Computational Neuropsychology and Bayesian Inference

Thomas Parr[1]*, Geraint Rees[1,2] and Karl J. Friston[1]

[1] Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, United Kingdom,
[2] Institute of Cognitive Neuroscience, University College London, London, United Kingdom

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 paper, we provide a narrative review of the body of computational research addressing neuropsychological syndromes, and focus on those that employ Bayesian frameworks. 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 characterized by the set of prior beliefs that would make a patient's behavior 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 emphasize 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.

Keywords: neuropsychology, active inference, predictive coding, computational phenotyping, precision

INTRODUCTION

The process of relating brain dysfunction to cognitive and behavioral 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 behavioral deficit (Husain and Nachev, 2007). The problem is further complicated by the distributed nature of 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 (Price et al., 2001; Carrera and Tononi, 2014). 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 (Rocca et al., 2007; Grefkes et al., 2008; Abutalebi et al., 2009; Mintzopoulos et al., 2009) in response to lesions.

In this paper, we consider a complementary approach that has started to gain traction in psychiatric research (Adams et al., 2013b, 2015; Corlett and Fletcher, 2014; Huys et al., 2016; Schwartenbeck and Friston, 2016; Friston et al., 2017c). This is the use of models that relate the computations performed by the brain to measurable behaviors (Krakauer and Shadmehr, 2007; Mirza et al., 2016; Testolin and Zorzi, 2016; Iglesias et al., 2017). 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.

Bayesian inference is the process of forming beliefs about the causes of sensory data. It relies upon the combination of prior beliefs about these causes, and beliefs about how these causes give rise to sensations. Using these two probabilities it is possible to calculate the probability, given a sensation, of its cause. This is known as a 'posterior' probability. This means that prior beliefs are updated by a sensory experience to become posterior beliefs. These posteriors can then be used as the prior for the next sensory experience. In short, Bayesian theories of brain function propose the brain encodes beliefs about the causes of sensory data, and that these beliefs are updated in response to new sensory evidence.

Our motivation for pursuing a Bayesian framework is that it captures many different types of behavior, including apparently suboptimal behaviors. According to an important result known as *the complete class theorem* (Wald, 1947; Daunizeau et al., 2010), there is always a set of a prior belief that renders an observed behavior Bayes optimal. This is fundamental for computational neuropsychology as it means we can cast even pathological behaviors as the result of processes that implement Bayesian inference (Schwartenbeck et al., 2015). 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 (Knill and Pouget, 2004; Doya, 2007), 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 between these models, and ensures they could all 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
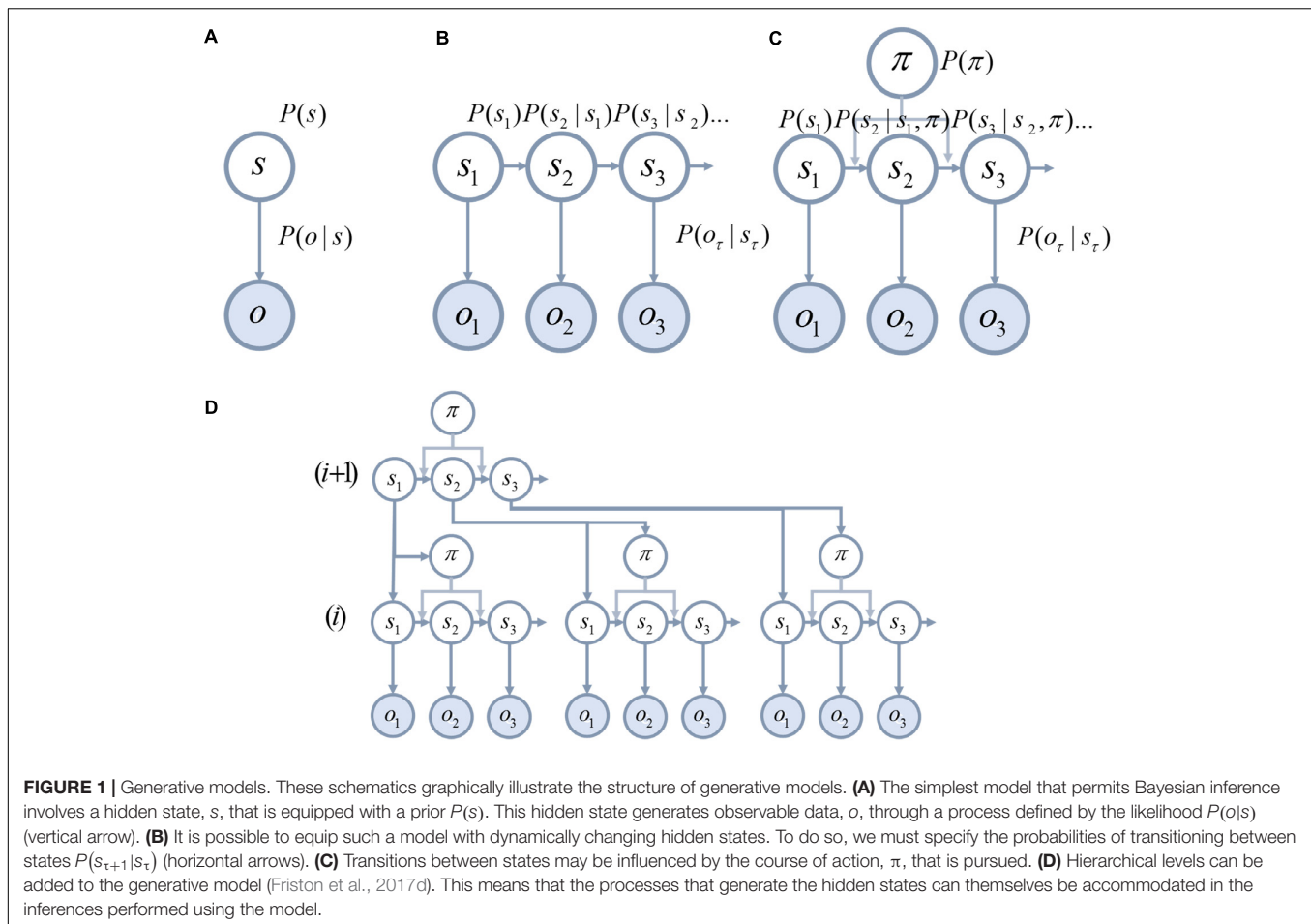
phrased in terms of two questions: 'what are the prior beliefs that would have to be held to make this behavior optimal?' and 'what are the biological substrates of these priors?'

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 behavior. 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., 2015). If pathological priors relate to the properties of the musculoskeletal system, we might expect motor disorders such as tremors or paralysis (Friston et al., 2010; Adams et al., 2013a). If abnormal priors relate to perceptual systems, the results may include sensory hallucinations (Fletcher and Frith, 2009; Adams et al., 2013b) or anesthesia. 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 (Geschwind, 1965a; Catani and Ffytche, 2005).

THE GENERATIVE MODEL

Bayesian Inference

Much work in theoretical neurobiology rests on the notion that the brain performs Bayesian inference (Knill and Pouget, 2004; Doya, 2007; Friston, 2010; 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 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 1B**). These dynamics can be subdivided into those that a subject has control

FIGURE 1 | Generative models. These schematics graphically illustrate the structure of generative models. **(A)** The simplest model that permits Bayesian inference involves a hidden state, *s*, that is equipped with a prior $P(s)$. This hidden state generates observable data, *o*, through a process defined by the likelihood $P(o|s)$ (vertical arrow). **(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). **(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., 2017d). This means that the processes that generate the hidden states can themselves be accommodated in the inferences performed using the model.

over (**Figure 1C**), such as muscle length, and environmental causes that they cannot directly influence.

Predictive Coding

Predictive coding is a prominent theory describing how the brain could perform Bayesian inference (Rao and Ballard, 1999; Friston and Kiebel, 2009; Bastos et al., 2012). 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. Under this theory, the messages passed through neuronal signaling are either predictions, or prediction errors. There are other local message passing schemes that can implement Bayesian inference (Winn and Bishop, 2005; Yedidia et al., 2005; Dauwels, 2007; Friston et al., 2017b), particularly for categorical (as opposed to continuous) inferences. Although we use the language of predictive coding in the following, we note that our discussion generalizes 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 (Teunisse et al., 1996; Menon et al., 2003), 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 (Frith et al., 2000; De Ridder et al., 2014). 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 1D**). As the hierarchy is ascended, causes tend to become more abstract, and have dynamics that play out over a longer time course (Kiebel et al.,

2008, 2009). An intuitive example is the kind of generative model required for reading (Friston et al., 2017d). 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 organization. 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., 2008, 2015; Kiebel et al., 2008; Murray et al., 2014). This is consistent with the structure of deep temporal generative models (Friston et al., 2017d) (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; Shipp, 2007; Markov et al., 2013). As illustrated in **Figure 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.

Ascending and Descending Messages

The parallel between the hierarchical structure of generative models and that of cortical organization 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 (Kiebel et al., 2008; Friston et al., 2017d). 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, and a summary is presented in **Table 1**. Under predictive coding, the signals carried by inter-areal connections have a clear interpretation (Shipp et al., 2013; Shipp, 2016). 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.

It has been argued that deficits in semantic knowledge can only be interpreted with reference to a hierarchically organized set of representations in the brain. This argument rests on observations that patients with agnosia, a failure to recognize
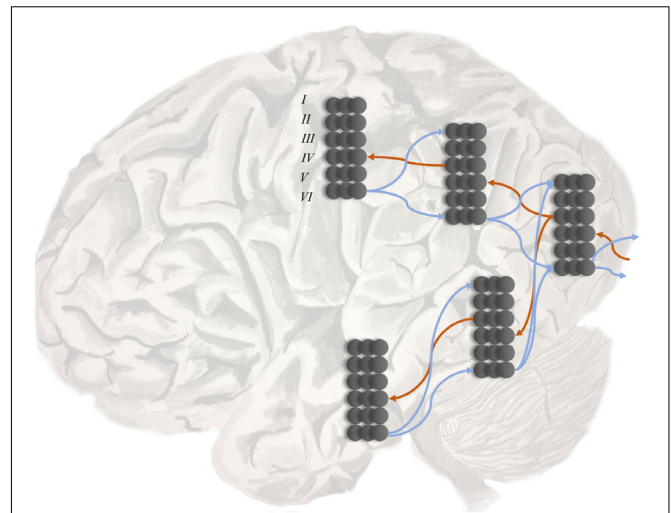

FIGURE 2 | Hierarchy in the cortex. This schematic illustrates two key features of cortical organization. The first is hierarchy, as defined by laminar specific projections. Projections from primary sensory areas, such as area VI, to higher cortical areas typically arise from layer III of a cortical 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 factorization 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).

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 the 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.

SENSORY STREAMS AND DISCONNECTION SYNDROMES

What and where?

Figure 2 illustrates an additional feature common to cortical architectures and inference methods. This is the factorization

TABLE 1 | Bayesian computational neuropsychology.

Syndrome	Abnormal prior	Neurobiology	Reference
Anosognosia	Low exteroceptive or interoceptive sensory precision	Insula lesions	Karnath et al., 2005; Fotopoulou et al., 2010; Fotopoulou, 2012
	Failure of *active* inference	Hemiplegia	
Apraxia	Disrupted likelihood (action to vision or command to action consequences)	Callosal disconnection Left frontoparietal disconnection	Geschwind, 1965b
Autism	High sensory precision	↑Cholinergic transmission?	Dayan and Yu, 2006; Lawson et al., 2014; Marshall et al., 2016; Lawson et al., 2017
	Secondary to high volatility	↑Noradrenergic transmission	
Complex visual hallucinations (Lewy body dementia, Charles Bonnet syndrome)	Low sensory precision	↓Cholinergic transmission	Collerton et al., 2005; Reichert et al., 2013; O'Callaghan et al., 2017
	Disrupted likelihood mapping	Retino-geniculate disconnection	
Conduction aphasia	Disrupted likelihood mapping (speech to proprioceptive consequences)	Arcuate fasciculus disconnection	Wernicke, 1969
Parkinson's disease	Low prior precision over policies	↓Dopaminergic transmission	Friston et al., 2013
Visual agnosia	Disrupted likelihood ('what' to sensory data)	Ventral visual stream disconnection	Geschwind, 1965b
Visual neglect	Disrupted likelihood (fixation to 'what') mapping	SLF II disconnection	Karnath et al., 2002; Thiebaut de Schotten et al., 2005; Bartolomeo et al., 2007
	Biased outcome prior	Pulvinar lesion	
	Biased policy prior	Putamen lesion	

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 factorization of visual hierarchies into ventral and dorsal 'what' and 'where' streams (Ungerleider and Mishkin, 1982; Ungerleider and Haxby, 1994) appears to be an example of this. A closely related factorization separates the dorsal and ventral attention networks (Corbetta and Shulman, 2002). This factorization 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, 1988; Warrington and Taylor, 1973). Lesions to contralateral (left hemispheric) homologs are more likely to give rise to difficulties in naming objects (Kirshner, 2003).
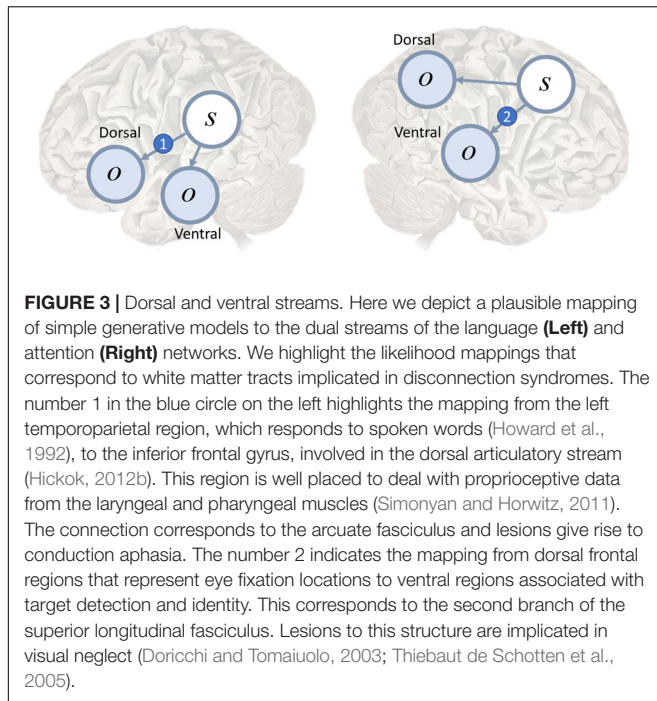
The bilateral representation of space has an important consequence when we frame neuronal processing as probabilistic inference. Following an inference that a stimulus is likely to be on one side of space, it must be the case that it is less likely to be on the contralateral side. If neuronal activities in each hemisphere represent these probabilities, this induces a form of interhemispheric competition (Vuilleumier et al., 1996; Rushmore et al., 2006; Dietz et al., 2014). An important role of commissural fiber pathways may be to enforce the normalization 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 et al., 2007; Bartolomeo, 2014). Any intrahemispheric lesion that induces a bias toward one side of space necessarily alters the interhemispheric balance of activity (Parr and Friston, 2017b).

Disconnections and Likelihoods

The factorization 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 factorization 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,b), 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

FIGURE 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).

as in **Figure 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, harmonizes 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) generalizes 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 behaviors. This implies a disconnection between visual and motor regions (Goldenberg, 2003; Catani and Ffytche, 2005). 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 (Geschwind, 1965a; Catani and Ffytche, 2005) visual agnosia, caused by disruption of connections in the ventral visual stream, and visual neglect (Doricchi and Tomaiuolo, 2003; Bartolomeo et al., 2007; He et al., 2007; Ciaraffa et al., 2013). Neglect can be a consequence of frontoparietal disconnections (**Figure 3**, right), leading to an impaired awareness of stimuli on the left despite intact early visual processing (Rees et al., 2000). We consider the behavioral manifestations of visual neglect in a later section. Before we do so, we turn from disconnections to a subtler form of computational pathology.

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 noisy, the gain of the prediction error signal is turned down. A very 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.

The generative models depicted in **Figure 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 (Friston and Kiebel, 2009; Feldman and Friston, 2010). 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; Yu and Dayan, 2002; Moran et al., 2013), 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 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 optimize sequences of actions into the future (O'Reilly
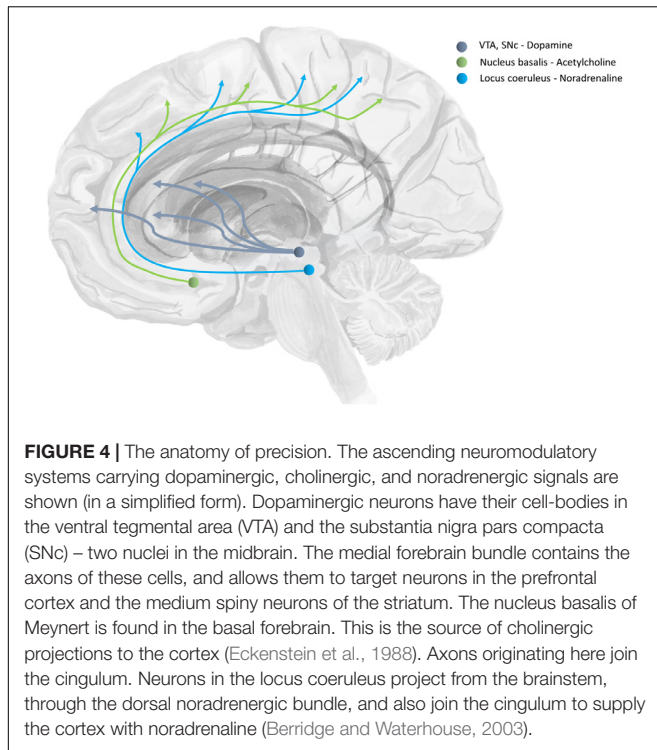
FIGURE 4 | The anatomy of precision. The ascending neuromodulatory systems carrying dopaminergic, cholinergic, and noradrenergic signals are shown (in a simplified form). Dopaminergic neurons have their cell-bodies in the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNc) – two 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).

and Frank, 2006). Deficient cholinergic signaling has been implicated in the complex visual hallucinations associated with some neurodegenerative conditions (Collerton et al., 2005).

Precision and Autism

One condition that has received considerable attention using Bayesian formulations is autism (Pellicano and Burr, 2012; Lawson et al., 2014). 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 (Geisler and Kersten, 2002; Brown and Friston, 2012). 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 'noisiness' (or 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,b). 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 to an active process (Gibson, 1966; Ognibene and Baldassare, 2015; 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 (Lindley, 1956; Friston et al., 2016b; Clark, 2017). A mathematical

formulation of this imperative (Friston et al., 2015) 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., 2017d). 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, 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 characterized by a failure to attend to the left side of space. This rightward lateralization may be a consequence of the mean-field factorization 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 behavioral manifestation of this disorder is a bias in saccadic sampling (Husain et al., 2001; Fruhmann Berger et al., 2008; 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 behavior optimal. We will discuss three possibilities (Parr and Friston, 2017b), and consider their biological bases (**Figure 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 (Ungerleider and Christensen, 1979; Robinson and Petersen, 1992; Kanai et al., 2015). 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 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
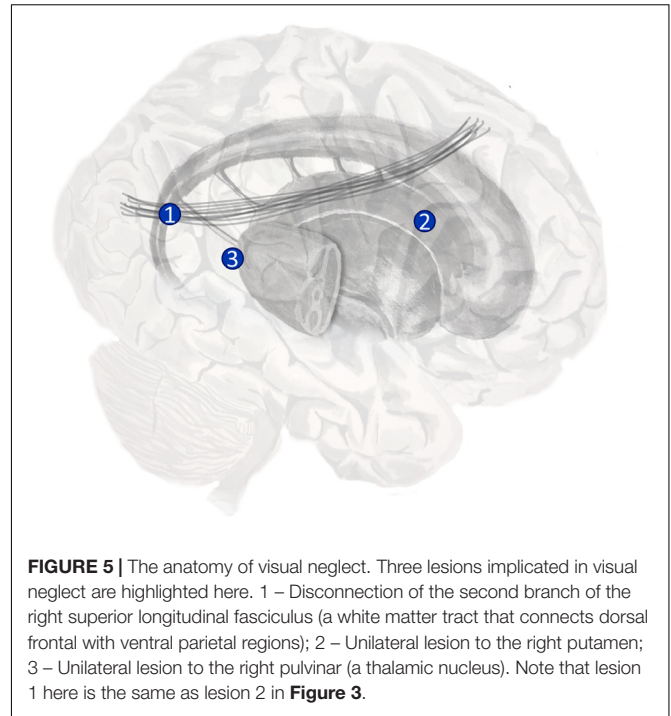

FIGURE 5 | The anatomy of visual neglect. Three lesions implicated in visual neglect are highlighted here. 1 – Disconnection of the second branch of the right superior longitudinal fasciculus (a white matter tract that connects dorsal frontal with ventral parietal regions); 2 – Unilateral lesion to the 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 3**.

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 emphasizes proprioceptive (rather than visual) consequences of action (Committeri et al., 2007). In these circumstances, optimal behavior would require a greater frequency of rightward saccades. Lesions to SLF II (Doricchi and Tomaiuolo, 2003; Thiebaut de Schotten et al., 2005; Lunven et al., 2015), and the regions it connects (Corbetta et al., 2000; Corbetta and Shulman, 2002, 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 (Yin and Knowlton, 2006; Graybiel and Grafton, 2015). Habits may be formalized 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 behavioral bias toward 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.

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 paralyzed 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). 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 (Karnath et al., 2005; Fotopoulou et al., 2010). This is a region often associated with interoceptive inference (Barrett and Simmons, 2015) that has substantial efferent connectivity to somatosensory cortex (Showers and Lauer, 1961; Mesulam and Mufson, 1982). 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 is generated from hidden states. 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 factorization, we noted that such disconnections are likely to have a hemispheric

asymmetry in the behaviors 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, behavior 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 behavior. 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, not discussed in the above, 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 toward 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 modeling 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 behavior. 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 behavior in patients is promising both scientifically and clinically. If individual patients can be uniquely characterized by subject-specific priors, this facilitates a precision medicine approach grounded in computational phenotyping (Adams et al., 2016; Schwartenbeck and Friston, 2016; Mirza et al., 2018). 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.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Abutalebi, J., Rosa, P. A. D., Tettamanti, M., Green, D. W., and Cappa, S. F. (2009). Bilingual aphasia and language control: a follow-up fMRI and intrinsic connectivity study. *Brain Lang.* 109, 141–156. doi: 10.1016/j.bandl.2009.03.003

Adams, R. A., Bauer, M., Pinotsis, D., and Friston, K. J. (2016). Dynamic causal modelling of eye movements during pursuit: confirming precision-encoding in V1 using MEG. *Neuroimage* 132, 175–189. doi: 10.1016/j.neuroimage.2016.02.055

Adams, R. A., Huys, Q. J., and Roiser, J. P. (2015). Computational psychiatry: towards a mathematically informed understanding of mental illness. *J. Neurol. Neurosurg. Psychiatry* 87, 53–63. doi: 10.1136/jnnp-2015-310737

Adams, R. A., Shipp, S., and Friston, K. J. (2013a). Predictions not commands: active inference in the motor system. *Brain Struct. Funct.* 218, 611–643. doi: 10.1007/s00429-012-0475-5

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., and Friston, K. J. (2013b). The computational anatomy of psychosis. *Front. Psychiatry* 4:47. doi: 10.3389/fpsyt.2013.00047

Allen, M., Fardo, F., Dietz, M. J., Hillebrandt, H., Friston, K. J., Rees, G., et al. (2016). Anterior insula coordinates hierarchical processing of tactile mismatch responses. *Neuroimage* 127(Suppl. C), 34–43. doi: 10.1016/j.neuroimage.2015.11.030

Barrett, L. F., and Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nat. Rev. Neurosci.* 16, 419–429. doi: 10.1038/nrn3950

Bartolomeo, P. (2014). "Spatially biased decisions: toward a dynamic interactive model of visual neglect," in *Cognitive Plasticity in Neurologic Disorders*, eds J. I. Tracy, B. Hampstead, and K. Sathian (Oxford: Oxford University Press), 299.

Bartolomeo, P., Thiebaut de Schotten, M., and Doricchi, F. (2007). Left unilateral neglect as a disconnection syndrome. *Cereb. Cortex* 17, 2479–2490. doi: 10.1093/cercor/bhl181

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion–symptom mapping. *Nat. Neurosci.* 6, 448–450. doi: 10.1038/nn1050

Beal, M. J. (2003). *Variational Algorithms for Approximate Bayesian Inference.* London: University of London.

Berridge, C. W., and Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Res. Rev.* 42, 33–84. doi: 10.1016/S0165-0173(03)00143-7

Binder, J. R. (2015). The Wernicke area: modern evidence and a reinterpretation. *Neurology* 85, 2170–2175. doi: 10.1212/WNL.0000000000002219

Boes, A. D., Prasad, S., Liu, H., Liu, Q., Pascual-Leone, A., Caviness, V. S., et al. (2015). Network localization of neurological symptoms from focal brain lesions. *Brain* 138, 3061–3075. doi: 10.1093/brain/awv228

Brown, H., and Friston, K. J. (2012). Free-energy and illusions: the cornsweet effect. *Front. Psychol.* 3:43. doi: 10.3389/fpsyg.2012.00043

Carrera, E., and Tononi, G. (2014). Diaschisis: past, present, future. *Brain* 137, 2408–2422. doi: 10.1093/brain/awu101

Catani, M., and Ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain* 128, 2224–2239. doi: 10.1093/brain/awh622

Catani, M., and Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex* 44, 953–961. doi: 10.1016/j.cortex.2008.04.002

Ciaraffa, F., Castelli, G., Parati, E. A., Bartolomeo, P., and Bizzi, A. (2013). Visual neglect as a disconnection syndrome? A confirmatory case report. *Neurocase* 19, 351–359. doi: 10.1080/13554794.2012.667130

Clark, A. (2017). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenol. Cogn. Sci.* 1–14.

Collerton, D., Perry, E., and McKeith, I. (2005). Why people see things that are not there: a novel perception and attention deficit model for recurrent complex visual hallucinations. *Behav. Brain Sci.* 28, 737–757. doi: 10.1017/S0140525X05000130

Committeri, G., Pitzalis, S., Galati, G., Patria, F., Pelle, G., Sabatini, U., et al. (2007). Neural bases of personal and extrapersonal neglect in humans. *Brain* 130, 431–441. doi: 10.1093/brain/awl265

Conant, R. C., and Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* 1, 89–97. doi: 10.1080/00207727008920220

Cooray, G. K., Sengupta, B., Douglas, P., Englund, M., Wickstrom, R., and 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

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., and Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3, 292–297. doi: 10.1038/73009

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Corbetta, M., and Shulman, G. L. (2011). Spatial neglect and attention networks. *Annu. Rev. Neurosci.* 34, 569–599. doi: 10.1146/annurev-neuro-061010-113731

Corlett, P. R., and Fletcher, P. C. (2014). Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *Lancet Psychiatry* 1, 399–402. doi: 10.1016/S2215-0366(14)70298-6

Daunizeau, J., den Ouden, H. E. M., Pessiglione, M., Kiebel, S. J., Stephan, K. E., and Friston, K. J. (2010). Observing the observer (I): meta-bayesian models of learning and decision-making. *PLoS One* 5:e15554. doi: 10.1371/journal.pone.0015554

Dauwels, J. (2007). "On variational message passing on factor graphs," in *Proceedings of the IEEE International Symposium Information Theory (ISIT), 24–29 June 2007*, Nice, 2546–2550. doi: 10.1109/ISIT.2007.4557602

Dayan, P., and Yu, A. J. (2001). *ACh, Uncertainty, and Cortical Inference.* Montreal, QC: NIPS.

Dayan, P., and Yu, A. J. (2006). Phasic norepinephrine: a neural interrupt signal for unexpected events. *Network* 17, 335–350. doi: 10.1080/09548980600104024

De Ridder, D., Vanneste, S., and Freeman, W. (2014). The Bayesian brain: phantom percepts resolve sensory uncertainty. *Neurosci. Biobehav. Rev.* 44(Suppl. C), 4–15. doi: 10.1016/j.neubiorev.2012.04.001

Desimone, R., Schein, S. J., Moran, J., and Ungerleider, L. G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Res.* 25, 441–452. doi: 10.1016/0042-6989(85)90069-0

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180. doi: 10.1007/BF00230027

Dietz, M. J., Friston, K. J., Mattingley, J. B., Roepstorff, A., and Garrido, M. I. (2014). Effective connectivity reveals right-hemisphere dominance in audiospatial perception: implications for models of spatial neglect. *J. Neurosci.* 34, 5003–5011. doi: 10.1523/JNEUROSCI.3765-13.2014

Doricchi, F., and Tomaiuolo, F. (2003). The anatomy of neglect without hemianopia: a key role for parietal–frontal disconnection? *Neuroreport* 14, 2239–2243.

Doya, K. (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding.* Cambridge, MA: MIT Press.

Dronkers, N. F., and Baldo, J. V. (2009). "Language: aphasia A2 - squire," in *Encyclopedia of Neuroscience,* ed. R. Larry (Oxford: Academic Press), 343–348.

Eckenstein, F. P., Baughman, R. W., and Quinn, J. (1988). An anatomical study of cholinergic innervation in rat cerebral cortex. *Neuroscience* 25, 457–474. doi: 10.1016/0306-4522(88)90251-5

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., and Friston, K. J. (2012). A Bayesian account of 'hysteria'. *Brain* 135, 3495–3512. doi: 10.1093/brain/aws129

Feldman, H., and Friston, K. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1

Ffytche, D. H., and Howard, R. J. (1999). The perceptual consequences of visual loss: 'positive' pathologies of vision. *Brain* 122, 1247–1260. doi: 10.1093/brain/122.7.1247

FitzGerald, T. H. B., Dolan, R. J., and Friston, K. J. (2014). Model averaging, optimal inference, and habit formation. *Front. Hum. Neurosci.* 8:457. doi: 10.3389/fnhum.2014.00457

Fletcher, P. C., and Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

Fornito, A., Zalesky, A., and 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. *J. Neuropsychol.* 8, 1–19. doi: 10.1111/jnp.12010

Fotopoulou, A., Pernigo, S., Maeda, R., Rudd, A., and Kopelman, M. A. (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., and 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., and Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* 36, 1133–1140. doi: 10.1016/S0028-3932(98)00006-2

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science* 306, 1940–1943. doi: 10.1126/science.1102941

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K., Adams, R. A., Perrinet, L., and Breakspear, M. (2012a). Perceptions as hypotheses: saccades as experiments. *Front. Psychol.* 3:151. doi: 10.3389/fpsyg.2012.00151

Friston, K., Brown, H. R., Siemerkus, J., and Stephan, K. E. (2016a). The dysconnection hypothesis 2016). *Schizophr. Res.* 176, 83–94. doi: 10.1016/j.schres.2016.07.014

Friston, K., and Buzsáki, G. (2016). The functional anatomy of time: what and when in the brain. *Trends Cogn. Sci.* 20, 500–511. doi: 10.1016/j.tics.2016.05.001

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J., and Pezzulo, G. (2016b). Active inference and learning. *Neurosci. Biobehav. Rev.* 68, 862–879. doi: 10.1016/j.neubiorev.2016.06.022

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214. doi: 10.1080/17588928.2015.1020053

Friston, K., Samothrakis, S., and Montague, R. (2012b). Active inference and agency: optimal control without cost functions. *Biol. Cybern.* 106, 523–541.

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., and Dolan, R. (2013). The anatomy of choice: active inference and agency. *Front. Hum. Neurosci.* 7:598. doi: 10.3389/fnhum.2013.00598

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130481. doi: 10.1098/rstb.2013.0481

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern.* 102, 227–260. doi: 10.1007/s00422-010-0364-z

Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017a). Active inference: a process theory. *Neural Comput.* 29, 1–49. doi: 10.1162/NECO_a_00912

Friston, K. J., Parr, T., and Vries, B. D. (2017b). The graphical brain: belief propagation and active inference. *Netw. Neurosci.* 1, 381–414. doi: 10.1162/NETN_a_00018

Friston, K. J., Redish, A. D., and Gordon, J. A. (2017c). Computational nosology and precision psychiatry. *Comput. Psychiatry* 1, 2–23. doi: 10.1162/CPSY_a_00001

Friston, K. J., Rosch, R., Parr, T., Price, C., and Bowman, H. (2017d). Deep temporal models and active inference. *Neurosci. Biobehav. Rev.* 77, 388–402. doi: 10.1016/j.neubiorev.2017.04.009

Frith, C. D., Blakemore, S. J., and Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philos. Trans. R. Soc. B Biol. Sci.* 355, 1771–1788. doi: 10.1098/rstb.2000.0734

Frith, C. D., and Frith, U. (1999). Interacting minds—a biological basis. *Science* 286, 1692–1695. doi: 10.1126/science.286.5445.1692

Fruhmann Berger, M., Johannsen, L., and Karnath, H.-O. (2008). Time course of eye and head deviation in spatial neglect. *Neuropsychology* 22, 697–702. doi: 10.1037/a0013351

Gallant, J., Braun, J., and Van Essen, D. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259, 100–103. doi: 10.1126/science.8418487

Geisler, W. S., and Kersten, D. (2002). Illusions, perception and Bayes. *Nat. Neurosci.* 5, 508–510. doi: 10.1038/nn0602-508

Geschwind, N. (1965a). Disconnexion syndromes in animals and man1. *Brain* 88, 237–294. doi: 10.1093/brain/88.2.237

Geschwind, N. (1965b). Disconnexion syndromes in animals and man. II. *Brain* 88, 585–644. doi: 10.1093/brain/88.3.585

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature* 521, 452–459. doi: 10.1038/nature14541

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems.* Boston, MA: Houghton Mifflin Company.

Glickstein, M., and Berlucchi, G. (2008). Classical disconnection studies of the corpus callosum. *Cortex* 44, 914–927. doi: 10.1016/j.cortex.2008.04.001

Goldenberg, G. (2003). Apraxia and beyond: life and work of Hugo Liepmann. *Cortex* 39, 509–524. doi: 10.1016/S0010-9452(08)70261-2

Graybiel, A. M., and Grafton, S. T. (2015). The striatum: where skills and habits meet. *Cold Spring Harb. Perspect. Biol.* 7:a021691. doi: 10.1101/cshperspect.a021691

Grefkes, C., Nowak, D. A., Eickhoff, S. B., Dafotakis, M., Küst, J., Karbe, H., et al. (2008). Cortical connectivity after subcortical stroke assessed with functional magnetic resonance imaging. *Ann. Neurol.* 63, 236–246. doi: 10.1002/ana.21228

Gregory, R. L. (1970). *The Intelligent Eye.* London: Weidenfeld and Nicolson.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197. doi: 10.1098/rstb.1980.0090

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *J. Neurophysiol.* 35, 96–111. doi: 10.1152/jn.1972.35.1.96

Gurney, K., Prescott, T. J., and Redgrave, P. (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biol. Cybern.* 84, 401–410. doi: 10.1007/PL00007984

Halligan, P. W., and Marshall, J. C. (1998). Neglect of awareness. *Conscious. Cogn.* 7, 356–380. doi: 10.1006/ccog.1998.0362

Happé, F. G. E. (1996). Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. A research note. *J. Child Psychol. Psychiatry* 37, 873–877. doi: 10.1111/j.1469-7610.1996.tb01483.x

Hasson, U., Chen, J., and Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn. Sci.* 19, 304–313. doi: 10.1016/j.tics.2015.04.006

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci* 28, 2539–2550. doi: 10.1523/JNEUROSCI.5487-07.2008

He, B. J., Snyder, A. Z., Vincent, J. L., Epstein, A., Shulman, G. L., and Corbetta, M. (2007). Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron* 53, 905–918. doi: 10.1016/j.neuron.2007.02.013

Hickok, G. (2012a). Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* 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. *J. Commun. Disord.* 45, 393–402. doi: 10.1016/j.jcomdis.2012.06.004

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113

Hikosaka, O., Takikawa, Y., and Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.* 80, 953–978. doi: 10.1152/physrev.2000.80.3.953

Howard, D., Patterson, K., Wise, R., Brown, W. D., Friston, K., Weiller, C., et al. (1992). The cortical localization of the lexicons: positron emission tomography evidence. *Brain* 115, 1769–1782. doi: 10.1093/brain/115.6.1769

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148, 574–591. doi: 10.1113/jphysiol.1959.sp006308

Humphreys, G. W., and Forde, E. M. (2001). Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *Behav. Brain Sci.* 24, 453–476.

Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J., and Kennard, C. (2001). Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect. *Brain* 124(Pt 5), 941–952. doi: 10.1093/brain/124.5.941

Husain, M., and Nachev, P. (2007). Space and the parietal cortex. *Trends Cogn. Sci.* 11, 30–36. doi: 10.1016/j.tics.2006.10.011

Huys, Q. J., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* 19, 404–413. doi: 10.1038/nn.4238

Iglesias, S., Tomiello, S., Schneebeli, M., and Stephan, K. E. (2017). Models of neuromodulation for computational psychiatry. *Wiley Interdiscip. Rev. Cogn. Sci.* 8:e1420. doi: 10.1002/wcs.1420

Jahanshahi, M., Obeso, I., Rothwell, J. C., and Obeso, J. A. (2015). A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition. *Nat. Rev. Neurosci.* 16, 719–732. doi: 10.1038/nrn4038

Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20140169. doi: 10.1098/rstb.2014.0169

Karnath, H.-O., Baier, B., and Nägele, T. (2005). Awareness of the functioning of one's own limbs mediated by the insular cortex? *J. Neurosci.* 25, 7134–7138. doi: 10.1523/JNEUROSCI.1590-05.2005

Karnath, H. O., Himmelbach, M., and Rorden, C. (2002). The subcortical anatomy of human spatial neglect: putamen, caudate nucleus and pulvinar. *Brain* 125, 350–360. doi: 10.1093/brain/awf032

Karnath, H.-O., and Rorden, C. (2012). The anatomy of spatial neglect. *Neuropsychologia* 50, 1010–1017. doi: 10.1016/j.neuropsychologia.2011.06.027

Karnath, H.-O., and Smith, D. V. (2014). The next step in modern brain lesion analysis: multivariate pattern analysis. *Brain* 137, 2405–2407. doi: 10.1093/brain/awu180

Keysers, C., and Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends Cogn. Sci.* 8, 501–507. doi: 10.1016/j.tics.2004.09.005

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2009). Perception and hierarchical dynamics. *Front. Neuroinform.* 3:20. doi: 10.3389/neuro.11.020.2009

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166. doi: 10.1007/s10339-007-0170-2

Kinsbourne, M. (1970). A model for the mechanism of unilateral neglect of space. *Trans. Am. Neurol. Assoc.* 95, 143–146.

Kirshner, H. S. (2003). "Chapter 140 - speech and language disorders A2 - Samuels, Martin A," in *Office Practice of Neurology,* 2nd Edn, ed. S. K. Feske (Philadelphia, PA: Churchill Livingstone), 890–895.

Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007

Koss, M. C. (1986). Pupillary dilation as an index of central nervous system α2-adrenoceptor activation. *J. Pharmacol. Methods* 15, 1–19. doi: 10.1016/0160-5402(86)90002-1

Krakauer, J. W., and Shadmehr, R. (2007). Towards a computational neuropsychology of action. *Prog. Brain Res.* 165, 383–394. doi: 10.1016/S0079-6123(06)65024-3

Krause, A. (2008). *Optimizing Sensing: Theory and Applications.* Pittsburgh, PA: Carnegie Mellon University.

Lawson, R. P., Mathys, C., and Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* 20, 1293–1299. doi: 10.1038/nn.4615

Lawson, R. P., Rees, G., and Friston, K. J. (2014). An aberrant precision account of autism. *Front. Hum. Neurosci.* 8:302. doi: 10.3389/fnhum.2014.00302

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Stat.* 27, 986–1005. doi: 10.1214/aoms/1177728069

Lunven, M., Thiebaut De Schotten, M., Bourlon, C., Duret, C., Migliaccio, R., Rode, G., et al. (2015). White matter lesional predictors of chronic visual neglect: a longitudinal study. *Brain* 138(Pt 3), 746–760. doi: 10.1093/brain/awu389

Mah, Y.-H., Husain, M., Rees, G., and Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain* 137, 2522–2531. doi: 10.1093/brain/awu164

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., et al. (2013). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* 522, 225–259. doi: 10.1002/cne.23458

Marshall, L., Mathys, C., Ruge, D., de Berker, A. O., Dayan, P., Stephan, K. E., et al. (2016). Pharmacological fingerprints of contextual uncertainty. *PLoS Biol.* 14:e1002575. doi: 10.1371/journal.pbio.1002575

Mathys, C. D. (2012). *Hierarchical Gaussian Filtering.* Doctoral dissertation, ETH Zurich, Zürich.

Menon, G. J., Rahman, I., Menon, S. J., and Dutton, G. N. (2003). Complex visual hallucinations in the visually impaired: the Charles bonnet syndrome. *Surv. Ophthalmol.* 48, 58–72. doi: 10.1016/S0039-6257(02)00414-9

Mesulam, M. M., and Mufson, E. J. (1982). Insula of the old world monkey. III: efferent cortical output and comments on function. *J. Comp. Neurol.* 212, 38–52. doi: 10.1002/cne.902120104

Mintzopoulos, D., Astrakas, L. G., Khanicheh, A., Konstas, A. A., Singhal, A., Moskowitz, M. A., et al. (2009). Connectivity alterations assessed by combining fMRI and MR-compatible hand robots in chronic stroke. *Neuroimage* 47(Suppl. 2), T90–T97. doi: 10.1016/j.neuroimage.2009.03.007

Mirza, M. B., Adams, R. A., Mathys, C., and Friston, K. J. (2018). Human visual exploration reduces uncertainty about the sensed world. *PLoS One* 13:e0190429. doi: 10.1371/journal.pone.0190429

Mirza, M. B., Adams, R. A., Mathys, C. D., and Friston, K. J. (2016). Scene construction, visual foraging, and active inference. *Front. Comput. Neurosci.* 10:56. doi: 10.3389/fncom.2016.00056

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., and Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. *J. Neurosci.* 33, 8227–8236. doi: 10.1523/JNEUROSCI.4255-12.2013

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663. doi: 10.1038/nn.3862

Nachev, P. (2015). The first step in modern lesion-deficit analysis. *Brain* 138(Pt 6):e354. doi: 10.1093/brain/awu275

Nachev, P., and Hacker, P. (2014). The neural antecedents to voluntary action: a conceptual analysis. *Cogn. Neurosci.* 5, 193–208. doi: 10.1080/17588928.2014.934215

O'Callaghan, C., Hall, J. M., Tomassini, A., Muller, A. J., Walpola, I. C., Moustafa, A. A., et al. (2017). Visual hallucinations are characterized by impaired sensory evidence accumulation: insights from hierarchical drift diffusion modeling in Parkinson's disease. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 2, 680–688. doi: 10.1016/j.bpsc.2017.04.007

Ochipa, C., Rothi, L. J., and Heilman, K. M. (1994). Conduction apraxia. *J. Neurol. Neurosurg. Psychiatry* 57, 1241–1244. doi: 10.1136/jnnp.57.10.1241

Ognibene, D., and Baldassarre, G. (2015). Ecological active vision: four bioinspired principles to integrate bottom–up and adaptive top–down attention tested with a simple camera-arm robot. *IEEE Trans. Auton. Ment. Dev.* 7, 3–25. doi: 10.1109/TAMD.2014.2341351

O'Reilly, J. X., Jbabdi, S., and Behrens, T. E. (2012). How can a Bayesian approach inform neuroscience? *Eur. J. Neurosci.* 35, 1169–1179. doi: 10.1111/j.1460-9568.2012.08010.x

O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science* 314, 91–94. doi: 10.1126/science.1127242

O'Reilly, R. C., and Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.* 18, 283–328. doi: 10.1162/089976606775093909

Parr, T., and Friston, K. J. (2017a). The active construction of the visual world. *Neuropsychologia* 104, 92–101. doi: 10.1016/j.neuropsychologia.2017.08.003

Parr, T., and Friston, K. J. (2017b). The computational anatomy of visual neglect. *Cereb. Cortex* 28, 777–790. doi: 10.1093/cercor/bhx316

Parr, T., and Friston, K. J. (2017c). Uncertainty, epistemics and active inference. *J. R. Soc. Interface* 14:20170376. doi: 10.1098/rsif.2017.0376

Pellicano, E., and Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* 16, 504–510. doi: 10.1016/j.tics.2012.08.009

Price, C., Warburton, E., Moore, C., Frackowiak, R., and Friston, K. (2001). Dynamic diaschisis: anatomically remote and context-sensitive human brain lesions. *J. Cogn. Neurosci.* 13, 419–429. doi: 10.1162/08989290152001853

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., and Driver, J. (2000). Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain* 123, 1624–1633. doi: 10.1093/brain/123.8.1624

Reichert, D. P., Seriès, P., and Storkey, A. J. (2013). Charles bonnet syndrome: evidence for a generative model in the cortex? *PLoS Comput. Biol.* 9:e1003134. doi: 10.1371/journal.pcbi.1003134

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060

Robinson, D. L., and Petersen, S. E. (1992). The pulvinar and visual salience. *Trends Neurosci.* 15, 127–132. doi: 10.1016/0166-2236(92)90354-B

Rocca, M. A., Pagani, E., Absinta, M., Valsasina, P., Falini, A., Scotti, G., 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

Rushmore, R. J., Valero-Cabre, A., Lomber, S. G., Hilgetag, C. C., and Payne, B. R. (2006). Functional circuitry underlying visual neglect. *Brain* 129, 1803–1821. doi: 10.1093/brain/awl140

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., et al. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040. doi: 10.1073/pnas.0805234105

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., et al. (2015). Optimal inference with suboptimal models: addiction and active Bayesian inference. *Med. Hypotheses* 84, 109–117. doi: 10.1016/j.mehy.2014.12.007

Schwartenbeck, P., and Friston, K. (2016). Computational phenotyping in psychiatry: a worked example. *eNeuro* 3:ENEURO.0049-16.2016. doi: 10.1523/ENEURO.0049-16.2016

Shah, A., and Frith, U. (1983). An islet of ability in autistic children: a research note. *J. Child Psychol. Psychiatry* 24, 613–620. doi: 10.1111/j.1469-7610.1983.tb00137.x

Shipp, S. (2003). The functional logic of cortico-pulvinar connections. *Philos. Trans. R. Soc. B Biol. Sci.* 358, 1605–1624. doi: 10.1098/rstb.2002.1213

Shipp, S. (2004). The brain circuitry of attention. *Trends Cogn. Sci.* 8, 223–230. doi: 10.1016/j.tics.2004.03.004

Shipp, S. (2007). Structure and function of the cerebral cortex. *Curr. Biol.* 17, R443–R449. doi: 10.1016/j.cub.2007.03.044

Shipp, S. (2016). Neural elements for predictive coding. *Front. Psychol.* 7:1792. doi: 10.3389/fpsyg.2016.01792

Shipp, S., Adams, R. A., and Friston, K. J. (2013). Reflections on agranular architecture: predictive coding in the motor cortex. *Trends Neurosci.* 36, 706–716. doi: 10.1016/j.tins.2013.09.004

Showers, M. J., and Lauer, E. W. (1961). Somatovisceral motor patterns in the insula. *J. Comp. Neurol.* 117, 107–115. doi: 10.1002/cne.901170109

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., and Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Res.* 49, 2705–2739. doi: 10.1016/j.visres.2009.08.005

Simonyan, K., and Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *Neuroscientist* 17, 197–208. doi: 10.1177/1073858410386727

Smith, A. T., Singh, K. D., Williams, A. L., and Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb. Cortex* 11, 1182–1190. doi: 10.1093/cercor/11.12.1182

Testolin, A., and Zorzi, M. (2016). Probabilistic models and generative neural networks: towards an unified framework for modeling normal and impaired neurocognitive functions. *Front. Comput. Neurosci.* 10:73. doi: 10.3389/fncom.2016.00073

Teunisse, R. J., Zitman, F. G., Cruysberg, J. R., Hoefnagels, W. H., and Verbeek, A. L. (1996). Visual hallucinations in psychologically normal people: Charles Bonnet's syndrome. *Lancet* 347, 794–797. doi: 10.1016/S0140-6736(96)90869-7

Thiebaut de Schotten, M., Urbanski, M., Duffau, H., Volle, E., Lévy, R., Dubois, B., et al. (2005). Direct evidence for a parietal-frontal pathway subserving spatial awareness in humans. *Science* 309, 2226–2228. doi: 10.1126/science.1116251

Ungerleider, L. G., and Christensen, C. A. (1979). Pulvinar lesions in monkeys produce abnormal scanning of a complex visual array. *Neuropsychologia* 17, 493–501. doi: 10.1016/0028-3932(79)90056-3

Ungerleider, L. G., and Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165. doi: 10.1016/0959-4388(94)90066-3

Ungerleider, L. G., and Mishkin, M. (1982). "Two cortical visual systems," in *Analysis of Visual Behavior,* eds D. Ingle, M. A. Goodale, and J. W. Mansfield (Cambridge, MA: MIT Press), 549–586.

Valdez, A. B., Papesh, M. H., Treiman, D. M., Smith, K. A., Goldinger, S. D., and Steinmetz, P. N. (2015). Distributed representation of visual objects by single neurons in the human brain. *J. Neurosci.* 35, 5180–5186. doi: 10.1523/JNEUROSCI.1958-14.2015

von der Heydt, R., and Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J. Neurosci.* 9, 1731–1748.

Vuilleumier, P., Hester, D., Assal, G., and Regli, F. (1996). Unilateral spatial neglect recovery after sequential strokes. *Neurology* 46, 184–189. doi: 10.1212/WNL.46.1.184

Wald, A. (1947). An essentially complete class of admissible decision functions. *Ann. Math. Stat.* 18, 549–555. doi: 10.1214/aoms/1177730345

Warrington, E. K. (1975). The selective impairment of semantic memory. *Q. J. Exp. Psychol.* 27, 635–657. doi: 10.1080/14640747508400525

Warrington, E. K., and James, M. (1967). Disorders of visual perception in patients with localised cerebral lesions. *Neuropsychologia* 5, 253–266. doi: 10.1016/0028-3932(67)90040-1

Warrington, E. K., and James, M. (1988). Visual apperceptive agnosia: a clinico-anatomical study of three cases. *Cortex* 24, 13–32. doi: 10.1016/S0010-9452(88)80014-5

Warrington, E. K., and Shallice, T. (1984). Category specific semantic impairments. *Brain* 107, 829–853. doi: 10.1093/brain/107.3.829

Warrington, E. K., and Taylor, A. M. (1973). The contribution of the right parietal lobe to object recognition. *Cortex* 9, 152–164. doi: 10.1016/S0010-9452(73)80024-3

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604. doi: 10.1038/nn0602-858

Wernicke, C. (1969). "The symptom complex of aphasia," in *Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968*, eds R. S. Cohen and M. W. Wartofsky (Dordrecht: Springer), 34–97. doi: 10.1007/978-94-010-3378-7_2

Winn, J., and Bishop, C. M. (2005). Variational message passing. *J. Mach. Learn. Res.* 6, 661–694.

Yang, S. C.-H., Lengyel, M., and Wolpert, D. M. (2016a). Active sensing in the categorization of visual patterns. *eLife* 5:e12215. doi: 10.7554/eLife.12215

Yang, S. C.-H., Wolpert, D. M., and Lengyel, M. (2016b). Theoretical perspectives on active sensing. *Curr. Opin. Behav. Sci.* 11, 100–108. doi: 10.1016/j.cobeha.2016.06.009

Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans. Inform. Theory* 51, 2282–2312. doi: 10.1109/TIT.2005.850085

Yin, H. H., and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat. Rev. Neurosci.* 7, 464–476. doi: 10.1038/nrn1919

Yu, A. J., and Dayan, P. (2002). Acetylcholine in cortical inference. *Neural Netw.* 15, 719–730. doi: 10.1016/S0893-6080(02)00058-8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.