

## ORIGINAL ARTICLE

# The Computational Anatomy of Visual Neglect

Thomas Parr and Karl J. Friston

Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3BG, UK

Address correspondence to Thomas Parr, The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK.  
Email: thomas.parr.12@ucl.ac.uk

## Abstract

Visual neglect is a debilitating neuropsychological phenomenon that has many clinical implications and—in cognitive neuroscience—offers an important lesion deficit model. In this article, 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. We introduce a Bayes optimal model based upon Markov decision processes that reproduces the visual searches induced by the line cancellation task (used to characterize visual neglect at the bedside). We then consider 3 distinct ways in which the model could be lesioned to reproduce neuropsychological (visual search) deficits. Crucially, these 3 levels of pathology map nicely onto the neuroanatomy of saccadic eye movements and the systems implicated in visual neglect.

**Key words:** active inference, Bayesian, visual neglect, neuropsychology, saccades

## Introduction

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 (Ho et al. 2003; Andrade et al. 2010) diseases. It has also been observed as a feature of seizure activity (Heilman and Howell 1980; Turtzo et al. 2008; Schomer and Drislane 2015), 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 (Kamath et al. 2002) insults. There is some evidence that the heterogeneity of the causes of visual neglect map on to distinct behavioral phenotypes (Hillis et al. 2005; Grimsen et al. 2008; Medina et al. 2009; Verdon et al. 2009), and this has the potential to be exploited clinically and scientifically.

Eye tracking provides one way to characterize behavioral 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 with leftward saccades. This occurs both spontaneously (Fruhmann Berger et al. 2008; Kamath and Rorden 2012) and during search tasks (Husain et al. 2001). While these biases will form the main subject of this article, 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 (Ladavas et al. 1997; Bourgeois et al. 2015). In this article, 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 behavior 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 behavior optimal is mandated by the complete class theorems (Wald 1947; Daunizeau et al. 2010). This means that we can characterize 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). The article concludes by asking the question whether the different sorts of (saccadic) behavior induced by distinct sorts of lesions this sufficient to identify the locus of the lesion. We address this question using in silico neuropsychology and Bayesian model selection.

The purpose of this paper is to describe the active inference scheme and establish its predictive validity in (simulated) visual neglect. In subsequent papers, 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 characterization of neuronal computations in patients with visual neglect.

## Active Inference

The formal or normative framework used to characterize hemineglect calls on the notion of active inference. Active inference provides a Bayes optimal account of perception and action by appealing to some fundamental (variational) principles that apply to any system that has evolved to maintain an adaptive exchange with its environment. In brief, to sustain their integrity, adaptive systems must minimize the dispersion of their states (Friston et al. 2006). Mathematically, this dispersion corresponds to entropy, and is equivalent to surprise averaged over time (Friston 2009). Surprise, in the information theoretic sense used here, is the negative log probability of making a particular observation. A surprising observation is one that is unlikely under the prior beliefs possessed by an adaptive system, or subject. An intuitive example is that of blood pressure control. Animals are compelled to keep their blood pressure within narrow bounds in order to survive. The baroreceptor system implicitly “expects” blood pressure to be within this range with a high probability, so it is surprising when pressures outside this range are sensed. It is clear from this example that surprise is something to be avoided, if life is to pursue its familiar course. To interact with the environment, without incurring potentially fatal surprises, it is necessary to possess a generative model that describes how observations are generated. In most situations, the processes generating observable outcomes are sufficiently complex that it becomes intractable to compute surprise explicitly. Instead, a quantity called free energy can be computed (Dayan et al. 1995; Beal 2003; Friston 2003). This furnishes an upper bound on surprise, through Jensen’s inequality (The average of a log is less than or equal to the log of an average. This is a consequence of the concavity of the logarithmic function.).

$$\underbrace{-\ln P(\bar{o})}_{\text{Surprise}} = \underbrace{-\ln \sum_x \left[ \frac{P(\bar{o}, x)Q(x)}{Q(x)} \right]}_{\text{Jensen's inequality}} \leq \underbrace{-\sum_x Q(x) \ln \left[ \frac{P(\bar{o}, x)}{Q(x)} \right]}_{\text{Free energy}} \quad (1)$$

Minimizing the free energy therefore minimizes surprise (when the bound is tight). In this equation,  $x$  represents hidden (latent) variables, and  $\bar{o}$  represents the sequence of observations made over time. The particular form of the hidden variables depends upon the generative model.  $P(\bar{o}, x)$  describes the joint probability distribution of observations (i.e., consequences) and hidden variables (i.e., causes) under the generative model.  $Q(x)$  is an arbitrary distribution which becomes an approximate posterior as the free energy is minimized. Note that the first equality holds simply because  $Q(x)$  on the right hand side can be canceled, resulting in a marginalization of the joint distribution over all hidden variables.

The above demonstrates the (implicit or explicit) free energy minimization in adaptive systems. A system which does not minimize its free energy will fail to bound its entropy and, over time, will cease to exist; that is, will dissipate and decay. Active inference takes this further, by equipping an agent with beliefs about the policy (sequence of actions) it will pursue (Friston, Samothrakis, et al. 2012). A consequence of the imperative to minimize free energy is that, a priori, agents must believe they will minimize the free energy expected and allowable policies. 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.

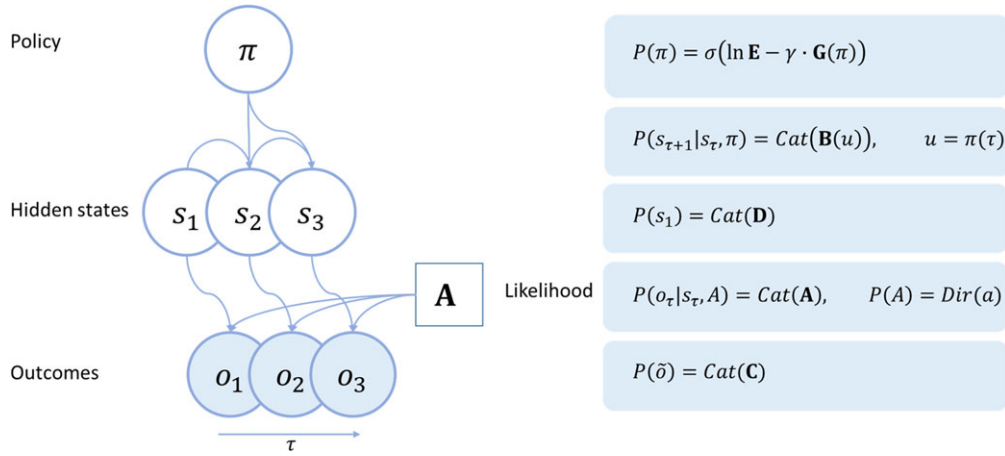
## Generative Model

### Markov Decision Processes

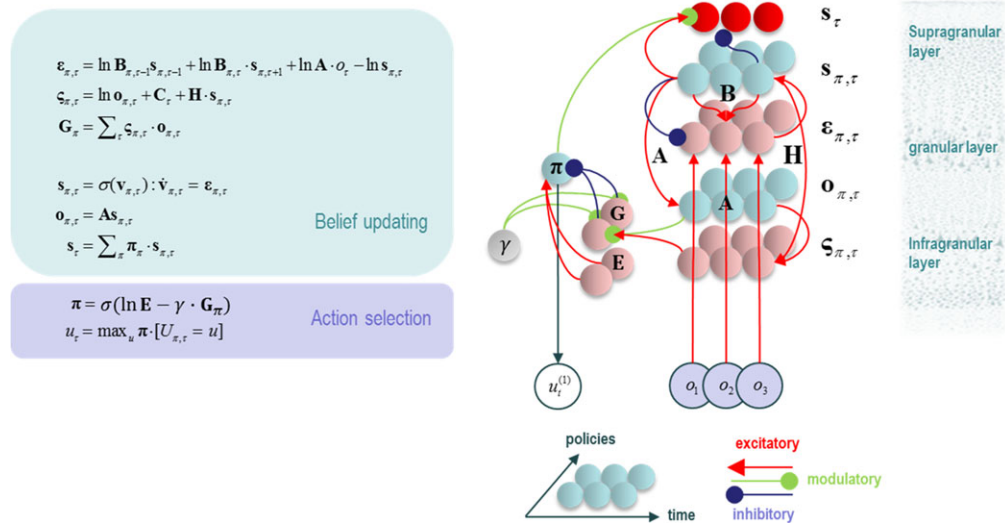
Given the discrete, serial, nature of saccadic sampling, an appropriate model structure—for the purposes of this article—is a Markov decision process (MDP) (Mirza et al. 2016). These models are defined in terms of a discrete state space, with observations made at discrete time points. The generic structure of an MDP is shown in Figure 1. The hidden variables in this model are the hidden states,  $s_t$ , the parameters of the likelihood mapping,  $A$ , and the policy  $\pi$ . The free energy can be expressed in terms of these unknown or hidden quantities (The notation  $E_Q[\cdot]$  means the expectation under the distribution  $Q$ ).

$$F = E_{Q(\bar{s}, A, \pi)} [\ln Q(\bar{s}, A, \pi) - \ln P(\bar{o}, \bar{s}, A, \pi)] \quad (2)$$

An MDP is structured such that observable outcomes depend only on the hidden states. The probabilistic mapping from hidden states to outcomes is expressed in the matrix  $A$ , in which  $A_{ij} = P(o_t = i | s_t = j)$ . The hidden states depend only on the previous hidden state, and on the transition matrix  $B$ , which is a function of the policy. Preferences are specified in terms of the prior beliefs an agent has about the outcomes they will observe, and these are contained in the matrix  $C$ .  $D$  determines the probabilities of the initial states. The vectors  $E$  and  $G$  correspond to prior expectations about policies and expected free energy respectively (please see below).



**Figure 1.** Markov decision process. The Bayesian network shown on the left describes the conditional dependencies in a Markov decision process (MDP). Each variable is shown in a circle, with shaded circles being observed variables. Arrows from one circle to another indicate that the probability of the second variable is conditioned on the first. The forms of these conditional distributions are given in the panel on the right. Cat means a categorical distribution, while Dir is a Dirichlet distribution.  $\sigma$  is a softmax (normalized exponential) function. Please see main text for a fuller explanation of the variables.



**Figure 2.** Variational update equations. The equations in the left panel can be obtained through a minimization of the variational free energy with respect to posterior expectations of each variable in the MDP model described in Figure 1. Each expectation is expressed as a vector, each component of which corresponds to the approximate ( $Q$ ) distribution for a particular value of that variable.  $\mathbf{H}$  is the entropy of the likelihood matrix. On the right, variables have been assigned to neuronal populations, and the connections between these populations determined from the update equations. The result closely resembles a cortical column with a corticosubcortical loop. The modulation of this loop by the precision ( $\gamma$ ) unit, which is thought to correspond to dopaminergic activity (Friston et al. 2013) is suggestive of a nigrostriatal influence. The 2 opposing inputs to the policy unit could reflect the direct and indirect basal ganglia pathways. Notably,  $\mathbf{E}$  (prior expectations about policies) does not depend on current beliefs about states, while  $\mathbf{G}$  (expected free energy) is context dependent. Correspondingly, indirect pathway neurons in the striatum have relatively small dendritic trees with limited cortical input, while direct pathway neurons have very large dendritic trees (Gertler et al. 2008).

This generative model allows the factorization of  $P(\delta, \tilde{s}, A, \pi)$  into conditionally independent factors. Using a “mean field approximation,” we can additionally factorize  $Q(\tilde{s}, A, \pi)$  into approximately independent factors. It is then possible to derive update equations for each factor of  $Q$ , by taking the derivative of the (variational) free energy with respect to that factor, and setting the result to zero (see Appendix for the derivation of the hidden state updates). In doing so, the update equations shown in Figure 2 are obtained (Friston, FitzGerald, Rigoli, Schwartenbeck and Pezzulo 2016). Reassuringly, when the variables in these equations are mapped out in terms of the influence each has over the others, the emergent structure closely resembles the architecture of cortical microcircuits, cortical hierarchies, and

even corticosubcortical loops involved in policy evaluation. This loop is consistent with the structure and function of the basal ganglia (Jahanshahi et al. 2015). See Figure 2.

### Memory and Short-Term Plasticity

Given that the probability distributions are specified as categorical distributions, the appropriate conjugate prior for the likelihood  $\mathbf{A}$  matrix is a Dirichlet distribution. This means that the probability can be represented simply in terms of Dirichlet concentration parameters. For each state,  $s_{\tau} = j$ , there are a set of Dirichlet parameters,  $a_{ij}$ , one for each outcome,  $o_{\tau} = i$ , 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 outcome  $i$  given state  $j$  is as follows:

$$\mathbf{A}_{ij} = \frac{a_{ij}}{\sum_k a_{kj}} \quad (3)$$

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 an agent to remember the observations they sampled when they believed they were in a particular state (Friston, FitzGerald, Rigoli, Schwartenbeck, O’Doherty, et al. 2016). The notion that the mapping between representations of 2 variables should be increased when the 2 are simultaneously active is strikingly similar to Hebbian plasticity (Hebb 1949; Brown et al. 2009). This analysis suggests that this form of memory could be implemented by short-term changes in synaptic efficacy.

An important consequence of the Dirichlet parameterization 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,  $\mathbf{A}$ . If the parameters are very small, an observation can trigger one-shot learning, suggesting a rapid short-term plasticity effect. Such effects have been proposed as one mechanism underlying working memory (Mongillo et al. 2008). This behavior is of particular interest in the current context, as will become apparent in the next section, where the form of the MDP used to model hemineglect is described.

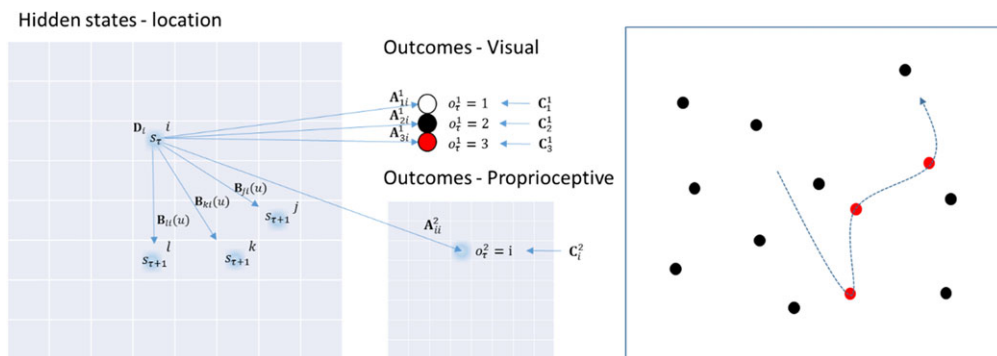
## Saccadic Cancellation Task

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  $8 \times 8$  grid. The task is to look at each of the targets until all targets have been sampled (i.e., canceled). When a target has been fixated, it changes color from black to red (see the right panel in Fig. 3), indicating that it has been seen. The model used to emulate this behavior is shown in Figure 3, 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 3 possible outcomes: no target (white), target (black), and canceled 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 canceled, and does not expect to see targets that have already been canceled. The subject begins with (almost) uniform beliefs about the  $\mathbf{A}$  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 canceled that location. This implements a synaptic form of spatial working memory (Mongillo et al. 2008). 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 ( $\mathbf{B}$ ) 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 and how salient 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



**Figure 3.** 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. See Fig. 1 for the definitions of the symbols used. The hidden states in this model are the locations in the visual field that are fixated. These are indicated by the  $8 \times 8$  grid on the left of this figure. The start location, is specified by  $D$ . The agent may saccade to any location on the grid (three possible saccades are shown), and the particular saccade is defined by  $u$ , which selects the appropriate  $\mathbf{B}$  matrix. Each component of this matrix defines the probability of a saccade to a given location ( $j$ ,  $k$ , or  $l$  in the figure), given a current location ( $i$  in the figure). There are 2  $\mathbf{A}$  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  $\mathbf{C}$  matrices, which are defined 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 canceled.

would have to possess to render their behavior Bayes optimal. For visual neglect, we consider the abnormal patterns of saccadic eye movements in patients (Husain et al. 2001; Fruhmann Berger et al. 2008; Bays et al. 2010; 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. We can express this formally as follows:

$$P(\pi) = \sigma(-\gamma \cdot \mathbf{G}(\pi)) \quad (4)$$

In this equation,  $\sigma(\cdot)$  is a softmax function, which ensures the resulting distribution will sum to one, making it a proper probability distribution. The scale parameter  $\gamma$  is an inverse temperature parameter, which acts as a precision over policy priors.  $G(\pi)$  is the expected free energy associated with each policy  $\pi$ . This is defined as the sum of expected free energies for each future time point.

$$G(\pi) = \sum_{\tau > t} G(\pi, \tau) \quad (5)$$

The expected free energy has a similar form to that of the variational free energy. However, there are 2 key differences. The first is that it must be conditioned on the policy pursued, and the second is that, by definition, future observations have not yet been made. This means the expectation in Equation 2 must now include beliefs about future outcomes. Defining  $\tilde{Q}(o_\tau, s_\tau, A|\pi) = Q(s_\tau, A|\pi)P(o_\tau|s_\tau, A)$  allows us to express expected free energy as follows:

$$G(\pi, \tau) = E_Q[\ln Q(s_\tau, A|\pi) - \ln P(o_\tau, s_\tau, A|\pi)] \quad (6)$$

Rearranging this, we can separate out the key terms that influence policy selection in the generative model described above.

$$\begin{aligned} G(\pi, \tau) = & \underbrace{-E_Q[\ln P(o_\tau)]}_{\text{Expected cost}} \\ & - E_{Q(o_\tau|\pi)} \underbrace{[D_{\text{KL}}[P(s_\tau|o_\tau, \pi) \| Q(s_\tau|\pi)]]}_{\text{Saliency}} \\ & + D_{\text{KL}} \underbrace{[P(A|o_\tau, \pi) \| Q(A|\pi)]}_{\text{Novelty}} \end{aligned} \quad (7)$$

The second (saliency) 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 no certainty. This means there is no information gain or epistemic value that would otherwise resolve uncertainty about the hidden state. The key terms that determine policy selection are the first and third. The former implies that a policy which is expected to fulfill 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,  $P(A|o_\tau, \pi)$ , from beliefs prior to seeing counterfactual outcomes,  $Q(A|\pi)$ , 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; that is, 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 2 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 minimize 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 follows:

$$\ln P(\pi) = \ln \mathbf{E} - \gamma \cdot \mathbf{G}(\pi) \quad (8)$$

Here,  $\mathbf{E}$  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  $\mathbf{E}$  corresponds to the y-intercept and precision is the sensitivity or slope.

In summary, the above formal considerations have led us to identify 3 possible synthetic lesions which could give rise to visual neglect. These are changes in the priors over policy  $\mathbf{E}$ , the Dirichlet parameters of the beliefs about  $\mathbf{A}^1$ , and the priors concerning proprioceptive outcomes, contained in the matrix  $\mathbf{G}^2$ . In the following section, we review plausible neurobiological substrates for each of these computational pathologies.

## The Neuroanatomy of Hemineglect

### The Dorsal Attention Network

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 (Künzle and Akert 1977; Berson and McIlwain 1983; Fries 1984, 1985; Shook et al. 1990; Gaymard et al. 2003). 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 (Makris et al. 2004; Bartolomeo et al. 2012). 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).

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  $\mathbf{A}^2$  (Fig. 3). 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 behaviors (Ungerleider and Christensen 1979). The connections from this region would then encode the  $\mathbf{C}^2$  matrix.

### The Ventral Attention Network

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, 2011). These regions are the constituents of the ventral attention network, and are connected by the third branch of the superior longitudinal fasciculus (Rushworth et al. 2005; Bartolomeo et al. 2012). 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 (Duncan 2001; Dosenbach et al. 2006) and in target detection (Stevens et al. 2005). This suggests that the third branch of the superior longitudinal fasciculus is the anatomical substrate of  $\mathbf{C}^1$ .

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 2017). There is evidence to suggest that this unilateral representation of identity is right lateralised (Warrington and James 1967, 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 2 networks (He et al. 2007). The 2 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  $\mathbf{A}^1$ . 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.

### Subcortical Structures

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 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 behavior, it makes intuitive sense that pathological biasing of policies would take place within a structure which is involved in habit formation; that is, 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. 2017). 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 (Weller et al. 2002; Behrens et al. 2003). 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 (Shipp 2003; Kanai et al. 2015) and eye movements (Petersen et al. 1985), and as a “salience map” (Robinson and Petersen 1992; 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 (Schiller et al. 1980, 1987; Hikosaka and Wurtz 1985a), 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 prioritized the lesions corresponding to the white matter tract that connects the dorsal and ventral attention networks, in addition

to 2 common subcortical lesions; the putamen and pulvinar. These closely resemble the theoretically motivated lesions of  $A^1$ ,  $E$ , and  $C^2$ .

### Interhemispheric Interactions

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 (Vuilleumier et al. 1996; Rushmore et al. 2006; Dietz et al. 2014) that neglect involves interhemispheric imbalances in addition to intrahemispheric disruptions (Bartolomeo et al. 2007; Bartolomeo 2014). This is a key feature of an existing model of neglect (Kinsbourne 1970). Fortunately for our framework, the 2 are inherently linked. Examination of the equations in Figure 2 (and the Appendix) reveals 2 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 2 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 interhemispheric competition—a competition that is won by

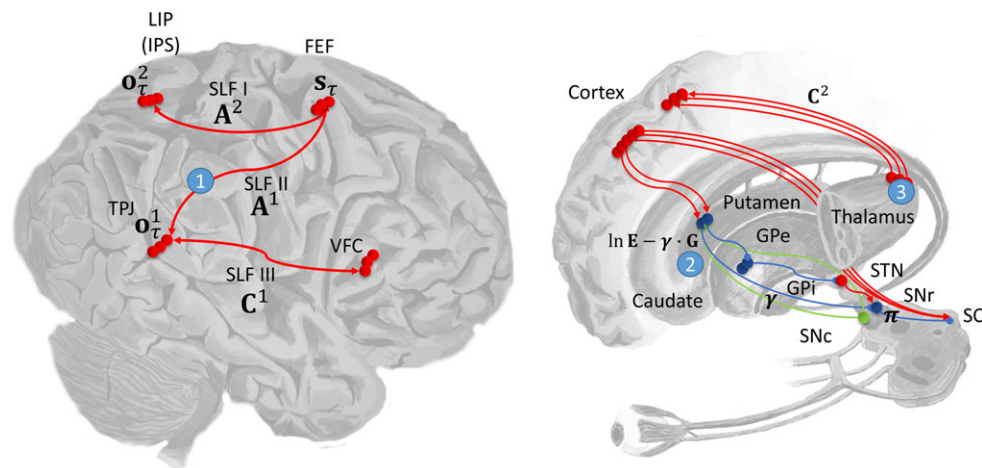
the left hemisphere if any of the lesions described in the previous section bias policies towards rightward saccades (Fig. 4).

## Simulating Hemineglect

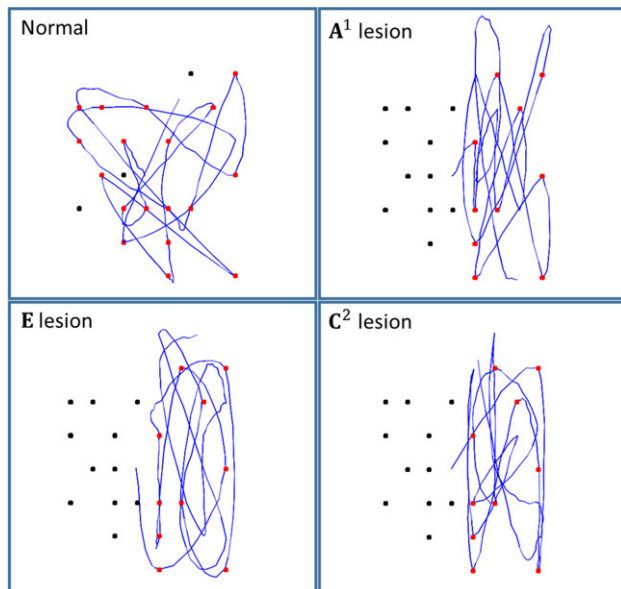
### Heterogeneous Pathology to Homogenous Syndrome

Figure 5 shows the results of running the simulation for 20 saccades, under different prior beliefs (i.e., lesions). Strikingly, all 3 lesioned models produce very similar behavioral patterns. This heterogeneity of functional lesions is consistent with the diverse set of anatomical lesions known to cause visual neglect. While the nonlesioned model samples both sides of space, all 3 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 (contained in  $E$ ), that result in a slightly different pattern of saccadic search than that depicted in Figure 5. 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  $A^1$  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 Schneider 2010) and attention (Shulman et al. 2009). The simulated pulvinar lesion causes the agent to fulfill 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 favor of saccades in this direction.



**Figure 4.** Computational anatomy and lesion sites. This schematic illustrates the proposed mapping from the computational entities implicated by the model (Figs 2 and 3) 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 ( $A^2$ ). 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 ( $C^2$ ). Visual outcomes are assumed to be represented in the TPJ, which is the mapping from hidden states to visual outcomes ( $A^1$ ), 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  $C^2$ . In addition, the pathways through the basal ganglia are also shown. The policy evaluation processes shown in Figure 2 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, in proportion to their precision  $\gamma$ , and the output of the putamen is transformed by the substantia nigra pars reticulata into a distribution over policies. The simulated lesions we considered are numbered: 1—SLF II; 2—Putamen; 3—Pulvinar. As in the previous figure, red connections are excitatory, blue inhibitory, and green modulatory.



**Figure 5.** 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  $A^1$  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.

While mechanistically distinct, the behavioral profiles of each of these lesions do not appear to lend themselves to precise diagnoses in terms of observable behavior. 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 behavior.

### Representing Visual 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; Zeki and Shipp 1988, 1989; Nealey and Maunsell 1994). 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 2 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 (Driver and Pouget 2000; Corbetta and Shulman 2011). If both are considered to take place in the same reference frame, the 2 behavioral patterns would be consistent with visual neglect operating at a coarse spatial scale in the first case, and a finer scale in the second.

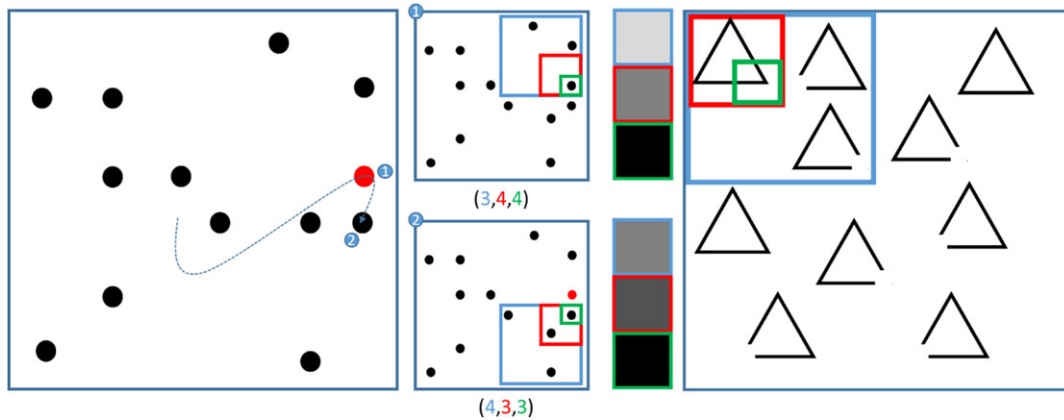
Equipping the model with a multiscale representation is simple to do in our generative model; instead of representing each of the 64 locations at a high resolution, we can encode each location using 3 levels (i.e., factors) of resolution, each level divided into 4 quadrants that, collectively, specify  $4^3 = 64$  locations. Technically, this means the  $A$  matrix now becomes 3 matrices encoding the likelihood mappings at low, intermediate, and high levels of resolution. Functionally, this means that the subject perceives visual input at 3 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 6 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—that is, quadrant enclosed by the blue frame—an egocentric behavioral pattern of saccadic sampling would be expected. However, if induced at a finer scale (green frame), neglect would cause an allocentric pattern. Figure 7 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 2 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).

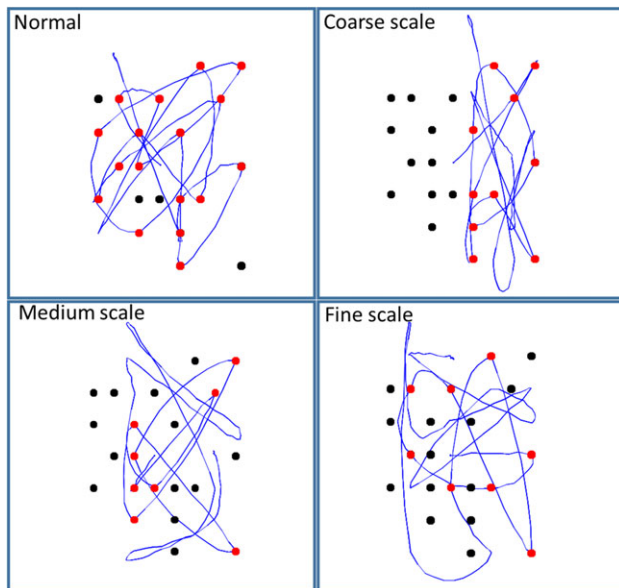
A simplification we have made in the generative model we have used is that we have assumed the head position is stationary. This allows us to treat the coordinate transform, performed by the first branch of the superior longitudinal fasciculus, as an identity transformation. If we did not make this assumption, 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, 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 article. Importantly, deficits that are classically described as "object-centered" 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





**Figure 6.** Multiscale representations of space. In the illustration on the left, 2 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 center left, this state space is shown for each fixation point. This specifies a location in an  $8 \times 8$  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 3 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 center 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. Canceled targets appear red at this level—lower resolutions are considered to be color-blind; consistent with the properties of the magnocellular system (Hubel and Livingstone 1987). As a saccade is made from a quadrant containing 3 targets to one containing 4, 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 2. 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 7.** Lesions at different spatial scales. By changing the number of the initial Dirichlet parameters, we have simulated hemineglect at 3 resolutions. As can be seen in the above, the coarse scale representation biases saccades to the right side of the array, similar to the patterns seen in Figure 5. The medium scale representation biases saccades to the right side within each of the 4 quadrants of visual space. Hemineglect at the finest scale biases saccades to the right of each subquadrant (comprising 4 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.

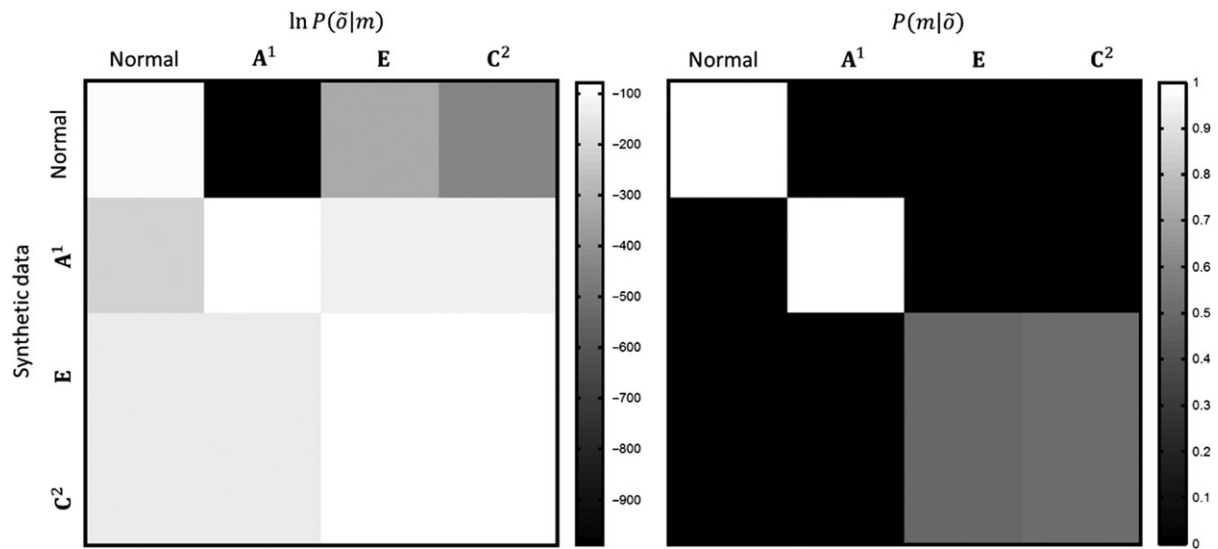
movements—and the deficits in epistemic foraging seen in patients with neglect. In the final section, we turn to the practical issues of using this sort of model to make inferences about lesions on the basis of saccadic eye movements.

## Computational Lesion Deficit Analysis

We have established that, just as with anatomical lesions, there are several functional lesions that can induce very similar behavior. This raises an important question. Is the mapping from lesion to behavior truly a many-to-one mapping? In other words, is it possible, given the (simulated) behavioral 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 behavior 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 behavior (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 8. It is clear from the confusion matrix shown in the figure that one can reliably disambiguate between health and pathology. Furthermore, the lesions in  $A^1$  (i.e., a synthetic disconnection between dorsal and ventral attentional systems), although visually very similar to those of the other 2 lesion models, produce a characteristic behavioral pattern, allowing the lesion identity to be recovered. We aim to use this disambiguation to provide an empirical test of our anatomical model.



**Figure 8.** Confusion matrices constructed from 40 saccades. The matrix on the left shows the (log) model evidence  $\ln P(\bar{\delta}|m)$  for each model,  $m$  (columns), given synthetic eye tracking data,  $\bar{\delta}$  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  $P(m|\bar{\delta})$ . This is obtained from the matrix on the left, using a softmax function applied to the log evidence in each row (i.e., for different models of each synthetic dataset).

If we use transcranial magnetic stimulation to disrupt the communication between dorsal frontal and right temporoparietal regions, we expect to find that a lesion deficit analysis using eye movements will find greater evidence for an A1 lesion than for any of the other lesion models. Lesions of **E** and **C<sup>2</sup>** 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; that is, **A<sup>1</sup>**. This suggests that distinguishing between the 2 may require an additional data modality, such as reaction time, pupillometry, or electrophysiology.

## Theoretical Neurobiology

The work presented in this paper closely relates to a number of recent advances in theoretical neurobiology. We have built upon previous formulations of visual exploration under active inference (Friston, Adams, et al. 2012; Mirza et al. 2016), but there are number of important distinctions between these accounts and the current work. The first is that we have used active inference to address the impact of (computational) lesions, and to demonstrate how neuropsychological disorders can be described functionally in terms of pathological priors.

The second difference is subtle: the formulations mentioned above involved the selection of saccadic targets to minimize uncertainty. This is shared with the current work, and with earlier theories of visual salience (Itti and Baldi 2006), but the quantities about which the simulated agent is uncertain differ. Previously, we have emphasized uncertainty about hidden states in the environment. Here, we focus on the uncertainty about the relationship between hidden states and their sensory consequences. It is this important difference that facilitates the analysis of disconnection syndromes, as these can be formulated as disruption of sensorimotor contingencies.

Previous models have addressed attentional processes in general (Bundesen 1998; Heinke and Humphreys 2005), and neglect specifically (Kinsbourne 1970; Heinke and Humphreys 2003). 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 behaviors (FitzGerald et al. 2014; Moutoussis et al. 2014; Friston et al. 2015), physiological responses (Friston et al. 2014; Schwartenbeck, FitzGerald, Mathys, Dolan and Friston 2015), and pathologies (Schwartenbeck, FitzGerald, Mathys, Dolan, Wurst, et al. 2015), emphasizing 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, FitzGerald, Rigoli, Schwartenbeck and Pezzulo 2016), 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 modeling for evoked responses (David et al. 2006) provides a means to test this hypothesis experimentally. The simulation of eye movements adds to this, as we can use this behavioral data to complement imaging data as in previous experimental work in this area (Adams et al. 2016).

## Conclusion

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 3, 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 behavior 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 behaviors 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.

## Funding

Rosetrees Trust (Award Number 173346) to T.P. K.J.F. is a Wellcome Principal Research Fellow (Ref: 088130/Z/09/Z).

## Notes

*Conflict of Interest:* The authors have no disclosures or conflict of interest.

## References

- 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.
- Albert ML. 1973. A simple test of visual neglect. *Neurology*. 23:658.
- Andersen RA, Essick GK, Siegel RM. 1985. Encoding of spatial location by posterior parietal neurons. *Science*. 230:456.
- Andrade K, Samri D, Sarazin M, de Souza LC, Cohen L, Thiebaut de Schotten M, Dubois B, Bartolomeo P. 2010. Visual neglect in posterior cortical atrophy. *BMC Neurol*. 10:68.
- Auclair L, Siérouff E, Kocer S. 2008. A case of spatial neglect dysgraphia in Wilson's Disease. *Arch Clin Neuropsychol*. 23:47–62.
- Bartolomeo P. 2014. Spatially biased decisions: toward a dynamic interactive model of visual neglect. In: Tracy J, Hampstead B, Sathian K, editors. *Cognitive plasticity in neurologic disorders*. Oxford: Oxford University Press. p. 299.
- Bartolomeo P, Thiebaut de Schotten M, Chica AB. 2012. Brain networks of visuospatial attention and their disruption in visual neglect. *Front Hum Neurosci*. 6:110.
- Bartolomeo P, Thiebaut de Schotten M, Doricchi F. 2007. Left unilateral neglect as a disconnection syndrome. *Cereb Cortex*. 17:2479–2490.
- 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. *J Neurosci*. 30:5968.
- Beal MJ. 2003. *Variational algorithms for approximate Bayesian inference*. United Kingdom: University of London.
- Behrens TEJ, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CAM, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, et al. 2003. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat Neurosci*. 6:750–757.
- Berson DM, McIlwain JT. 1983. Visual cortical inputs to deep layers of cat's superior colliculus. *J Neurophysiol*. 50:1143.
- Blei DM, Ng AY, Jordan MI. 2003. Latent Dirichlet allocation. *J Mach Learn Res*. 3:993–1022.
- Bourgeois A, Chica AB, Migliaccio R, Bayle DJ, Duret C, Pradat-Diehl P, Lunven M, Pouget P, Bartolomeo P. 2015. Inappropriate rightward saccades after right hemisphere damage: oculomotor analysis and anatomical correlates. *Neuropsychologia*. 73:1–11.
- Brown TH, Zhao Y, Leung V. 2009. Hebbian plasticity A2—squire. In: Larry R, editor. *Encyclopedia of neuroscience*. Oxford: Academic Press. p. 1049–1056.
- Bruce CJ, Goldberg ME, Bushnell MC, Stanton GB. 1985. Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *J Neurophysiol*. 54:714–734.
- Bundesden C. 1998. A computational theory of visual attention. *Philos Trans R Soc London B Biol Sci*. 353:1271–1281.
- Conant RC, Ashby WR. 1970. Every good regulator of a system must be a model of that system. *Int J Syst Sci*. 1:89–97.
- Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van Essen DC, et al. 1998. A common network of functional areas for attention and eye movements. *Neuron*. 21:761–773.
- 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. *J Cogn Neurosci*. 14:508–523.
- 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. *Annu Rev Neurosci*. 34:569–599.
- 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.
- 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.
- Dayan P, Hinton GE, Neal RM, Zemel RS. 1995. The Helmholtz machine. *Neural Comput*. 7:889–904.
- Di Stefano F, Floris G, Vacca M, Serra G, Cannas A, Borghero G, Marrosu MG, Marrosu F. 2013. Transient unilateral spatial neglect during aura in a woman with sporadic hemiplegic migraine. *Cephalgia*. 33:1194–1197.
- 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. *J Neurosci*. 34:5003–5011.
- Doricchi F, Tomaiuolo F. 2003. The anatomy of neglect without hemianopia: a key role for parietal-frontal disconnection? *Neuroreport*. 14:2239–2243.
- Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE. 2006. A Core System for the implementation of task sets. *Neuron*. 50:799–812.
- Driver J, Baylis GC, Goodrich SJ, Rafal RD. 1994. Axis-based neglect of visual shapes. *Neuropsychologia*. 32:1353–1356.
- Driver J, Pouget A. 2000. Object-centered visual neglect, or relative egocentric neglect? *J Cogn Neurosci*. 12:542–545.
- Duncan J. 2001. An adaptive coding model of neural function in prefrontal cortex. *Nat Rev Neurosci*. 2:820–829.
- Ferber S, Karnath H-O. 2001. How to assess spatial neglect-line bisection or cancellation tasks? *J Clin Exp Neuropsychol*. 23: 599–607.

- FitzGerald T, Dolan R, Friston K. 2014. Model averaging, optimal inference, and habit formation. *Front Hum Neurosci*. doi:10.3389/fnhum.2014.00457.
- Fries W. 1984. Cortical projections to the superior colliculus in the macaque monkey: a retrograde study using horseradish peroxidase. *J Comp Neurol*. 230:55–76.
- Fries W. 1985. Inputs from motor and premotor cortex to the superior colliculus of the macaque monkey. *Behav Brain Res*. 18:95–105.
- Friston K. 2003. Learning and inference in the brain. *Neural Netw*. 16:1325–1352.
- Friston K. 2009. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci*. 13:293–301.
- Friston K, Adams RA, Perrinet L, Breakspear M. 2012. Perceptions as hypotheses: saccades as experiments. *Front Psychol*. 3:151.
- Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, O'Doherty J, Pezzulo G. 2016. Active inference and learning. *Neurosci Biobehav Rev*. 68:862–879.
- Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. 2016. Active inference: a process theory. *Neural Comput*. 29:1–49.
- Friston K, Kilner J, Harrison L. 2006. A free energy principle for the brain. *J Physiol Paris*. 100:70–87.
- Friston K, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G. 2015. Active inference and epistemic value. *Cogn Neurosci*. 6:187–214.
- Friston K, Samothrakis S, Montague R. 2012. Active inference and agency: optimal control without cost functions. *Biol Cybern*. 106:523–541.
- Friston K, Schwartenbeck P, Fitzgerald T, Moutoussis M, Behrens T, Dolan R. 2013. The anatomy of choice: active inference and agency. *Front Hum Neurosci*. 7:598.
- Friston K, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ. 2014. The anatomy of choice: dopamine and decision-making. *Philos Trans R Soc B Biol Sci*. 369:20130481.
- Friston KJ, Rosch R, Parr T, Price C, Bowman H. 2017. Deep temporal models and active inference. *Neurosci Biobehav Rev*. 77:388–402.
- Fruhmann Berger M, Johannsen L, Karnath H-O. 2008. Time course of eye and head deviation in spatial neglect. *Neuropsychologia*. 22:697–702.
- Fullerton KJ, McSherry D, Stout RW. 1986. Albert's test: a neglected test of perceptual neglect. *Lancet*. 327:430–432.
- Gaymard B, Lynch J, Ploner CJ, Condy C, Rivaud-Péchoix S. 2003. The parieto-collicular pathway: anatomical location and contribution to saccade generation. *Eur J Neurosci*. 17:1518–1526.
- Gertler TS, Chan CS, Surmeier DJ. 2008. Dichotomous anatomical properties of adult striatal medium spiny neurons. *J Neurosci*. 28:10814.
- Gilad R, Sadeh M, Boaz M, Lampl Y. 2006. Visual spatial neglect in multiple sclerosis. *Cortex*. 42:1138–1142.
- Goodale MA, Milner AD. 1992. Separate visual pathways for perception and action. *Trends Neurosci*. 15:20–25.
- 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.
- Haber SN. 2003. The primate basal ganglia: parallel and integrative networks. *J Chem Neuroanat*. 26:317–330.
- Halligan PW, Marshall JC. 1998. Neglect of awareness. *Conscious Cogn*. 7:356–380.
- 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.
- Hebb DO. 1949. The first stage of perception: growth of the assembly. *The Organization of Behavior*: pp. 60–78.
- Heilman KM, Howell GJ. 1980. Seizure-induced neglect. *J Neurosurg 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). *Psychol Rev*. 110:29–87.
- Heinke D, Humphreys GW. 2005. Computational models of visual selective attention: a review. *Connect Models Cogn Psychol*. 1:273–312.
- Hikosaka O, Wurtz RH. 1983. Visual and oculomotor functions of monkey substantia nigra pars reticulata. IV. Relation of substantia nigra to superior colliculus. *J Neurophysiol*. 49:1285.
- Hikosaka O, Wurtz RH. 1985a. Modification of saccadic eye movements by GABA-related substances. I. Effect of muscimol and bicuculline in monkey superior colliculus. *J Neurophysiol*. 53:266.
- 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. *J Neurosci*. 25:3161.
- Ho AK, Manly T, Nestor PJ, Sahakian BJ, Bak TH, Robbins TW, Rosser AE, Barker RA. 2003. A case of unilateral neglect in Huntington's disease. *Neurocase*. 9:261–273.
- Hubel DH, Livingstone MS. 1987. Segregation of form, color, and stereopsis in primate area 18. *J Neurosci*. 7:3378.
- 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.
- Itti L, Baldi P. 2006. Bayesian surprise attracts human attention. *Adv Neural Inf Process Syst*. 18:547.
- 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.
- Kanai R, Komura Y, Shipp S, Friston K. 2015. Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos Trans R Soc B Biol Sci*. 370:1668.
- Karnath H-O, Rorden C. 2012. The anatomy of spatial neglect. *Neuropsychologia*. 50:1010–1017.
- Karnath HO, Himmelmach M, Rorden C. 2002. The subcortical anatomy of human spatial neglect: putamen, caudate nucleus and pulvinar. *Brain*. 125:350–360.
- 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. *J Neurosci*. 15:912.
- Kinsbourne M. 1970. A model for the mechanism of unilateral neglect of space. *Trans Am Neurol Assoc*. 95:143–146.
- Kirshner HS. 2003. Chapter 140—speech and language disorders A2—Samuels, Martin A. In: Feske SK, editor. *Office practice of neurology*. 2nd ed. Philadelphia: Churchill Livingstone. p. 890–895.
- 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. *J Neurosci*. 15:928.
- Künzle H, Akert K. 1977. Efferent connections of cortical, area 8 (frontal eye field) in *Macaca fascicularis*. A reinvestigation using the autoradiographic technique. *J Comp Neurol*. 173:147–163.
- Ladavas E, Zeloni G, Zaccara G, Gangemi P. 1997. Eye movements and orienting of attention in patients with visual neglect. *J Cogn Neurosci*. 9:67–74.

- Li D, Karnath H-O, Rorden C. 2014. Egocentric representations of space co-exist with allocentric representations: evidence from spatial neglect. *Cortex*. 58:161–169.
- Livingstone M, Hubel D. 1988. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*. 240:740.
- 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. *Cereb Cortex*. 15:854–869.
- Medina J, Kannan V, Pawlak MA, Kleinman JT, Newhart M, Davis C, Heidler-Gary JE, Herskovits EH, Hillis AE. 2009. Neural substrates of visuospatial processing in distinct reference frames: evidence from unilateral spatial neglect. *J Cogn Neurosci*. 21:2073–2084.
- Mirza MB, Adams RA, Mathys CD, Friston KJ. 2016. Scene construction, visual foraging, and active inference. *Front Comput Neurosci*. 10:56.
- Mongillo G, Barak O, Tsodyks M. 2008. Synaptic theory of working memory. *Science*. 319:1543–1546.
- Moutoussis M, Trujillo-Barreto NJ, El-Deredy W, Dolan RJ, Friston KJ. 2014. A formal model of interpersonal inference. *Front Hum Neurosci*. 8:160.
- Nealey TA, Maunsell JH. 1994. Magnocellular and parvocellular contributions to the responses of neurons in macaque striate cortex. *J Neurosci*. 14:2069.
- 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.
- Parr T, Friston KJ. 2017. The active construction of the visual world. *Neuropsychologia*. 104:92–101.
- Petersen SE, Robinson DL, Keys W. 1985. Pulvinar nuclei of the behaving rhesus monkey: visual responses and their modulation. *J Neurophysiol*. 54:867.
- Pouget A, Sejnowski TJ. 1997. A new view of hemineglect based on the response properties of parietal neurones. *Philos Trans R Soc B Biol Sci*. 352:1449–1459.
- Pouget A, Sejnowski TJ. 2001. Simulating a lesion in a basis function model of spatial representations: comparison with hemineglect. *Psychol Rev*. 108:653.
- Ptak R, Schnider A. 2010. The dorsal attention network mediates orienting toward behaviorally relevant stimuli in spatial neglect. *J Neurosci*. 30:12557.
- Raybourn MS, Keller EL. 1977. Colliculoreticular organization in primate oculomotor system. *J Neurophysiol*. 40:861.
- Robinson DL, Petersen SE. 1992. The pulvinar and visual salience. *Trends Neurosci*. 15:127–132.
- Rorden C, Hjalton H, Fillmore P, Fridriksson J, Kjartansson O, Magnusdottir S, Karnath H-O. 2012. Allocentric neglect strongly associated with egocentric neglect. *Neuropsychologia*. 50:1151–1157.
- 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. *Cereb Cortex*. 16:1418–1430.
- 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. *Cereb Cortex*. 25:3932–3952.
- Schiller PH, Sandell JH, Maunsell JH. 1987. The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. *J Neurophysiol*. 57:1033.
- Schiller PH, True SD, Conway JL. 1980. Deficits in eye movements following frontal eye-field and superior colliculus ablations. *J Neurophysiol*. 44:1175.
- Schomer AC, Drislane FW. 2015. Severe hemispatial neglect as a manifestation of seizures and nonconvulsive status epilepticus: utility of prolonged EEG monitoring. *J Clin Neurophysiol*. 32:e4–e7.
- Schwartenbeck P, FitzGerald TH, Mathys C, Dolan R, Friston K. 2015. The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cereb Cortex*. 25:3434–3445.
- Schwartenbeck P, FitzGerald TH, Mathys C, Dolan R, Wurst F, Kronbichler M, Friston K. 2015. Optimal inference with sub-optimal models: addiction and active Bayesian inference. *Med Hypotheses*. 84:109–117.
- Schwartenbeck P, Friston K. 2016. Computational phenotyping in psychiatry: a worked example. *eNeuro*. 3:ENEURO.0049-0016.2016.
- Shipp S. 2003. The functional logic of cortico-pulvinar connections. *Philos Trans R Soc B Biol Sci*. 358:1605–1624.
- Shipp S. 2004. The brain circuitry of attention. *Trends Cogn Sci*. 8:223–230.
- Shomstein S, Lee J, Behrmann M. 2010. Top-down and bottom-up attentional guidance: investigating the role of the dorsal and ventral parietal cortices. *Exp Brain Res*. 206:197–208.
- Shook BL, Schlag-Rey M, Schlag J. 1990. Primate supplementary eye field: I. Comparative aspects of mesencephalic and pontine connections. *J Comp Neurol*. 301:618–642.
- 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. *J Neurosci*. 29:4392–4407.
- Stevens MC, Calhoun VD, Kiehl KA. 2005. Hemispheric differences in hemodynamics elicited by auditory oddball stimuli. *NeuroImage*. 26:782–792.
- 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.
- 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.
- Turtzo LC, Kleinman JT, Llinas RH. 2008. Capgras syndrome and unilateral spatial neglect in nonconvulsive status epilepticus. *Behav Neurol*. 20:61–64.
- Ungerleider LG, Christensen CA. 1979. Pulvinar lesions in monkeys produce abnormal scanning of a complex visual array. *Neuropsychologia*. 17:493–501.
- Ungerleider LG, Haxby JV. 1994. 'What' and 'where' in the human brain. *Curr Opin Neurobiol*. 4:157–165.
- Veale R, Hafed ZM, Yoshida M. 2017. How is visual salience computed in the brain? Insights from behaviour, neurobiology and modelling. *Philos Trans R Soc B Biol Sci*. 372:1714.
- 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.
- Vossel S, Weidner R, Driver J, Friston KJ, Fink GR. 2012. Deconstructing the architecture of dorsal and ventral attention systems with dynamic causal modeling. *J Neurosci*. 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. *Ann Math Stat*. 18:549–555.
- Warrington EK, James M. 1967. Disorders of visual perception in patients with localised cerebral lesions. *Neuropsychologia*. 5:253–266.
- Warrington EK, James M. 1988. Visual apperceptive agnosia: a clinico-anatomical study of three cases. *Cortex*. 24:13–32.
- Warrington EK, Taylor AM. 1973. The contribution of the right parietal lobe to object recognition. *Cortex*. 9:152–164.
- Weller RE, Steele GE, Kaas JH. 2002. Pulvinar and other subcortical connections of dorsolateral visual cortex in monkeys. *J Comp Neurol*. 450:215–240.
- Yin HH, Knowlton BJ. 2006. The role of the basal ganglia in habit formation. *Nat Rev Neurosci*. 7:464–476.
- Yue Y, Song W, Huo S, Wang M. 2012. Study on the occurrence and neural bases of hemispatial neglect with different reference frames. *Arch Phys Med Rehabil*. 93:156–162.
- 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. *Eur J Neurosci*. 1: 494–506.

## Appendix

### Update equations

Using the mean field approximation, and factorizing the joint distribution according to the generative model, the free energy can be rewritten as follows:

$$F = E_{Q(A|\pi)Q(\pi)} \prod_{\tau} Q(s_{\tau}|\pi) \left[ \ln Q(A|\pi)Q(\pi) \prod_{\tau} Q(s_{\tau}|\pi) - \ln P(\pi)P(A)P(s_1)P(o_1|s_1) \prod_{\tau=2} P(s_{\tau}|s_{\tau-1}, \pi) P(o_{\tau}|s_{\tau}) \right]$$

To find the variational solution to the approximate posterior for  $s_t|\pi$ , we can omit all terms which are constant with respect to this variable. This gives:

$$F(\pi) = E_{Q(s_t|\pi)}[\ln Q(s_t|\pi)] - E_{Q(A|\pi)Q(s_t|\pi)}[\ln P(o_t|s_t, A)] - E_{Q(s_{t-1}|\pi)Q(s_t|\pi)}[\ln P(s_t|s_{t-1}, \pi)] - E_{Q(s_t|\pi)Q(s_{t+1}|\pi)}[\ln P(s_{t+1}|s_t, \pi)]$$

Taking the functional derivative with respect to  $Q(s_t|\pi)$

$$\frac{\delta F(\pi)}{\delta Q(s_t|\pi)} = \ln Q(s_t|\pi) - E_{Q(A|\pi)}[\ln P(o_t|s_t, A)] - E_{Q(s_{t-1}|\pi)}[\ln P(s_t|s_{t-1}, \pi)] - E_{Q(s_{t+1}|\pi)}[\ln P(s_{t+1}|s_t, \pi)]$$

Setting equal to zero and rearranging gives the variational solution

$$\begin{aligned} \ln Q(s_t|\pi) &= E_{Q(A|\pi)}[\ln P(o_t|s_t, A)] \\ &+ E_{Q(s_{t-1}|\pi)}[\ln P(s_t|s_{t-1}, \pi)] \\ &+ E_{Q(s_{t+1}|\pi)}[\ln P(s_{t+1}|s_t, \pi)] \\ &= \ln \mathbf{A} \cdot \mathbf{o}_t + \ln \mathbf{B}_{\pi, t-1} \mathbf{s}_{\pi, t-1} + \ln \mathbf{B}_{\pi, t} \cdot \mathbf{s}_{\pi, t+1} \end{aligned}$$

This can be used to set up a biologically plausible gradient descent scheme by defining an error term, which expresses the difference between this and the current belief

$$\epsilon_{\pi, t} = \ln \mathbf{A} \cdot \mathbf{o}_t + \ln \mathbf{B}_{\pi, t-1} \mathbf{s}_{\pi, t-1} + \ln \mathbf{B}_{\pi, t} \cdot \mathbf{s}_{\pi, t+1} - \ln \mathbf{s}_{\pi, t}$$

When this error is positive,  $\mathbf{s}_{\pi, t}$  should increase. The gradient descent on variational free energy then corresponds to an ascent of the form:

$$\mathbf{s}_{\pi, t} = \sigma(\nu_{\pi, t}) : \dot{\nu}_{\pi, t} = \epsilon_{\pi, t}$$

These are the dynamics for the state updates in Figure 2.