ &\textit{subject to} \\ \mathbf{S}^n_{\pi\tau}\mathbf{1} &= \mathbf{s}^n_{\pi\tau} \\ \mathbf{S}^n_{\pi\tau}\cdot\mathbf{1} &= \mathbf{s}^n_{\pi\tau-1} \end{aligned}\right\} \Leftrightarrow \mathbf{s}^n_{\pi\tau} = e^{\sum_m \ln \mathbf{A}\mathbf{s}^{\backslash n}_{\pi\tau} \cdot o^m_\tau} \circ e^{-\boldsymbol{\lambda}^{\tau}_{\tau-1} - \boldsymbol{\lambda}^{\tau}_{\tau+1}}$$

Substituting the singleton (**s**) distribution into the expression for the pairwise (**S**) distribution, and summing the pairwise distributions over each of their dimensions then gives:

$$\begin{aligned} \mathbf{S}^n_{\pi\tau} &= \mathbf{B}_\pi \circ \left(\left(\mathbf{s}^n_{\pi\tau} \circ e^{\boldsymbol{\lambda}^{\tau}_{\tau-1}} \right) \otimes \left(e^{\sum_m \ln \mathbf{A}\mathbf{s}^{\backslash n}_{\pi\tau-1} \cdot o^m_{\tau-1}} \circ e^{-\boldsymbol{\lambda}^{\tau-1}_{\tau-2}} \right) \right) \\ \mathbf{s}^n_{\pi\tau} &= \mathbf{S}^n_{\pi\tau}\cdot\mathbf{1} = \mathbf{s}^n_{\pi\tau} \circ e^{\boldsymbol{\lambda}^{\tau}_{\tau-1}} \circ \mathbf{B}_\pi \left(e^{\sum_m \ln \mathbf{A}\mathbf{s}^{\backslash n}_{\pi\tau-1} \cdot o^m_{\tau-1}} \circ e^{-\boldsymbol{\lambda}^{\tau-1}_{\tau-2}} \right) \\ &\Rightarrow e^{-\boldsymbol{\lambda}^{\tau}_{\tau-1}} = \mathbf{B}_\pi \left(e^{\sum_m \ln \mathbf{A}\mathbf{s}^{\backslash n}_{\pi\tau-1} \cdot o^m_{\tau-1}} \circ e^{-\boldsymbol{\lambda}^{\tau-1}_{\tau-2}} \right) \end{aligned} \tag{A9}$$

$$\begin{aligned} \mathbf{S}^n_{\pi\tau} &= \mathbf{B}_\pi \circ \left(\left(e^{\sum_m \ln \mathbf{A}\mathbf{s}^{\backslash n}_{\pi\tau} \cdot o^m_\tau} \circ e^{-\boldsymbol{\lambda}^{\tau}_{\tau+1}} \right) \otimes \left(\mathbf{s}^n_{\pi\tau-1} \circ e^{\boldsymbol{\lambda}^{\tau-1}_{\tau}} \right) \right) \\ \mathbf{s}^n_{\pi\tau-1} &= \mathbf{S}^n_{\pi\tau}\mathbf{1} = \mathbf{s}^n_{\pi\tau-1} \circ e^{\boldsymbol{\lambda}^{\tau-1}_{\tau}} \circ \mathbf{B}_\pi \left(e^{\sum_m \ln \mathbf{A}\mathbf{s}^{\backslash n}_{\pi\tau} \cdot o^m_\tau} \circ e^{-\boldsymbol{\lambda}^{\tau}_{\tau+1}} \right) \\ &\Rightarrow e^{-\boldsymbol{\lambda}^{\tau-1}_{\tau}} = \mathbf{B}_\pi \cdot \left(e^{\sum_m \ln \mathbf{A}\mathbf{s}^{\backslash n}_{\pi\tau} \cdot o^m_\tau} \circ e^{-\boldsymbol{\lambda}^{\tau}_{\tau+1}} \right) \end{aligned}$$

Defining the following messages, this gives rise to the following belief-propagation scheme:

$$\mathbf{s}^{n}_{\pi\tau} = \boldsymbol{\alpha}_{\tau} \circ \vec{\boldsymbol{\beta}}_{\tau} \circ \overleftarrow{\boldsymbol{\beta}}_{\tau}$$

$$\begin{aligned} \boldsymbol{\alpha}_{\tau} &\triangleq e^{\sum_{m} \ln \mathbf{A}\mathbf{s}^{\setminus n}_{\pi\tau} \cdot o^{m}_{\tau}} \\ \vec{\boldsymbol{\beta}}_{\tau} &\triangleq \mathbf{B}_{\pi}(\boldsymbol{\alpha}_{\tau-1} \circ \vec{\boldsymbol{\beta}}_{\tau-1}) = e^{-\boldsymbol{\lambda}^{\tau}_{\tau-1}} \\ \overleftarrow{\boldsymbol{\beta}}_{\tau} &\triangleq \mathbf{B}_{\pi} \cdot (\boldsymbol{\alpha}_{\tau+1} \circ \overleftarrow{\boldsymbol{\beta}}_{\tau+1}) = e^{-\boldsymbol{\lambda}^{\tau}_{\tau+1}} \end{aligned} \tag{A10}$$

This may be implemented as a gradient ascent through the following scheme (used in the simulations in Figure 2.3):

$$\begin{aligned} \mathbf{s}^{n}_{\pi\tau} &= \sigma(\mathbf{v}^{n}_{\pi\tau}) \\ \dot{\mathbf{v}}^{n}_{\pi\tau} &= \ln\left(\mathbf{B}_{\pi}(\vec{\boldsymbol{\beta}}_{\tau-1} \circ \boldsymbol{\alpha}_{\tau-1})\right) + \ln\left(\mathbf{B}_{\pi} \cdot (\overleftarrow{\boldsymbol{\beta}}_{\tau+1} \circ \boldsymbol{\alpha}_{\tau+1})\right) + \ln \boldsymbol{\alpha}_{\tau} \\ \vec{\boldsymbol{\beta}}_{\tau} &= \sigma(\ln \mathbf{s}^{n}_{\pi\tau} - \ln \overleftarrow{\boldsymbol{\beta}}_{\tau} - \ln \boldsymbol{\alpha}_{\tau}) \\ \overleftarrow{\boldsymbol{\beta}}_{\tau} &= \sigma(\ln \mathbf{s}^{n}_{\pi\tau} - \ln \vec{\boldsymbol{\beta}}_{\tau} - \ln \boldsymbol{\alpha}_{\tau}) \end{aligned} \tag{A11}$$

Note the similarity between this and the simpler (but less accurate) Equation A4.

A.3 – Expected free energy

This appendix motivates the role of the expected free energy in policy selection based upon the more technical accounts in (Friston 2019; Parr et al. In Press). Because these derivations were originally formulated in continuous time, we do not explicitly index time *steps* in what follows. The key idea is that we can write down a KL-Divergence between two distributions, one of which depends upon the policy, and one that does not. By definition, this will be greater than or equal to zero.

$$D_{KL}[P(o,s \mid \pi) \,||\, P(o,s)] \geq 0 \tag{A12}$$

Unpacking this divergence, we get:

$$
\begin{aligned}
&E_{P(o,s|\pi)}[\ln P(o,s \mid \pi) - \ln P(o,s)] \geq 0 \\
&\qquad \Rightarrow G(\pi) \geq -E_{P(o|\pi)}[\ln P(o \mid \pi)] \\
&G(\pi) \triangleq E_{P(o,s|\pi)}[\ln P(s \mid o,\pi) - \ln P(o,s)]
\end{aligned}
\tag{A13}
$$

Equation A13 says that the average surprise, or negative log evidence, associated with a policy is upper-bounded by the expected difference between the log posterior under a policy, and the log joint (steady-state) distribution over outcomes and states. This has the same form as the variational free energy (Equation 2.2), except that the expectation now includes a predictive distribution over outcomes, so is termed the 'expected free energy' (G). Given that this bounds the negative expected evidence, the negative expected free energy may be used to score the plausibility of each policy. Note that the smaller the expected free energy, the tighter the bound.

One problem we face here is that the quantities used to calculate the expected free energy may not be tractable to compute. However, as detailed in Chapter 2, we can calculate variational approximations to these quantities that become increasingly accurate as free energy is minimised. Appealing to these approximations, we express the expected free energy as follows:

$$
\begin{aligned}
F(\pi) &= D_{KL}[Q(s \mid \pi) \,||\, P(s \mid o,\pi)] - \ln P(o \mid \pi) \\
&= D_{KL}[Q(s \mid \pi) \,||\, P(s \mid \pi)] - E_{Q(s|\pi)}[\ln P(o \mid s)]
\end{aligned}
$$

$$
\begin{aligned}
&\delta_{Q(s|\pi)} F(\pi) = 0 \Rightarrow Q(s \mid \pi) \approx P(s \mid o,\pi) \\
&\qquad \Rightarrow G(\pi) \approx E_{P(o|s)P(s|\pi)}[\ln Q(s \mid \pi) - \ln P(o,s)] \\
&\delta_{P(s|\pi)} F(\pi) = 0 \Rightarrow P(s \mid \pi) \approx Q(s \mid \pi) \\
&\qquad \Rightarrow G(\pi) \approx E_{P(o|s)Q(s|\pi)}[\ln Q(s \mid \pi) - \ln P(o,s)]
\end{aligned}
\tag{A14}
$$

The first of these approximations is the same variational approximation introduced in Chapter 2. This says that, once we have found the Q that minimises free energy, we can use this as an approximation to the exact posterior distribution. The second (using the posterior as the new

prior[19]) lets us express the expected free energy in terms of a KL-Divergence between posteriors and priors (and a conditional entropy):

$$G(\pi) \approx D_{KL}[Q(s \mid \pi) \| P(s)] + E_{Q(s|\pi)}[H[P(o \mid s)]] \quad \text{(A15}$$

A.4 – Inferring uncertainty

This appendix outlines a derivation for Bayesian belief updating of precision parameters using discrete-state space models (Parr and Friston 2017c). We assume that the precision parameters are distributed as gamma distributions, and follow a similar line of reasoning to that used to derive updates for policy precisions in previous papers. The prior distribution over the precision parameters is then:

$$\begin{aligned} p(\zeta) &\propto \beta_{\zeta} \exp\left(-\beta_{\zeta}\zeta\right) \\ p(\omega) &\propto \beta_{\omega} \exp\left(-\beta_{\omega}\omega\right) \\ p(\gamma) &\propto \beta_{\gamma} \exp\left(-\beta_{\gamma}\gamma\right) \end{aligned} \quad \text{(A16}$$

The approximate posterior distributions have the same (gamma distribution) form and we will use a bold beta hyper-parameter to distinguish between the sufficient statistics of the posterior and prior above. A useful property of the gamma distribution, when parameterised in this way, is the following:

$$\begin{aligned} \boldsymbol{\zeta} &= E_Q[\zeta] = \boldsymbol{\beta}_{\zeta}^{-1} \\ \boldsymbol{\omega} &= E_Q[\omega] = \boldsymbol{\beta}_{\omega}^{-1} \\ \boldsymbol{\gamma} &= E_Q[\gamma] = \boldsymbol{\beta}_{\gamma}^{-1} \end{aligned} \quad \text{(A17}$$

[19] This use of 'yesterday's posterior as today's prior' is sometimes referred to as 'Bayesian belief updating'. Note that this assumes we have already minimised free energy w.r.t *Q*.

Having defined these distributions, we can write the variational free energy:

$$\begin{aligned} F &= E_Q[F(\pi,\zeta,\omega) + D_{KL}[Q(\pi)\,||\,P(\pi\,|\,\gamma)]] \\ &+ D_{KL}[Q(\gamma)\,||\,P(\gamma)] + D_{KL}[Q(\omega)\,||\,P(\omega)] + D_{KL}[Q(\zeta)\,||\,P(\zeta)] \end{aligned} \tag{A18}$$

Which can be expressed in terms of sufficient statistics (omitting constants),

$$\begin{aligned} F &= \boldsymbol{\pi}\cdot(\mathbf{F} + \ln\boldsymbol{\pi} + \boldsymbol{\gamma}\cdot\mathbf{G} + \ln\mathbf{Z}(\boldsymbol{\gamma})) \\ &+ \ln\boldsymbol{\beta}_\gamma + \ln\boldsymbol{\beta}_\omega + \ln\boldsymbol{\beta}_\zeta - \ln\beta_\gamma - \ln\beta_\omega - \ln\beta_\zeta \\ &+ \boldsymbol{\gamma}\beta_\gamma + \boldsymbol{\omega}\beta_\omega + \boldsymbol{\zeta}\beta_\zeta \\ \mathbf{F}_\pi &\approx -\sum\nolimits_\tau \mathbf{s}_{\pi\tau}\cdot\left(\ln\mathbf{A}^\zeta\cdot o_\tau + \ln\mathbf{B}^{\boldsymbol{\omega}}_\pi\mathbf{s}_{\pi\tau-1} - \ln\mathbf{Z}(\boldsymbol{\zeta})\cdot o_\tau - \ln\mathbf{Z}(\boldsymbol{\omega})\mathbf{s}_{\pi\tau-1}\right) \end{aligned} \tag{A19}$$

To simplify the maths, we have approximated the free energy with that we would have obtained using a mean-field approximation for states over time. This is subtly different to that used for the marginal message passing described in Chapter 2. In Equation A19, **Z** represent partition functions (i.e. normalising constants) given by:

$$\begin{aligned} \mathbf{Z}(\boldsymbol{\zeta})_j &= \sum\nolimits_i \mathbf{A}^{\boldsymbol{\zeta}}_{ij} \\ \mathbf{Z}(\boldsymbol{\omega})_j &= \sum\nolimits_i \mathbf{B}^{\boldsymbol{\omega}}_{\pi ij} \\ \mathbf{Z}(\boldsymbol{\gamma}) &= \sum\nolimits_\pi \exp(-\boldsymbol{\gamma}\cdot\mathbf{G}_\pi) \\ &\Rightarrow \\ \partial_\zeta \ln\mathbf{Z}(\boldsymbol{\zeta})\mathbf{s}_\tau &= \mathbf{o}^\zeta_\tau\cdot\ln\mathbf{A} \\ \partial_{\boldsymbol{\omega}} \ln\mathbf{Z}(\boldsymbol{\omega})\mathbf{s}_{\pi\tau-1} &= \mathbf{s}^{\boldsymbol{\omega}}_{\pi\tau}\cdot\ln\mathbf{B}_\pi \\ \partial_\gamma \ln\mathbf{Z}(\boldsymbol{\gamma}) &= -\boldsymbol{\pi}_0\cdot\mathbf{G} \\ \mathbf{o}^\zeta_\tau &\triangleq \boldsymbol{\pi}\cdot\left(\tfrac{\mathbf{A}^\zeta}{\mathbf{Z}(\boldsymbol{\zeta})}\mathbf{s}_{\pi\tau}\right) \\ \mathbf{s}^{\boldsymbol{\omega}}_{\pi\tau} &\triangleq \tfrac{\mathbf{B}^{\boldsymbol{\omega}}_\pi}{\mathbf{Z}(\boldsymbol{\omega})}\mathbf{s}_{\pi\tau-1} \\ \boldsymbol{\pi}_0 &\triangleq \sigma(-\boldsymbol{\gamma}\cdot\mathbf{G}) \end{aligned} \tag{A20}$$

Taking the partial derivative with respect to the expected precisions gives:

$$\begin{Bmatrix} \partial_{\zeta} F \\ \partial_{\boldsymbol{\omega}} F \\ \partial_{\gamma} F \end{Bmatrix} = 0 \Leftrightarrow \begin{Bmatrix} \boldsymbol{\beta}_{\zeta} \\ \boldsymbol{\beta}_{\omega} \\ \boldsymbol{\beta}_{\gamma} \end{Bmatrix} = \begin{Bmatrix} \sum_{\tau} (\mathbf{o}_{\tau}^{\zeta} - o_{\tau}) \cdot \ln \mathbf{A} + \beta_{\zeta} \\ \sum_{\tau} \boldsymbol{\pi} \cdot (\mathbf{s}_{\pi\tau}^{\boldsymbol{\omega}} - \mathbf{s}_{\pi\tau}) \cdot \ln \mathbf{B}_{\pi} \mathbf{s}_{\pi\tau-1} + \beta_{\omega} \\ (\boldsymbol{\pi} - \boldsymbol{\pi}_0) \cdot \mathbf{G} + \beta_{\gamma} \end{Bmatrix} \quad \text{(A21}$$

Expressing these updates as biologically plausible gradient descents, the resulting equations are:

$$\begin{Bmatrix} \dot{\boldsymbol{\beta}}_{\zeta} \\ \dot{\boldsymbol{\beta}}_{\omega} \\ \dot{\boldsymbol{\beta}}_{\gamma} \end{Bmatrix} = \begin{Bmatrix} \sum_{\tau} (\mathbf{o}_{\tau}^{\zeta} - o_{\tau}) \cdot \ln \mathbf{A} + \beta_{\zeta} - \boldsymbol{\beta}_{\zeta} \\ \sum_{\tau} \boldsymbol{\pi} \cdot (\mathbf{s}_{\pi\tau}^{\boldsymbol{\omega}} - \mathbf{s}_{\pi\tau}) \cdot \ln \mathbf{B}_{\pi} \mathbf{s}_{\pi\tau-1} + \beta_{\omega} - \boldsymbol{\beta}_{\omega} \\ (\boldsymbol{\pi} - \boldsymbol{\pi}_0) \cdot \mathbf{G} + \beta_{\gamma} - \boldsymbol{\beta}_{\gamma} \end{Bmatrix} \quad \text{(A22}$$

Note that the dimensionality implies a vector of precisions for **A**, where each state (column of **A**) is associated with its own precision parameter.

A.5 – Novelty

This appendix derives the form used to calculate the novelty (i.e. expected information gain) associated with the outcomes expected under a given policy. The KL-Divergence between two Dirichlet distributions (before and after experiencing a state-outcome pair) is as follows:

$$P(A) = Dir(a)$$
$$P(A \mid o, s) = Dir(\mathbf{a})$$

$$\begin{aligned} & D_{KL}[P(A \mid o, s) \,\|\, P(A)] \\ & = \sum_{j} \Bigg(\ln \Gamma(\textstyle\sum_{k} \mathbf{a}_{kj}) - \sum_{k} \ln \Gamma(\mathbf{a}_{kj}) - \ln \Gamma(\textstyle\sum_{k} a_{kj}) + \\ & \qquad \sum_{k} \ln \Gamma(a_{kj}) + \sum_{k} (\mathbf{a}_{kj} - a_{kj})(\psi(\mathbf{a}_{kj}) - \psi(\textstyle\sum_{k} \mathbf{a}_{kj})) \Bigg) \end{aligned} \quad \text{(A23}$$

If a_{ij} is the Dirichlet parameter associated with the prior probability of observing i given state j, we add +1 to this on making this observation. This implies $\mathbf{a}_{ij}$ would be equal a_{ij} +1. We can compute the information gain (**W**) that would be associated with a given combination of states and outcomes[20]:

$$\begin{aligned}
\mathbf{W}_{ij} &= \underbrace{\left(\ln\Gamma(a_{ij}) - \ln\Gamma(a_{ij}+1)\right)}_{-\ln a_{ij}} + \underbrace{\left(\ln\Gamma(a_{0j}+1) - \ln\Gamma(a_{0j})\right)}_{+\ln a_{0j}} \\
&+\psi(a_{ij}+1) - \psi(a_{0j}+1) \\
&= \ln\frac{a_{0j}}{a_{ij}} + \frac{\partial_{a_{ij}}\Gamma(a_{ij}+1)}{\Gamma(a_{ij}+1)} - \frac{\partial_{a_{0j}}\Gamma(a_{0j}+1)}{\Gamma(a_{0j}+1)} \\
&= \ln\frac{a_{0j}}{a_{ij}} + \frac{\partial_{a_{ij}}(a_{ij}\Gamma(a_{ij}))}{a_{ij}\Gamma(a_{ij})} - \frac{\partial_{a_{0j}}(a_{0j}\Gamma(a_{0j}))}{a_{0j}\Gamma(a_{0j})} \\
&= \ln\frac{a_{0j}}{a_{ij}} + \frac{1}{a_{ij}} + \frac{\partial_{a_{ij}}\Gamma(a_{ij})}{\Gamma(a_{ij})} - \frac{1}{a_{0j}} - \frac{\partial_{a_{0j}}\Gamma(a_{0j})}{\Gamma(a_{0j})} \\
&= \ln\frac{a_{0j}}{a_{ij}} + \frac{1}{a_{ij}} - \frac{1}{a_{0j}} + \psi(a_{ij}) - \psi(a_{0j})
\end{aligned} \tag{A24}$$

We can then use an expansion of the digamma function:

$$\psi(x) \approx \ln x - \frac{1}{2x} - \ldots \tag{A25}$$

Giving the following simple expression for the information gain:

[20] This uses the following identities:
$x\Gamma(x) = \Gamma(x+1)$
$\psi(x)\Gamma(x) = \partial_x\Gamma(x)$

This also uses the shorthand:
$a_{0j} \triangleq \sum_k a_{kj}$

$$\mathbf{W}_{ij} \approx \frac{1}{2a_{ij}} - \frac{1}{2a_{0j}} \quad \text{(A26}$$

The expected information gain, under a given policy, is then:

$$E_{Q(o,s|\pi)}[D_{KL}[P(A|o,s) \,||\, P(A)]] = \mathbf{o}_{\pi} \cdot \mathbf{W}\mathbf{s}_{\pi} \quad \text{(A27}$$

References

Abutalebi J, Rosa PAD, Tettamanti M, Green DW, Cappa SF (2009) Bilingual aphasia and language control: A follow-up fMRI and intrinsic connectivity study Brain and Language 109:141-156 doi:https://doi.org/10.1016/j.bandl.2009.03.003

Adams RA, Bauer M, Pinotsis D, Friston KJ (2016) Dynamic causal modelling of eye movements during pursuit: Confirming precision-encoding in V1 using MEG Neuroimage 132:175-189

Adams RA, Huys QJ, Roiser JP (2015) Computational Psychiatry: towards a mathematically informed understanding of mental illness J Neurol Neurosurg Psychiatry:jnnp-2015-310737

Adams RA, Perrinet LU, Friston K (2012) Smooth Pursuit and Visual Occlusion: Active Inference and Oculomotor Control in Schizophrenia PLOS ONE 7:e47502 doi:10.1371/journal.pone.0047502

Adams RA, Shipp S, Friston KJ (2013a) Predictions not commands: active inference in the motor system Brain Structure & Function 218:611-643 doi:10.1007/s00429-012-0475-5

Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013b) The computational anatomy of psychosis Frontiers in psychiatry 4:47-47 doi:10.3389/fpsyt.2013.00047

Aghajanian GK, Marek GJ (1999) Serotonin, via 5-HT2A receptors, increases EPSCs in layer V pyramidal cells of prefrontal cortex by an asynchronous mode of glutamate release Brain Research 825:161-171 doi:https://doi.org/10.1016/S0006-8993(99)01224-X

Albano JE, Wurtz RH (1982) Deficits in eye position following ablation of monkey superior colliculus, pretectum, and posterior-medial thalamus Journal of Neurophysiology 48:318

Albert ML (1973) A simple test of visual neglect Neurology 23:658 doi:10.1212/wnl.23.6.658

Alexander GE, Crutcher MD (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing Trends in Neurosciences 13:266-271 doi:https://doi.org/10.1016/0166-2236(90)90107-L

Allen M, Fardo F, Dietz MJ, Hillebrandt H, Friston KJ, Rees G, Roepstorff A (2016) Anterior insula coordinates hierarchical processing of tactile mismatch responses NeuroImage 127:34-43 doi:https://doi.org/10.1016/j.neuroimage.2015.11.030

Allen M, Levy A, Parr T, Friston KJ (2019) In the Body's Eye: The Computational Anatomy of Interoceptive Inference bioRxiv:603928 doi:10.1101/603928

Allport DA (1968) PHENOMENAL SIMULTANEITY AND THE PERCEPTUAL MOMENT HYPOTHESIS British Journal of Psychology 59:395-406 doi:10.1111/j.2044-8295.1968.tb01154.x

Andersen RA, Essick GK, Siegel RM (1985) Encoding of spatial location by posterior parietal neurons Science 230:456
Anderson RW, Keller EL, Gandhi NJ, Das S (1998) Two-Dimensional Saccade-Related Population Activity in Superior Colliculus in Monkey Journal of Neurophysiology 80:798
Anderson TJ, MacAskill MR (2013) Eye movements in patients with neurodegenerative disorders Nat Rev Neurol 9:74-85
Andrade K et al. (2010) Visual neglect in posterior cortical atrophy BMC Neurology 10:68 doi:10.1186/1471-2377-10-68
Andreopoulos A, Tsotsos J (2013) A computational learning theory of active object recognition under uncertainty International journal of computer vision 101:95-142
Antonini M, Barlaud M, Mathieu P, Daubechies I (1992) Image coding using wavelet transform IEEE Transactions on image processing 1:205-220
Arnsten AFT (2011) Catecholamine Influences on Dorsolateral Prefrontal Cortical Networks Biological Psychiatry 69:e89-e99 doi:https://doi.org/10.1016/j.biopsych.2011.01.027
Arnsten AFT, Goldman-Rakic PS (1984) Selective prefrontal cortical projections to the region of the locus coeruleus and raphe nuclei in the rhesus monkey Brain Research 306:9-18 doi:https://doi.org/10.1016/0006-8993(84)90351-2
Arnsten AFT, Li B-M (2005) Neurobiology of Executive Functions: Catecholamine Influences on Prefrontal Cortical Functions Biological Psychiatry 57:1377-1384 doi:https://doi.org/10.1016/j.biopsych.2004.08.019
Attias H Planning by Probabilistic Inference. In: Proc. of the 9th Int. Workshop on Artificial Intelligence and Statistics, 2003.
Auclair L, Siéroff E, Kocer S (2008) A case of spatial neglect dysgraphia in Wilson's Disease Archives of Clinical Neuropsychology 23:47-62 doi:http://dx.doi.org/10.1016/j.acn.2007.08.011
Avery MC, Krichmar JL (2017) Neuromodulatory Systems and Their Interactions: A Review of Models, Theories, and Experiments Frontiers in neural circuits 11:108-108 doi:10.3389/fncir.2017.00108
Badre D (2008) Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes Trends in Cognitive Sciences 12:193-200 doi:https://doi.org/10.1016/j.tics.2008.02.004
Badre D, D'Esposito M (2009) Is the rostro-caudal axis of the frontal lobe hierarchical? Nature Reviews Neuroscience 10:659 doi:10.1038/nrn2667

https://www.nature.com/articles/nrn2667#supplementary-information
Baker R, Highstein SM (1978) Vestibular projections to medial rectus subdivision of oculomotor nucleus Journal of Neurophysiology 41:1629
Baltieri M, Buckley LC (2019) PID Control as a Process of Active Inference with Linear Generative Models Entropy 21 doi:10.3390/e21030257
Barak O, Tsodyks M, Romo R (2010) Neuronal Population Coding of Parametric Working Memory The Journal of Neuroscience 30:9424
Barlow H (1961) Possible principles underlying the transformations of sensory messages. In: Rosenblith W (ed) Sensory Communication. MIT Press, Cambridge, MA, pp 217-234
Barlow HB (1974) Inductive inference, coding, perception, and language Perception 3:123-134
Barrett LF, Simmons WK (2015) Interoceptive predictions in the brain Nat Rev Neurosci 16:419-429 doi:10.1038/nrn3950

http://www.nature.com/nrn/journal/v16/n7/abs/nrn3950.html#supplementary-information
Bartolomeo P (2014) Spatially biased decisions: toward a dynamic interactive model of visual neglect Cognitive Plasticity in Neurologic Disorders:299
Bartolomeo P, Chokron S (2002) Orienting of attention in left unilateral neglect Neuroscience & Biobehavioral Reviews 26:217-234

Bartolomeo P, Thiebaut de Schotten M, Chica AB (2012) Brain networks of visuospatial attention and their disruption in visual neglect Frontiers in Human Neuroscience 6:110 doi:10.3389/fnhum.2012.00110
Bartolomeo P, Thiebaut de Schotten M, Doricchi F (2007) Left Unilateral Neglect as a Disconnection Syndrome Cerebral Cortex 17:2479-2490 doi:10.1093/cercor/bhl181
Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding Neuron 76:695-711 doi:10.1016/j.neuron.2012.10.038
Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF (2003) Voxel-based lesion–symptom mapping Nature neuroscience 6:448-450
Bays PM, Singh-Curry V, Gorgoraptis N, Driver J, Husain M (2010) Integration of Goal- and Stimulus-Related Visual Signals Revealed by Damage to Human Parietal Cortex The Journal of Neuroscience 30:5968
Beal MJ (2003) Variational algorithms for approximate Bayesian inference. University of London United Kingdom,
Behrens TEJ et al. (2003) Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging Nat Neurosci 6:750-757
Berretta S, Bosco G, Giaquinta G, Smecca G, Perciavalle V (1993) Cerebellar influences on accessory oculomotor nuclei of the rat: A neuroanatomical, immunohistochemical, and electrophysiological study The Journal of Comparative Neurology 338:50-66 doi:10.1002/cne.903380105
Berridge CW, Waterhouse BD (2003) The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes Brain Research Reviews 42:33-84 doi:https://doi.org/10.1016/S0165-0173(03)00143-7
Berson DM, McIlwain JT (1983) Visual cortical inputs to deep layers of cat's superior colliculus Journal of Neurophysiology 50:1143
Biehl M, Guckelsberger C, Salge C, Smith SC, Polani D (2018) Expanding the Active Inference Landscape: More Intrinsic Motivations in the Perception-Action Loop Frontiers in neurorobotics 12:45-45 doi:10.3389/fnbot.2018.00045
Binder JR (2015) The Wernicke area: Modern evidence and a reinterpretation Neurology 85:2170-2175 doi:10.1212/WNL.0000000000002219
Bittencourt PR, Wade P, Smith AT, Richens A (1981) The relationship between peak velocity of saccadic eye movements and serum benzodiazepine concentration British journal of clinical pharmacology 12:523-533
Blatow M, Caputi A, Burnashev N, Monyer H, Rozov A (2003) Ca2+ Buffer Saturation Underlies Paired Pulse Facilitation in Calbindin-D28k-Containing Terminals Neuron 38:79-88 doi:http://dx.doi.org/10.1016/S0896-6273(03)00196-X
Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation J Mach Learn Res 3:993-1022
Blier P, El Mansari M (2007) The importance of serotonin and noradrenaline in anxiety International Journal of Psychiatry in Clinical Practice 11:16-23 doi:10.1080/13651500701388310
Boes AD, Prasad S, Liu H, Liu Q, Pascual-Leone A, Caviness VS, Fox MD (2015) Network localization of neurological symptoms from focal brain lesions Brain 138:3061-3075 doi:10.1093/brain/awv228
Botvinick M, Toussaint M (2012) Planning as inference Trends Cogn Sci 16:485-488
Bourgeois A et al. (2015) Inappropriate rightward saccades after right hemisphere damage: Oculomotor analysis and anatomical correlates Neuropsychologia 73:1-11 doi:https://doi.org/10.1016/j.neuropsychologia.2015.04.013
Bowers D, Heilman KM (1980) Pseudoneglect: Effects of hemispace on a tactile line bisection task Neuropsychologia 18:491-498 doi:http://dx.doi.org/10.1016/0028-3932(80)90151-7
Bozis A, Moschovakis AK (1998) Neural network simulations of the primate oculomotor system III. An one-dimensional, one-directional model of the superior colliculus Biological Cybernetics 79:215-230 doi:10.1007/s004220050472

Bridgeman B, Hendry D, Stark L (1975) Failure to detect displacement of the visual world during saccadic eye movements Vision Research 15:719-722 doi:http://dx.doi.org/10.1016/0042-6989(75)90290-4
Brockmann D, Geisel T (1999) Are human scanpaths Levy flights? IET Conference Proceedings:263-268
Brown H, Friston KJ (2012) Free-Energy and Illusions: The Cornsweet Effect Frontiers in Psychology 3:43 doi:10.3389/fpsyg.2012.00043
Brown TH, Zhao Y, Leung V (2009) Hebbian Plasticity A2 - Squire, Larry R. In: Encyclopedia of Neuroscience. Academic Press, Oxford, pp 1049-1056. doi:http://dx.doi.org/10.1016/B978-008045046-9.00796-8
Bruce CJ, Goldberg ME, Bushnell MC, Stanton GB (1985) Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements Journal of Neurophysiology 54:714-734
Bruce NDB, Tsotsos JK (2009) Saliency, attention, and visual search: An information theoretic approach Journal of Vision 9:5-5 doi:10.1167/9.3.5
Bruineberg J (2017) Active inference and the primacy of the 'I can'
Bruineberg J, Kiverstein J, Rietveld E (2016) The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective Synthese:1-28 doi:10.1007/s11229-016-1239-1
Bruineberg J, Kiverstein J, Rietveld E (2018) The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective Synthese 195:2417-2444 doi:10.1007/s11229-016-1239-1
Büchel C, Josephs O, Rees G, Turner R, Frith CD, Friston KJ (1998) The functional anatomy of attention to visual motion. A functional MRI study Brain 121:1281-1294
Buckley CL, Kim CS, McGregor S, Seth AK (2017) The free energy principle for action and perception: A mathematical review Journal of Mathematical Psychology 81:55-79 doi:https://doi.org/10.1016/j.jmp.2017.09.004
Bundesen C (1998) A computational theory of visual attention Philosophical Transactions of the Royal Society of London B: Biological Sciences 353:1271-1281
Burak Y, Fiete IR (2012) Fundamental limits on persistent activity in networks of noisy neurons Proceedings of the National Academy of Sciences of the United States of America 109:17645-17650 doi:10.1073/pnas.1117386109
Buschman T, Miller E (2010a) Shifting the Spotlight of Attention: Evidence for Discrete Computations in Cognition Frontiers in Human Neuroscience 4 doi:10.3389/fnhum.2010.00194
Buschman TJ, Miller EK (2007) Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices Science 315:1860
Buschman TJ, Miller EK (2009) Serial, Covert, Shifts of Attention during Visual Search are Reflected by the Frontal Eye Fields and Correlated with Population Oscillations Neuron 63:386-396 doi:10.1016/j.neuron.2009.06.020
Buschman TJ, Miller EK (2010b) Shifting the Spotlight of Attention: Evidence for Discrete Computations in Cognition Frontiers in Human Neuroscience 4:194 doi:10.3389/fnhum.2010.00194
Büttner-Ennever JA, Büttner U (1978) A cell group associated with vertical eye movements in the rostral mesencephalic reticular formation of the monkey Brain Research 151:31-47 doi:http://dx.doi.org/10.1016/0006-8993(78)90948-4
Büttner-Ennever JA, Büttner U (1988) Neuroanatomy of the oculomotor system. The reticular formation Rev Oculomot Res 2:119-176
Büttner-Ennever JA, Cohen B, Pause M, Fries W (1988) Raphe nucleus of the pons containing omnipause neurons of the oculomotor system in the monkey, and Its homologue in man The Journal of Comparative Neurology 267:307-321 doi:10.1002/cne.902670302
Büttner U, Büttner-Ennever JA (2006) Present concepts of oculomotor organization. In: Büttner-Ennever JA (ed) Progress in Brain Research, vol Volume 151. Elsevier, pp 1-42. doi:http://dx.doi.org/10.1016/S0079-6123(05)51001-X

Büttner U, Helmchen C, Brandt T (1999) Diagnostic criteria for central versus peripheral positioning nystagmus and vertigo: a review Acta oto-laryngologica 119:1-5
Buzsaki G (2006) Rhythms of the Brain. Oxford University Press,
Buzsáki G (2005) Theta rhythm of navigation: Link between path integration and landmark navigation, episodic and semantic memory Hippocampus 15:827-840 doi:10.1002/hipo.20113
Campbell N et al. (2009) The cognitive impact of anticholinergics: a clinical review Clinical interventions in aging 4:225-233
Candy JM, Perry RH, Perry EK, Irving D, Blessed G, Fairbairn AF, Tomlinson BE (1983) Pathological changes in the nucleus of meynert in Alzheimer's and Parkinson's diseases Journal of the Neurological Sciences 59:277-289 doi:https://doi.org/10.1016/0022-510X(83)90045-X
Cannon WB (1929) ORGANIZATION FOR PHYSIOLOGICAL HOMEOSTASIS Physiological Reviews 9:399-431 doi:10.1152/physrev.1929.9.3.399
Carrera E, Tononi G (2014) Diaschisis: past, present, future Brain 137:2408-2422 doi:10.1093/brain/awu101
Catani M, ffytche DH (2005) The rises and falls of disconnection syndromes Brain 128:2224-2239 doi:10.1093/brain/awh622
Catani M, Mesulam M (2008) The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state Cortex 44:953-961 doi:https://doi.org/10.1016/j.cortex.2008.04.002
Chadwick MJ, Hassabis D, Weiskopf N, Maguire EA (2010) Decoding Individual Episodic Memory Traces in the Human Hippocampus Current Biology 20:544-547 doi:https://doi.org/10.1016/j.cub.2010.01.053
Chan F, Armstrong IT, Pari G, Riopelle RJ, Munoz DP (2005) Deficits in saccadic eye-movement control in Parkinson's disease Neuropsychologia 43:784-796 doi:10.1016/j.neuropsychologia.2004.06.026
Chelazzi L, Miller EK, Duncan J, Desimone R (1993) A neural basis for visual search in inferior temporal cortex Nature 363:345-347
Chica AB, Bartolomeo P, Valero-Cabré A (2011) Dorsal and Ventral Parietal Contributions to Spatial Orienting in the Human Brain The Journal of Neuroscience 31:8143
Christoff K, Keramatian K, Gordon AM, Smith R, Mädler B (2009) Prefrontal organization of cognitive control according to levels of abstraction Brain Research 1286:94-105 doi:https://doi.org/10.1016/j.brainres.2009.05.096
Ciaraffa F, Castelli G, Parati EA, Bartolomeo P, Bizzi A (2013) Visual neglect as a disconnection syndrome? A confirmatory case report Neurocase 19:351-359
Cisek P, Kalaska JF (2005) Neural Correlates of Reaching Decisions in Dorsal Premotor Cortex: Specification of Multiple Direction Choices and Final Selection of Action Neuron 45:801-814 doi:http://dx.doi.org/10.1016/j.neuron.2005.01.027
Clark A (2017) A nice surprise? Predictive processing and the active pursuit of novelty Phenomenology and the Cognitive Sciences:1-14
Cocchi L et al. (2016) A hierarchy of timescales explains distinct effects of local inhibition of primary visual cortex and frontal eye fields 5 doi:10.7554/eLife.15252
Cohen B, Komatsuzaki A, Bender MB (1968) Electrooculographic syndrome in monkeys after pontine reticular formation lesions Archives of Neurology 18:78-92 doi:10.1001/archneur.1968.00470310092008
Collerton D, Perry E, McKeith I (2005) Why people see things that are not there: A novel Perception and Attention Deficit model for recurrent complex visual hallucinations Behavioral and Brain Sciences 28:737-757 doi:10.1017/S0140525X05000130
Committeri G et al. (2007) Neural bases of personal and extrapersonal neglect in humans Brain 130:431-441 doi:10.1093/brain/awl265
Compte A (2006) Computational and in vitro studies of persistent activity: Edging towards cellular and synaptic mechanisms of working memory Neuroscience 139:135-151 doi:10.1016/j.neuroscience.2005.06.011

Conant RC, Ashby WR (1970) Every good regulator of a system must be a model of that system International journal of systems science 1:89-97
Consonni G, Marin J-M (2007) Mean-field variational approximate Bayesian inference for latent variable models Computational Statistics & Data Analysis 52:790-798 doi:https://doi.org/10.1016/j.csda.2006.10.028
Cooper S, Daniel PM (1949) MUSCLE SPINDLES IN HUMAN EXTRINSIC EYE MUSCLES Brain 72:1-24 doi:10.1093/brain/72.1.1
Cooper S, Daniel PM, Whitteridge D (1951) Afferent impulses in the oculomotor nerve, from the extrinsic eye muscles The Journal of Physiology 113:463-474 doi:10.1113/jphysiol.1951.sp004588
Cooray GK, Sengupta B, Douglas P, Englund M, Wickstrom R, Friston K (2015) Characterising seizures in anti-NMDA-receptor encephalitis with dynamic causal modelling Neuroimage 118:508-519 doi:10.1016/j.neuroimage.2015.05.064
Corallo CE, Whitfield A, Wu A (2009) Anticholinergic syndrome following an unintentional overdose of scopolamine Therapeutics and clinical risk management 5:719-723
Corbetta M et al. (1998) A Common Network of Functional Areas for Attention and Eye Movements Neuron 21:761-773 doi:10.1016/S0896-6273(00)80593-0
Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex Nat Neurosci 3:292-297
Corbetta M, Kincade JM, Shulman GL (2002) Neural Systems for Visual Orienting and Their Relationships to Spatial Working Memory Journal of Cognitive Neuroscience 14:508-523 doi:10.1162/089892902317362029
Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain Nat Rev Neurosci 3:201-215
Corbetta M, Shulman GL (2011) SPATIAL NEGLECT AND ATTENTION NETWORKS Annual review of neuroscience 34:569-599 doi:10.1146/annurev-neuro-061010-113731
Corlett PR, Fletcher PC (2014) Computational psychiatry: a Rosetta Stone linking the brain to mental illness The Lancet Psychiatry 1:399-402
Cowan LS, Walker ID (2013) The Importance of Continuous and Discrete Elements in Continuum Robots International Journal of Advanced Robotic Systems 10:165 doi:10.5772/55270
Cox DR, Miller HD (1965) The theory of stochastic processes
Craighero L, Carta A, Fadiga L (2001) Peripheral oculomotor palsy affects orienting of visuospatial attention NeuroReport 12:3283-3286
Crick F, Koch C (1998) Constraints on cortical and thalamic projections: the no-strong-loops hypothesis Nature 391:245 doi:10.1038/34584
Daunizeau J, den Ouden HEM, Pessiglione M, Kiebel SJ, Stephan KE, Friston KJ (2010) Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making PLOS ONE 5:e15554 doi:10.1371/journal.pone.0015554
Dautan D, Huerta-Ocampo I, Witten IB, Deisseroth K, Bolam JP, Gerdjikov T, Mena-Segovia J (2014) A major external source of cholinergic innervation of the striatum and nucleus accumbens originates in the brainstem The Journal of neuroscience : the official journal of the Society for Neuroscience 34:4509-4518 doi:10.1523/JNEUROSCI.5071-13.2014
Dauwels J On variational message passing on factor graphs. In: Information Theory, 2007. ISIT 2007. IEEE International Symposium on, 2007. IEEE, pp 2546-2550
David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ (2006) Dynamic causal modeling of evoked responses in EEG and MEG NeuroImage 30:1255-1272
Daw ND, Doya K (2006) The computational neurobiology of learning and reward Curr Opin Neurobiol 16:199-204
Dayan P, Hinton GE, Neal RM, Zemel RS (1995) The Helmholtz machine Neural computation 7:889-904
Dayan P, Yu AJ ACh, uncertainty, and cortical inference. In: NIPS, 2001. pp 189-196

Dayan P, Yu AJ (2006) Phasic norepinephrine: A neural interrupt signal for unexpected events Network: Computation in Neural Systems 17:335-350 doi:10.1080/09548980601004024
De Ridder D, Vanneste S, Freeman W (2014) The Bayesian brain: Phantom percepts resolve sensory uncertainty Neuroscience & Biobehavioral Reviews 44:4-15 doi:https://doi.org/10.1016/j.neubiorev.2012.04.001
de Visser SJ, van der Post J, Pieters MS, Cohen AF, van Gerven JM (2001) Biomarkers for the effects of antipsychotic drugs in healthy volunteers British journal of clinical pharmacology 51:119-132 doi:10.1111/j.1365-2125.2001.01308.x
de Visser SJ, van der Post JP, de Waal PP, Cornet F, Cohen AF, van Gerven JMA (2003) Biomarkers for the effects of benzodiazepines in healthy volunteers British journal of clinical pharmacology 55:39-50 doi:10.1046/j.1365-2125.2002.t01-10-01714.x
de Vries B, Friston KJ (2017) A Factor Graph Description of Deep Temporal Active Inference Frontiers in Computational Neuroscience 11 doi:10.3389/fncom.2017.00095
Demirdjian D, Taycher L, Shakhnarovich G, Grauman K, Darrell T Avoiding the "streetlight effect": tracking by exploring likelihood modes. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 17-21 Oct. 2005 2005. pp 357-364 Vol. 351. doi:10.1109/ICCV.2005.41
Deng P-Y, Klyachko VA (2011) The diverse functions of short-term plasticity components in synaptic computations Communicative & Integrative Biology 4:543-548 doi:10.4161/cib.4.5.15870
Denzler J, Brown CM (2002) Information theoretic sensor data selection for active object recognition and state estimation IEEE Transactions on Pattern Analysis and Machine Intelligence 24:145-157 doi:10.1109/34.982896
Desimone R, Schein SJ, Moran J, Ungerleider LG (1985) Contour, color and shape analysis beyond the striate cortex Vision Research 25:441-452 doi:10.1016/0042-6989(85)90069-0
Deubel H, Schneider WX (1996) Saccade target selection and object recognition: Evidence for a common attentional mechanism Vision Research 36:1827-1837 doi:http://dx.doi.org/10.1016/0042-6989(95)00294-4
Di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: a neurophysiological study Experimental brain research 91:176-180
Di Russo F, Aprile T, Spitoni G, Spinelli D (2007) Impaired visual processing of contralesional stimuli in neglect patients: a visual-evoked potential study Brain 131:842-854 doi:10.1093/brain/awm281
Di Stefano F et al. (2013) Transient unilateral spatial neglect during aura in a woman with sporadic hemiplegic migraine Cephalalgia 33:1194-1197 doi:10.1177/0333102413487446
Dietz MJ, Friston KJ, Mattingley JB, Roepstorff A, Garrido MI (2014) Effective connectivity reveals right-hemisphere dominance in audiospatial perception: implications for models of spatial neglect Journal of Neuroscience 34:5003-5011
Disney AA, Aoki C, Hawken MJ (2007) Gain Modulation by Nicotine in Macaque V1 Neuron 56:701-713 doi:10.1016/j.neuron.2007.09.034
Donaldson IM (2000) The functions of the proprioceptors of the eye muscles Philosophical Transactions of the Royal Society B: Biological Sciences 355:1685-1754
Doricchi F, Tomaiuolo F (2003) The anatomy of neglect without hemianopia: a key role for parietal–frontal disconnection? Neuroreport 14:2239-2243
Dosenbach NUF et al. (2006) A Core System for the Implementation of Task Sets Neuron 50:799-812 doi:http://dx.doi.org/10.1016/j.neuron.2006.04.031
Doya K (2002) Metalearning and neuromodulation Neural Netw 15:495-506
Doya K (2007) Bayesian brain: Probabilistic approaches to neural coding. MIT press,
Doya K (2008) Modulators of decision making Nat Neurosci 11:410-416
Drewes J, VanRullen R (2011) This Is the Rhythm of Your Eyes: The Phase of Ongoing Electroencephalogram Oscillations Modulates Saccadic Reaction Time The Journal of Neuroscience 31:4698

Driver J, Baylis GC, Goodrich SJ, Rafal RD (1994) Axis-based neglect of visual shapes Neuropsychologia 32:1353-1356 doi:http://dx.doi.org/10.1016/0028-3932(94)00068-9
Driver J, Pouget A (2000) Object-Centered Visual Neglect, or Relative Egocentric Neglect? Journal of Cognitive Neuroscience 12:542-545 doi:10.1162/089892900562192
Dronkers NF, Baldo JV (2009) Language: Aphasia A2 - Squire, Larry R. In: Encyclopedia of Neuroscience. Academic Press, Oxford, pp 343-348. doi:https://doi.org/10.1016/B978-008045046-9.01876-3
Dugué L, Marque P, VanRullen R (2011) The Phase of Ongoing Oscillations Mediates the Causal Relation between Brain Excitation and Visual Perception The Journal of Neuroscience 31:11889
Duncan J (2001) An adaptive coding model of neural function in prefrontal cortex Nat Rev Neurosci 2:820-829
Duncan J, Ward R, Shapiro K (1994) Direct measurement of attentional dwell time in human vision Nature 369:313-315
Eckenstein FP, Baughman RW, Quinn J (1988) An anatomical study of cholinergic innervation in rat cerebral cortex Neuroscience 25:457-474 doi:https://doi.org/10.1016/0306-4522(88)90251-5
Edwards MJ, Adams RA, Brown H, Pareés I, Friston KJ (2012) A Bayesian account of 'hysteria' Brain 135:3495-3512 doi:10.1093/brain/aws129
Elliott MC, Tanaka PM, Schwark RW, Andrade R (2018) Serotonin Differentially Regulates L5 Pyramidal Cell Classes of the Medial Prefrontal Cortex in Rats and Mice eNeuro 5:ENEURO.0305-0317.2018 doi:10.1523/ENEURO.0305-17.2018
Ellison A, Schindler I, Pattison LL, Milner AD (2004) An exploration of the role of the superior temporal gyrus in visual search and spatial perception using TMS Brain 127:2307-2315 doi:10.1093/brain/awh244
Epstein R, Harris A, Stanley D, Kanwisher N (1999) The Parahippocampal Place Area: Recognition, Navigation, or Encoding? Neuron 23:115-125 doi:https://doi.org/10.1016/S0896-6273(00)80758-8
Ergenoglu T, Demiralp T, Bayraktaroglu Z, Ergen M, Beydagi H, Uresin Y (2004) Alpha rhythm of the EEG modulates visual detection performance in humans Cognitive Brain Research 20:376-383 doi:https://doi.org/10.1016/j.cogbrainres.2004.03.009
Etienne P, Robitaille Y, Wood P, Gauthier S, Nair NPV, Quirion R (1986) Nucleus basalis neuronal loss, neuritic plaques and choline acetyltransferase activity in advanced Alzheimer's disease Neuroscience 19:1279-1291 doi:https://doi.org/10.1016/0306-4522(86)90142-9
Faisal AA, Selen LPJ, Wolpert DM (2008) Noise in the nervous system Nature Reviews Neuroscience 9:292-303 doi:10.1038/nrn2258
Feldman AG (2009) New insights into action-perception coupling Exp Brain Res 194:39-58
Feldman AG, Levin MF (2009) The Equilibrium-Point Hypothesis – Past, Present and Future. In: Sternad D (ed) Progress in Motor Control: A Multidisciplinary Perspective. Springer US, Boston, MA, pp 699-726. doi:10.1007/978-0-387-77064-2_38
Feldman H, Friston K (2010) Attention, Uncertainty, and Free-Energy Frontiers in Human Neuroscience 4 doi:10.3389/fnhum.2010.00215
Felleman DJ, Van DE (1991) Distributed hierarchical processing in the primate cerebral cortex Cerebral cortex (New York, NY: 1991) 1:1-47
Felleman DJ, Van Essen DC (1991) Distributed Hierarchical Processing in the Primate Cerebral Cortex Cerebral Cortex 1:1-47
Ferber S, Karnath H-O (2001) How to assess spatial neglect-line bisection or cancellation tasks? Journal of clinical and experimental neuropsychology 23:599-607
ffytche DH, Howard RJ (1999) The perceptual consequences of visual loss: `positive' pathologies of vision Brain 122:1247-1260 doi:10.1093/brain/122.7.1247
Fibiger HC, Mason ST (1978) The effects of dorsal bundle injections of 6-hydroxydopamine on avoidance responding in rats British Journal of Pharmacology 64:601-605

FitzGerald T, Dolan R, Friston K (2014) Model averaging, optimal inference, and habit formation Front Hum Neurosci:doi: 10.3389/fnhum.2014.00457
Fletcher PC, Frith CD (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia Nat Rev Neurosci 10:48-58
Forney Jr GD, Vontobel PO (2011) Partition functions of normal factor graphs arXiv preprint arXiv:11020316
Fornito A, Zalesky A, Breakspear M (2015) The connectomics of brain disorders Nat Rev Neurosci 16:159-172 doi:10.1038/nrn3901
Fotopoulou A (2012) Illusions and delusions in anosognosia for hemiplegia: from motor predictions to prior beliefs Brain 135:1344-1346 doi:10.1093/brain/aws094
Fotopoulou A (2014) Time to get rid of the 'Modular' in neuropsychology: A unified theory of anosognosia as aberrant predictive coding Journal of Neuropsychology 8:1-19 doi:10.1111/jnp.12010
Fotopoulou A, Pernigo S, Maeda R, Rudd A, Kopelman MA (2010) Implicit awareness in anosognosia for hemiplegia: unconscious interference without conscious re-representation Brain 133:3564-3577 doi:10.1093/brain/awq233
Fotopoulou A, Tsakiris M, Haggard P, Vagopoulou A, Rudd A, Kopelman M (2008) The role of motor intention in motor awareness: an experimental study on anosognosia for hemiplegia Brain 131:3432-3442 doi:10.1093/brain/awn225
Fourneret P, Jeannerod M (1998) Limited conscious monitoring of motor performance in normal subjects Neuropsychologia 36:1133-1140 doi:https://doi.org/10.1016/S0028-3932(98)00006-2
Frank MJ (2005) Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational Account of Cognitive Deficits in Medicated and Nonmedicated Parkinsonism Journal of Cognitive Neuroscience 17:51-72 doi:10.1162/0898929052880093
Frank MJ, Loughry B, O'Reilly RC (2001) Interactions between frontal cortex and basal ganglia in working memory: A computational model Cognitive, Affective, & Behavioral Neuroscience 1:137-160 doi:10.3758/CABN.1.2.137
Frank MJ, Seeberger LC, O'Reilly RC (2004) By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism Science 306:1940-1943 doi:10.1126/science.1102941
Freeze BS, Kravitz AV, Hammack N, Berke JD, Kreitzer AC (2013) Control of Basal Ganglia Output by Direct and Indirect Pathway Projection Neurons The Journal of Neuroscience 33:18531-18539 doi:10.1523/JNEUROSCI.1278-13.2013
Freund TF, Powell JF, Smith AD (1984) Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines Neuroscience 13:1189-1215 doi:https://doi.org/10.1016/0306-4522(84)90294-X
Fries W (1984) Cortical projections to the superior colliculus in the macaque monkey: A retrograde study using horseradish peroxidase The Journal of Comparative Neurology 230:55-76 doi:10.1002/cne.902300106
Fries W (1985) Inputs from motor and premotor cortex to the superior colliculus of the macaque monkey Behavioural Brain Research 18:95-105 doi:10.1016/0166-4328(85)90066-X
Friston K (2008) Hierarchical Models in the Brain PLOS Computational Biology 4:e1000211 doi:10.1371/journal.pcbi.1000211
Friston K (2010) The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11:127-138
Friston K (2013) Life as we know it Journal of The Royal Society Interface 10
Friston K (2019) A free energy principle for a particular physics arXiv preprint arXiv:190610184
Friston K, Adams RA, Perrinet L, Breakspear M (2012a) Perceptions as Hypotheses: Saccades as Experiments Frontiers in Psychology 3:151 doi:10.3389/fpsyg.2012.00151
Friston K, Ao P (2012) Free Energy, Value, and Attractors Computational and Mathematical Methods in Medicine 2012:27 doi:10.1155/2012/937860

Friston K, Brown HR, Siemerkus J, Stephan KE (2016a) The dysconnection hypothesis (2016) Schizophrenia Research 176:83-94 doi:https://doi.org/10.1016/j.schres.2016.07.014
Friston K, Buzsaki G (2016) The Functional Anatomy of Time: What and When in the Brain Trends Cogn Sci doi:10.1016/j.tics.2016.05.001
Friston K, Buzsáki G (2016) The Functional Anatomy of Time: What and When in the Brain Trends in Cognitive Sciences 20:500-511 doi:10.1016/j.tics.2016.05.001
Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, O'Doherty J, Pezzulo G (2016b) Active inference and learning Neuroscience & Biobehavioral Reviews 68:862-879 doi:http://dx.doi.org/10.1016/j.neubiorev.2016.06.022
Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G (2017a) Active Inference: A Process Theory Neural Comput 29:1-49 doi:10.1162/NECO_a_00912
Friston K et al. (2008) Multiple sparse priors for the M/EEG inverse problem NeuroImage 39:1104-1120 doi:https://doi.org/10.1016/j.neuroimage.2007.09.048
Friston K, Herreros I (2016) Active Inference and Learning in the Cerebellum Neural Computation 28:1812-1839 doi:10.1162/NECO_a_00863
Friston K, Kiebel S (2009) Predictive coding under the free-energy principle Philosophical Transactions of the Royal Society B: Biological Sciences 364:1211
Friston K, Kilner J, Harrison L (2006) A free energy principle for the brain Journal of Physiology-Paris 100:70-87 doi:10.1016/j.jphysparis.2006.10.001
Friston K, Levin M, Sengupta B, Pezzulo G (2015a) Knowing one's place: a free-energy approach to pattern regulation Journal of The Royal Society Interface 12:20141383 doi:10.1098/rsif.2014.1383
Friston K, Mattout J, Kilner J (2011) Action understanding and active inference Biological cybernetics 104:137-160 doi:10.1007/s00422-011-0424-z
Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W (2007) Variational free energy and the Laplace approximation Neuroimage 34:220-234
Friston K, Parr T, Zeidman P (2018) Bayesian model reduction arXiv preprint arXiv:180507092
Friston K, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G (2015b) Active inference and epistemic value Cognitive Neuroscience 6:187-214 doi:10.1080/17588928.2015.1020053
Friston K, Samothrakis S, Montague R (2012b) Active inference and agency: optimal control without cost functions Biological Cybernetics 106:523-541 doi:10.1007/s00422-012-0512-8
Friston K, Schwartenbeck P, Fitzgerald T, Moutoussis M, Behrens T, Dolan R (2013) The anatomy of choice: active inference and agency Frontiers in Human Neuroscience 7 doi:10.3389/fnhum.2013.00598
Friston K, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ (2014) The anatomy of choice: dopamine and decision-making Philosophical Transactions of the Royal Society B: Biological Sciences 369:20130481 doi:10.1098/rstb.2013.0481
Friston K, Stephan K, Li B, Daunizeau J (2010a) Generalised filtering Mathematical Problems in Engineering 2010
Friston K, Thornton C, Clark A (2012c) Free-Energy Minimization and the Dark-Room Problem Frontiers in Psychology 3 doi:10.3389/fpsyg.2012.00130
Friston KJ, Daunizeau J, Kilner J, Kiebel SJ (2010b) Action and behavior: a free-energy formulation Biological Cybernetics 102:227-260 doi:10.1007/s00422-010-0364-z
Friston KJ, Frith CD (2015) Active inference, communication and hermeneutics() Cortex; a Journal Devoted to the Study of the Nervous System and Behavior 68:129-143 doi:10.1016/j.cortex.2015.03.025
Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling NeuroImage 19:1273-1302 doi:https://doi.org/10.1016/S1053-8119(03)00202-7
Friston KJ, Lin M, Frith CD, Pezzulo G, Hobson JA, Ondobaka S (2017b) Active inference, curiosity and insight Neural Computation

Friston KJ et al. (2016c) Bayesian model reduction and empirical Bayes for group (DCM) studies NeuroImage 128:413-431 doi:https://doi.org/10.1016/j.neuroimage.2015.11.015
Friston KJ, Parr T, Vries Bd (2017c) The graphical brain: belief propagation and active inference Network Neuroscience 0:1-78 doi:10.1162/NETN_a_00018
Friston KJ, Penny W (2003) Posterior probability maps and SPMs NeuroImage 19:1240-1249 doi:https://doi.org/10.1016/S1053-8119(03)00144-7
Friston KJ, Preller KH, Mathys C, Cagnan H, Heinzle J, Razi A, Zeidman P (2017d) Dynamic causal modelling revisited NeuroImage doi:https://doi.org/10.1016/j.neuroimage.2017.02.045
Friston KJ, Redish AD, Gordon JA (2017e) Computational Nosology and Precision Psychiatry Computational Psychiatry
Friston KJ, Rosch R, Parr T, Price C, Bowman H (2017f) Deep temporal models and active inference Neuroscience & Biobehavioral Reviews 77:388-402 doi:10.1016/j.neubiorev.2017.04.009
Frith CD, Blakemore SJ, Wolpert DM (2000) Abnormalities in the awareness and control of action Philosophical Transactions of the Royal Society B: Biological Sciences 355:1771-1788
Frith CD, Frith U (1999) Interacting minds—a biological basis Science 286:1692-1695 doi:10.1126/science.286.5445.1692
Fruhmann Berger M, Johannsen L, Karnath H-O (2008) Time course of eye and head deviation in spatial neglect Neuropsychology 22:697-702 doi:10.1037/a0013351
Fullerton KJ, McSherry D, Stout RW (1986) Albert's Test: a Neglected Test of Perceptual Neglect The Lancet 327:430-432 doi:http://dx.doi.org/10.1016/S0140-6736(86)92381-0
Funahashi S (2015) Functions of delay-period activity in the prefrontal cortex and mnemonic scotomas revisited Frontiers in Systems Neuroscience 9 doi:10.3389/fnsys.2015.00002
Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex Journal of Neurophysiology 61:331
Galea JM, Bestmann S, Beigi M, Jahanshahi M, Rothwell JC (2012) Action Reprogramming in Parkinson's Disease: Response to Prediction Error Is Modulated by Levels of Dopamine The Journal of Neuroscience 32:542
Gallant J, Braun J, Van Essen D (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex Science 259:100-103 doi:10.1126/science.8418487
Gandhi NJ, Keller EL (1997) Spatial Distribution and Discharge Characteristics of Superior Colliculus Neurons Antidromically Activated From the Omnipause Region in Monkey Journal of Neurophysiology 78:2221
Gaymard B, Lynch J, Ploner CJ, Condy C, Rivaud-Péchoux S (2003) The parieto-collicular pathway: anatomical location and contribution to saccade generation European Journal of Neuroscience 17:1518-1526 doi:10.1046/j.1460-9568.2003.02570.x
Geisler WS, Kersten D (2002) Illusions, perception and Bayes 5:508 doi:10.1038/nn0602-508
Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex The Journal of Neuroscience 2:1527
Geschwind N (1965a) DISCONNEXION SYNDROMES IN ANIMALS AND MAN1 Brain 88:237-237 doi:10.1093/brain/88.2.237
Geschwind N (1965b) Disconnexion syndromes in animals and man. II Brain: a journal of neurology 88:585
Ghahramani Z (2015) Probabilistic machine learning and artificial intelligence Nature 521:452-459 doi:10.1038/nature14541
Gibson JJ (1966) The senses considered as perceptual systems

Gil Z, Connors BW, Amitai Y (1997) Differential Regulation of Neocortical Synapses by Neuromodulators and Activity Neuron 19:679-686 doi:https://doi.org/10.1016/S0896-6273(00)80380-3
Gilad R, Sadeh M, Boaz M, Lampl Y (2006) Visual spatial neglect in multiple sclerosis Cortex 42:1138-1142
Glickstein M, Berlucchi G (2008) Classical disconnection studies of the corpus callosum Cortex 44:914-927 doi:https://doi.org/10.1016/j.cortex.2008.04.001
Goldenberg G (2003) Apraxia and Beyond: Life and Work of Hugo Liepmann Cortex 39:509-524 doi:https://doi.org/10.1016/S0010-9452(08)70261-2
Goldman-Rakic PS (1995) Cellular basis of working memory Neuron 14:477-485 doi:http://dx.doi.org/10.1016/0896-6273(95)90304-6
Goldman-Rakic PS, Castner SA, Svensson TH, Siever LJ, Williams GV (2004) Targeting the dopamine D1 receptor in schizophrenia: insights for cognitive dysfunction Psychopharmacology 174:3-16 doi:10.1007/s00213-004-1793-y
Goldman-Rakic PS (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory Comprehensive Physiology
Goodale MA, Milner AD (1992) Separate visual pathways for perception and action Trends in Neurosciences 15:20-25 doi:http://dx.doi.org/10.1016/0166-2236(92)90344-8
Gören MZ, Cabadak H (2014) Noradrenaline and Post-traumatic Stress Disorder. In: Martin CR, Preedy VR, Patel VB (eds) Comprehensive Guide to Post-Traumatic Stress Disorder. Springer International Publishing, Cham, pp 1-16. doi:10.1007/978-3-319-08613-2_26-1
Grandoso L, Pineda J, Ugedo L (2004) Comparative study of the effects of desipramine and reboxetine on locus coeruleus neurons in rat brain slices Neuropharmacology 46:815-823 doi:https://doi.org/10.1016/j.neuropharm.2003.11.033
Graybiel AM, Grafton ST (2015) The striatum: where skills and habits meet Cold Spring Harbor perspectives in biology 7:a021691
Grefkes C, Nowak DA, Eickhoff SB, Dafotakis M, Küst J, Karbe H, Fink GR (2008) Cortical connectivity after subcortical stroke assessed with functional magnetic resonance imaging Annals of Neurology 63:236-246 doi:10.1002/ana.21228
Gregory RL (1970) The intelligent eye
Gregory RL (1980) Perceptions as Hypotheses Philosophical Transactions of the Royal Society of London B, Biological Sciences 290:181
Griffin CE, 3rd, Kaye AM, Bueno FR, Kaye AD (2013) Benzodiazepine pharmacology and central nervous system-mediated effects The Ochsner journal 13:214-223
Grimsen C, Hildebrandt H, Fahle M (2008) Dissociation of egocentric and allocentric coding of space in visual search after right middle cerebral artery stroke Neuropsychologia 46:902-914 doi:http://dx.doi.org/10.1016/j.neuropsychologia.2007.11.028
Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the Macaque Journal of Neurophysiology 35:96-111
Guillery RW, Sherman SM (2002) Thalamic Relay Functions and Their Role in Corticocortical Communication Neuron 33:163-175 doi:10.1016/S0896-6273(01)00582-7
Gurney K, Prescott TJ, Redgrave P (2001) A computational model of action selection in the basal ganglia. I. A new functional anatomy Biological Cybernetics 84:401-410 doi:10.1007/PL00007984
Haber SN (2003) The primate basal ganglia: parallel and integrative networks Journal of Chemical Neuroanatomy 26:317-330 doi:10.1016/j.jchemneu.2003.10.003
Haber SN, Calzavara R (2009) The cortico-basal ganglia integrative network: The role of the thalamus Brain Research Bulletin 78:69-74 doi:https://doi.org/10.1016/j.brainresbull.2008.09.013
Haeusler S, Maass W (2007) A Statistical Analysis of Information-Processing Properties of Lamina-Specific Cortical Microcircuit Models Cerebral Cortex 17:149-162 doi:10.1093/cercor/bhj132

Halassa MM, Kastner S (2017) Thalamic functions in distributed cognitive control Nature Neuroscience 20:1669-1679 doi:10.1038/s41593-017-0020-1
Hall JC (1999) GABAergic inhibition shapes frequency tuning and modifies response properties in the auditory midbrain of the leopard frog Journal of Comparative Physiology A 185:479-491 doi:10.1007/s003590050409
Halliday G, Cullen K, Harding A (1993) Neuropathological correlates of memory dysfunction in the Wernicke-Korsakoff syndrome Alcohol and alcoholism (Oxford, Oxfordshire) Supplement 2:245-251
Halligan PW, Marshall JC (1998) Neglect of Awareness Consciousness and Cognition 7:356-380 doi:http://dx.doi.org/10.1006/ccog.1998.0362
Hanes DP, Wurtz RH (2001) Interaction of the Frontal Eye Field and Superior Colliculus for Saccade Generation Journal of Neurophysiology 85:804
Hanslmayr S, Volberg G, Wimber M, Dalal Sarang S, Greenlee Mark W (2013) Prestimulus Oscillatory Phase at 7 Hz Gates Cortical Information Flow and Visual Perception Current Biology 23:2273-2278 doi:https://doi.org/10.1016/j.cub.2013.09.020
Happé FGE (1996) Studying Weak Central Coherence at Low Levels: Children with Autism do not Succumb to Visual Illusions. A Research Note Journal of Child Psychology and Psychiatry 37:873-877 doi:10.1111/j.1469-7610.1996.tb01483.x
Harding AJ, Broe GA, Halliday GM (2002) Visual hallucinations in Lewy body disease relate to Lewy bodies in the temporal lobe Brain 125:391-403 doi:10.1093/brain/awf033
Hassabis D, Kumaran D, Vann SD, Maguire EA (2007) Patients with hippocampal amnesia cannot imagine new experiences Proceedings of the National Academy of Sciences 104:1726-1731 doi:10.1073/pnas.0610561104
Hasson U, Chen J, Honey CJ (2015) Hierarchical process memory: memory as an integral component of information processing Trends in cognitive sciences 19:304-313 doi:10.1016/j.tics.2015.04.006
Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A Hierarchy of Temporal Receptive Windows in Human Cortex The Journal of neuroscience : the official journal of the Society for Neuroscience 28:2539-2550 doi:10.1523/JNEUROSCI.5487-07.2008
He BJ, Snyder AZ, Vincent JL, Epstein A, Shulman GL, Corbetta M (2007) Breakdown of Functional Connectivity in Frontoparietal Networks Underlies Behavioral Deficits in Spatial Neglect Neuron 53:905-918 doi:http://dx.doi.org/10.1016/j.neuron.2007.02.013
Hebb DO (1949) The first stage of perception: Growth of the assembly The Organization of Behavior:60-78
Heilman KM, Howell GJ (1980) Seizure-induced neglect Journal of Neurology, Neurosurgery, and Psychiatry 43:1035-1040
Heinke D, Humphreys GW (2003) Attention, spatial representation, and visual neglect: Simulating emergent attention and spatial memory in the selective attention for identification model (SAIM) Psychological Review 110:29-87 doi:10.1037/0033-295X.110.1.29
Heinke D, Humphreys GW (2005) Computational models of visual selective attention: A review Connectionist models in cognitive psychology 1:273-312
Heitz C et al. (2015) Neural correlates of visual hallucinations in dementia with Lewy bodies Alzheimer's Research & Therapy 7:6 doi:10.1186/s13195-014-0091-0
Hempel CM, Hartman KH, Wang XJ, Turrigiano GG, Nelson SB (2000) Multiple Forms of Short-Term Plasticity at Excitatory Synapses in Rat Medial Prefrontal Cortex Journal of Neurophysiology 83:3031
Henn V (1992) Pathophysiology of rapid eye movements in the horizontal, vertical and torsional directions Bailliere's clinical neurology 1:373-391
Herkenham M (1980) Laminar organization of thalamic projections to the rat neocortex Science 207:532
Hickok G (2012a) Computational neuroanatomy of speech production Nature reviews Neuroscience 13:135-145 doi:10.1038/nrn3158

Hickok G (2012b) The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model Journal of Communication Disorders 45:393-402 doi:https://doi.org/10.1016/j.jcomdis.2012.06.004
Hickok G, Poeppel D (2007) The cortical organization of speech processing Nat Rev Neurosci 8:393-402
Hikosaka O, Takikawa Y, Kawagoe R (2000) Role of the Basal Ganglia in the Control of Purposive Saccadic Eye Movements Physiological Reviews 80:953
Hikosaka O, Wurtz RH (1983) Visual and oculomotor functions of monkey substantia nigra pars reticulata. IV. Relation of substantia nigra to superior colliculus Journal of Neurophysiology 49:1285
Hikosaka O, Wurtz RH (1985) Modification of saccadic eye movements by GABA-related substances. I. Effect of muscimol and bicuculline in monkey superior colliculus Journal of Neurophysiology 53:266
Hikosaka O, Wurtz RH (1985b) Modification of saccadic eye movements by GABA-related substances. II. Effects of muscimol in monkey substantia nigra pars reticulata Journal of Neurophysiology 53:292
Hillis AE, Newhart M, Heidler J, Barker PB, Herskovits EH, Degaonkar M (2005) Anatomy of Spatial Attention: Insights from Perfusion Imaging and Hemispatial Neglect in Acute Stroke The Journal of Neuroscience 25:3161
Hillyard SA, Vogel EK, Luck SJ (1998) Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence Philosophical Transactions of the Royal Society B: Biological Sciences 353:1257-1270
Ho AK et al. (2003) A case of unilateral neglect in Huntington's disease Neurocase 9:261-273
Hoffman KL, Dragan MC, Leonard TK, Micheli C, Montefusco-Siegmund R, Valiante TA (2013) Saccades during visual exploration align hippocampal 3-8 Hz rhythms in human and non-human primates Frontiers in systems neuroscience 7:43-43 doi:10.3389/fnsys.2013.00043
Hohwy J (2007) The sense of self in the phenomenology of agency and perception Psyche 13:1-20
Hohwy J (2016) The Self-Evidencing Brain Noûs 50:259-285 doi:10.1111/nous.12062
Holzman PS, Levy DL (1977) Smooth pursuit eye movements and functional psychoses: A review Schizophrenia Bulletin 3:15
Honey CJ et al. (2012) Slow Cortical Dynamics and the Accumulation of Information over Long Timescales Neuron 76:423-434 doi:10.1016/j.neuron.2012.08.011
Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control Nat Neurosci 3:284-291
Howard D, Patterson K, Wise R, Brown WD, Friston K, Weiller C, Frackowiak R (1992) The cortical localization of the lexicons: positron emission tomography evidence Brain 115:1769-1782
Howes OD, Kapur S (2009) The Dopamine Hypothesis of Schizophrenia: Version III—The Final Common Pathway Schizophrenia Bulletin 35:549-562 doi:10.1093/schbul/sbp006
Hubel DH, Livingstone MS (1987) Segregation of form, color, and stereopsis in primate area 18 The Journal of Neuroscience 7:3378
Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex The Journal of Physiology 148:574-591 doi:10.1113/jphysiol.1959.sp006308
Hübener M, Bolz J (1988) Morphology of identified projection neurons in layer 5 of rat visual cortex Neuroscience Letters 94:76-81 doi:https://doi.org/10.1016/0304-3940(88)90273-X
Humphreys GW, Forde EM (2001) Hierarchies, similarity, and interactivity in object recognition:"Category-specific" neuropsychological deficits Behavioral and Brain Sciences 24:453-476
Husain M, Mannan S, Hodgson T, Wojciulik E, Driver J, Kennard C (2001) Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect Brain 124:941-952 doi:10.1093/brain/124.5.941

Husain M, Nachev P (2007) Space and the parietal cortex Trends Cogn Sci 11 doi:10.1016/j.tics.2006.10.011
Huys QJ, Maia TV, Frank MJ (2016) Computational psychiatry as a bridge from neuroscience to clinical applications Nature neuroscience 19:404-413
Hwa R (2004) Sample selection for statistical parsing Computational linguistics 30:253-276
Iglesias S, Tomiello S, Schneebeli M, Stephan KE (2017) Models of neuromodulation for computational psychiatry Wiley Interdisciplinary Reviews: Cognitive Science 8:e1420-n/a doi:10.1002/wcs.1420
Ijspeert AJ (2008) Central pattern generators for locomotion control in animals and robots: A review Neural Networks 21:642-653 doi:https://doi.org/10.1016/j.neunet.2008.03.014
Itti L, Baldi P (2006) Bayesian surprise attracts human attention Advances in neural information processing systems 18:547
Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention Vision Research 40:1489-1506 doi:10.1016/S0042-6989(99)00163-7
Itti L, Koch C (2001) Computational modelling of visual attention Nature Reviews Neuroscience 2:194-203 doi:10.1038/35058500
Jahanshahi M, Obeso I, Rothwell JC, Obeso JA (2015) A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition Nat Rev Neurosci 16:719-732 doi:10.1038/nrn4038
Javoy-Agid F et al. (1989) Distribution of monoaminergic, cholinergic, and GABAergic markers in the human cerebral cortex Neuroscience 29:251-259 doi:https://doi.org/10.1016/0306-4522(89)90055-9
Jenner P (1995) The rationale for the use of dopamine agonists in Parkinson's disease Neurology 45:S6-S12 doi:10.1212/WNL.45.3_Suppl_3.S6
Jewell G, McCourt ME (2000) Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks Neuropsychologia 38:93-110 doi:http://dx.doi.org/10.1016/S0028-3932(99)00045-7
Kaelbling LP, Littman ML, Cassandra AR (1998) Planning and acting in partially observable stochastic domains Artificial intelligence 101:99-134
Kalsbeek A, la Fleur S, Fliers E (2014) Circadian control of glucose metabolism Molecular Metabolism 3:372-383 doi:https://doi.org/10.1016/j.molmet.2014.03.002
Kanai R, Komura Y, Shipp S, Friston K (2015) Cerebral hierarchies: predictive processing, precision and the pulvinar Philosophical Transactions of the Royal Society B: Biological Sciences 370
Kapur S, Zipursky R, Jones C, Remington G, Houle S (2000) Relationship Between Dopamine D2 Occupancy, Clinical Response, and Side Effects: A Double-Blind PET Study of First-Episode Schizophrenia American Journal of Psychiatry 157:514-520 doi:10.1176/appi.ajp.157.4.514
Karnath H-O, Baier B, Nägele T (2005a) Awareness of the Functioning of One's Own Limbs Mediated by the Insular Cortex? The Journal of Neuroscience 25:7134
Karnath H-O, Rorden C (2012) The anatomy of spatial neglect Neuropsychologia 50:1010-1017 doi:10.1016/j.neuropsychologia.2011.06.027
Karnath H-O, Smith DV (2014) The next step in modern brain lesion analysis: multivariate pattern analysis Brain 137:2405-2407 doi:10.1093/brain/awu180
Karnath H-O, Zopf R, Johannsen L, Berger MF, Nägele T, Klose U (2005b) Normalized perfusion MRI to identify common areas of dysfunction: patients with basal ganglia neglect Brain 128:2462-2469 doi:10.1093/brain/awh629
Karnath HO, Himmelbach M, Rorden C (2002) The subcortical anatomy of human spatial neglect: putamen, caudate nucleus and pulvinar Brain 125:350-360 doi:10.1093/brain/awf032
Karson CN (1983) Spontaneous Eye-Blink Rates and Dopaminergic Systems Brain 106:643-653 doi:10.1093/brain/106.3.643
Kasper EM, Larkman AU, Lübke J, Blakemore C (1994) Pyramidal neurons in layer 5 of the rat visual cortex. I. Correlation among cell morphology, intrinsic electrophysiological

properties, and axon targets The Journal of Comparative Neurology 339:459-474 doi:10.1002/cne.903390402
Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation Neuron 22:751-761 doi:http://dx.doi.org/10.1016/S0896-6273(00)80734-5
Kato M, Miyashita N, Hikosaka O, Matsumura M, Usui S, Kori A (1995) Eye movements in monkeys with local dopamine depletion in the caudate nucleus. I. Deficits in spontaneous saccades The Journal of Neuroscience 15:912
Keysers C, Perrett DI (2004) Demystifying social cognition: a Hebbian perspective Trends in Cognitive Sciences 8:501-507 doi:https://doi.org/10.1016/j.tics.2004.09.005
Khundakar AA et al. (2016) Analysis of primary visual cortex in dementia with Lewy bodies indicates GABAergic involvement associated with recurrent complex visual hallucinations Acta Neuropathologica Communications 4:66 doi:10.1186/s40478-016-0334-3
Kiebel SJ, Daunizeau J, Friston KJ (2008) A Hierarchy of Time-Scales and the Brain PLoS Computational Biology 4:e1000209 doi:10.1371/journal.pcbi.1000209
Kiebel SJ, Daunizeau J, Friston KJ (2009) Perception and Hierarchical Dynamics Frontiers in Neuroinformatics 3:20 doi:10.3389/neuro.11.020.2009
Kiebel SJ, David O, Friston KJ (2006) Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization NeuroImage 30:1273-1284 doi:https://doi.org/10.1016/j.neuroimage.2005.12.055
Kilner JM, Friston KJ, Frith CD (2007) Predictive coding: an account of the mirror neuron system Cognitive processing 8:159-166 doi:10.1007/s10339-007-0170-2
Kim M, Barrett AM, Heilman KM (1998) Lateral Asymmetries of Pupillary Responses Cortex 34:753-762 doi:http://dx.doi.org/10.1016/S0010-9452(08)70778-0
Kinsbourne M (1970) A model for the mechanism of unilateral neglect of space Transactions of the American Neurological Association 95:143-146
Kirshner HS (2003) Chapter 140 - Speech and Language Disorders A2 - Samuels, Martin A. In: Feske SK (ed) Office Practice of Neurology (Second Edition). Churchill Livingstone, Philadelphia, pp 890-895. doi:http://dx.doi.org/10.1016/B0-44-306557-8/50142-8
Klein C, Fischer B, Fischer B, Hartnegg K (2002) Effects of methylphenidate on saccadic responses in patients with ADHD Experimental Brain Research 145:121-125 doi:10.1007/s00221-002-1105-x
Klein RM (2000) Inhibition of return Trends in Cognitive Sciences 4:138-147 doi:https://doi.org/10.1016/S1364-6613(00)01452-2
Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation TRENDS in Neurosciences 27:712-719
Kobayashi Y, Isa T (2002) Sensory-motor gating and cognitive control by the brainstem cholinergic system Neural Networks 15:731-741 doi:https://doi.org/10.1016/S0893-6080(02)00059-X
Kojima S, Kojima M, Goldman-Rakic PS (1982) Operant behavioral analysis of memory loss in monkeys with prefrontal lesions Brain Research 248:51-59 doi:10.1016/0006-8993(82)91146-5
Kori A, Miyashita N, Kato M, Hikosaka O, Usui S, Matsumura M (1995) Eye movements in monkeys with local dopamine depletion in the caudate nucleus. II. Deficits in voluntary saccades The Journal of Neuroscience 15:928
Koss MC (1986) Pupillary dilation as an index of central nervous system α2-adrenoceptor activation Journal of Pharmacological Methods 15:1-19 doi:http://dx.doi.org/10.1016/0160-5402(86)90002-1
Krakauer JW, Shadmehr R (2007) TOWARDS A COMPUTATIONAL NEUROPSYCHOLOGY OF ACTION Progress in brain research 165:383-394 doi:10.1016/S0079-6123(06)65024-3
Krause A (2008) Optimizing sensing: Theory and applications. Carnegie Mellon University

Krauzlis RJ, Lisberger SG (1989) A Control Systems Model of Smooth Pursuit Eye Movements with Realistic Emergent Properties Neural Computation 1:116-122 doi:10.1162/neco.1989.1.1.116
Krishnamurthy K, Nassar MR, Sarode S, Gold JI (2017) Arousal-related adjustments of perceptual biases optimize perception in dynamic environments Nature human behaviour 1:0107 doi:10.1038/s41562-017-0107
Kritzer MF, Goldman-Rakic PS (1995) Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey The Journal of Comparative Neurology 359:131-143 doi:10.1002/cne.903590109
Kuhl DE et al. (1996) In vivo mapping of cholinergic terminals in normal aging, Alzheimer's disease, and Parkinson's disease Annals of neurology 40:399-410
Künzle H, Akert K (1977) Efferent connections of cortical, area 8 (frontal eye field) in Macaca fascicularis. A reinvestigation using the autoradiographic technique The Journal of Comparative Neurology 173:147-163 doi:10.1002/cne.901730108
Kwon C, Ao P, Thouless DJ (2005) Structure of stochastic dynamics near fixed points Proceedings of the National Academy of Sciences of the United States of America 102:13029 doi:10.1073/pnas.0506347102
Laar Tvd, Vries Bd (2016) A Probabilistic Modeling Approach to Hearing Loss Compensation IEEE/ACM Transactions on Audio, Speech, and Language Processing 24:2200-2213 doi:10.1109/TASLP.2016.2599275
Ladavas E, Zeloni G, Zaccara G, Gangemi P (1997) Eye movements and orienting of attention in patients with visual neglect Journal of Cognitive Neuroscience 9:67-74
Lam B, Hollingdrake E, Kennedy JL, Black SE, Masellis M (2009) Cholinesterase inhibitors in Alzheimer's disease and Lewy body spectrum disorders: the emerging pharmacogenetic story Human Genomics 4:91 doi:10.1186/1479-7364-4-2-91
Lambe EK, Goldman-Rakic PS, Aghajanian GK (2000) Serotonin Induces EPSCs Preferentially in Layer V Pyramidal Neurons of the Frontal Cortex in the Rat Cerebral Cortex 10:974-980 doi:10.1093/cercor/10.10.974
Landau Ayelet N, Fries P (2012) Attention Samples Stimuli Rhythmically Current Biology 22:1000-1004 doi:https://doi.org/10.1016/j.cub.2012.03.054
Latto R, Cowey A (1971) Fixation changes after frontal eye-field lesions in monkeys Brain Research 30:25-36 doi:http://dx.doi.org/10.1016/0006-8993(71)90003-5
Lavín C, San Martín R, Rosales Jubal E (2013) Pupil dilation signals uncertainty and surprise in a learning gambling task Frontiers in Behavioral Neuroscience 7:218 doi:10.3389/fnbeh.2013.00218
Lavine N, Reuben M, Clarke P (1997) A population of nicotinic receptors is associated with thalamocortical afferents in the adult rat: laminal and areal analysis Journal of Comparative Neurology 380:175-190
Lawson RP, Mathys C, Rees G (2017) Adults with autism overestimate the volatility of the sensory environment Nat Neurosci 20:1293-1299 doi:10.1038/nn.4615

http://www.nature.com/neuro/journal/v20/n9/abs/nn.4615.html#supplementary-information
Lawson RP, Rees G, Friston KJ (2014) An aberrant precision account of autism Frontiers in Human Neuroscience 8:302 doi:10.3389/fnhum.2014.00302
LeDoux JE, Iwata J, Cicchetti P, Reis DJ (1988) Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear The Journal of Neuroscience 8:2517
Lee C, Rohrer WH, Sparks DL (1988) Population coding of saccadic eye movements by neurons in the superior colliculus Nature 332:357-360
Lenartowicz A, Escobedo-Quiroz R, Cohen JD (2010) Updating of context in working memory: An event-related potential study Cognitive, Affective, & Behavioral Neuroscience 10:298-315 doi:10.3758/CABN.10.2.298
Lennie P (2003) The Cost of Cortical Computation Current Biology 13:493-497 doi:http://dx.doi.org/10.1016/S0960-9822(03)00135-0

Lewis DD, Gale WA A sequential algorithm for training text classifiers. In: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, 1994. Springer-Verlag New York, Inc., pp 3-12
Li B-M, Mao Z-M, Wang M, Mei Z-T (1999) Alpha-2 Adrenergic Modulation of Prefrontal Cortical Neuronal Activity Related to Spatial Working Memory in Monkeys Neuropsychopharmacology 21:601 doi:10.1016/S0893-133X(99)00070-6
Li D, Karnath H-O, Rorden C (2014) Egocentric representations of space co-exist with allocentric representations: evidence from spatial neglect Cortex; a journal devoted to the study of the nervous system and behavior 58:161-169 doi:10.1016/j.cortex.2014.06.012
Liao H-I, Yoneya M, Kidani S, Kashino M, Furukawa S (2016) Human Pupillary Dilation Response to Deviant Auditory Stimuli: Effects of Stimulus Properties and Voluntary Attention Frontiers in Neuroscience 10:43 doi:10.3389/fnins.2016.00043
Linden R, Perry VH (1983) Massive retinotectal projection in rats Brain Research 272:145-149 doi:https://doi.org/10.1016/0006-8993(83)90371-2
Linderman S, Johnson M, Miller A, Adams R, Blei D, Paninski L (2017) Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems. Paper presented at the Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research,
Lindley DV (1956) On a Measure of the Information Provided by an Experiment Ann Math Statist 27:986-1005 doi:10.1214/aoms/1177728069
Lipton RB, Levy DL, Holzman PS, Levin S (1983) Eye movement dysfunctions in psychiatric patients: A review Schizophrenia Bulletin 9:13-32
Litvak V et al. (2011) EEG and MEG data analysis in SPM8 Computational intelligence and neuroscience 2011:852961 doi:10.1155/2011/852961
Liu AKL, Chang RC-C, Pearce RKB, Gentleman SM (2015) Nucleus basalis of Meynert revisited: anatomy, history and differential involvement in Alzheimer's and Parkinson's disease Acta Neuropathologica 129:527-540 doi:10.1007/s00401-015-1392-5
Livingstone M, Hubel D (1988) Segregation of form, color, movement, and depth: anatomy, physiology, and perception Science 240:740
Lobotesis K et al. (2001) Occipital hypoperfusion on SPECT in dementia with Lewy bodies but not AD Neurology 56:643-649 doi:10.1212/wnl.56.5.643
Loeliger HA (2004) An introduction to factor graphs IEEE Signal Processing Magazine 21:28-41 doi:10.1109/MSP.2004.1267047
Loeliger HA, Dauwels J, Hu J, Korl S, Ping L, Kschischang FR (2007) The Factor Graph Approach to Model-Based Signal Processing Proceedings of the IEEE 95:1295-1322 doi:10.1109/JPROC.2007.896497
Lombardo S, Maskos U (2015) Role of the nicotinic acetylcholine receptor in Alzheimer's disease pathology and treatment Neuropharmacology 96, Part B:255-262 doi:https://doi.org/10.1016/j.neuropharm.2014.11.018
Lovejoy LP, Krauzlis RJ (2010) Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments Nature neuroscience 13:261-266
Lukas JR, Aigner M, Blumer R, Heinzl H, Mayr R (1994) Number and distribution of neuromuscular spindles in human extraocular muscles Investigative Ophthalmology & Visual Science 35:4317-4327
Lukashin AV, Amirikian BR, Mozhaev VL, Wilcox GL, Georgopoulos AP (1996) Modeling motor cortical operations by an attractor network of stochastic neurons Biological Cybernetics 74:255-261 doi:10.1007/BF00652226
Lunven M, Thiebaut de Schotten M, Bourlon C, Duret C, Migliaccio R, Rode G, Bartolomeo P (2015) White matter lesional predictors of chronic visual neglect: a longitudinal study Brain 138:746-760
Lynch G, King DJ, Green JF, Byth W, Wilson-Davis K (1997) The effects of haloperidol on visual search, eye movements and psychomotor performance Psychopharmacology 133:233-239 doi:10.1007/s002130050396

Ma TP, Graybiel AM, Wurtz RH (1991) Location of saccade-related neurons in the macaque superior colliculus Experimental Brain Research 85:21-35 doi:10.1007/BF00229983
MacKay DJC (1992) Information-Based Objective Functions for Active Data Selection Neural Computation 4:590-604 doi:10.1162/neco.1992.4.4.590
Mah Y-H, Husain M, Rees G, Nachev P (2014) Human brain lesion-deficit inference remapped Brain 137:2522-2531 doi:10.1093/brain/awu164
Mair R, Anderson CD, Langlais P, McEntee W (1985) Thiamine deficiency depletes cortical norepinephrine and impairs learning processes in the rat Brain Research 360:273-284
Makris N, Kennedy DN, McInerney S, Sorensen AG, Wang R, Caviness JVS, Pandya DN (2004) Segmentation of Subcomponents within the Superior Longitudinal Fascicle in Humans: A Quantitative, In Vivo, DT-MRI Study Cerebral Cortex 15:854-869 doi:10.1093/cercor/bhh186
Malhotra P et al. (2005) Spatial working memory capacity in unilateral neglect Brain 128:424-435 doi:10.1093/brain/awh372
Malhotra P, Mannan S, Driver J, Husain M (2004) Impaired Spatial Working Memory: One Component of the Visual Neglect Syndrome? Cortex 40:667-676 doi:http://dx.doi.org/10.1016/S0010-9452(08)70163-1
Malhotra PA, Parton AD, Greenwood R, Husain M (2006) Noradrenergic modulation of space exploration in visual neglect Annals of Neurology 59:186-190 doi:10.1002/ana.20701
Mannan SK, Mort DJ, Hodgson TL, Driver J, Kennard C, Husain M (2005) Revisiting Previously Searched Locations in Visual Neglect: Role of Right Parietal and Frontal Lesions in Misjudging Old Locations as New Journal of Cognitive Neuroscience 17:340-354 doi:10.1162/0898929053124983
Marchetti G (2014) Attention and working memory: two basic mechanisms for constructing temporal experiences Frontiers in Psychology 5:880 doi:10.3389/fpsyg.2014.00880
Marder E, Thirumalai V (2002) Cellular, synaptic and network effects of neuromodulation Neural Networks 15:479-493 doi:https://doi.org/10.1016/S0893-6080(02)00043-6
Marek R, Strobel C, Bredy TW, Sah P (2013) The amygdala and medial prefrontal cortex: partners in the fear circuit The Journal of Physiology 591:2381-2391 doi:10.1113/jphysiol.2012.248575
Markov NT et al. (2013) Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex The Journal of Comparative Neurology 522:225-259 doi:10.1002/cne.23458
Marshall L, Mathys C, Ruge D, de Berker AO, Dayan P, Stephan KE, Bestmann S (2016) Pharmacological Fingerprints of Contextual Uncertainty PLOS Biology 14:e1002575 doi:10.1371/journal.pbio.1002575
Mason ST, Fibiger HC (1977) Altered exploratory behaviour after 6-OHDA lesion to the dorsal noradrenergic bundle Nature 269:704-705
Mathys CD (2012) Hierarchical Gaussian filtering
Mattingley JB, Bradshaw JL, Phillips JG (1992) Impairments of movement initiation and execution in unilateral neglect Brain 115:1849
Maurice N et al. (2015) Striatal Cholinergic Interneurons Control Motor Behavior and Basal Ganglia Function in Experimental Parkinsonism Cell Reports 13:657-666 doi:https://doi.org/10.1016/j.celrep.2015.09.034
Mays LE, Sparks DL (1980) Dissociation of visual and saccade-related responses in superior colliculus neurons Journal of Neurophysiology 43:207
McCann H, Stevens CH, Cartwright H, Halliday GM (2014) α-Synucleinopathy phenotypes Parkinsonism & Related Disorders 20:S62-S67 doi:https://doi.org/10.1016/S1353-8020(13)70017-8
McCloskey M, Rapp B (2000) Attention-referenced visual representations: evidence from impaired visual localization Journal of Experimental Psychology: Human Perception and Performance 26:917
McDougall SJ, Münzberg H, Derbenev AV, Zsombok A (2014) Central control of autonomic functions in health and disease Frontiers in Neuroscience 8:440 doi:10.3389/fnins.2014.00440

McFarland NR, Haber SN (2002) Thalamic Relay Nuclei of the Basal Ganglia Form Both Reciprocal and Nonreciprocal Cortical Connections, Linking Multiple Frontal Cortical Areas The Journal of Neuroscience 22:8117
McKeith I et al. (2004) Dementia with Lewy bodies The Lancet Neurology 3:19-28 doi:https://doi.org/10.1016/S1474-4422(03)00619-7
McLaughlin JT, McKie S (2016) Human brain responses to gastrointestinal nutrients and gut hormones Current Opinion in Pharmacology 31:8-12 doi:https://doi.org/10.1016/j.coph.2016.08.006
McSpadden A (1998) A Mathematical Model of Human Saccadic Eye Movement. Texas Tech University,
Medina J et al. (2009) Neural Substrates of Visuospatial Processing in Distinct Reference Frames: Evidence from Unilateral Spatial Neglect Journal of cognitive neuroscience 21:2073-2084 doi:10.1162/jocn.2008.21160
Menon GJ, Rahman I, Menon SJ, Dutton GN (2003) Complex Visual Hallucinations in the Visually Impaired: The Charles Bonnet Syndrome Survey of Ophthalmology 48:58-72 doi:https://doi.org/10.1016/S0039-6257(02)00414-9
Mesulam MM (1999) Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events Philosophical Transactions of the Royal Society B: Biological Sciences 354:1325-1346
Mesulam MM, Mufson EJ (1982) Insula of the old world monkey. III: Efferent cortical output and comments on function The Journal of Comparative Neurology 212:38-52 doi:10.1002/cne.902120104
Miller KD (2003) Understanding Layer 4 of the Cortical Circuit: A Model Based on Cat V1 Cerebral Cortex 13:73-82 doi:10.1093/cercor/13.1.73
Minka T (2005) Divergence measures and message passing. Technical report, Microsoft Research,
Minoshima S, Foster NL, Petrie EC, Albin RL, Frey KA, Kuhl DE (2002) Neuroimaging in Dementia with Lewy Bodies: Metabolism, Neurochemistry, and Morphology Journal of Geriatric Psychiatry and Neurology 15:200-209 doi:10.1177/089198870201500405
Minoshima S, Frey KA, Cross DJ, Kuhl DE (2004) Neurochemical imaging of dementias Seminars in Nuclear Medicine 34:70-82 doi:https://doi.org/10.1053/j.semnuclmed.2003.09.008
Minton NA, Murray VSG (1988) A Review of Organophosphate Poisoning Medical Toxicology and Adverse Drug Experience 3:350-375 doi:10.1007/BF03259890
Mintzopoulos D et al. (2009) Connectivity alterations assessed by combining fMRI and MR-compatible hand robots in chronic stroke NeuroImage 47S2:T90-T97 doi:10.1016/j.neuroimage.2009.03.007
Mirza MB, Adams RA, Mathys C, Friston KJ (2018) Human visual exploration reduces uncertainty about the sensed world PLOS ONE 13:e0190429 doi:10.1371/journal.pone.0190429
Mirza MB, Adams RA, Mathys CD, Friston KJ (2016) Scene Construction, Visual Foraging, and Active Inference Frontiers in Computational Neuroscience 10 doi:10.3389/fncom.2016.00056
Mongillo G, Barak O, Tsodyks M (2008) Synaptic Theory of Working Memory Science 319:1543-1546 doi:10.1126/science.1150769
Montoya A, Bruins R, Katzman MA, Blier P (2016) The noradrenergic paradox: implications in the management of depression and anxiety Neuropsychiatric disease and treatment 12:541-557 doi:10.2147/NDT.S91311
Moore T, Fallah M (2001) Control of eye movements and spatial attention Proceedings of the National Academy of Sciences 98:1273-1276
Moran R, Pinotsis DA, Friston K (2013a) Neural masses and fields in dynamic causal modeling Frontiers in Computational Neuroscience 7:57 doi:10.3389/fncom.2013.00057

Moran RJ, Campo P, Symmonds M, Stephan KE, Dolan RJ, Friston KJ (2013b) Free energy, precision and learning: the role of cholinergic neuromodulation The Journal of neuroscience : the official journal of the Society for Neuroscience 33:8227-8236 doi:10.1523/JNEUROSCI.4255-12.2013
Moret C, Briley M (2011) The importance of norepinephrine in depression Neuropsychiatric Disease and Treatment 7:9-13 doi:10.2147/NDT.S19619
Mori T, Ikeda M, Fukuhara R, Nestor PJ, Tanabe H (2006) Correlation of visual hallucinations with occipital rCBF changes by donepezil in DLB Neurology 66:935-937 doi:10.1212/01.wnl.0000203114.03976.b0
Moss J, Bolam JP (2008) A Dopaminergic Axon Lattice in the Striatum and Its Relationship with Cortical and Thalamic Terminals The Journal of Neuroscience 28:11221
Moutoussis M, Trujillo-Barreto NJ, El-Deredy W, Dolan RJ, Friston KJ (2014) A formal model of interpersonal inference Front Hum Neurosci 8:160 doi:10.3389/fnhum.2014.00160
Mukherjee P, Sabharwal A, Kotov R, Szekely A, Parsey R, Barch DM, Mohanty A (2016) Disconnection Between Amygdala and Medial Prefrontal Cortex in Psychotic Disorders Schizophrenia Bulletin 42:1056-1067 doi:10.1093/schbul/sbw012
Munoz DP, Istvan PJ (1998) Lateral Inhibitory Interactions in the Intermediate Layers of the Monkey Superior Colliculus Journal of Neurophysiology 79:1193
Munoz DP, Wurtz RH (1995a) Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells Journal of Neurophysiology 73:2313
Munoz DP, Wurtz RH (1995b) Saccade-related activity in monkey superior colliculus. II. Spread of activity during saccades Journal of Neurophysiology 73:2334
Murray JD et al. (2014) A hierarchy of intrinsic timescales across primate cortex Nature neuroscience 17:1661-1663 doi:10.1038/nn.3862
Nachev P (2015) The first step in modern lesion-deficit analysis Brain 138:e354-e354 doi:10.1093/brain/awu275
Nachev P, Hacker P (2014) The neural antecedents to voluntary action: A conceptual analysis Cognitive Neuroscience 5:193-208 doi:10.1080/17588928.2014.934215
Nadim F, Bucher D (2014) Neuromodulation of neurons and synapses Current Opinion in Neurobiology 29:48-56 doi:https://doi.org/10.1016/j.conb.2014.05.003
Naicker P, Anoopkumar-Dukie S, Grant GD, Kavanagh JJ (2017) Medications influencing central cholinergic neurotransmission affect saccadic and smooth pursuit eye movements in healthy young adults Psychopharmacology 234:63-71 doi:10.1007/s00213-016-4436-1
Nambu A (2004) A new dynamic model of the cortico-basal ganglia loop. In: Progress in Brain Research, vol Volume 143. Elsevier, pp 461-466. doi:http://dx.doi.org/10.1016/S0079-6123(03)43043-4
Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems Nature neuroscience 15:1040-1046 doi:10.1038/nn.3130
Nealey TA, Maunsell JH (1994) Magnocellular and parvocellular contributions to the responses of neurons in macaque striate cortex The Journal of Neuroscience 14:2069
Ness J, Hoth A, Barnett MJ, Shorr RI, Kaboli PJ (2006) Anticholinergic medications in community-dwelling older veterans: Prevalence of anticholinergic symptoms, symptom burden, and adverse drug events The American Journal of Geriatric Pharmacotherapy 4:42-51 doi:https://doi.org/10.1016/j.amjopharm.2006.03.008
Nobre AC, Gitelman DR, Dias EC, Mesulam MM (2000) Covert Visual Spatial Orienting and Saccades: Overlapping Neural Systems NeuroImage 11:210-216 doi:http://dx.doi.org/10.1006/nimg.2000.0539
O'Reilly RC (2006) Biologically Based Computational Models of High-Level Cognition Science 314:91
O'Reilly RC, Frank MJ (2006) Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia Neural Computation 18:283-328 doi:10.1162/089976606775093909

O'Callaghan C et al. (2017) Visual Hallucinations Are Characterized by Impaired Sensory Evidence Accumulation: Insights From Hierarchical Drift Diffusion Modeling in Parkinson's Disease Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 2:680-688 doi:https://doi.org/10.1016/j.bpsc.2017.04.007
O'Driscoll GA, Dépatie L, Holahan A-LV, Savion-Lemieux T, Barr RG, Jolicoeur C, Douglas VI (2005) Executive Functions and Methylphenidate Response in Subtypes of Attention-Deficit/Hyperactivity Disorder Biological Psychiatry 57:1452-1460 doi:https://doi.org/10.1016/j.biopsych.2005.02.029
O'Reilly JX, Jbabdi S, Behrens TEJ (2012) How can a Bayesian approach inform neuroscience? European Journal of Neuroscience 35:1169-1179 doi:10.1111/j.1460-9568.2012.08010.x
Ochipa C, Rothi LJ, Heilman KM (1994) Conduction apraxia Journal of Neurology, Neurosurgery, and Psychiatry 57:1241-1244
Ognibene D, Baldassarre G Ecological Active Vision: Four Bio-Inspired Principles to Integrate Bottom-Up and Adaptive Top-Down Attention Tested With a Simple Camera-Arm Robot. In: IEEE Transactions onAutonomous Mental Development, 2014. IEEE, p 99
Ojima H, Murakami K, Kishi K (1996) Dual termination modes of corticothalamic fibers originating from pyramids of layers 5 and 6 in cat visual cortical area 17 Neuroscience Letters 208:57-60 doi:https://doi.org/10.1016/0304-3940(96)12538-6
Oke A, Keller R, Mefford I, Adams RN (1978) Lateralization of norepinephrine in human thalamus Science 200:1411
Oliva GA, Bucci MP, Fioravanti R (1993) Impairment of Saccadic Eye Movements by Scopolamine Treatment Perceptual and Motor Skills 76:159-167 doi:10.2466/pms.1993.76.1.159
Ondobaka S, Kilner J, Friston K (2017) The role of interoceptive inference in theory of mind Brain and Cognition 112:64-68 doi:https://doi.org/10.1016/j.bandc.2015.08.002
Ota H, Fujii T, Suzuki K, Fukatsu R, Yamadori A (2001) Dissociation of body-centered and stimulus-centered representations in unilateral neglect Neurology 57:2064-2069 doi:10.1212/wnl.57.11.2064
Oudeyer P-Y, Kaplan F (2007) What is Intrinsic Motivation? A Typology of Computational Approaches Frontiers in neurorobotics 1:6-6 doi:10.3389/neuro.12.006.2007
Owens AP, Friston KJ, Low DA, Mathias CJ, Critchley HD (2018) Investigating the relationship between cardiac interoception and autonomic cardiac control using a predictive coding framework Autonomic Neuroscience: Basic and Clinical doi:10.1016/j.autneu.2018.01.001
Paré M, Crommelinck M, Guitton D (1994) Gaze shifts evoked by stimulation of the superior colliculus in the head-free cat conform to the motor map but also depend on stimulus strength and fixation activity Experimental brain research 101:123-139
Parr T, Benrimoh D, Vincent P, Friston K (2018a) Precision and false perceptual inference Front Integr Neurosci doi:10.3389/fnint.2018.00039
Parr T, Corcoran AW, Friston KJ, Hohwy J (2019a) Perceptual awareness and active inference Neuroscience of Consciousness 2019 doi:10.1093/nc/niz012
Parr T, Da Costa L, Friston K (In Press) Markov blankets, information geometry and stochastic thermodynamics Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences doi:DOI 10.1098/rsta.2019.0159
Parr T, Friston KJ (2017a) The active construction of the visual world Neuropsychologia 104:92-101 doi:http://dx.doi.org/10.1016/j.neuropsychologia.2017.08.003
Parr T, Friston KJ (2017b) The Computational Anatomy of Visual Neglect Cerebral Cortex:1-14 doi:10.1093/cercor/bhx316
Parr T, Friston KJ (2017c) Uncertainty, epistemics and active inference Journal of The Royal Society Interface 14
Parr T, Friston KJ (2017d) Working memory, attention, and salience in active inference Scientific reports 7:14678 doi:10.1038/s41598-017-15249-0

Parr T, Friston KJ (2018a) Active inference and the anatomy of oculomotion Neuropsychologia doi:https://doi.org/10.1016/j.neuropsychologia.2018.01.041
Parr T, Friston KJ (2018b) The Anatomy of Inference: Generative Models and Brain Structure Frontiers in Computational Neuroscience 12 doi:10.3389/fncom.2018.00090
Parr T, Friston KJ (2018c) The Discrete and Continuous Brain: From Decisions to Movement—and Back Again Neural computation:2319–2347
Parr T, Friston KJ (2019a) Attention or salience? Current Opinion in Psychology 29:1-5 doi:https://doi.org/10.1016/j.copsyc.2018.10.006
Parr T, Friston KJ (2019b) The computational pharmacology of oculomotion Psychopharmacology doi:10.1007/s00213-019-05240-0
Parr T, Friston KJ (2019c) Generalised free energy and active inference Biological Cybernetics doi:10.1007/s00422-019-00805-w
Parr T, Markovic D, Kiebel SJ, Friston KJ (2019b) Neuronal message passing using Mean-field, Bethe, and Marginal approximations Scientific reports 9:1889 doi:10.1038/s41598-018-38246-3
Parr T, Mirza MB, Cagnan H, Friston KJ (2019c) Dynamic causal modelling of active vision The Journal of Neuroscience:2459-2418 doi:10.1523/JNEUROSCI.2459-18.2019
Parr T, Rees G, Friston KJ (2018b) Computational Neuropsychology and Bayesian Inference Frontiers in Human Neuroscience 12 doi:10.3389/fnhum.2018.00061
Paus T (1996) Location and function of the human frontal eye-field: a selective review Neuropsychologia 34:475-483
Peck CK, Baro JA, Warder SM (1993) Chapter 9 Sensory integration in the deep layers of superior colliculus. In: T.P. Hicks SM, Ono T (eds) Progress in Brain Research, vol Volume 95. Elsevier, pp 91-102. doi:http://dx.doi.org/10.1016/S0079-6123(08)60360-X
Pellicano E, Burr D (2012) When the world becomes 'too real': a Bayesian explanation of autistic perception Trends in Cognitive Sciences 16:504-510 doi:https://doi.org/10.1016/j.tics.2012.08.009
Perrinet LU, Adams RA, Friston KJ (2014) Active inference, eye movements and oculomotor delays Biological Cybernetics 108:777-801 doi:10.1007/s00422-014-0620-8
Perry EK et al. (1994) Neocortical cholinergic activities differentiate Lewy body dementia from classical Alzheimer's disease Neuroreport 5:747-749
Perry EK et al. (1991) Topography, Extent, and Clinical Relevance of Neurochemical Deficits in Dementia of Lewy Body Type, Parkinson's Disease, and Alzheimer's Disease Annals of the New York Academy of Sciences 640:197-202 doi:10.1111/j.1749-6632.1991.tb00217.x
Perry RJ, Zeki S (2000) The neurology of saccades and covert shifts in spatial attentionAn event-related fMRI study Brain 123:2273-2288 doi:10.1093/brain/123.11.2273
Pertzov Y, Bays PM, Joseph S, Husain M (2013) Rapid forgetting prevented by retrospective attention cues Journal of Experimental Psychology: Human Perception and Performance 39:1224-1231 doi:10.1037/a0030947
Peters A, McEwen BS, Friston K (2017) Uncertainty and stress: Why it causes diseases and how it is mastered by the brain Progress in Neurobiology 156:164-188 doi:https://doi.org/10.1016/j.pneurobio.2017.05.004
Petersen SE, Robinson DL, Keys W (1985) Pulvinar nuclei of the behaving rhesus monkey: visual responses and their modulation Journal of Neurophysiology 54:867
Pierrot-Deseilligny C, Rivaud S, Gaymard B, Müri R, Vermersch A-I (1995) Cortical control of saccades Annals of Neurology 37:557-567 doi:doi:10.1002/ana.410370504
Platz T, Schüttauf J, Aschenbach J, Mengdehl C, Lotze M (2016) Effects of inhibitory theta burst TMS to different brain sites involved in visuospatial attention–a combined neuronavigated cTBS and behavioural study Restorative neurology and neuroscience 34:271-285
Posner MI, Rafal RD, Choate LS, Vaughan J (1985) Inhibition of return: Neural basis and function Cognitive Neuropsychology 2:211-228 doi:10.1080/02643298508252866

Pouget A, Beck JM, Ma WJ, Latham PE (2013) Probabilistic brains: knowns and unknowns Nature neuroscience 16:1170-1178 doi:10.1038/nn.3495
Pouget A, Sejnowski TJ (1997) A new view of hemineglect based on the response properties of parietal neurones Philosophical Transactions of the Royal Society B: Biological Sciences 352:1449-1459
Pouget A, Sejnowski TJ (2001) Simulating a lesion in a basis function model of spatial representations: comparison with hemineglect Psychological review 108:653
Price C, Warburton E, Moore C, Frackowiak R, Friston K (2001) Dynamic diaschisis: anatomically remote and context-sensitive human brain lesions Journal of Cognitive Neuroscience 13:419-429
Ptak R, Schnider A (2010) The Dorsal Attention Network Mediates Orienting toward Behaviorally Relevant Stimuli in Spatial Neglect The Journal of Neuroscience 30:12557
Quaia C, Aizawa H, Optican LM, Wurtz RH (1998) Reversible Inactivation of Monkey Superior Colliculus. II. Maps of Saccadic Deficits Journal of Neurophysiology 79:2097
Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects Nature neuroscience 2:79-87
Rasmussen J (1985) The role of hierarchical knowledge representation in decisionmaking and system management IEEE Transactions on Systems, Man, and Cybernetics SMC-15:234-243 doi:10.1109/TSMC.1985.6313353
Raybourn MS, Keller EL (1977) Colliculoreticular organization in primate oculomotor system Journal of Neurophysiology 40:861
Redish AD, Elga AN, Touretzky DS (1996) A coupled attractor model of the rodent head direction system Network: Computation in Neural Systems 7:671-685 doi:10.1088/0954-898X_7_4_004
Rees G, Wojciulik E, Clarke K, Husain M, Frith C, Driver J (2000) Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction Brain 123:1624-1633
Reichert DP, Seriès P, Storkey AJ (2013) Charles Bonnet Syndrome: Evidence for a Generative Model in the Cortex? PLoS Computational Biology 9:e1003134 doi:10.1371/journal.pcbi.1003134
Reilly JL, Lencer R, Bishop JR, Keedy S, Sweeney JA (2008) Pharmacological treatment effects on eye movement control Brain and cognition 68:415-435 doi:10.1016/j.bandc.2008.08.026
Ressler KJ (2010) Amygdala Activity, Fear, and Anxiety: Modulation by Stress Biological psychiatry 67:1117-1119 doi:10.1016/j.biopsych.2010.04.027
Richert M, Nageswaran JM, Sokol S, Szatmary B, Petre C, Piekniewski F, Izhikevich E (2013) A spiking model of superior colliculus for bottom-up saliency BMC Neuroscience 14:P185 doi:10.1186/1471-2202-14-s1-p185
Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding and imitation of action Nat Rev Neurosci 2:661-670
Rizzolatti G, Riggio L, Dascola I, Umiltá C (1987) Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention Neuropsychologia 25:31-40 doi:10.1016/0028-3932(87)90041-8
Roberts JA, Wallis G, Breakspear M (2013) Fixational eye movements during viewing of dynamic natural scenes Frontiers in Psychology 4:797 doi:10.3389/fpsyg.2013.00797
Robinson D (1964) The mechanics of human saccadic eye movement The Journal of physiology 174:245-264
Robinson DA (1968) The oculomotor control system: A review Proceedings of the IEEE 56:1032-1049
Robinson DL, Petersen SE The pulvinar and visual salience Trends in Neurosciences 15:127-132 doi:10.1016/0166-2236(92)90354-B
Rocca MA et al. (2007) Altered functional and structural connectivities in patients with MS: A 3-T study Neurology 69:2136-2145 doi:10.1212/01.wnl.0000295504.92020.ca

Rockland KS (1998) Convergence and branching patterns of round, type 2 corticopulvinar axons The Journal of Comparative Neurology 390:515-536 doi:10.1002/(SICI)1096-9861(19980126)390:4<515::AID-CNE5>3.0.CO;2-3
Roh E, Song DK, Kim M-S (2016) Emerging role of the brain in the homeostatic regulation of energy and glucose metabolism Experimental &Amp; Molecular Medicine 48:e216 doi:10.1038/emm.2016.4
Rorden C, Hjaltason H, Fillmore P, Fridriksson J, Kjartansson O, Magnusdottir S, Karnath H-O (2012) Allocentric neglect strongly associated with egocentric neglect Neuropsychologia 50:1151-1157 doi:10.1016/j.neuropsychologia.2012.03.031
Rushmore RJ, Valero-Cabre A, Lomber SG, Hilgetag CC, Payne BR (2006) Functional circuitry underlying visual neglect Brain 129:1803-1821
Rushworth MFS, Behrens TEJ, Johansen-Berg H (2005) Connection Patterns Distinguish 3 Regions of Human Parietal Cortex Cerebral Cortex 16:1418-1430 doi:10.1093/cercor/bhj079
Ruskell GL (1989) The fine structure of human extraocular muscle spindles and their potential proprioceptive capacity Journal of Anatomy 167:199-214
Rycroft N, Hutton SB, Rusted JM (2006) The antisaccade task as an index of sustained goal activation in working memory: modulation by nicotine Psychopharmacology 188:521-529 doi:10.1007/s00213-006-0455-7
Sahin M, Bowen WD, Donoghue JP (1992) Location of nicotinic and muscarinic cholinergic and μ-opiate receptors in rat cerebral neocortex: evidence from thalamic and cortical lesions Brain Research 579:135-147 doi:https://doi.org/10.1016/0006-8993(92)90752-U
Sajad A, Sadeh M, Keith GP, Yan X, Wang H, Crawford JD (2015) Visual–Motor Transformations Within Frontal Eye Fields During Head-Unrestrained Gaze Shifts in the Monkey Cerebral Cortex 25:3932-3952 doi:10.1093/cercor/bhu279
Sales AC, Friston KJ, Jones MW, Pickering AE, Moran RJ (2018) Locus Coeruleus tracking of prediction errors optimises cognitive flexibility: an Active Inference model bioRxiv:340620
Samuel W, Terry RD, DeTeresa R, Butters N, Masliah E (1994) Clinical correlates of cortical and nucleus basalis pathology in alzheimer dementia Archives of Neurology 51:772-778 doi:10.1001/archneur.1994.00540200048015
Sara SJ, Hervé-Minvielle A (1995) Inhibitory influence of frontal cortex on locus coeruleus neurons Proceedings of the National Academy of Sciences of the United States of America 92:6032-6036
Saur D et al. (2008) Ventral and dorsal pathways for language Proceedings of the National Academy of Sciences 105:18035-18040 doi:10.1073/pnas.0805234105
Sawaguchi T, Matsumura M, Kubota K (1990) Catecholaminergic effects on neuronal activity related to a delayed response task in monkey prefrontal cortex Journal of Neurophysiology 63:1385-1400 doi:10.1152/jn.1990.63.6.1385
Schaal S (2006) Dynamic movement primitives-a framework for motor control in humans and humanoid robotics. In: Adaptive motion of animals and machines. Springer, pp 261-280
Schiller PH, Sandell JH, Maunsell JH (1987) The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey Journal of Neurophysiology 57:1033
Schiller PH, Stryker M (1972) Single-unit recording and stimulation in superior colliculus of the alert rhesus monkey Journal of Neurophysiology 35:915
Schiller PH, True SD, Conway JL (1980) Deficits in eye movements following frontal eye-field and superior colliculus ablations Journal of Neurophysiology 44:1175
Schomer AC, Drislane FW (2015) Severe Hemispatial Neglect as a Manifestation of Seizures and Nonconvulsive Status Epilepticus: Utility of Prolonged EEG Monitoring Journal of Clinical Neurophysiology 32:e4-e7

Schwartenbeck P, FitzGerald TH, Mathys C, Dolan R, Wurst F, Kronbichler M, Friston K (2015a) Optimal inference with suboptimal models: addiction and active Bayesian inference Medical hypotheses 84:109-117 doi:10.1016/j.mehy.2014.12.007
Schwartenbeck P, FitzGerald THB, Mathys C, Dolan R, Friston K (2015b) The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes Cerebral Cortex 25:3434-3445 doi:10.1093/cercor/bhu159
Schwartenbeck P, FitzGerald THB, Mathys C, Dolan R, Wurst F, Kronbichler M, Friston K (2015c) Optimal inference with suboptimal models: Addiction and active Bayesian inference Medical hypotheses 84:109-117 doi:10.1016/j.mehy.2014.12.007
Schwartenbeck P, Friston K (2016) Computational Phenotyping in Psychiatry: A Worked Example eNeuro 3:ENEURO.0049-0016.2016 doi:10.1523/ENEURO.0049-16.2016
Schwöbel S, Kiebel S, Marković D (2018) Active Inference, Belief Propagation, and the Bethe Approximation Neural computation:1-38
Serences JT, Shomstein S, Leber AB, Golay X, Egeth HE, Yantis S (2005) Coordination of Voluntary and Stimulus-Driven Attentional Control in Human Cortex Psychological Science 16:114-122 doi:10.1111/j.0956-7976.2005.00791.x
Sereno AB, Holzman PS (1995) Antisaccades and smooth pursuit eye movements in schizophrenia Biological psychiatry 37:394-401
Seth AK (2013) Interoceptive inference, emotion, and the embodied self Trends in Cognitive Sciences 17:565-573 doi:https://doi.org/10.1016/j.tics.2013.09.007
Seth AK, Friston KJ (2016) Active interoceptive inference and the emotional brain Philosophical Transactions of the Royal Society B: Biological Sciences 371:20160007 doi:10.1098/rstb.2016.0007
Seung HS (1998) Continuous attractors and oculomotor control Neural Networks 11:1253-1258 doi:10.1016/S0893-6080(98)00064-1
Seung HS, Lee DD, Reis BY, Tank DW (2000) Stability of the Memory of Eye Position in a Recurrent Network of Conductance-Based Model Neurons Neuron 26:259-271 doi:https://doi.org/10.1016/S0896-6273(00)81155-1
Shah A, Frith U (1983) AN ISLET OF ABILITY IN AUTISTIC CHILDREN: A RESEARCH NOTE Journal of Child Psychology and Psychiatry 24:613-620 doi:10.1111/j.1469-7610.1983.tb00137.x
Shallice T (1964) THE DETECTION OF CHANGE AND THE PERCEPTUAL MOMENT HYPOTHESIS British Journal of Statistical Psychology 17:113-135 doi:10.1111/j.2044-8317.1964.tb00254.x
Sheliga BM, Riggio L, Rizzolatti G (1994) Orienting of attention and eye movements Experimental Brain Research 98:507-522 doi:10.1007/BF00233988
Sheliga BM, Riggio L, Rizzolatti G (1995) Spatial attention and eye movements Experimental Brain Research 105:261-275 doi:10.1007/BF00240962
Sherman SM (2007) The thalamus is more than just a relay Current Opinion in Neurobiology 17:417-422 doi:https://doi.org/10.1016/j.conb.2007.07.003
Sherr JD, Myers C, Avila MT, Elliott A, Blaxton TA, Thaker GK (2002) The effects of nicotine on specific eye tracking measures in schizophrenia Biological Psychiatry 52:721-728 doi:https://doi.org/10.1016/S0006-3223(02)01342-2
Sherrington CS (1893) Further Experimental Note on the Correlation of Action of Antagonistic Muscles British Medical Journal 1:1218-1218
Shewry MC, Wynn HP (1987) Maximum entropy sampling Journal of Applied Statistics 14:165-170 doi:10.1080/02664768700000020
Shipp S (2003) The functional logic of cortico-pulvinar connections Philosophical Transactions of the Royal Society B: Biological Sciences 358:1605-1624 doi:10.1098/rstb.2002.1213
Shipp S (2004) The brain circuitry of attention Trends in Cognitive Sciences 8:223-230 doi:http://dx.doi.org/10.1016/j.tics.2004.03.004
Shipp S (2007) Structure and function of the cerebral cortex Current Biology 17:R443-R449 doi:10.1016/j.cub.2007.03.044

Shipp S (2016) Neural Elements for Predictive Coding Frontiers in Psychology 7 doi:10.3389/fpsyg.2016.01792
Shipp S (2017) The functional logic of corticostriatal connections Brain Structure & Function 222:669-706 doi:10.1007/s00429-016-1250-9
Shipp S, Adams RA, Friston KJ (2013) Reflections on agranular architecture: predictive coding in the motor cortex Trends in Neurosciences 36:706-716 doi:https://doi.org/10.1016/j.tins.2013.09.004
Shomstein S, Lee J, Behrmann M (2010) Top-down and bottom-up attentional guidance: investigating the role of the dorsal and ventral parietal cortices Experimental Brain Research 206:197-208 doi:10.1007/s00221-010-2326-z
Shook BL, Schlag-Rey M, Schlag J (1990) Primate supplementary eye field: I. Comparative aspects of mesencephalic and pontine connections The Journal of Comparative Neurology 301:618-642 doi:10.1002/cne.903010410
Showers MJC, Lauer EW (1961) Somatovisceral motor patterns in the insula The Journal of Comparative Neurology 117:107-115 doi:10.1002/cne.901170109
Shulman GL, Astafiev SV, Franke D, Pope DLW, Snyder AZ, McAvoy MP, Corbetta M (2009) Interaction of stimulus-driven reorienting and expectation in ventral and dorsal fronto-parietal and basal ganglia-cortical networks The Journal of neuroscience : the official journal of the Society for Neuroscience 29:4392-4407 doi:10.1523/JNEUROSCI.5609-08.2009
Simmons DR, Robertson AE, McKay LS, Toal E, McAleer P, Pollick FE (2009) Vision in autism spectrum disorders Vision research 49:2705-2739
Simonyan K, Horwitz B (2011) Laryngeal Motor Cortex and Control of Speech in Humans The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry 17:197-208 doi:10.1177/1073858410386727
Smith AT, Singh KD, Williams AL, Greenlee MW (2001) Estimating Receptive Field Size from fMRI Data in Human Striate and Extrastriate Visual Cortex Cerebral Cortex 11:1182-1190 doi:10.1093/cercor/11.12.1182
Smith DT, Schenk T (2012) The Premotor theory of attention: Time to move on? Neuropsychologia 50:1104-1114 doi:http://dx.doi.org/10.1016/j.neuropsychologia.2012.01.025
Smith Y, Wichmann T, Factor SA, DeLong MR (2012) Parkinson's disease therapeutics: new developments and challenges since the introduction of levodopa Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 37:213-246 doi:10.1038/npp.2011.212
Sparks DL (1986) Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus Physiological Reviews 66:118
Sparks DL (2002) The brainstem control of saccadic eye movements Nat Rev Neurosci 3:952-964
Sparks DL, Mays LE (1990) Signal transformations required for the generation of saccadic eye movements Annual review of neuroscience 13:309-336
Squire LR, Stark CEL, Clark RE (2004) THE MEDIAL TEMPORAL LOBE Annual Review of Neuroscience 27:279-306 doi:10.1146/annurev.neuro.27.070203.144130
Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli Journal of Cognitive Neuroscience 1:12-24 doi:10.1162/jocn.1989.1.1.12
Stevens MC, Calhoun VD, Kiehl KA (2005) Hemispheric differences in hemodynamics elicited by auditory oddball stimuli NeuroImage 26:782-792 doi:10.1016/j.neuroimage.2005.02.044
Strassman A, Highstein SM, McCrea RA (1986) Anatomy and physiology of saccadic burst neurons in the alert squirrel monkey. I. Excitatory burst neurons The Journal of Comparative Neurology 249:337-357 doi:10.1002/cne.902490303
Stroud JM (1967) THE FINE STRUCTURE OF PSYCHOLOGICAL TIME Annals of the New York Academy of Sciences 138:623-631 doi:10.1111/j.1749-6632.1967.tb55012.x

Suzuki TW, Tanaka M (2017) Causal Role of Noradrenaline in the Timing of Internally Generated Saccades in Monkeys Neuroscience 366:15-22 doi:https://doi.org/10.1016/j.neuroscience.2017.10.003
Szczepanski SM, Kastner S (2013) Shifting attentional priorities: control of spatial attention through hemispheric competition Journal of Neuroscience 33:5411-5421
Szczepanski SM, Konen CS, Kastner S (2010) Mechanisms of spatial attention control in frontal and parietal cortex Journal of Neuroscience 30:148-160
Szczepanski SM, Pinsk MA, Douglas MM, Kastner S, Saalmann YB (2013) Functional and structural architecture of the human dorsal frontoparietal attention network Proceedings of the National Academy of Sciences 110:15806-15811
Tait DS, Brown VJ, Farovik A, Theobald DE, Dalley JW, Robbins TW (2007) Lesions of the dorsal noradrenergic bundle impair attentional set-shifting in the rat European Journal of Neuroscience 25:3719-3724 doi:10.1111/j.1460-9568.2007.05612.x
Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND (2011) How to grow a mind: statistics, structure, and abstraction Science 331:1279-1285 doi:10.1126/science.1192788
Testolin A, Zorzi M (2016) Probabilistic Models and Generative Neural Networks: Towards an Unified Framework for Modeling Normal and Impaired Neurocognitive Functions Frontiers in Computational Neuroscience 10 doi:10.3389/fncom.2016.00073
Teunisse RJ, Zitman FG, Cruysberg JRM, Hoefnagels WHL, Verbeek ALM (1996) Visual hallucinations in psychologically normal people: Charles Bonnet's syndrome The Lancet 347:794-797 doi:https://doi.org/10.1016/S0140-6736(96)90869-7
Thaker GK, Ross DE, Cassady SL, et al. (1998) Smooth pursuit eye movements to extraretinal motion signals: Deficits in relatives of patients with schizophrenia Archives of General Psychiatry 55:830-836 doi:10.1001/archpsyc.55.9.830
Thiebaut de Schotten M, Dell'Acqua F, Forkel SJ, Simmons A, Vergani F, Murphy DGM, Catani M (2011) A lateralized brain network for visuospatial attention Nat Neurosci 14:1245-1246 doi:10.1038/nn.2905

http://www.nature.com/neuro/journal/v14/n10/abs/nn.2905.html#supplementary-information
Thiebaut de Schotten M et al. (2014) Damage to white matter pathways in subacute and chronic spatial neglect: a group study and 2 single-case studies with complete virtual “in vivo” tractography dissection Cerebral Cortex 24:691-706
Thiebaut de Schotten M, Urbanski M, Duffau H, Volle E, Lévy R, Dubois B, Bartolomeo P (2005) Direct Evidence for a Parietal-Frontal Pathway Subserving Spatial Awareness in Humans Science 309:2226
Thomson AM, West DC, Wang Y, Bannister AP (2002) Synaptic Connections and Small Circuits Involving Excitatory and Inhibitory Neurons in Layers 2–5 of Adult Rat and Cat Neocortex: Triple Intracellular Recordings and Biocytin Labelling In Vitro Cerebral Cortex 12:936-953 doi:10.1093/cercor/12.9.936
Tiraboschi P et al. (2000) Cholinergic dysfunction in diseases with Lewy bodies Neurology 54:407-407 doi:10.1212/wnl.54.2.407
Tomlinson RD, Schwarz DWF (1977) Response of oculomotor neurons to eye muscle stretch Canadian Journal of Physiology and Pharmacology 55:568-573 doi:10.1139/y77-079
Trappenberg TP, Dorris MC, Munoz DP, Klein RM (2001) A Model of Saccade Initiation Based on the Competitive Integration of Exogenous and Endogenous Signals in the Superior Colliculus Journal of Cognitive Neuroscience 13:256-271 doi:10.1162/089892901564306
Troost BT (1989) Nystagmus: a clinical review Rev Neurol (Paris) 145:417-428
Tsuboi Y, Dickson DW (2005) Dementia with Lewy bodies and Parkinson's disease with dementia: Are they different? Parkinsonism & Related Disorders 11:S47-S51 doi:https://doi.org/10.1016/j.parkreldis.2004.10.014
Tsukada H, Fujii H, Aihara K, Tsuda I (2015) Computational model of visual hallucination in dementia with Lewy bodies Neural Networks 62:73-82 doi:https://doi.org/10.1016/j.neunet.2014.09.001

Turtzo LC, Kleinman JT, Llinas RH (2008) Capgras Syndrome and Unilateral Spatial Neglect in Nonconvulsive Status Epilepticus Behavioural Neurology 20 doi:10.3233/ben-2008-0210
Ungerleider LG, Christensen CA (1979) Pulvinar lesions in monkeys produce abnormal scanning of a complex visual array Neuropsychologia 17:493-501 doi:http://dx.doi.org/10.1016/0028-3932(79)90056-3
Ungerleider LG, Haxby JV (1994) 'What' and 'where' in the human brain Current Opinion in Neurobiology 4:157-165 doi:http://dx.doi.org/10.1016/0959-4388(94)90066-3
Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: .Ingle D, Goodale MA, Mansfield RJW (eds) Analysis of Visual Behavior. MIT Press, Cambridge, MA, pp 549-586
Valdez AB, Papesh MH, Treiman DM, Smith KA, Goldinger SD, Steinmetz PN (2015) Distributed Representation of Visual Objects by Single Neurons in the Human Brain The Journal of Neuroscience 35:5180-5186 doi:10.1523/JNEUROSCI.1958-14.2015
van de Laar TW, de Vries B (2019) Simulating Active Inference Processes by Message Passing Frontiers in Robotics and AI 6 doi:10.3389/frobt.2019.00020
van Dijk H, Schoffelen J-M, Oostenveld R, Jensen O (2008) Prestimulus Oscillatory Activity in the Alpha Band Predicts Visual Discrimination Ability The Journal of Neuroscience 28:1816
VanRullen R (2013) Visual Attention: A Rhythmic Process? Current Biology 23:R1110-R1112 doi:https://doi.org/10.1016/j.cub.2013.11.006
VanRullen R (2016) Perceptual Cycles Trends in Cognitive Sciences 20:723-735 doi:https://doi.org/10.1016/j.tics.2016.07.006
VanRullen R, Koch C (2003) Is perception discrete or continuous? Trends in Cognitive Sciences 7:207-213 doi:https://doi.org/10.1016/S1364-6613(03)00095-0
Veale R, Hafed ZM, Yoshida M (2017) How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling Philosophical Transactions of the Royal Society B: Biological Sciences 372
Verdon V, Schwartz S, Lovblad K-O, Hauert C-A, Vuilleumier P (2009) Neuroanatomy of hemispatial neglect and its functional components: a study using voxel-based lesion-symptom mapping Brain 133:880-894 doi:10.1093/brain/awp305
Vernon D (2008) Cognitive vision: The case for embodied perception Image and Vision Computing 26:127-140 doi:https://doi.org/10.1016/j.imavis.2005.08.009
Vincent P, Parr T, Benrimoh D, Friston K (In press) With an eye on uncertainty: modelling pupillary responses to environmental volatility PLOS Computational Biology
Virgo JD, Plant GT (2017) Internuclear ophthalmoplegia Practical Neurology 17:149
von der Heydt R, Peterhans E (1989) Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity The Journal of Neuroscience 9:1731
Von Helmholtz H (1867) Handbuch der physiologischen Optik vol 9. Voss,
Vossel S, Bauer M, Mathys C, Adams RA, Dolan RJ, Stephan KE, Friston KJ (2014) Cholinergic Stimulation Enhances Bayesian Belief Updating in the Deployment of Spatial Attention The Journal of Neuroscience 34:15735
Vossel S, Weidner R, Driver J, Friston KJ, Fink GR (2012) Deconstructing the Architecture of Dorsal and Ventral Attention Systems with Dynamic Causal Modeling The Journal of Neuroscience 32:10637
Vuilleumier P, Hester D, Assal G, Regli F (1996) Unilateral spatial neglect recovery after sequential strokes Neurology 46:184-189
Wald A (1947) An Essentially Complete Class of Admissible Decision Functions The Annals of Mathematical Statistics:549-555 doi:10.1214/aoms/1177730345
Wandell BA, Dumoulin SO, Brewer AA (2007) Visual Field Maps in Human Cortex Neuron 56:366-383 doi:http://dx.doi.org/10.1016/j.neuron.2007.10.012
Wang Y, Markram H, Goodman PH, Berger TK, Ma J, Goldman-Rakic PS (2006) Heterogeneity in the pyramidal network of the medial prefrontal cortex Nat Neurosci 9:534-542 doi:http://www.nature.com/neuro/journal/v9/n4/suppinfo/nn1670_S1.html

Wansard M, Meulemans T, Gillet S, Segovia F, Bastin C, Toba MN, Bartolomeo P (2014) Visual neglect: is there a relationship between impaired spatial working memory and re-cancellation? Experimental Brain Research 232:3333-3343 doi:10.1007/s00221-014-4028-4
Warrington EK (1975) The selective impairment of semantic memory The Quarterly journal of experimental psychology 27:635-657
Warrington EK, James M (1967) Disorders of visual perception in patients with localised cerebral lesions Neuropsychologia 5:253-266 doi:http://dx.doi.org/10.1016/0028-3932(67)90040-1
Warrington EK, James M (1988) Visual Apperceptive Agnosia: A Clinico-Anatomical Study of Three Cases Cortex 24:13-32 doi:http://dx.doi.org/10.1016/S0010-9452(88)80014-5
Warrington EK, Shallice T (1984) CATEGORY SPECIFIC SEMANTIC IMPAIRMENTS Brain 107:829-853 doi:10.1093/brain/107.3.829
Warrington EK, Taylor AM (1973) The Contribution of the Right Parietal Lobe to Object Recognition Cortex 9:152-164 doi:http://dx.doi.org/10.1016/S0010-9452(73)80024-3
Weil RS, Schrag AE, Warren JD, Crutch SJ, Lees AJ, Morris HR (2016) Visual dysfunction in Parkinson's disease Brain 139:2827-2843 doi:10.1093/brain/aww175
Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts Nat Neurosci 5:598-604 doi:http://www.nature.com/neuro/journal/v5/n6/suppinfo/nn858_S1.html
Weller RE, Steele GE, Kaas JH (2002) Pulvinar and other subcortical connections of dorsolateral visual cortex in monkeys The Journal of Comparative Neurology 450:215-240 doi:10.1002/cne.10298
Wendlandt S, File SE (1979) Behavioral effects of lesions of the locus ceruleus noradrenaline system combined with adrenalectomy Behavioral and Neural Biology 26:189-201 doi:https://doi.org/10.1016/S0163-1047(79)92579-2
Wernicke C (1969) The Symptom Complex of Aphasia. In: Cohen RS, Wartofsky MW (eds) Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968. Springer Netherlands, Dordrecht, pp 34-97. doi:10.1007/978-94-010-3378-7_2
Whitehouse PJ, Price DL, Clark AW, Coyle JT, DeLong MR (1981) Alzheimer disease: Evidence for selective loss of cholinergic neurons in the nucleus basalis Annals of Neurology 10:122-126 doi:10.1002/ana.410100203
Wimmer K, Nykamp DQ, Constantinidis C, Compte A (2014) Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory Nat Neurosci 17:431-439 doi:10.1038/nn.3645
Winn J, Bishop CM (2005) Variational message passing Journal of Machine Learning Research 6:661-694
Winter DA (1984) Biomechanics of human movement with applications to the study of human locomotion Crit Rev Biomed Eng 9:287-314
Wurtz RH, McAlonan K, Cavanaugh J, Berman RA (2011) Thalamic pathways for active vision Trends Cogn Sci 5:177-184
Wurtz RH, Mohler CW (1976) Organization of monkey superior colliculus: enhanced visual response of superficial layer cells Journal of Neurophysiology 39:745
Yager LM, Garcia AF, Wunsch AM, Ferguson SM (2015) The ins and outs of the striatum: Role in drug addiction Neuroscience 301:529-541 doi:10.1016/j.neuroscience.2015.06.033
Yang SC-H, Lengyel M, Wolpert DM (2016a) Active sensing in the categorization of visual patterns eLife 5:e12215 doi:10.7554/eLife.12215
Yang SC-H, Wolpert DM, Lengyel M (2016b) Theoretical perspectives on active sensing Current Opinion in Behavioral Sciences 11:100-108 doi:http://dx.doi.org/10.1016/j.cobeha.2016.06.009
Yedidia JS, Freeman WT, Weiss Y (2005) Constructing free-energy approximations and generalized belief propagation algorithms IEEE Transactions on Information Theory 51:2282-2312

Yin HH, Knowlton BJ (2006) The role of the basal ganglia in habit formation Nat Rev Neurosci 7:464-476
Yoshida K, Iwamoto Y, Chimoto S, Shimazu H (2001) Disynaptic Inhibition of Omnipause Neurons Following Electrical Stimulation of the Superior Colliculus in Alert Cats Journal of Neurophysiology 85:2639
Yu AJ, Dayan P (2002) Acetylcholine in cortical inference Neural Networks 15:719-730
Yu AJ, Dayan P (2005) Uncertainty, Neuromodulation, and Attention Neuron 46:681-692 doi:http://doi.org/10.1016/j.neuron.2005.04.026
Yue Y, Song W, Huo S, Wang M (2012) Study on the occurrence and neural bases of hemispatial neglect with different reference frames Archives of physical medicine and rehabilitation 93:156-162
Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? Trends in Cognitive Sciences 10:301-308 doi:https://doi.org/10.1016/j.tics.2006.05.002
Zaborszky L, Gaykema RP, Swanson DJ, Cullinan WE (1997) Cortical input to the basal forebrain Neuroscience 79:1051-1078 doi:https://doi.org/10.1016/S0306-4522(97)00049-3
Zeki S, Shipp S (1988) The functional logic of cortical connections Nature 335:311-317
Zeki S, Shipp S (1989) Modular connections between areas V2 and V4 of macaque monkey visual cortex European Journal of Neuroscience 1:494-506
Zelinsky GJ, Bisley JW (2015) The what, where, and why of priority maps and their interactions with visual working memory Annals of the New York Academy of Sciences 1339:154-164 doi:10.1111/nyas.12606
Zhang H, Yi Z, Zhang L (2008) Continuous attractors of a class of recurrent neural networks Computers & Mathematics with Applications 56:3130-3137 doi:http://dx.doi.org/10.1016/j.camwa.2008.07.035
Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory The Journal of Neuroscience 16:2112
Zhang Z, Cordeiro Matos S, Jego S, Adamantidis A, Séguéla P (2013) Norepinephrine Drives Persistent Activity in Prefrontal Cortex via Synergistic α1 and α2 Adrenoceptors PLOS ONE 8:e66122 doi:10.1371/journal.pone.0066122
Zimmermann E, Lappe M (2016) Visual Space Constructed by Saccade Motor Maps Frontiers in Human Neuroscience 10 doi:10.3389/fnhum.2016.00225