Precision modulation in predictive coding hierarchies: theoretical, behavioural and neuroimaging investigations

Submitted by Harriet Ruth Brown to University College London (UCL) as a thesis for the degree of Doctor of Philosophy

July 2014

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

Signed:

Date:

Abstract

Estimation of uncertainty is an important aspect of perception and a prerequisite for effective action. This thesis explores the implementation of uncertainty estimation as precision modulation within a predictive coding hierarchy, optimised within a neurbiologically-plausible message-passing scheme via the minimisation of free-energy.

This thesis consists of six chapters. The first presents a new model of a classic visual illusion, the Cornsweet illusion, which demonstrates that the Cornsweet illusion is a natural consequence of Bayes-optimal perception under the free-energy principle, and demonstrates that increasing contrast can be modelled by increasing signal-to-noise ratio. The second chapter describes dynamic causal modelling of EEG data collected from participants viewing the Cornsweet illusion, demonstrating that a reduction in precision, or superficial pyramidal cell gain, in lower visual hierarchical levels, is sufficient to explain contrast-dependent changes in ERPs. The third describes a model of a simple attentional paradigm – the Posner paradigm – recasting attention as the optimal modulation of precision in sensory channels. The fourth describes an MEG study of the Posner paradigm, using Bayesian model selection to explore the role of changes in backwards and modulatory connections and changes in local superficial pyramidal cell gain in producing the electrophysiological and behavioural correlates of the Posner paradigm. The fifth chapter recasts the Posner paradigm in the motor domain to investigate the level (intrinsic vs. extrinsic) of precision modulation by motor cues. The sixth describes a new model of sensory attenuation based on using precision modulation to balance the imperatives to act and perceive.

I hope to demonstrate that precision modulation within predictive coding hierarchies, under the free-energy principle, is a flexible and powerful way of describing and explaining both behavioural and neuroimaging data

Acknowledgements

First and foremost I would like to sincerely thank my supervisor, Professor Karl Friston, for his support throughout my PhD; for being patient with explanations, generous with his time, encouraging when I was disheartened and enthusiastic when I found new directions of my own.

I'd also like to thank Professor Peter Dayan, Dr Mark Edwards and Dr Sven Bestmann for their invaluable supervisory support and help in stimulating and developing the ideas and experiments in this thesis.

Dr Vladimir Litvak and Dr Gareth Barnes both gave a great deal of time and expertise to help with the EEG and MEG analyses.

David Bradbury, Alphonso Reid and Letty Manyade provided many hours of technical support for both the EEG and MEG scans, as well as for the psychophysical testing.

May thanks also to Dr Rick Adams for being an inspiring and meticulous collaborator and colleague.

I also thank my husband, David Brown, and my father, Nick Feldman, for their unwavering confidence and support throughout the last four years.

Table of Contents

Introduction

1. Background

Living organisms are remarkable in their apparent ability to violate the second law of thermodynamics; that is, to resist the tendency to disorder over time (Schrödinger, 1944). Humans, for example, manage to keep their blood sugar, temperature, extracellular ion concentrations, blood pressure and many other such physiological variables within tightly controlled limits at all times. This is managed through elaborate systems of interoceptive and exteroceptive sensory receptors and reflex arcs which can monitor changes in these variables and act to return them to physiological limits. In essence, this involves reducing the entropy of the distributions of states occupied – ensuring that a small number of states are occupied most of the time, with the rest being very rarely occupied. Assuming that the organism and the system containing it are not changing over time ('ergodic assumptions'), entropy is just the long-term average of surprise, and instead of minimising entropy the organism can maintain its homeostasis by reducing surprising sensations – a body temperature of 41^{o}C, or a blood pH of 5.5, would be surprising (as well as deadly) to a human, for example.

Although this process is remarkable in living organisms, it is by no means uniquely human. An earthworm or a bee also has systems that allow maintenance of extracellular sodium concentration, or temperature. However, humans are vastly better at doing this under a wider variety of circumstances than, for example, earthworms or bees. If I were to go for a cycle ride tomorrow, I would pack a bottle of water, knowing that my osmolarity is likely

to drop as the day goes on. If I stayed out too late and it started to get dark, I would adjust my route to arrive home more quickly, knowing that I didn't have enough clothing with me to maintain my body temperature overnight.

What I am doing in these circumstances is building a model of my expected physiological variables, which is both hierarchical – it acknowledges lower level causes, such as not having drunk enough water, as well as higher level ones, such as not having packed any in my panniers– and dynamic, in that it extrapolates future states based on the current trajectory. There are a number of new variables required by this model, to represent these higher-order causes of interoceptive data such as the time of day, my location, what's in my panniers etc. – from now on, I will refer to these variables as 'causes'; in the sense that they cause (potentially surprising) sensory consequences.

Both interoceptive sensory receptors and exteroceptive sensory receptors provide only indirect information about the causes of data in the outside world. For example, true core temperature is a 'cause', which is inferred or perceived by the brain based on discharges in core temperature receptors, perhaps calibrated based on past experience through long-term synaptic plasticity. Likewise, complex causes such as 'location' are inferred from complex patterns of sensory data such as photoreceptor discharges in the retina giving visual (e.g. landmark recognition) and semantic (e.g. signposts) information, as well as past memory and experience.

So, given that organisms cannot directly perceive causes of surprising states, as they can only infer them through sensory data, they must instead reduce the surprisingness of these data. However, in order to use indirect information about the likely future trajectory of interoceptive causes, gleaned from estimates of higher-level causes, the surprise of the sensory data pertaining to the causes must be minimised too. For example, if I believe it to be 12pm, the visual data caused by twilight would be surprising, and it would also likely lead to surprising interoceptive states, such as a drop in body temperature as I found myself far from home at night.

Reducing surprise thus entails building models of the causes of sensory data in the world which match the actual sensory data perceived as closely as possible. As explained above, these models will be hierarchical, incorporating the hierarchical causal structure of the world, and dynamic, explaining how current states and causes are linked to past and future states and causes.

A ‘good’ model of the world is one which reduces surprise – but how is this evaluated? Surprise itself cannot be calculated because it is mathematically intractable for all but the most basic of probability distributions. However, there is a quantity – the free-energy – which is an upper bound on surprise. Free-energy will always be greater than surprise, such that an organism which minimises free energy will also minimise surprise (Friston, 2006). Crucially, Free Energy is a function of just sensory data and predictions, both of which are available to the agent. This thesis is based around the assumption than the brain uses free-energy minimisation to reduce its long-term surprise.

How might these models be adjusted and refined? The answer lies in a signal-processing method first used to process and transmit speech: predictive coding. Predictive coding decomposes a signal into parts that can be ‘predicted’ based on information already held, and parts that cannot – the ‘prediction error’. For example, in processing speech, predictive coding schemes calculate the factors of the signal that are constant to a speaker, and thus can be ‘predicted’ across an entire segment of speech, and the actual vowels and consonants produced, the ‘prediction error’. The constant, speaker-specific factors need only be transmitted once and the entire sound segment can be reconstructed. Note that in this thesis, ‘prediction’ refers to the ability to reproduce current sensory data rather than forecast the future.

Linear predictive coding has been successfully applied to explaining extraclassical receptive field effects (Rao & Ballard, 1998), trichromacy and colour opposition (Buchsbaum & Gottschalk, 1983) and retinal processing (Srinivasen et al., 1982). Predictive coding also arises naturally from both an assumption of hierarchical processing (it is inherently hierarchical) and the free-energy principle – under certain assumptions, prediction error is exactly equivalent to free-energy (Friston & Kiebel, 2009).

This thesis examines a particular aspect of predictive coding – the inclusion of precision. The classical linear predictive coding schemes mentioned above are obviously too limited to be of much use in all but the most rarified situations. In the real world there will be many thousands of prediction errors arising at any one moment and many of them will compete with each other for explanation – for instance, they may represent competing estimates of the value of a cause, or competing imperatives. To reconcile these competitions, prediction errors must be weighted by their **precision** (or inverse-variance, or signal-to-noise), with high-precision prediction errors having a greater impact on processing at higher levels of the hierarchy than those with low precision. The precision can encode a number of stimulus properties including task-relevance, attention, basic

signal-to-noise ratio, semantic content, or plausibility based on external factors. Put simply, precision is the reliability (inverse variability) – or confidence afforded – prediction errors and may be a crucial aspect of confidence-sensitive predictive coding in the brain.

This thesis explores the implementation and explanatory power of precision-weighted prediction error in the perception of visual illusions, directed attention and two motor paradigms: motor preparation and sensory attenuation. I hope to demonstrate that precision modulation within predictive coding hierarchies, under the free-energy principle, is a flexible and powerful way of describing and explaining both behavioural and neuroimaging data.

2. Free energy, predictive coding and Bayes-optimal perception

The following section provides a more mathematical description of how Bayes-optimal perception and action might be implemented in the brain. A more detailed derivation of the following equations can be found in Appendices 1-3. The outcome of this normative treatment is a set of differential equations (Equation 0.3) that describe neuronal activity and ensuing action, which we then use to demonstrate the necessary role of sensory attenuation and the illusory phenomena that it entails. The equations may appear a bit complicated but they are based on just three assumptions:

- The brain minimises the free energy of sensory inputs defined by a generative model.
- The generative model used by the brain is hierarchical, nonlinear and dynamic.
- Neuronal firing rates encode the expected state of the world, under this model.

The first assumption is the free energy principle, which leads to active inference in the embodied context of action. This provides a principled (Bayes optimal) explanation for action and perception, in which both minimise a free energy bound on the (negative) Bayesian log evidence for a generative model of the sensorium. This means that minimising free energy maximises Bayesian model evidence. The second assumption – about the nature of the models entailed by neuronal circuits – is motivated easily by noting that the world is both dynamic and nonlinear and that hierarchical causal structure emerges inevitably from a separation of temporal scales (Ginzburg & Landau, 1950; Haken, 1983), and the observation that the functional anatomy of the brain is hierarchical. The final assumption is the Laplace assumption that, in terms of neural codes, leads to the

Laplace code that is arguably the simplest and most flexible of all neural codes (Friston, 2009).

Given these assumptions, one can simulate a whole variety of neuronal processes by specifying the particular equations that constitute a generative model. The resulting perception and action are specified completely by the above assumptions and can be implemented in a biologically plausible way as described below. In brief, these simulations use differential equations that minimise the free energy of sensory input using a generalised gradient descent (Friston, Stephan, Li, & Daunizeau, 2010).

$$\begin{aligned}\dot{\tilde{\mu}}(t) &= \mathcal{D}\tilde{\mu}(t) - \partial_{\tilde{\mu}} F(\tilde{s}, \tilde{\mu}) \\ \dot{a}(t) &= -\partial_a F(\tilde{s}, \tilde{\mu})\end{aligned} \tag{0.1}$$

These coupled differential equations describe perception and action respectively and just say that neuronal activity encoding conditional expectations $\tilde{\mu} \equiv (\mu, \mu', \mu'', \ldots)$ and action a change to reduce free energy, where free energy $F(\tilde{s}, \tilde{\mu})$ is a function of sensory inputs $\tilde{s} = (s, s', s'', \ldots)$ and conditional expectations encoded by neuronal activity. The first differential equation has the same form as Bayesian (e.g., Kalman-Bucy) filters used in time series analysis. The first term is a prediction based upon a differential matrix operator D that returns the generalised motion of the expectation. The second (correction) term is usually expressed as a mixture of prediction errors that ensures the changes in conditional expectations are Bayes-optimal predictions about hidden states of the world.

The second differential equation says that action also minimises free energy. The differential equations above are coupled because sensory input depends upon action, which depends upon perception through the conditional expectations. This circular dependency leads to a sampling of sensory input that is both predicted and predictable, thereby minimising free energy and prediction errors.

To perform neuronal simulations under this scheme, it is only necessary to integrate or solve Equation 0.1 to simulate the neuronal dynamics that encode conditional expectations and the ensuing action.

2.1 Hierarchical Dynamic Models

Conditional expectations depend upon the brain's generative model of the world, which we assume has the following (flexible and general) hierarchical form:

$$
\begin{aligned}
s &= g^{(1)}(x^{(1)}, v^{(1)}) + \omega_v^{(1)} \\
\dot{x}^{(1)} &= f^{(1)}(x^{(1)}, v^{(1)}) + \omega_x^{(1)} \\
&\vdots \\
v^{(i-1)} &= g^{(i)}(x^{(i)}, v^{(i)}) + \omega_v^{(i)} \\
\dot{x}^{(i)} &= f^{(i)}(x^{(i)}, v^{(i)}) + \omega_x^{(i)} \\
&\vdots
\end{aligned}
\qquad (0.2)
$$

$$
\begin{aligned}
\omega_x^{(i)} &\sim N(0, \Pi_x^{(i)-1}) \\
\omega_v^{(i)} &\sim N(0, \Pi_v^{(i)-1}) \\
\Pi_x^{(i)} &= diag(\exp(\pi_x^{(i)}(x^{(i)}, v^{(i)}))) \\
\Pi_v^{(i)} &= diag(\exp(\pi_v^{(i)}(x^{(i)}, v^{(i)})))
\end{aligned}
$$

This equation is just a way of specifying a generative model in terms of a probability density over the sensory and hidden states, where the hidden states have been divided into hidden states and causes $(x^{(i)}, v^{(i)})$. Here, $(g^{(i)}, f^{(i)})$ are nonlinear functions of hidden states that generate sensory inputs at the first level. Random fluctuations $(\omega_x^{(i)}, \omega_v^{(i)})$ in the hidden causes and motion of states enter each level of the hierarchy. Gaussian assumptions about these random fluctuations make the model probabilistic – they play the role of sensory noise at the first level and induce uncertainty at higher levels. The amplitudes of these random fluctuations are quantified by their precisions $(\Pi_x^{(i)}, \Pi_v^{(i)})$ that may depend upon the hidden states or causes through log-precisions $(\pi_x^{(i)}, \pi_v^{(i)})$. Hidden causes link hierarchical levels, whereas hidden states link dynamics over time. Hidden states and causes are abstract quantities (like the motion of an object in the field of view) that the brain uses to explain or predict sensations.

2.2 Perception and predictive coding

Given the form of the generative model (Equation 0.2) we can now write down the differential equations (Equation 0.1) describing neuronal dynamics in terms of (precision-weighted) prediction errors on the hidden causes and states. These errors represent the

difference between conditional expectations and predicted values, under the generative model (using $A \times B := A^T B$, omitting higher-order terms and where subscripts denote derivatives):

$$\dot{\tilde{\mu}}_x^{(i)} = \mathrm{D}\tilde{\mu}_x^{(i)} + \left(\frac{\partial \tilde{g}^{(i)}}{\partial \tilde{\mu}_x^{(i)}} - \tfrac{1}{2}\tilde{\varepsilon}_v^{(i)} \cdot \frac{\partial \tilde{\Omega}_v^{(i)}}{\partial \tilde{\mu}_x^{(i)}} \right) \cdot \xi_v^{(i)} + \left(\frac{\partial \tilde{f}^{(i)}}{\partial \tilde{\mu}_x^{(i)}} - \tfrac{1}{2}\tilde{\varepsilon}_x^{(i)} \cdot \frac{\partial \tilde{\Omega}_x^{(i)}}{\partial \tilde{\mu}_x^{(i)}} \right) \cdot \xi_x^{(i)} + \frac{\partial tr(\tilde{\Omega}_v^{(i)} + \tilde{\Omega}_x^{(i)})}{\partial \tilde{\mu}_x^{(i)}} - \mathrm{D} \cdot \xi_x^{(i)}$$

$$\dot{\tilde{\mu}}_v^{(i)} = \mathrm{D}\tilde{\mu}_v^{(i)} + \left(\frac{\partial \tilde{g}^{(i)}}{\partial \tilde{\mu}_v^{(i)}} - \tfrac{1}{2}\tilde{\varepsilon}_v^{(i)} \cdot \frac{\partial \tilde{\Omega}_v^{(i)}}{\partial \tilde{\mu}_v^{(i)}} \right) \cdot \xi_v^{(i)} + \left(\frac{\partial \tilde{f}^{(i)}}{\partial \tilde{\mu}_x^{(i)}} - \tfrac{1}{2}\tilde{\varepsilon}_x^{(i)} \cdot \frac{\partial \tilde{\Omega}_x^{(i)}}{\partial \tilde{\mu}_v^{(i)}} \right) \cdot \xi_x^{(i)} + \frac{\partial tr(\tilde{\Omega}_v^{(i)} + \tilde{\Omega}_x^{(i)})}{\partial \tilde{\mu}_v^{(i)}} - \xi_v^{(i+1)}$$

$$\xi_x^{(i)} = \tilde{\Pi}_x^{(i)} \tilde{\varepsilon}_x^{(i)} = \Pi_x^{(i)} (\mathrm{D}\tilde{\mu}_x^{(i)} - \tilde{f}^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}))$$
$$\xi_v^{(i)} = \Pi_v^{(i)} \tilde{\varepsilon}_v^{(i)} = \Pi_v^{(i)} (\tilde{\mu}_v^{(i-1)} - \tilde{g}^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}))$$

$$\Omega_x^{(i)} = diag(\pi_x^{(i)}(\mu_x^{(i)}, \mu_v^{(i)}))$$
$$\Omega_v^{(i)} = diag(\pi_v^{(i)}(\mu_x^{(i)}, \mu_v^{(i)}))$$

$$\Pi_x^{(i)} = diag(\exp(\pi_x^{(i)}(\mu_x^{(i)}, \mu_v^{(i)})))$$
$$\Pi_v^{(i)} = diag(\exp(\pi_v^{(i)}(\mu_x^{(i)}, \mu_v^{(i)})))$$

(0.3)

Equation 0.3 can be derived fairly easily by computing the free energy for the hierarchical model in Equation 0.2 and inserting its gradients into Equation 0.1. This produces a relatively simple update scheme, in which conditional expectations are driven by a mixture of prediction errors, where prediction errors are defined by the equations of the generative model.

The top two lines of equation 0.3 describe how the initial value of the expectations $\mathrm{D}\tilde{\mu}^{(i)}$ is updated to minimise free energy. Precision-weighted prediction errors $\xi_v^{(i)}, \xi_x^{(i)}$ influence the update in proportion to the connection strengths $\left(\frac{\partial(\tilde{g}^{(i)}, \tilde{f}^{(i)})}{\partial \tilde{\mu}^{(i)}} - \tfrac{1}{2}\tilde{\varepsilon}^{(i)} \cdot \frac{\partial \tilde{\Omega}^{(i)}}{\partial \tilde{\mu}^{(i)}} \right)$. These precision-weighted prediction errors are calculated in the third and fourth lines. The difference between the current expectations $\mathrm{D}\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i-1)}$ and the values as predicted by the generative model $\tilde{f}^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}), \tilde{g}^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)})$ are weighted by the precision $\Pi^{(i)}$, which is itself a function of causes and states (seventh and eighth lines). With this particular form for the precisions, the term $tr(\tilde{\Omega}_v^{(i)} + \tilde{\Omega}_x^{(i)})$ is

constant for causes and states that affect the log-precisions linearly and zero if they have no effect.

It is difficult to overstate the generality of Equation 0.3: its solutions grandfather nearly every known statistical estimation scheme, under parametric assumptions about additive or multiplicative noise (Friston, 2008). These range from ordinary least squares to advanced variational deconvolution schemes. The scheme is called *generalised Bayesian filtering* or predictive coding (Friston et al., 2010): see also (Rao & Ballard, 1999). In neural network terms, Equation 0.3 says that error-units receive predictions from the same level and the level above. Conversely, conditional expectations (encoded by the activity of state units) are driven by prediction errors from the same level and the level below. These constitute bottom-up and lateral messages that drive conditional expectations towards a better prediction to reduce the prediction error in the level below. This is the essence of recurrent message passing between hierarchical levels to optimise free energy or suppress prediction error: see (Friston & Kiebel, 2009) and Appendix 1 for a more detailed discussion. In neurobiological implementations of this scheme, the sources of bottom-up prediction errors are thought to be superficial pyramidal cells that send forward connections to higher cortical areas. Conversely, predictions are conveyed from deep pyramidal cells, by backward connections, to target (polysynaptically) the superficial pyramidal cells encoding prediction error (Mumford, 1992; Friston & Kiebel, 2009).

In the present context, the key thing about this predictive coding scheme is that the precisions at each level in the hierarchy depend on the expected hidden causes and states in the level above. It is this dependency we have proposed mediates attention or selection in hierarchical inference (Feldman & Friston, 2010; Friston, et al., 2012). Equation 0.3 tells us that the state-dependent precisions modulate the responses of the error-units to their presynaptic inputs. This modulation depends on the conditional expectations about the states and suggests something intuitive – attention is mediated by activity-dependent modulation of the synaptic gain of principal cells that convey sensory information (prediction error) from one cortical level to the next. This translates into a top-down control of synaptic gain in principal (superficial pyramidal) cells elaborating prediction errors and fits comfortably with the modulatory effects of top-down connections in cortical hierarchies that have been associated with attention and action selection.

The numerics of the integration scheme used to simulate inference (Equation 0.3) are provided in Appendix 4.

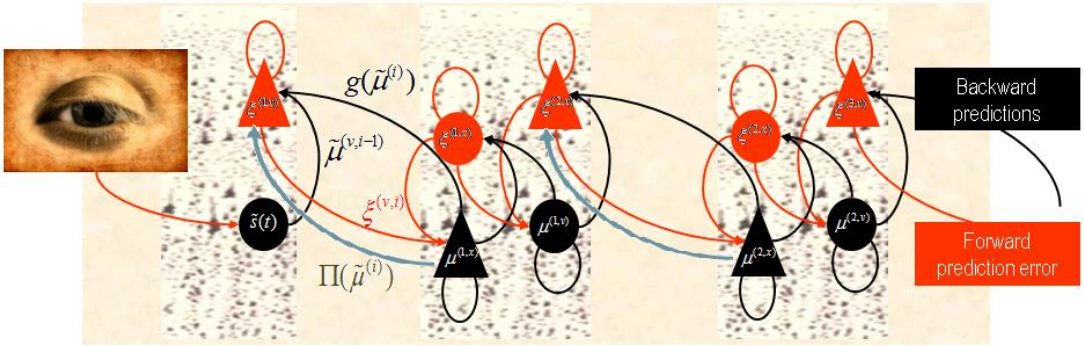

Figure 0.1: Predictive coding in the brain. Schematic detailing the neuronal architecture that might implement the generalised predictive coding described in Equation 0.3. This shows the speculative cells of origin of forward driving connections that convey prediction error from a lower area to a higher area and the backward connections that construct predictions (Mumford et al 1992; Friston 2008). These predictions try to explain away prediction error in lower levels. In this scheme, the sources of forward and backward connections are superficial and deep pyramidal cells – represented by red and black triangles – respectively. State-units are in black and error-units in red. Here, neuronal populations are deployed hierarchically within three cortical areas (or macro-columns). Blue arrows represent backwards modulatory connections, the mechanism whereby activity in higher levels can provide contextual or attentional modulation of the gain in lower levels.

2.3 Action

In active inference, conditional expectations elicit behaviour by sending top-down predictions down the hierarchy that are unpacked into proprioceptive predictions at the level of the cranial nerve nuclei and spinal-cord. These engage classical reflex arcs to suppress proprioceptive prediction errors and produce the predicted motor trajectory

$$\dot{a} = -\frac{\partial}{\partial a}F = -\frac{\partial \tilde{s}}{\partial a}.\xi_v^{(1)} \tag{0.4}$$

The reduction of action to classical reflexes follows because the only way that action can minimise free energy is to change sensory (proprioceptive) prediction errors by changing sensory signals; cf., the equilibrium point formulation of motor control (Feldman & Levin, 1995). In short, active inference can be regarded as equipping a generalised predictive coding scheme with classical reflex arcs: see (Friston et al., 2010; Friston, Daunizeau, & Kiebel, 2009) for details. The actual movements produced clearly depend upon top-down predictions that can have a rich and complex structure.

2.4 Neurobiological implementation of predictive coding

Figure 0.1 details the neuronal architecture that might implement the generalised predictive coding described in Equation 0.3. This shows the speculative cells of origin of forward, driving connections that convey prediction error from a lower area to a higher area and the backward connections that construct predictions (Mumford et al 1992; Friston 2008). Forwards prediction errors are encoded by superficial pyramidal cells (red triangles) while backwards predictions are conveyed by deep pyramidal cells (black triangles). Backwards modulatory connections, which adjust the precision of superficial pyramidal cells, are also signalled by deep pyramidal cells. Expectations about causes and states are signalled by interneurons. This architecture was refined and expanded based on anatomical and electrophysiological data by (Bastos et al., 2012), to yield a scheme which can be used to model neuroimaging data (chapters 2 and 4).

3. Outline of the Thesis

The first two chapters investigate the relationship between precision, sensory signal-to-noise ratio and superficial pyramidal cell gain. The paradigm used is a very simple visual illusion – the Cornsweet effect (O'Brien 1959; Craik 1966; Cornsweet 1970).

In chapter 1, I review the nature of illusions using the free-energy formulation of Bayesian perception described above. Following the work of previous authors (Purves et al., 1999; Lotto & Purves, 2001) I reiterate the notion that illusory percepts are, in fact, Bayes-optimal and represent the most likely explanation for ambiguous sensory input. By using plausible prior beliefs about the spatial gradients of illuminance and reflectance in visual scenes, I show that the Cornsweet effect emerges as a natural consequence of Bayes-optimal perception. Furthermore, the appearance of secondary illusory percepts (Mach bands) is simulated as a function of stimulus contrast. The contrast-dependent emergence of the Cornsweet effect and subsequent appearance of Mach bands were simulated using a simple but plausible generative model of the general form described above. Because this generative model was inverted using the neurobiologically plausible scheme described in Equation 0.3, we could use the inversion as a simulation of neuronal processing and implicit inference. Finally, the qualitative and quantitative predictions of this Bayes-optimal simulation are verified psychophysically, using stimuli presented briefly to normal subjects at different contrast levels, in the context of a fixed alternative forced choice paradigm.

In chapter 2, we use the conclusions of chapter 1 – that increasing contrast can be modelled by increasing sensory precision – to test the hypothesis that increasing visual contrast increased the precision encoded in early visual areas by the gain or excitability of superficial pyramidal cells. This hypothesis was investigated using electroencephalography and dynamic causal modelling (DCM) (Garrido et al., 2007a; Bastos et al., 2012); a biologically constrained modelling of the cortical processes underlying EEG activity. We presented Cornsweet stimuli of varying contrast to participants while recording from high-density EEG. Source localisation identified the electromagnetic sources of visually evoked responses and DCM was used to characterise the coupling among these sources. Bayesian model selection was used to select the most likely connectivity pattern and contrast-dependent changes in connectivity. As predicted, the model with the highest evidence entailed increased superficial pyramidal cell gain in higher-contrast trials. This increase in gain was most pronounced in early visual areas and lessened as the visual hierarchy was ascended. These results demonstrate that increased signal-to-noise ratio in sensory signals produce (or are represented by) increased superficial pyramidal cell gain, and that synaptic parameters encoding statistical properties like sensory precision can be quantified using EEG and dynamic causal modelling.

Chapters 1 and 2 looked at static encoding of precision; Chapters 3 and 4 change track to examine attention and look at the dynamic, online modulation of precision.

In chapter 3, I expand upon the suggestion (Friston, 2009) that attention can be understood as inferring the level of uncertainty or precision during inference using a hierarchical predictive coding scheme. This chapter tries to substantiate this claim using neuronal simulations of directed spatial attention and biased competition. Using the Bayesian formulations of perception and predictive coding described above, we demonstrate that if the precision depends dynamically on the values of the states, one can explain many aspects of attention. This is illustrated in the context of the Posner paradigm (Posner, 1980), using the simulations to generate both psychophysical and electrophysiological responses. These simulated responses are consistent with psychophysical (Posner et al., 1978) and physiological (Mangun & Hillyard, 1991) data on directed attention. Furthermore, if we present both attended and non-attended stimuli simultaneously, biased competition (Desimone & Duncan, 1995) for neuronal representation emerges as a principled and straightforward property of Bayes-optimal perception.

In chapter 4, I examine the assertions of the model described in chapter 3 – that the behavioural and electrophysiological phenomenology of the Posner paradigm can be explained in terms of a cue-dependent setting of precision or gain on the sensory channels reporting anticipated target locations, which is updated selectively by invalid targets. Taking advantage of the biologically plausible theory described above, where precision is encoded by the gain of superficial pyramidal cells reporting prediction error, I use dynamic causal modelling to assess the evidence in magnetoencephalographic responses for cue-dependent and top-down updating of superficial pyramidal cell gain. Bayesian model comparison suggested that it is almost certain that differences in superficial pyramidal cells gain – and its top-down modulation – contribute to observed responses; and we could be more than 80% certain that anticipatory effects on postsynaptic gain are limited to visual (extrastriate) sources. These empirical results speak to the role of attention in optimising perceptual inference and its formulation in terms of predictive coding.

In chapters 5 and 6 I turn my attention to the motor domain, to examine if the principles of precision modulation within predictive coding hierarchies expounded in the first four chapters can help to explain motor preparation and sensory attenuation. In chapter 5, I pursue the idea that action planning (motor preparation) is an attentional phenomenon

directed towards kinaesthetic signals. This rests on a view of motor control as active inference, where predictions of proprioceptive signals are fulfilled by peripheral motor reflexes, as described above. If valid, active inference suggests that attention should not be limited to the optimal biasing of perceptual signals in the exteroceptive (e.g. visual) domain but should also bias proprioceptive signals during movement. Here, this idea is investigated using the Posner paradigm recast in a motor setting. Specifically, reaction time changes when movements were preceded by valid relative to invalid cues were examined. Furthermore, the hierarchical level at which putative attentional effects were expressed is addressed by independently cueing the nature of the movement and the hand used to execute it. A significant interaction between the validity of movement and effector cues on reaction times was found. This suggests that attentional bias might be mediated at a low level in the motor hierarchy, in an intrinsic frame of reference. This finding is consistent with attentional enabling of top-down predictions of proprioceptive input and may rely upon the same synaptic mechanisms that mediate directed spatial attention in the visual system.

Chapter 6 applies active inference to a phenomenon where previous theoretical explanations were incomplete – sensory attenuation (Weiskrantz et al., 1971). Minisming prediction error can be achieved by changing our predictions to explain sensory input through *perception*, or by actively change sensory input to fulfil our predictions via *action*. However, this creates a conflict between action and perception, in that self-generated movements require predictions to override the sensory evidence that one is not actually moving. However, ignoring sensory evidence means that externally generated sensations will not be perceived. Conversely, attending to (proprioceptive and somatosensory) sensations enables the detection of externally generated events but precludes generation of actions. This conflict can be resolved by attenuating the precision of sensory evidence during movement or, equivalently, attending away from the consequences of self-made acts. We propose that this Bayes optimal withdrawal of precise sensory evidence during movement is the cause of psychophysical sensory attenuation. Furthermore, it explains the force-matching illusion (Shergill et al., 2003) and reproduces empirical results almost exactly (Shergill et al., 2005). Finally, if attenuation is removed, the force-matching illusion disappears and false (delusional) inferences about agency emerge. This is important, given the negative correlation between sensory attenuation and delusional beliefs in normal subjects (Teufel et al., 2010) – and the reduction in the magnitude of the illusion in schizophrenia (Shergill et al., 2005). Active inference therefore links the neuromodulatory optimisation of precision to sensory attenuation and illusory phenomena during the attribution of agency in normal subjects. It also provides a functional account of deficits in

syndromes characterised by false inference and impaired movement – like schizophrenia and Parkinsonism – syndromes that implicate abnormal modulatory neurotransmission.

Chapter 1

Free Energy and Illusions: The Cornsweet effect

1. Introduction

Illusions are often regarded as 'failures' of perception; however, Bayesian considerations often provide a principled explanation for apparent failures of inference in terms of prior beliefs. This chapter is about the nature of illusions and their relationship to Bayes-optimal perception. The main point made in this chapter is that illusory percepts are optimal in the sense of explaining sensations in terms of their most likely cause. In brief, illusions occur when the experimenter generates stimuli in an implausible or unlikely way. From the subject's perspective, these stimuli are ambiguous and could be explained by different underlying causes. This ambiguity is resolved in a Bayesian setting, by choosing the most likely explanation, given prior beliefs about the hidden causes of the percept. This key point has been made by many authors (e.g., Purves et al. 1999). Here, I develop it under biologically realistic simulations of Bayes-optimal perception and try to make some quantitative predictions about how subjects should make perceptual decisions. I then try to establish the scheme's validity by showing that these predictions are largely verified by experimental data from normal subjects viewing the same stimuli.

The example chosen here is the Cornsweet illusion, which has a long history, dating back to the days of Helmholtz (Mach, 1865; O'Brien 1959; Craik 1966; Cornsweet 1970). This is particularly relevant given this formulation of the Bayesian brain is based upon the idea that the brain is a Helmholtz or inference machine (Helmholtz, 1867; Barlow, 1974; Dayan

et al., 1995: Friston 2005). In other words, the brain is trying to infer the hidden causes and states of the world generating sensory information, using predictions based upon a generative model that includes prior beliefs. I hoped to show that the Cornsweet effect can be explained in a parsimonious way by some simple prior beliefs about the way that visual information is generated at different spatial and temporal scales.

1.1 The Cornsweet effect and the nature of illusions

Figure 1.1 provides an illustration of the Cornsweet illusion. The illusion is the false percept that the peripheral regions of a stimulus have a different brightness, despite the fact they are physically isoluminant. This illusion is induced by a biphasic luminance "edge" in the centre of the field of view (shown in the right hand column of Figure 1.1). The four rows of Figure 1.1 show the Cornsweet effect increasing in magnitude as the contrast of the stimulus is increased. Interestingly, at high levels of contrast, secondary illusions - Mach bands (Mach, 1865; Lotto et al. 1999) – appear at the para-central points of inflection of the true luminance profile. It is this contrast-dependent emergence of the Cornsweet effect and subsequent Mach bands that is simulated here, under the assumption that perception is Bayes-optimal.

The Bayesian aspect of perception becomes crucial when considering the nature of illusions. Bayesian theories of perception describe how sensory data (that have a particular likelihood) are combined with prior beliefs (a prior distribution) to create a percept (a posterior distribution). One can regard illusory percepts as those that are induced by ambiguous stimuli, which can be caused in different ways – in other words, the probability of the data given different causes or explanations is the same. When faced with these stimuli, the prior distribution can be used to create a unimodal posterior and an unambiguous percept. If the percept or inference about the hidden causes of sensory information (the posterior distribution) is different from the true causes used to generate stimuli, the inference is said to be illusory or false. However, with illusory stimuli the mapping of hidden causes to their sensory consequences is ill-posed (degenerate or many to one), such that a stimulus can have more than one cause. Thus, from the point of view of the observer, there can be no 'false' inference unless the true causes are known. The perceptual inference can be optimal in a Bayesian sense, but is still illusory. However, not all possible causes of sensory input will be equally likely, so there will be an optimal inference in relation to prior beliefs about their causes. Prior beliefs can be learnt or innate: priors that are learnt depend upon experience while innate priors can be associated with architectural features of the visual brain, such as the complex

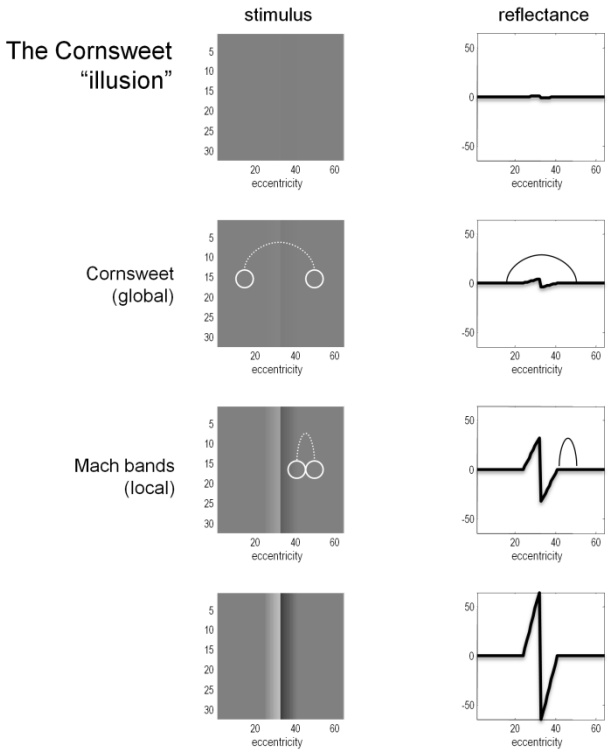

Figure 1.1: The Cornsweet Illusion and Mach Bands. The Cornsweet illusion is the false perception that the peripheral regions of a Cornsweet stimulus have different reflectance values. The magnitude of the effect increases as the contrast of the stimulus increases. At higher levels of contrast, the secondary illusion – Mach bands – appear. The Mach bands are situated at the point of inflection of the luminance gradient.

arrangement of blobs, interblobs and stripes in V1, that may reflect priors on the statistical structure of visual information selected by evolutionary pressure.

Prior beliefs are essential when resolving ambiguity or the ill-posed nature of perceptual inverse problems. There will always be an optimal posterior estimate of what caused a sensation that rests upon prior beliefs. The example in Figure 1.2 illustrates this: the central panel shows an ambiguous stimulus (luminance profile) that is formally similar to the sort of stimulus that induces the Cornsweet illusion. However, this stimulus can be caused in an infinite number of ways. Two plausible causes are shown here by assuming the stimulus is the product of (non-negative) illuminance and reflectance profiles. The lower two panels show the 'true' causes generating stimuli for the Cornsweet illusion. Here, the stimulus has a reflectance profile that reproduces the Cornsweet stimulus and is illuminated with a uniform illuminant. An alternative explanation for exactly the same stimulus is provided in the upper two panels, in which two isoreflectant surfaces are viewed under a smooth gradient of ambient illumination. In this example, both the illuminant and reflectance are made non-negative by applying an exponential transform before multiplying them to generate the stimulus.

The key point made by Figure 1.2 is that there are many possible gradients of illuminance and reflectance that can produce the same pattern of sensory input (luminance). These different explanations for a particular stimulus can only be distinguished by priors on the spatial and temporal characteristics of the reflectance and illuminance. In this example, the ambiguity about what caused the stimulus can be resolved if the observer believes, *a priori*, that the visual world is composed of isoreflectant surfaces, as opposed to surfaces that (implausibly) get brighter or darker nearer their edges or occlusions (as in the lower panels). Under this prior assumption, an observer who infers the presence of spatially extensive isoreflectant surfaces, and explains the edge at the centre with an illuminance gradient, would be inferring its most likely cause. The Cornsweet 'illusion' is thus only an illusion because the experimenter has chosen an unlikely combination of illuminance and reflectance profiles. In what follows, plausible priors on the spatial composition and generation of visual input are used to simulate the Cornsweet effect and the emergence of Mach bands.

The Bayesian approach to visual perception has been explored in previous work (Yuille, Geiger & Bulthoff, 1991; Knill & Pouget, 2004). In addition, several other visual illusions have been explained using Bayesian principles, including motion illusions (Weiss et al., 2002), the sound-induced flash illusion (Shams, Ma & Beierholm, 2005) and the Chubb

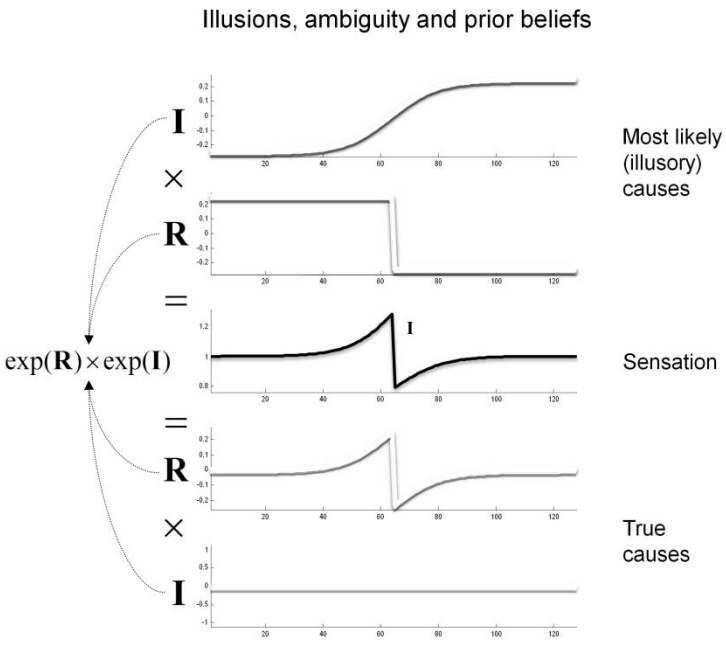

Figure 1.2: Contributions to luminance. Luminance values reaching the retina are modelled as a multiplication of illuminance from light sources and reflectance from the surfaces in the environment. The factorisation of luminance is thus an ill-posed problem. One toy example of this degeneracy is shown here; the same stimulus can be produced by (at least) two possible combinations of illuminance and reflectance. Prior beliefs about the likelihood of these causes can be used to pick the most likely percept.

illusion (Lotto & Purves, 2001). Additionally, Purves et al. (1999) demonstrated the Bayesian nature of the Cornsweet illusion: when presented in a context implying an illuminance gradient and reflectance step, the Cornsweet illusion is elicited more easily.

In terms of the neuronal systems mediating the Cornsweet illusion; some authors have implicated subcortical structures: for example, Anderson, Dakin & Rees (2009) found that BOLD signal in the lateral geniculate nuclei (LGN) best correlated with perception of the Cornsweet illusion, although correlations were also seen in visual cortex. Furthermore, the illusion could be abolished if the stimulus was not presented binocularly, suggesting an origin before V1. Mach bands similarly have been attributed to retinal mechanisms (e.g. Ratliff 1965); however, Lotto et al. (1999) have suggested a high-level contextual explanation for their appearance. Irrespective of the cortical or subcortical systems involved, it is assumed that the same Bayesian principles operate and, crucially, rest on a hierarchical generative model that necessarily implicates distributed neuronal processing at the subcortical and cortical levels.

1.2 Overview

This chapter comprises three sections. The first describes a simple generative model of visual input that entails prior beliefs about how visual stimuli are generated and can be used to infer their causes. This model is used in the second section to simulate the perception of the Cornsweet illusion and contrast-dependent emergence of Mach bands. In the third section, the predictions of the simulations are tested using a psychophysics study of normal subjects.

2. A generative model for the Cornsweet effect

These simulations are based upon the free-energy formulation of Bayes-optimal perception described in the introduction and elsewhere in the literature (Friston et al, 2006; Friston, 2009; Feldman & Friston 2010). Free-energy is a function of sensory samples and a probabilistic representation of what caused those samples. This representation can be cast in terms of the most likely or expected states of the world, under a generative model of how they conspire to produce sensory inputs. In brief, once the agent's generative model is known, one can use the free-energy principle to predict its behaviour and perception. In the present context, the focus will be on perception and the role of prior beliefs that are an inherent part of the generative model. In what follows, I

describe the model and then use it to simulate perceptual inference and electrophysiological responses.

2.1 The generative model

The generative model used here is straightforward: sensory input (luminance) is the product of reflectance and illuminance, where illuminance varies smoothly over space but can fluctuate with a high frequency over time. Conversely, the reflectance profile of the visual world is caused by isoreflectant fields or surfaces that fluctuate smoothly in time. Crucially, the spatial scales over which these fluctuations occur have a scale-free nature, of the sort found in natural images (Field, 1987; Burton & Moorhead, 1987; Tolhurst et al., 1992; Ruderman & Bialek 1994; Ruderman, 1997). To ensure positivity of the illuminant and reflectance an exponential transform to the two factors is applied before multiplying them (as in Figure 1.2). Equivalently, the underlying causes (reflectance and illuminance) can be envisaged as being composed additively in log-space. This model is shown schematically in Figure 1.3, in terms of hidden causes and states. Mathematically, this model can be expressed as:

$$
\begin{aligned}
s &= g(x,v) = \exp\left(\mathbf{R}\cdot x + \mathbf{I}\cdot v^{I}\right) + \omega_s \\
\dot{x} &= f(x,v) = v^{R} - x + \omega_x \\
\\
\omega_s &\sim \mathrm{N}\,(0, e^{-6}) \\
\omega_x &\sim \mathrm{N}\,(0, e^{-12})
\end{aligned}
\qquad (1.1)
$$

Here, s are sensory signals generated from hidden states x and causes v plus some random fluctuations ω. Luminance is modelled as the product (or sum of the exponents) of illuminance and reflectance. $\mathbf{R}$ and $\mathbf{L}$ are collections of spatial basis functions; inverting the model to find v means determining what weights to give to each of these basis functions to best explain the image.

The basis functions for reflectance are a set of Haar wavelets (simple step functions) thinned at the edges of the visual field because of the increasing size of classical receptive fields in the periphery. For simplicity (and ease of reporting the results), these simulations are restricted to a one-dimensional visual field. Because Haar wavelets afford local linear approximations to continuous reflectance profiles, the resulting reflectance has to be a mixture of isoreflectant surfaces at different spatial scales. To impose the

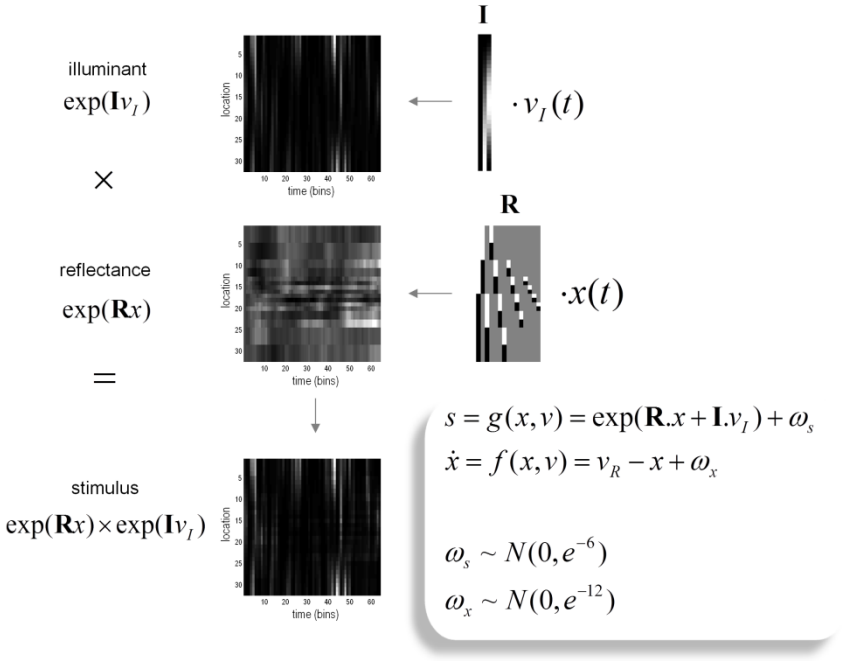

Figure 1.3: The generative model. The generative model employed in this chapter models illuminance as a discrete cosine function and reflectance as a Harr wavelet function with peripheral high-frequency wavelets removed. In addition, illumination is allowed to change quickly over time, whereas reflectance varies more slowly. This is achieved by making the coefficients on the reflectance basis functions hidden states, which accumulate hidden causes to generate changes in reflectance. Inversion of the model provides conditional estimates of the hidden causes and states responsible for sensory input as a function of time. See main text for an explanation of the variables in this figure.

scale-free aspect, I decrease the variance or, equivalently, increase the precision of the reflectance wavelet coefficients or hidden states in proportion to the order or spatial scale of their wavelet. This is implemented by placing a prior on the wavelet coefficients with the form $p(x_k) = \mathrm{N}\ (0, e^{-3k})$, where k is the order of the wavelet. Neuronally, these basis functions could stand in for a filling-in process such as that described by Grossberg & Hong (2006).

Conversely, the illuminant is modelled as a mixture of smoothly varying cosine functions with a low spatial frequency. This is easily motivated by the fact that most sources of illumination are point sources, which results in smooth illuminance profiles. These were modelled here with the first three components of a discrete cosine transform (see Figure 1.4 for a graphical representation of the basis functions and how they are used to generate a stimulus).

By construction, this generative model of visual signals separates the spatial scales or frequencies of the illuminance and reflectance such that all the high frequency components are in the reflectance profile, while the low frequency components are in illuminance profile. Temporal persistence of reflectance is assured because the reflectance coefficients $x(t) \in \Re^{16\times 1}$ are hidden states that accumulate hidden causes $v_R(t) \in \Re^{16\times 1}$. This persistence reflects the prior belief that surfaces move in a continuous fashion. For simplicity, the hidden causes controlling illuminance $v_L(t) \in \Re^{3\times 1}$ are mapped directly to the stimulus (although this is not an important feature of the model). This can be thought of as accommodating rapid changes in illuminance of the sort that might be produced by a flickering candle.

Equation 1.1 defines the generative model in terms of the joint probability over sensory information and the hidden variables producing fluctuations in reflectance and illuminance. The fluctuations in the hidden causes are assumed to be Gaussian with a precision (inverse variants) of one, while the fluctuations in the motion of the hidden states are assumed to have a log precision of twelve. Finally, sensory fluctuations or noise are given a log-precision of six. In the next section, manipulating the log-precision of the sensory noise is used as a proxy for changing the contrast of the stimulus.

Figure 1.4 shows a snapshot of the sort of visual signals this generative model produces. Here, the outer product of the discrete transforms above is used to generate a two-dimensional stimulus. This is not designed to be a veridical model of the real visual world.

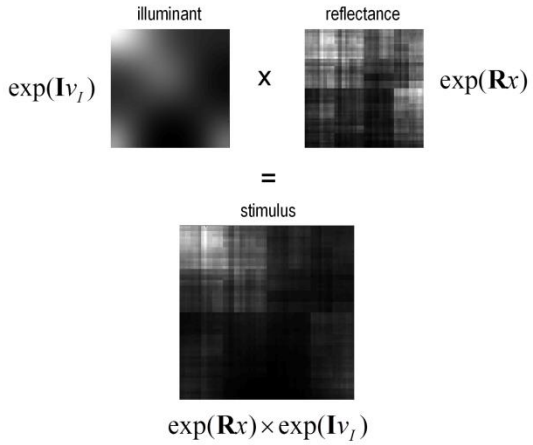

Figure 1.4: 2-D example. The top panels show 2-D examples of luminance (left) and reflectance (right), created form the generative model. The resulting stimulus is the product of the two (bottom panel).

However, it is sufficient to explain the Cornsweet illusion and related effects by incorporating simple and plausible priors on the spatial scales over which illuminance and reflectance change. In the next section, this generative model is used to simulate perceptual and physiological responses to a stimulus, under the free-energy principle. This reduces to a Bayesian deconvolution of sensory input that tries to discover the most likely hidden causes and states generating that input.

To simulate these responses, Equation 0.1 (discussed in the introduction), is integrated or solved using the functions $g(x,v)$ and $f(x,v)$ specified by a generative model in Equation 1.1. These functions map hidden variables to sensory input and encode prior beliefs about the dynamics of hidden states. In short, by plugging the equations of the generative model in Figure 1.3 into the predictive coding scheme of Figure 0.1 in the introduction, Bayes-optimal inference about the causes of sensations can be simulated. Crucially, the posterior or conditional beliefs about these causes can then be reconstructed and associated with percepts. In particular, the posterior belief about any mixture of hidden variables can be assessed. This process is used to quantify the Cornsweet and Mach band percepts, in terms of reflectance differences among different parts of the visual field. Note that the predictive coding scheme in Figure 0.1 weights the prediction errors by precision matrices, which encode the expected amplitude of random fluctuations.

2.2 Simulated responses

The simulated responses in Figure 1.5 were obtained by presenting the Cornsweet stimulus under uniform illumination. Here, the stimulus was presented transiently by modulating the illumination with a Gaussian envelope over time (see image inset). The resulting predictions are shown in panel A as solid lines, while the red dotted lines correspond to the prediction error. These predictions are based upon the inferred hidden states and causes shown on the right and lower left respectively. The lines correspond to the posterior expectations and the grey regions correspond to 90% Bayesian confidence intervals. In terms of the underlying causes, the blue curve in panel B is an estimate of the (log) amplitude of uniform illumination. This should have a roughly quadratic form (given the Gaussian envelope), peaking at around bin 30, which indeed it does. The remaining causes that deviate from zero (panel C) are the perturbations to the hidden states explaining or predicting changes in reflectance. These drive increases or decreases in the conditional expectations of the hidden states shown on the right. The green line is the coefficient of the second-order basis function splitting the visual field into an area of brightness on the left and darkness on the right. It can be seen that at the point of

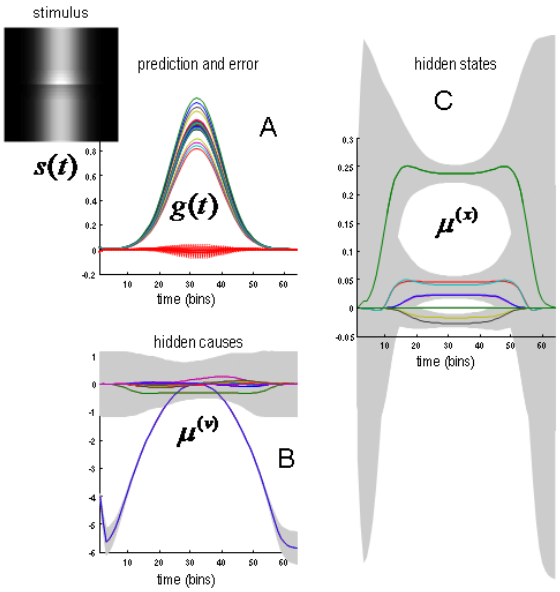

Figure 1.5: Predictions of the model. Panel C shows the estimates of the hidden states (coefficients of the reflectance basis functions) over time. The hidden state controlling the amplitude of the lowest-frequency basis function, which corresponds to the Cornsweet percept, contributes substantially to the overall perception of the stimulus (green line). The estimates for the hidden causes are shown in panel B. The grey areas are 90% confidence intervals. Panel A shows predictions (solid lines) of sensory input based on the estimated hidden causes and states and the resulting prediction error (bottom lines). The insert on the upper left shows the time-dependent luminance profile used in this simulation. Please see main text for further details.

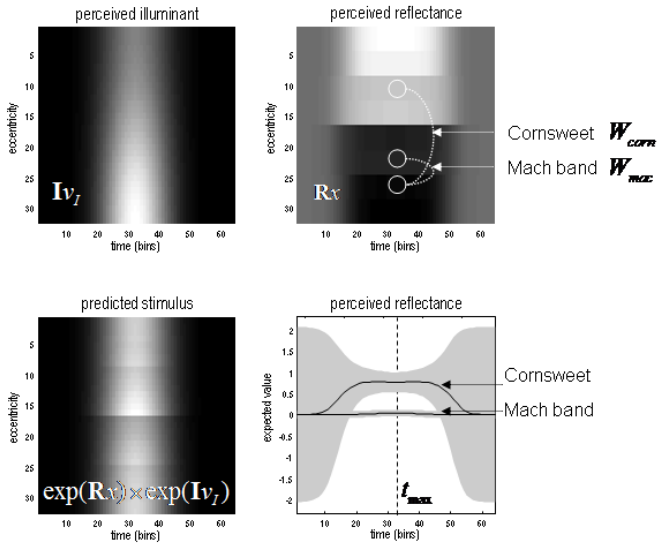

Figure 1.6: The model's 'perceptions'. The upper panels show the predicted illuminance (left) and reflectance profiles (right), reconstructed from the coefficients of the basis functions estimated from the model inversion shown in the previous figure. An inferred reflectance profile demonstrating the Cornsweet illusion is apparent, but at this level of contrast, Mach bands have not yet appeared. Please see main text further details.

maximum illumination, there is an extremely high degree of confidence that this hidden state is bigger than zero. This is the Cornsweet percept.

The corresponding percepts in sensory space are shown in Figure 1.6 as a function of peristimulus time. The upper panels show the implicit reflectance and illuminance profiles encoded by the conditional expectations of hidden variables respectively. After an exponential transform (and multiplication) these produce the sensory predictions shown on the lower left. By taking a weighted mixture of the perceived reflectance in different regions of the visual field (shown by the white circles) one can estimate the conditional certainty about both the Cornsweet effect (differences in perceived reflectance on different sides) and the appearance of Mach bands (differences in perceived reflectance on the same side). The weights used to evaluate these mixtures are denoted as W_{corn} and W_{mac} for the Cornsweet and Mach band effects respectively. The conditional expectation of these mixtures or effects $\mu_{mac} = W_{mac} \cdot \mu^{(x)}$ and their confidence intervals are shown on the lower right. At this level of visual contrast or precision (a log-precision of six), the Cornsweet effect is clearly evident with a high degree of certainty, while the confidence interval for the Mach band effect always contains zero. In other words, at this contrast (sensory precision) there is a Cornsweet effect but no Mach band effect. In the next section, the simulation above is repeated and the conditional expectations (and confidences) about illusory effects at the point of maximum illumination for different levels of contrast are recorded.

3. Contrast- or precision-dependent illusory percepts

Using the generative model and inversion scheme described above, the simulations are repeated over different levels of sensory precision. This can be regarded as a manipulation of contrast in the following sense: If it is assumed that the brain uses divisive normalisation (Weber, 1846; Fechner, 1860; Craik, 1938; Geisler & Albrecht 1992; Carandini & Heeger, 1994), the key change in sensory information, following an increase in contrast, is an increase in signal to noise; in other words, its precision increases. Therefore, a manipulation of the log-precision of sensory noiseis used to emulate changes in visual contrast. It should be noted that sensory noise was not actually added to the stimuli. The key quantity here is the level of precision assumed by the agent which, in these simulations, was changed explicitly. In more realistic simulations, the log-precision would be itself optimised with respect to free-energy (see Chapter 3 and Feldman & Friston 2010 for an example of this in the modelling of attention).

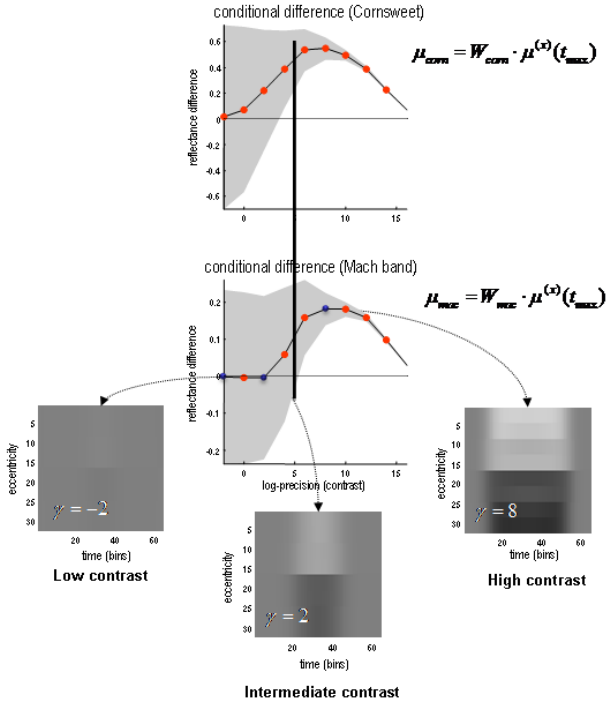

Figure 1.7: The effect of contrast. In the results presented here the precision of the sensory data was used to model stimulus contrast. As precision increases, the strength of both the Cornsweet and Mach effects increases until, at very high levels of precision (contrast) the true luminance profile is perceived and the illusory percepts fade. Crucially, the Mach bands appear at higher levels of contrast than the Cornsweet illusion. The inserts in the lower panels show the inferred reflectance is at different levels of contrast (indicated by the blue dots in the lower graph). The prediction errors associated with the processing of stimuli and these three levels are shown in the next figure. Please see main text for further details.

Figure 1.7 shows the results of perceptual inference under the Bayesian scheme described above. The only thing that was changed was the log-precision of the sensory input, from minus two (low) through to intermediate levels and ending with a very high log precision of sixteen. The two graphs show the conditional expectation and 90% confidence intervals for the Cornsweet effect (upper panel) and Mach bands (lower panel) respectively, at the point of maximum illumination. It can be seen in both instances that under low levels of contrast (sensory precision) both effects are very small and inferred with a large degree of uncertainty. However, as contrast increases, conditional uncertainty reduces and, at a critical level, produces a confident inference that the effect is greater than zero (or some small threshold). Crucially, the point at which this happens for the Cornsweet effect is at a lower level of contrast than for the Mach bands. In other words, the Cornsweet illusion occurs first and then the Mach bands appear as contrast continues to rise. The explanation for this is straightforward; the Mach band illusion rests upon higher spatial frequencies in the generative model, which have a higher prior precision (encoding prior beliefs about the statistical – scale free – structure of natural visual scenes). This means that there needs to be precise sensory evidence to change them from their prior expectation of zero. In short, at high levels of contrast or sensory precision, more and more fine detail in the posterior percept is recruited to provide the optimum explanation for the stimulus. Interestingly, as the contrast or sensory precision reaches very high levels, the veridical reflectance and illuminant profiles are inferred and, quantitatively, both the Cornsweet and Mach band effects disappear. The three images show exemplar percepts, at low, intermediate and high levels of contrast respectively. The key difference in the spatial banding that underlies the Cornsweet and Mach band effects is evident in the difference between the intermediate and high levels of contrast.

The key prediction of these simulations is that subjects categorize their percepts differently at different levels of contrast, following a brief exposure to a Cornsweet stimulus. At low levels of contrast, the stimulus would be categorised as uniformly flat. At intermediate levels of contrast, the stimulus would be categorised as a Cornsweet percept, with isoluminant and uniform differences in the right and left hand parts of the visual field; while at higher levels of contrast one would expect the Mach bands to dominate and the stimuli would be categorized as possessing para-central bands. In principle, at very high levels of contrast, the subject should perceive the veridical stimulus. However, whether this level of contrast can be attained empirically is an open question.

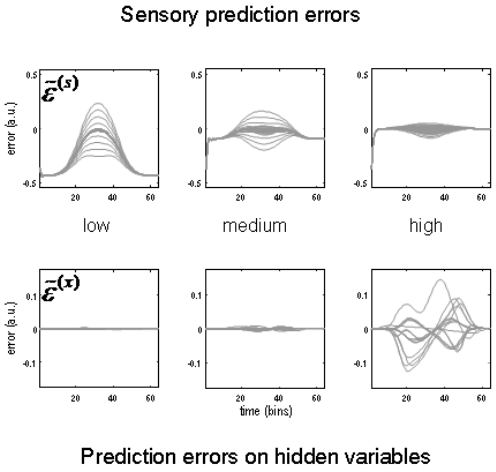

Figure 1.8: Contrast and prediction error. As stimulus contrast increases, prediction error is redistributed from sensory input to hidden variables. At high levels of precision, sensory information induces prediction errors at higher levels, which in turn explain away prediction error at the sensory level. The higher-level prediction errors at high precision reflect increasing confidence that the reflectance is different from the prior expectation of zero. Please see main text for further details.

3.1 Simulating Neural Responses

Figure 1.8 shows the prediction errors at low, intermediate and high levels of contrast. These are shown at the sensory level (upper row) and at the higher levels of the hidden causes and states (lower row). The key thing to note here is that as contrast increases and the spatial detail of the posterior predictions increases, the sensory prediction error falls. This is at the expense of inducing prediction errors at the higher level, which increase in proportion to the precision of sensory information. These higher prediction errors are simply the difference between the posterior and prior expectations and reflect an increasing departure from a prior expectation of zero as contrast (the log precision of sensory noise) increases. Although these results are interesting in themselves, they can also be regarded as a simulation of event related potentials. The reason that we can associate prediction error with observed electromagnetic brain responses is that it is usually assumed that prediction errors are encoded by the activity of superficial pyramidal cells (see Figure 0.1). It is these cells that are thought to contribute primarily to local field potentials and non-invasive EEG signals.

In the high-contrast condition, the prediction errors at the lower level are suppressed by the prediction of the presence of a Cornsweet stimulus. This sort of phenomenon has been demonstrated using fMRI (Alink et al., 2010, den Ouden et al., 2010); predictable stimuli cause less activation in stimulus-specific areas than unpredictable stimuli. However, the process simulated here is likely to produce more complicated neurophysiological correlates because of the confounding effect of precision; increased predictability (through increasing conditional confidence about the stimulus) will also increase estimates of precision. Since we believe that the prediction errors reported by superficial pyramidal cells are precision-weighted, decreasing prediction error in lower sensory areas may be masked by the increasing precision of those errors. We will return to this and related issues chapter 2. Here, we focus on psychophysical correlates.

4. A psychophysical test of theoretical predictions

In this section, I report a psychophysics study of normal subjects exposed to the same stimuli used in the simulations above. Instead of using normal procedures for assessing illusions (which usually involve matching intensity differences), I use a forced-choice paradigm. This is because I wanted to present stimuli briefly: Firstly, because brief presentation avoids the confounding effects of saccadic eye movements, and secondly, a

forced-choice paradigm places constraints on the subject's choices that map directly to the model predictions. In what follows, I describe the paradigm and interpret the results quantitatively, in relation to the predictions of the simulations above.

4.1 Subjects and experimental paradigm

27 healthy subjects were studied in accordance with guidelines established by the local ethical committee and after obtaining informed consent. 8 participants (4 female) completed the Mach band paradigm; 19 (12 female) completed the Cornsweet paradigm.

Experiment 1 (Cornsweet paradigm): The Cornsweet illusion was assessed using a 2-interval forced choice procedure. Subjects were presented with a (set contrast) Cornsweet stimulus and real luminance step for 200ms (a Gaussian temporal envelope was not used), separated by an interval of 200ms. One stimulus appeared to the left of fixation and one to the right; this was randomised across trials, as was the order of the stimuli. Subjects were asked to report the side on which the stimulus with the greatest contrast had appeared (Figure 1.9).

Six blocks were completed, using Cornsweet stimuli with Weber contrasts of 0.0073 to 0.0734. A Quest procedure (Watson & Pelli 1983) was used to select each step stimulus for comparison. The mean of the psychometric function was taken as the point of subjective equality between the Cornsweet stimulus and a real luminance. There were 200 presentations per block.

Experiment 2 (Mach band paradigm): The same method could not be used to identify the strength of the Mach bands percept, as there is no non-illusory stimulus that can be used for matching. Consequently, a two-alternative forced choice paradigm was used: A single Cornsweet stimulus was displayed for 200ms to the left or right of fixation and participants were asked to report if the stimulus contained Mach bands or not. Each participant completed six runs of 200 presentations at ten Weber contrast levels from 0.0204-0.2038. The probability of reporting a Mach band was assessed as the relative frequency of reporting its presence over trials, within subject (Figure 1.9).

In both experiments, stimuli were displayed on an LCD monitor under ambient room lighting. Subjects were seated 60cm away from the monitor, such that stimuli subtended an angle of 14.21^{o} vertically and 6.10^{o} horizontally, at 2.96^{o} -7.06^{o} eccentricity. The luminance ramp of the Cornsweet stimulus profile occupied 2.42 degrees. Only the lowest

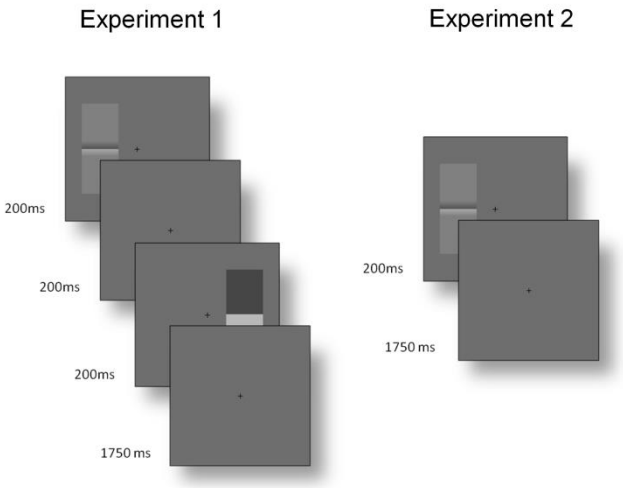

Figure 1.9: Time courses of trials. Experiment 1: Before the start of each trial, participants fixated a central cross. One Cornsweet and one real luminance step-stimulus appeared for 200ms each, with a 200ms interval between them. The order of the stimuli, their orientation and the side on which each appeared were randomized (although they were constrained to appear on opposite sides within each trial). Participants then had 1750ms to report the side on which the stimulus with the greatest contrast had appeared (using the arrow keys of the keyboard). Experiment 2: Each trial stared with fixation. A Cornsweet stimulus then appeared for 200ms with a random orientation to the left or right of fixation. Participants had 1750ms to report, with the 'Y' and 'N' keys, if they perceived Mach bands.

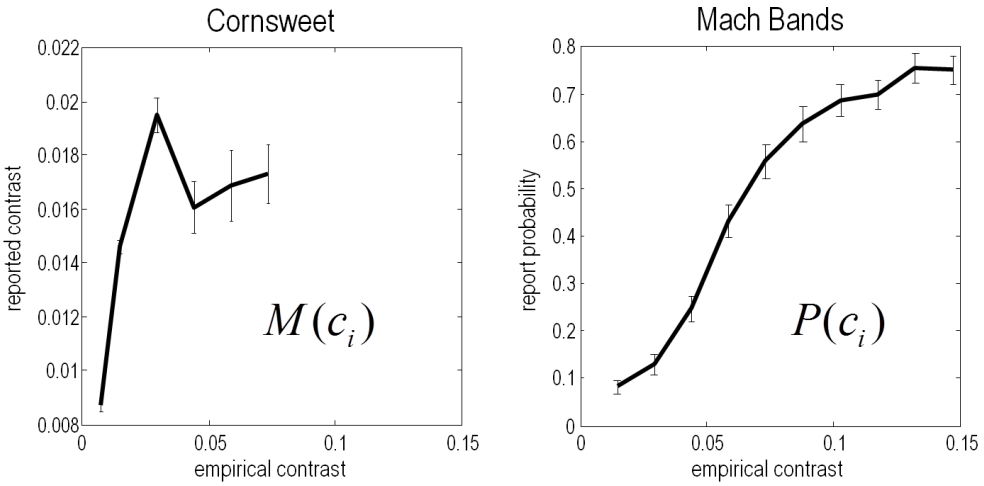

Figure 1.10: Experimental results. This figure shows the results of the empirical psychophysical study for the Cornsweet illusion (left panel) and the Mach band illusion (Right panel). Both report the average effect over subjects and the standard errors (bars). The Cornsweet illusion is measured in terms of the subjective contrast level (quantified in terms of subjective equivalence using psychometric functions). The Mach band illusion is quantified in terms of probability that the illusion is reported to the present. Both results are shown as functions of empirical (Weber) stimulus contrast levels. The same range has been used for both graphs so that the dependency on contrast levels can be compared. The key thing to note here is that the Cornsweet illusion peaks before the Mach band illusion (vertical line).

levels of contrast the monitor was able to produce were employed; thus, luminance values were linearised post hoc.

4.2 Results and discussion

The results of the psychophysics experiments are shown in Figure 1.10, as a function of empirical (Weber) contrast levels. These results are expressed as the mean over all subjects and associated standard errors. The reported Cornsweet effect (as indexed by the point of subjective equality) peaked, on average over subjects, at a contrast of about 0.025. At higher levels of contrast, as in the simulations, the effect falls before plateauing at the highest contrast used in Experiment 1. Conversely, the probability of reporting a Mach band increased monotonically as a function of the empirical contrast, reaching about 75% at a Weber contrast of about 0.15.

Qualitatively, these empirical results compare well with the theoretical predictions shown in Figure 1.9: that is, the subjective or inferred Cornsweet effect emerged before the Mach bands, as contrast increases. The characteristic 'inverted U' dependency of the Cornsweet effect on contrast levels is also seen. The empirical profile is somewhat compromised by the small range of contrasts employed, however, this range was sufficient to disclose an unambiguous peak. Clearly, it would be nice to relate these empirical results quantitatively to the simulations shown in Figure 1.8. This presents an interesting challenge because the psychophysical data consist of reported levels of an effect and the probability of an effect for the Cornsweet and Mach Band illusions respectively. However, because these simulations provide a conditional or posterior probability over the effects reported, both sorts of reports can be simulated to see how well they explain the psychophysical data. This quantitative analysis is now considered in more detail.

4.3 A formal behavioural analysis

The simulations provide conditional expectations (and precisions) of both the Cornsweet and Mach band effects over a number of simulated (Weber) contrast levels, as modelled with the precision of sensory noise. This means that one can compute a psychometric function of contrast c that returns the behavioural predictions of both illusions respectively; namely, the level of the illusion and the probability of inferring a Mach band. To predict the reported level of the Cornsweet illusion the conditional expectation $\mu_{corn}(c)$ can simply be scaled by some (unknown) coefficient β_1. To predict the probability of reporting the presence of Mach bands, one can integrate the conditional

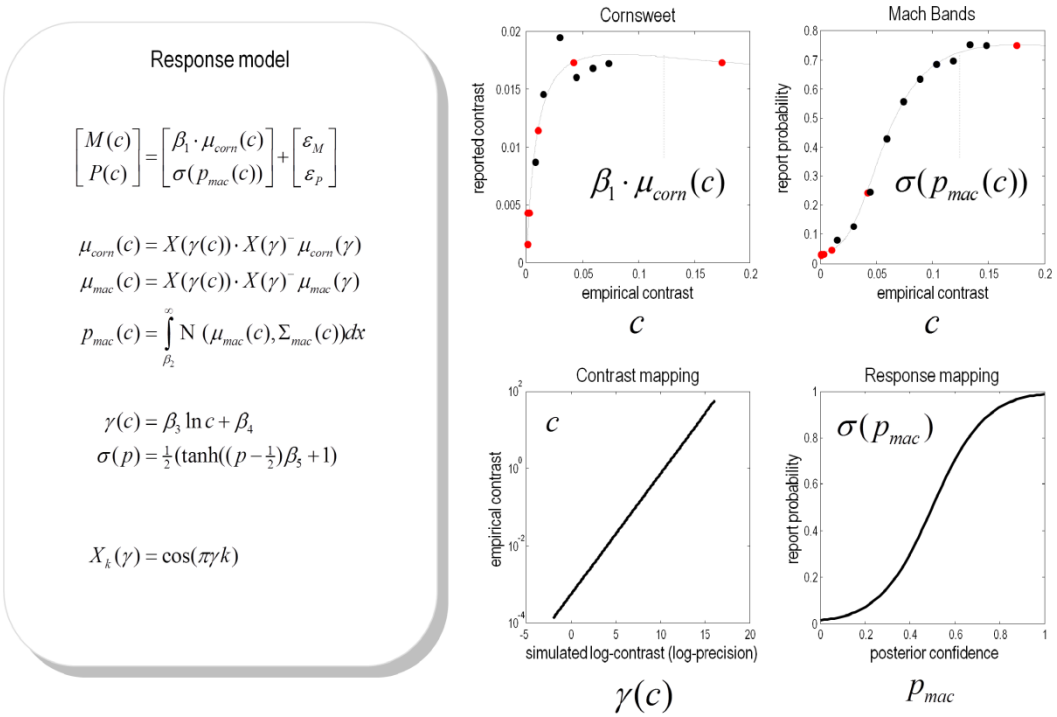

Figure 1.11: Predicted and experimental results. Theoretical predictions of the empirical results reported in the previous figure: These predictions are based upon a response model that maps from the conditional expectations and precisions in the simulations to the behavioural responses of subjects. This mapping rests on some unknown parameters or coefficients $\beta_i : i = 1, \ldots 5$ that control the relationship between the simulated $\gamma(c)$ and empirical contrast levels c used for stimuli and the relationship between the probabilities of reporting a Mach band $\sigma(p_{mac}(c))$ and the conditional probability that it exceeds some threshold $p_{mac}(c)$ at contrast level c. These relationships form the basis of a response model, whose equations are provided in the left panel (for simplicity, the expressions for conditional variance of the Mach band contrast have been omitted). Given empirical responses for the mean Cornsweet effect $M(c)$ and probability of reporting a Mach band $P(c)$, the coefficients β_i can be estimated under the assumption of additive prediction errors ε. The predicted responses following this estimation are shown in the graphs on the right-hand side. The upper panels show the empirical data superimposed upon conditional predictions from the model. The blue lines are the predicted psychometric functions, $\mu_{corn}(c)$ and $\sigma(p_{mac}(c))$. The red dots correspond to the predictions at levels of contrast used in the simulations (as shown in Figure 1.8), while the black dots correspond to the empirical responses: $M(c)$ and $P(c)$. The lower left panel show the relationship between the empirical and simulated contrast levels (as a semi-log plot of the empirical contrast against the log-precision of sensory noise). The lower right panel shows the relationship between the probability of reporting a Mach band is present and the underlying conditional probability that it is above threshold.

probability distribution over the Mach band effect above some (unknown) threshold β_2. However, to do this, the relationship between the simulated and empirical contrasts must be known:

As noted above, the log-precision of sensory noise γ is used to model log-contrast in accord with Weber's law. This means a linear relationship can be assumed between the empirical log-contrast and simulated log-precision. This induces two further unknown coefficients – the slope and intercept (β_3 and β_4) that parameterise the relationship between the simulated and empirical contrasts. Finally, the conditional probability of a suprathreshold Mach band effect must be related to the probability of reporting its presence. Here, a simple, monotonic sigmoid relationship is assumed, under the constraint that when the conditional probability was 50:50, the report probability was also 50:50. The precise form of this mapping is provided in Figure 1.11 (left panel) and has a single (unknown) slope coefficient β_5. These relationships provide a mapping between the results of the simulations and the observed responses averaged over subjects (under the assumptions of additive prediction errors). This is known as a response model and is detailed schematically in Figure 1.11. The predictions are based on the simulated responses in Figure 1.8, assuming a smooth psychometric function of contrast that was modelled as a linear mixture of cosine functions: $X_k(\gamma) = \cos(\pi\gamma k)$ for $k = 1,\ldots,6$. The coefficients of this discrete cosine set were estimated with ordinary least squares, using the responses of the model $(\mu_{corn}(\gamma), \mu_{mac}(\gamma))$ over different precision levels, at the time of maximum luminance.

Given the form of the relationships between the simulated and empirical contrasts and between the report probability and conditional probabilities for Mach bands, one can use the psychophysical data to estimate the unknown coefficients of these relationships: $\beta_i : i = 1,\ldots,5$. The results of this computationally informed response modelling are shown in the right panel of Figure 1.11. The upper panels show the same data as in Figure 1.10 placed over the theoretical psychometric functions based on the simulations of the previous section. These predictions are based on the mapping from simulated to empirical contrast levels (lower left) and the relationship between the probability of reporting a Mach band and the conditional confidence that it is present (lower right).

By construction, the relationship between the simulated and empirical contrasts is linear when plotted on a log-log scale. The slope of this plot suggests that the higher contrasts

used in the simulations are, practically, not realisable in an empirical setting. This is because as the simulated contrast increases the corresponding empirical contrast increases much more quickly. The implication of this is that the contrasts employed in the psychophysics study correspond to the first few levels of the simulated contrasts. This means that it may be difficult to demonstrate the theoretically predicted attenuation of the Cornsweet illusion at very high levels of contrast.

The relationship between the report and conditional probabilities suggests that subjects have a tendency to 'all or nothing' reporting; in the sense that a conditional confidence that the probability of reporting a Mach band is slightly greater than the conditional confidence it is above threshold. Conversely, subjects appear to report the absence of Mach bands with a probability that is slightly greater than the conditional probability it is below threshold. The resulting psychometric predictions (in the upper panels) show a remarkable agreement between the predicted and observed probabilities of reporting a Mach band. The correspondence between the predicted and empirical results for the Cornsweet illusion are less convincing but show that both asymptote to a peak level much more quickly than the probability of reporting a Mach band.

In summary, this analysis suggests that there is a reasonable quantitative agreement between the theoretical predictions and empirical results. Furthermore, in practical terms, it appears that the normal range of Weber contrasts that can be usefully employed corresponds to a relatively low level of sensory precision in the simulations. This means that it may be difficult to demonstrate the 'inverted U' behaviour for the Cornsweet illusion seen in Figure 1.8. This is because it may be difficult to present stimuli at the ultra high levels of contrast required on a standard LCD screen. Note that the empirical probability of reporting a Mach band does not decrease as a function of contrast level. This is to be anticipated from the theoretical predictions: increasing the contrast level increases the conditional precision about the inferred level of the Mach band effect, which means that the probability that is above threshold can still increase even if the conditional expectation decreases (as in Figure 1.8).

5. Conclusion

This chapter has reviewed the nature of illusions, in the context of Bayes-optimal perception. I reiterate the notion that illusory percepts are optimal in that they may represent the most likely explanation for ambiguous sensory input. This has been illustrated using the Cornsweet illusion. By using simple and plausible prior expectations

about the spatial deployment of illuminance and reflectance, the emergence of the Cornsweet effect as a natural consequence of Bayes-optimal perception has been demonstrated . Furthermore, a contrast-dependent emergence of the Cornsweet effect could be simulated and subsequent appearance of Mach bands that was verified psychophysically using a forced choice paradigm.

The work described in this chapter is published as:

Brown H, Friston KJ. (2012). Free-energy and illusions: the cornsweet effect. Front. Psychol. 3:43.

Chapter 2

Dynamic causal modelling of precision and synaptic gain in visual perception – an EEG study

1. Introduction

Predictive coding has proved helpful in explaining many visual phenomena, including extraclassical receptive field effects in V1 (Rao & Ballard, 1999), repetition suppression (Summerfield et al., 2008) and modulation of early cortical responses by attention (Rauss, Schwartz, & Pourtois, 2011). Friston & Kiebel (2009) suggested a neurobiologically plausible scheme by which predictive coding could be executed in the cortex; key to these proposals is the idea that superficial pyramidal cells pass prediction error forward to higher cortical areas.

Classical predictive coding schemes are linear; however, these schemes cannot accommodate state dependent changes in the precision (inverse variability) of sensory signals. An important generalisation of predictive coding was introduced by (Feldman & Friston, 2010), discussed in Chapter 3, to accommodate the fact that the precision of sensory signals is highly context sensitive and depends upon the (hidden) states of the world, generating sensory inputs. In generalised predictive coding, precision scales the prediction error such that precise prediction errors have more influence at higher levels in representational cortical hierarchies; effectively, precision represents the signal-to-noise ratio or salience of prediction error associated with bottom-up signals. This sort of scaling

is vitally important for sensory perception; for example, in multimodal integration (Ernst & Banks, 2002) and in reconciling information derived from sensory signals and prior knowledge (Rahnev, Lau, & de Lange, 2011). Precision also influences the perception of visual contrast – increasing the relevance and the probability of a visual contrast signal have dissociable effects on energy sensitivity (Wyart et al., 2012), and attention allows the adoption of more stringent (conservative) detection criteria during contrast detection (Rahnev et al., 2011), suggesting that contrast detection is dependent on the estimated precision of sensory information. The introduction of precision also explains away the apparent contradiction between biased competition, which boosts expected signals (prediction errors) and predictive coding, which attenuates them (Feldman & Friston, 2010). This has recently been demonstrated experimentally; (Kok et al., 2011) have shown, in an fMRI paradigm, that attention reverses the attenuation of BOLD signal seen in response to predictable stimuli. In generalised predictive coding, superficial pyramidal cells have been proposed to report precision-weighted prediction error, rather than pure prediction error (Friston & Kiebel, 2009). This sort of scheme is formally similar to those based upon adaptive resonance theory (Grossberg & Versace, 2008). See also (Spratling, 2008). Crucially, it provides a plausible mechanism for attentional modulation.

In chapter 1, I describe investigations into the role of precision in the context of visual illusions – in particular the Cornsweet illusion (Brown and Friston, 2012). This illusion is perhaps the simplest of visual illusions, where a visual 'edge' between two isoluminant regions creates the impression that the one region is brighter than the other. I will focus on this illusion because it provides clear psychophysical evidence that contrast affects the encoding of precision in predictive coding – in a way that motivates the electrophysiological hypotheses tested in this chapter. The Cornsweet illusion was simulated by assuming observers make predictions about their visual input using a generative model in which reflectance and illuminance interact to produce sensory signals. Crucially, the generative model included prior beliefs that luminance varied with low spatial frequency whereas reflectance varied with high spatial frequency. Inversion of this generative model under a generalised predictive coding framework replicated the illusory perception of the Cornsweet illusion stimulus. In short, it was sufficient to explain the Cornsweet illusion purely in terms of plausible prior beliefs about the spatial frequency structure of illuminance profiles and reflecting surfaces.

This is not the first explanation of the Cornsweet illusion to be based on Helmholtz's (Helmholtz, 1924) idea that the visual system must remove the effect of the illuminant to perceive. (Purves, Shimpi, & Lotto, 1999; Purves et al., 2004) have suggested a Bayesian

explanation for the Cornsweet illusion and demonstrated that contextual cues indicating two differently reflectant surfaces subject to an illuminance gradient increase the magnitude of the Cornsweet illusion. Mechanistically, this may be achieved by filling-in (Grossberg & Hong, 2006) - see also Reynolds & Heeger (2009) for a contemporary discussion in the context of the normalisation model.

In these simulations, the emergence of the Cornsweet illusion depends critically on the luminance contrast of the stimuli. To simulate the effects of changing contrast, the precision of the sensory inputs (the first hierarchical layer of the generative model), was manipulated; high precision corresponded to higher contrast and vice versa. This manipulation was based on Weber's law (Weber, 1846) and evidence that the brain uses divisive normalisation (Brady & Field, 2000; Carandini & Heeger, 1994), meaning that higher contrast stimuli have a higher signal-to-noise ratio and therefore higher precision. Associating stimulus contrast with precision enabled the generalised predictive coding scheme to accurately reproduce the effects of contrast on human observers; namely, the magnitude of the Cornsweet illusion increased to a plateau (Figure 1.10), and Mach bands appeared at higher contrast. These results suggest that increasing stimulus contrast increases the precision of sensory signals encoded by early visual areas.

In the context of the generalised predictive coding scheme outlined in the introduction, precision is encoded by the gain of superficial pyramidal cells. Gain in early visual areas is known to be important for contrast perception. Cells in visual area V1 are sensitive to contrast and, on a timescale that precludes neuronal adaptation, the firing rate of such cells generally increases linearly in response to increasing visual contrast, except at very high contrast levels (Albrecht, Farrar, & Hamilton, 1984; Ohzawa, Sclar, & Freeman, 1982). Although this pattern is consistent with a gain-based explanation of visual contrast coding, this is by no means the only explanation. In this study, I used EEG and dynamic causal modelling to investigate role of synaptic gain in contrast perception in early and higher cortical areas.

2. Methods

2.1 Participants

18 healthy right-handed subjects participated in the study (9 male; age 20-56). Ethical approval was obtained from the UCL Research Ethics Committee (no. 2715/002). Written informed consent was obtained from all subjects.

2.2 Experimental paradigm

Cornsweet illusion stimuli (Figure 2.1) were created by applying a bandpass filter to 1024x512 arrays of white noise to produce a random blob pattern with a fundamental frequency of 67 blobs/image (1 cycle/degree). This pattern was thresholded and convolved with a 2-D Laplacian-of-Gaussian filter to give a Cornsweet illusion stimulus. Stimuli were scaled to have 10%, 25% or 90% of the maximum contrast supported by the monitor. The stimuli occupied both lower quadrants of the screen, subtending approximately 32^{o} of visual angle. The central 2^{o} of visual angle were left blank. Stimuli were presented against a grey background on a gamma-corrected monitor. Average luminance was 48cd/m2.

Participants sat on a comfortable chair and rested their head on a chin rest. The stimuli were displayed on an LCD monitor 60cm from the subjects. During the task, subjects fixated on a central cross. One of the three Cornsweet illusion stimuli was presented on the bottom half of the screen for 400ms. Inter-trial interval was jittered between 600ms and 800ms. Three sessions of 1200 stimuli were presented, over about one hour's scanning time. During the task, the fixation cross changed to a circle and back again between stimuli, randomly with a probability of 0.01, to provide targets for an incidental task, used to maintain attentional set. Participants counted these events and reported the total to the experimenter after each session.

2.3 Data collection and processing

EEG data were recorded using a *Biosemi* system with 128 scalp electrodes at a sampling rate of 512 Hz. An average reference was used. Vertical and horizontal eye movements were monitored with electro-oculogram electrodes. Electrode positions were recorded with a Polhemus digitiser. Data were analysed using SPM8.

Data were down-sampled to 200Hz and bandpass-filtered between 0.5Hz and 45Hz to suppress very low frequencies (Luck 2005). Baseline-corrected epochs were extracted from the time series starting at 100ms before stimulus onset and ending at 400ms after stimulus onset. Blink and eye-movement artefacts were detected by simple thresholding of electro-oculogram channels; artefactual trials were removed from the analysis. 9.7% of trials were excluded (range across subjects 0.3%-32%). Three types of event related averages were taken – an average for each subject and contrast level, an average over contrast levels for each subject and an average for each contrast level over all subjects.

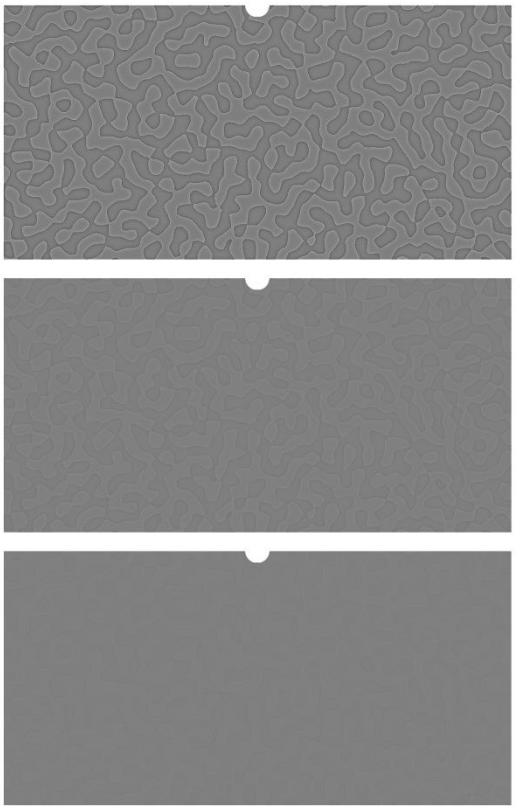

Figure 2.1: Cornsweet illusion stimuli used for this study. These stimuli were created by applying a bandpass filter to white noise to create a random blob pattern with a fundamental frequency of 67 blobs/image (1 cycle/degree). This pattern was thresholded and convolved with a 2-D Laplacian-of-Gaussian filter to produce a Cornsweet illusion stimulus. Stimuli were scaled to have 90% (top), 25% (middle) or 10% (bottom) of the maximum contrast supported by the monitor. The stimuli subtended approximately 32^{o} of visual angle. The central 2^{o} of visual angle were left blank. Stimuli were presented against a grey background on a gamma-corrected monitor.

2.4 Source Localisation

Using the event related potentials averaged over contrasts for each subject, source localisation was performed using multiple sparse priors and group constraints (Litvak & Friston, 2008). This localisation optimises prior covariance constraints on sources over subjects and provides maximum a posteriori estimates of activity at each source from a cortical mesh from 60ms post stimulus onset to 400ms post stimulus onset for each subject. These estimates were averaged over peristimulus time and projected to a three-dimensional source space, where they were smoothed to create an image of source activity for each subject. Individual subject images were averaged. This procedure was used to identify the location of four bilateral sources in each hemisphere (Figure 2.2). The sources were identified as the four bilateral peaks with the largest posterior estimates of evoked power (sum of squared source activity over peristimulus time – Litvak and Friston 2008).

2.5 DCM

Dynamic causal modelling as implemented in SPM8 was used to examine the changes in pyramidal cell gain due to changes in visual contrast (Kiebel et al., 2008). Dynamic causal modelling employs biophysically constrained models and a Bayesian inversion scheme to infer hidden variables relating to connectivity and synaptic efficacy by modelling EEG data as the response of a dynamic input-state-output system to experimental perturbations. The model comprises both a neuronal mass model that allows for directed coupling among hidden neuronal states and the electromagnetic forward model (used for source localisation above) that maps from hidden neuronal states to observed channel data.

The neuronal model employed in DCM consists of a number of discrete cortical sources, each comprising four cell populations – superficial and deep pyramidal cells, spiny stellate cells and inhibitory interneurons (figure 2.3). The activity of these populations is coupled by a series of differential equations, which are based on the intrinsic connectivity among cortical layers (Bastos et al., 2011). A series of parameters, (γ_1-γ_{10}) specifies the strength of intrinsic connectivity between populations; four of the intrinsic connections are optimised to fit the data, the others are fixed. One or more may be optimised in a condition-specific way.

Extrinsic connections link different sources. Extrinsic forward connections are excitatory, originate from superficial pyramidal cells and terminate on spiny stellate neurons.

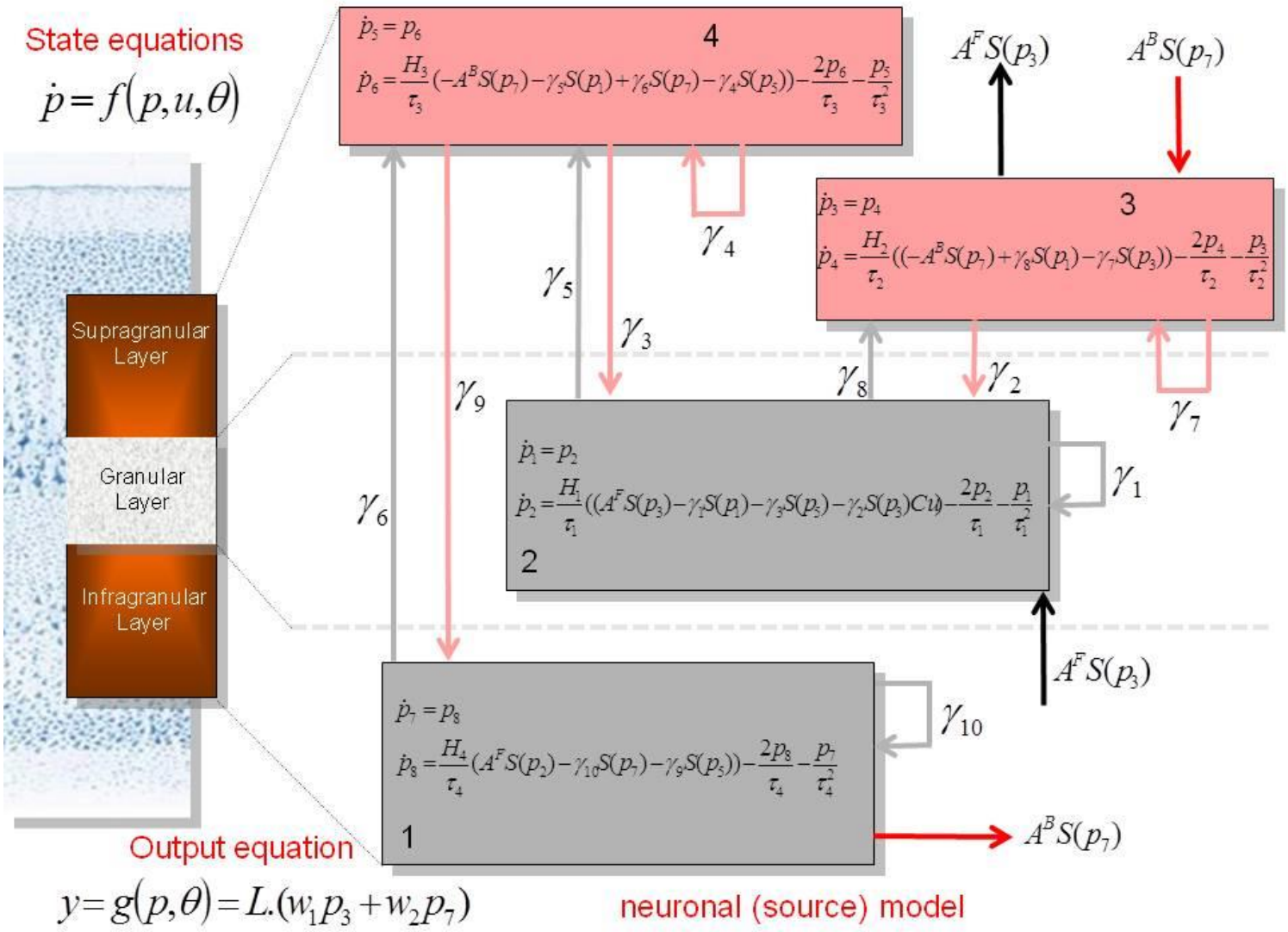

Figure 2.3: The structure of the neuronal mass model used to simulate activity in each cortical layer. The model consists of four cell populations: superficial (4) and deep pyramidal cells (1), spiny stellate cells (2) and inhibitory interneurons (3). The transmembrane potential of each cell is modelled by a set of differential equations which simulate the average voltage and current across a group of neurons (Jansen & Ritt, 1995). The voltage in each cell is transformed into a firing rate by the function S(p). Each cell can receive connections from other cells within the same cortical area, whose strength is modulated by γ parameters, or from specific cells in other cortical areas whose connection strengths are encoded in the matrices A_F and A_B.

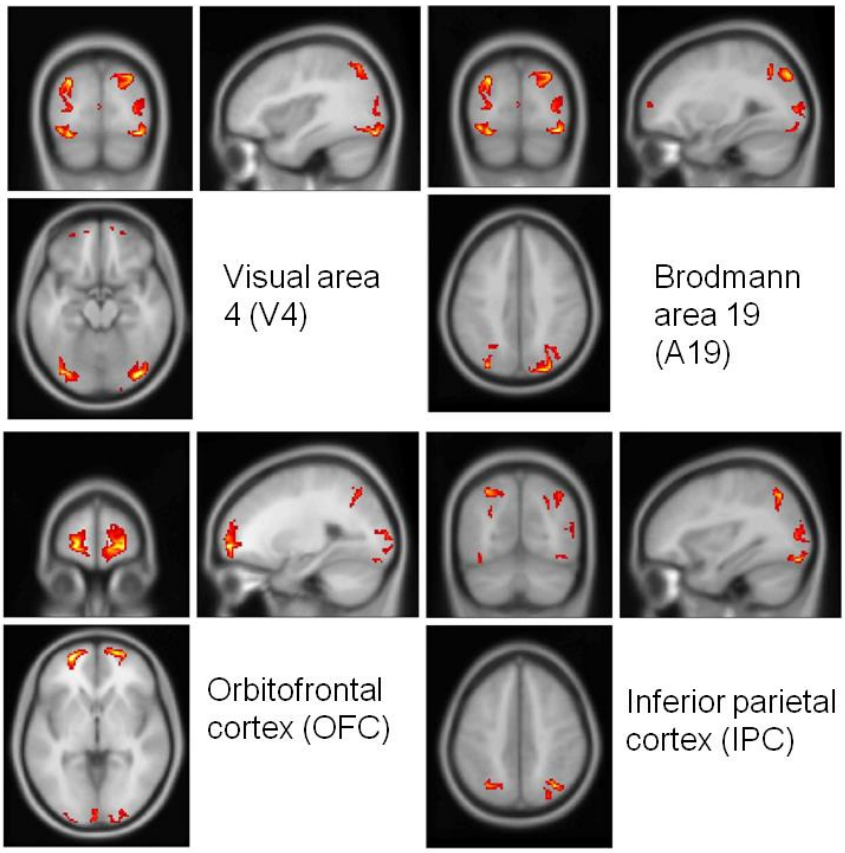

Figure 2.2: Source localisation. Sources located by source reconstruction using multiple sparse priors and group constraints. The figures show absolute source activity averaged across subjects; the maxima were used as source locations for DCM. Four locations emerged bilaterally: an area in the inferior occipital gyrus most closely corresponding to visual area V4, the inferior parietal cortex, Brodmann area 19 and the orbitofrontal cortex.

Extrinsic backward connections are inhibitory, originate from deep pyramidal cells and terminate on superficial pyramidal cells.

Under generalised predictive coding, superficial pyramidal cells are thought to signal precision-weighted prediction error (Friston & Kiebel, 2009). The precision is represented in the model by pyramidal cell self-connectivity, γ_7 (figure 2.3). As such, this parameter was altered to represent changes in precision on prediction errors. Anatomical evidence clearly shows that the superficial pyramidal cells are the origins of the majority of forward connections, which ascend the visual cortical hierarchy, and that these connections drive activity at the levels above (Bastos et al., 2012). As such, these cells are thought to signal prediction error, which must be passed to higher hierarchical levels within the predictive coding scheme. The prediction tested in this experiment is about the modulation of the precision, or gain, of the prediction error, and this superficial pyramidal cell self-connection is mathematically equivalent to a gain. One alternative hypothesis was that it was the precision of the predictions, signalled by deep pyramidal cells, that mediated contrast-dependent responses; as such, the level of the deep pyramidal cell gain was also fitted to contrast-specific responses and the model evidence of these two models was tested.
To generate predicted signals in sensor space, the superficial pyramidal cell activity and a small contribution from deep pyramidal cell activity (which is thought to represent most of the EEG signal) are multiplied by a lead field matrix which maps sources to sensors to produce simulated data (output equation, figure 2.3). This lead field matrix constitutes the conventional electromagnetic forward model.

The dynamic causal model is inverted using variational Bayesian procedures to obtain the posterior density of the free parameters given the data. As well as the four intrinsic connection strength parameters, the free parameters included the strength of all extrinsic connections. The posterior distributions were obtained using a standard Variational Laplace scheme as described in Friston et al., (2007).

To determine the connectivity of the areas identified by the source localisation, Bayesian model selection was first performed using the free energy, which is an approximation to log model evidence. Six plausible models were specified (Figure 2.4), representing both parallel and serial hierarchies, with and without inter-hemisphere connections. Each of the six models was fitted separately to the average response over all subjects for each contrast level. A fixed-effects model comparison was then performed. This stage of the analysis was conducted solely to find the most plausible model within which to test our

hypotheses about within-area parameter changes; we had no specific anatomical hypotheses.

The winning model was used for all subsequent analyses. Within this model, three sub-models of contrast-dependent effects were evaluated using subject-specific averages: a model with no contrast-dependent effects, a model with contrast-dependent changes in the self-connectivity of superficial pyramidal cells (γ_7) and a model allowing contrast-dependent changes in the self-connectivity of deep pyramidal cells (γ_{10}). A fixed effects Bayesian model comparison was then used to compare the final three models (contrast dependent effects upon the superficial, deep or no cells) by pooling their log evidences over subjects. This analysis tested our specific hypothesis: can the change in brain responses seen when visual contrast is altered be modelled only by changes in the gain of superficial pyramidal cells within each area?

2.6 Statistical Analysis

Statistical analysis of the parameter estimates from the winning model was performed in SPSS 20.0. The winning model had contrast-dependent changes in the γ_7 parameter (self inhibition of superficial pyramidal cells or negative gain). The maximum *a posteriori* estimates of the changes in these parameters were quantified using a classical summary statistic approach. Eight parameters changed in a contrast-specific way in each subject-specific model – one for each of four areas in both hemispheres. These parameters were entered into a two-way ANOVA with factors cortical source (with four levels) and hemisphere (with two levels). In addition, a one-way ANOVA with planned contrast testing for a (linear) change in gain with hierarchical level was performed, weighting the groups (from the bottom of the hierarchy to the top) as 4,3,2,1. Equal variance was assumed.

3. Results

Source localisation revealed four bilateral sources of activity (Figure 2.2): visual area 4 (V4), the inferior parietal cortex (IPC), Brodmann area 19 (A19) and the orbitofrontal cortex (OFC). These cortical areas have been implicated previously in the processing of visual form and the global (spatial) attributes of visual stimuli (Peterson et al., 1999; Podzebenko et al., 2005; Shikata et al., 2003). The locations of these sources were used in subsequent dynamic causal modelling of observed responses in sensor space. Note that

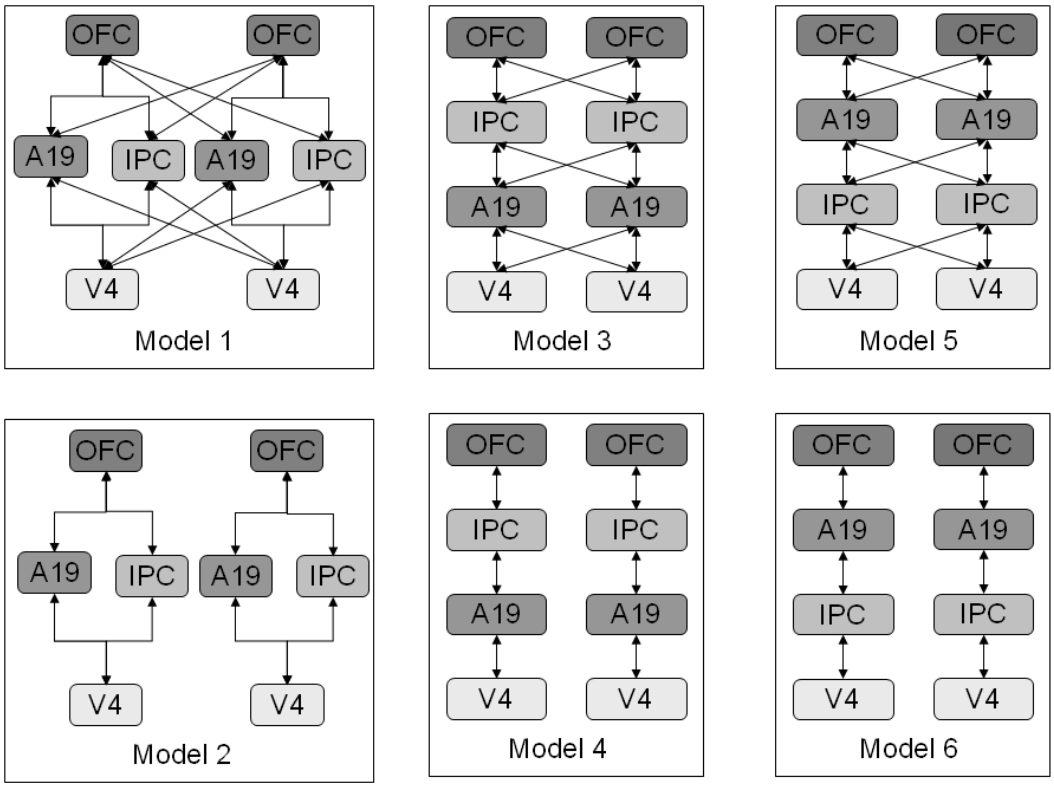

Figure 2.4: Model selection. Initial model selection was carried out to identify the extrinsic connectivity pattern on the sources identified. Only plausible models were tested; these models had the inferior orbital gyrus at the bottom of the hierarchy and the superior orbital gyrus at the top (Felleman & Van Essen, 1991). These models are distinguished by the deployment of forward and backward extrinsic connections, as determined by their level in the hierarchy. Here, this level corresponds to vertical position.

the anatomical designations are just mnemonic. Although the source reconstruction used a canonical template – and the sources can be associated with a Talairach and Tournoux location – the spatial precision of EEG source reconstruction means that anatomical localisation is very approximate.

Six dynamic causal models (Kiebel et al., 2008) employing a canonical microcircuit model of neural activity (Bastos et al., 2011) were fitted to the event related potentials averaged over all subjects for each level of contrast (Figure 2.4). Fixed effects Bayesian model comparison was used to compare the evidence for each model, pooled over subjects. The model with the greatest evidence was a simple hierarchy with diagonal interhemispheric connections (Figure 2.4, Figure 2.5). This model was then used to assess contrast-dependent changes in coupling for each subject.

This model was fitted to individual subject data with three possible models of contrast-dependent effects; one which allowed no changes in connectivity, a model with contrast-dependent changes in the self-connectivity of superficial pyramidal cells and a model allowing contrast-dependent changes in the self-connectivity of deep pyramidal cells. Fixed-effects Bayesian model comparison showed the model with contrast-dependent changes in the self-connectivity (gain) of the superficial pyramidal cells had the most evidence, with a log Bayes factor of 900 – compared to equivalent model with changes in the deep pyramidal cells. Both of these models had an overwhelming amount of evidence in relation to the null model, with no contrast dependent changes in gain (with Bayes factors of over 30,000). The model with contrast dependent changes in superficial pyramidal cell gain provided an excellent fit to the data (Figure 2.6).

Contrast-dependent changes in coupling under the winning model were assessed in a *post hoc* fashion, using classical inference. Two-way ANOVA (with factors cortical source and hemisphere) showed no effect of side ($F_{1,136} = 1.850$; $p = 0.073$), so parameters pertaining to left and right sources were analysed together subsequently. One-way ANOVA with planned testing for a (linear) change in gain with hierarchical level showed a significant trend for contrast-dependent increases in lower sources and smaller, or no, contrast-dependent increases in higher sources of the hierarchy ($t_{140} = -2.472$; $p = 0.015$) (Figure 2.7). The contrast-dependent changes in gain shown in Figure 2.6 produce a progressive attenuation of contrast-dependent effects at higher levels in the hierarchy. This can be seen easily in Figure 2.8, where solid lines represent the highest-contrast condition and dotted lines the lowest-contrast condition. The difference in responses to the different levels of contrast clearly decreases as the hierarchy is ascended.

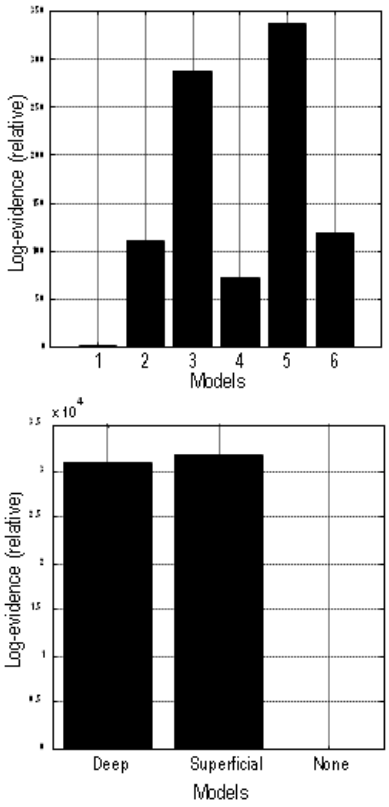

Figure 2.5: Results of fixed-effects Bayesian Model Selection. Upper panel: out of the different extrinsic connectivity models, Model 5, a serial hierarchy with interhemispheric connections, had the most evidence. This model was used for subsequent analyses.

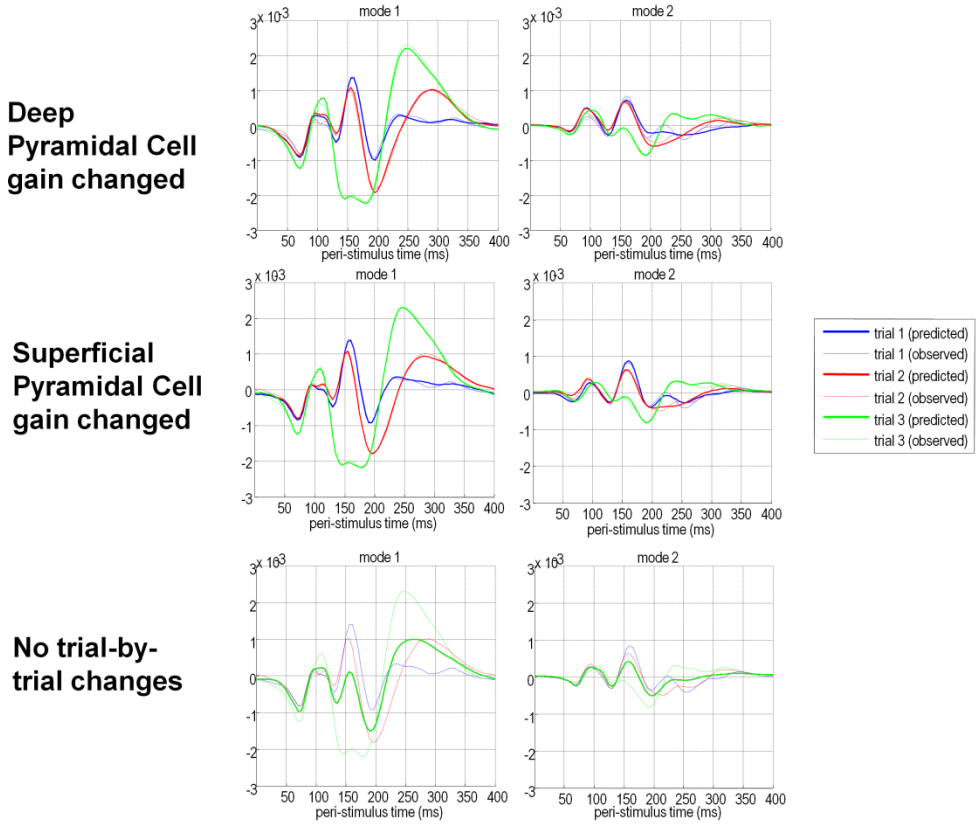

Figure 2.6: Model fits. The fits of the three models of contrast-dependent effects to event related potentials in sensor-space for an illustrative subject; these responses are summarised with the first two principal components or modes. The modes are used for data reduction – the data is projected onto the principal eigenvectors of the prior covariance of the data. In this chapter, eight modes are used in total. The dashed lines show the data modes and the solid lines the model predictions. In the best-fitting model (centre) these are almost superimposed, whereas in the less well-fitting models, substantial differences are evident.

4. Discussion

The results of this study suggest that the visual contrast of a stimulus increases the gain of superficial pyramidal cells in lower visual areas, relative to higher levels. This is entirely consistent with generalised predictive coding, where visual contrast determines the precision of sensory signals and the representation of that precision in terms of the gain or sensitivity of superficial pyramidal cells.

Generalised predictive coding suggests that forward connections in the brain (known to originate from superficial pyramidal cells (Felleman & van Essen, 1991; Maunsell & van Essen, 1983) convey precision-weighted prediction error. Theoretical work described above (Brown & Friston, in submission) has shown that increasing visual contrast corresponds to increasing the precision of sensory channels in accordance with Weber's law. In this study, I have shown that the changes in the EEG signal across levels of visual contrast can be modelled by changes in gain in superficial pyramidal cells. This gain is thought to represent the precision of the prediction error, which determines the signal-to-noise ratio associated with sensory input.

A technical issue – that deserves a brief comment – is that the cell populations, whose intrinsic gain best models visual contrast effects, are the same populations generating the observed EEG signal (in the model). One might ask whether this biases the model comparison, given that visually evoked responses generally increase with contrast (Polat and Norcia 1996). Although this is a possibility – in the sense that any inference in DCM pertains only to the models considered – the intrinsic connections between superficial and deep cells means that changes in the gain of deep cells could also easily explain the contrast dependent responses – through their influence on superficial cells. Furthermore, models with contrast dependent changes in extrinsic connections (targeting both superficial and deep populations) had substantially lower evidence than the models reported above (results not shown). In short, an increase in the gain of superficial pyramidal cells appears to be the best explanation for contrast dependent effects, within the set of alternative models that could be plausibly conceived of.

It should be noted that a contrast dependent increase in evoked responses could be modelled in many ways. In this sense, the use of DCM can be regarded as testing specific hypotheses about a limited number of competing explanations. I focused on the gain or intrinsic sensitivity of superficial and deep pyramidal cells because explanations in terms

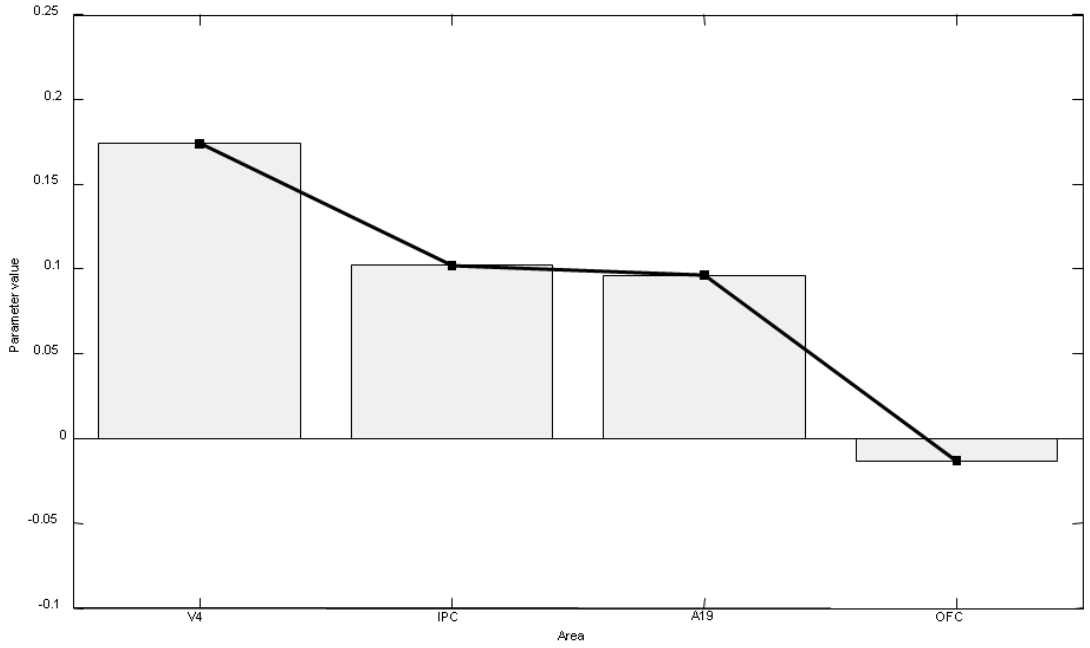

Figure 2.7: Contrast-dependent changes in the gain of superficial pyramidal cells. These are the average parameters, over subjects, controlling the contrast dependent changes in negative self-inhibition (gain) under the winning model of the previous figures. Note the progressive decrease in contrast-dependent effects at higher levels of the hierarchy. This is predicted theoretically, because the precision of prediction errors at the lowest (sensory) level has been manipulated through experimental manipulations of visual contrast.

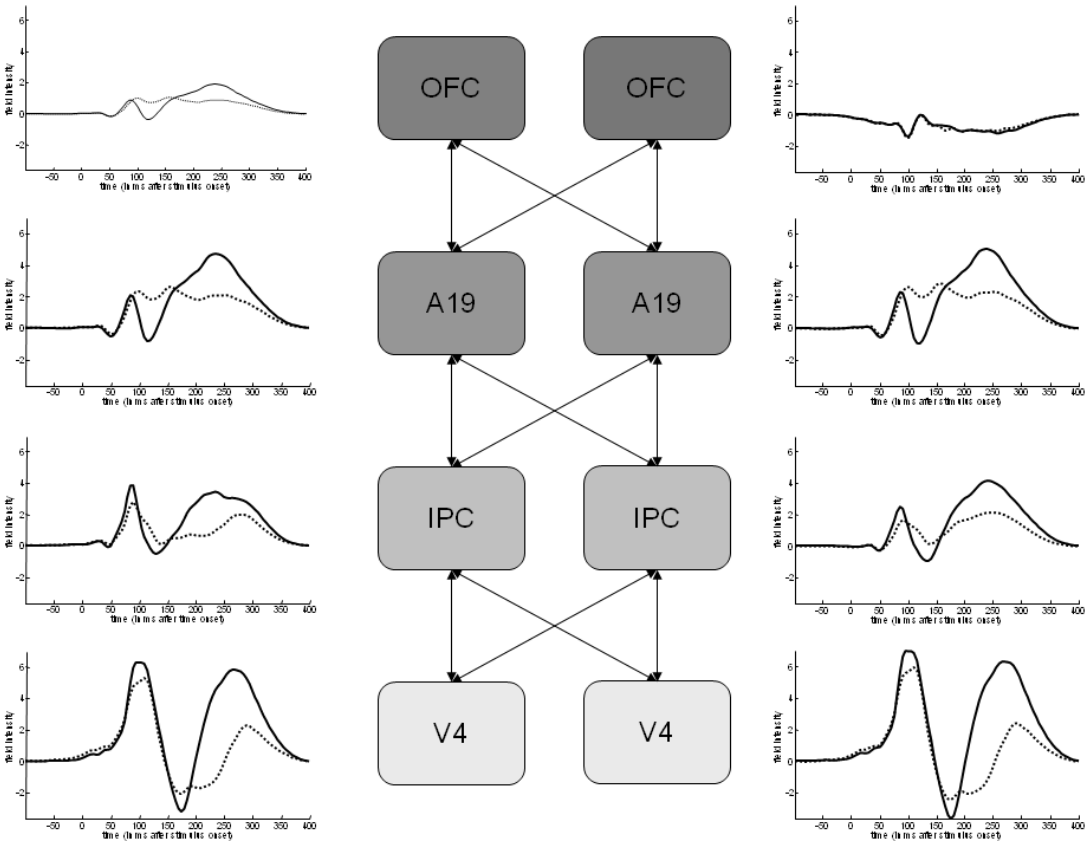

Figure 2.8: Prediction error in the cortical hierarchy. This figure shows the activity reconstructed at each of the sources used for DCM analysis (based on a DCM of the grand average event related potentials over subjects). These responses can be taken to be a rough proxy for prediction error, since superficial pyramidal cells contribute most of the EEG signal. The difference in signal between high-contrast and low-contrast clearly reduces as the hierarchy is ascended, reflecting the decreasing differences in the precision of prediction error.

of post-synaptic gain follow directly from predictive coding formulations of perceptual synthesis. This does not mean that other hypotheses could not be explored based upon alternative theoretical formulations. In brief, as with all dynamic causal modelling studies, the conclusions have to be qualified in relation to the hypotheses all models addressed.

On a general note, the conclusions of this chapter highlight the utility of dynamic causal modelling – in using experimental data to ask specific questions. Figure 2.6 shows that the effects of changing the gain of superficial and deep pyramidal cells are, qualitatively, very similar. This means that adjudicating between implicit explanations for contrast dependent effects is a very difficult problem. Note that this problem cannot be finessed experimentally – for example, there is no (non-invasive) experimental manipulation of contrast that selectively engages deep or superficial pyramidal cells. The solution offered by DCM is to place constraints on the way that data are explained and use Bayesian modelling to quantify the evidence for different hypotheses. Note that although the expression of the different hypotheses in figure 2.5 looks very similar, the evidence for contrast dependent changes in the gain of superficial cells is enormous (with a Bayes factor of over 900). This evidence could not be intuited by simply looking at the data: it is disclosed by careful and informed Bayesian modelling of those data. In short, dynamic causal modelling of this sort exploits prior knowledge to solve otherwise very difficult inference problems. However, this solution rests upon the specification of specific and well posed questions. In other words, the efficiency with which this sort of modelling adjudicates between different hypotheses depends on an efficient and careful experimental design.

The cells of early visual areas corresponding to human V1 have been studied extensively with electrophysiologically in cat and macaque. Although this study did not model V1 as a distinct source, the lack of spatial resolution with EEG means that the results from IOG can be regarded as representative of early visual responses. The supragranular superficial pyramidal cells in this dynamic causal model are located in the same cortical layers as complex cells in cat area 17, which predominate in layers 2 and 3 (Gilbert & Wiesel, 1979). Moreover, in Rao and Ballard's predictive coding model of visual cortex, prediction error units display complex cell-like behaviour in the presence and absence of feedback (Rao & Ballard, 1999). Subsequent examination of the contrast-dependent responses of such cells shows that, in the absence of adaptation, their firing rate generally increases linearly in response to increasing visual contrast, except at very high contrast levels (Albrecht et al., 1984; Ohzawa et al., 1982). The short stimulus duration and dim screen used in this study suggests adaptation of retinal or early cortical responses can be discounted and therefore

contrast-dependent responses at the cellular level should increase monotonically with contrast, which seems to be the case in figure 2.7. Dynamic causal modelling of the underlying synaptic mechanisms suggests that this increase is the result of increasing gain in superficial pyramidal cells.

What might be the mechanism behind these gain increases? In perceptual processing, acetylcholine signalling seems to be an important mechanism for contrast gain-control. Increasing endogenous acetylcholine reduces contrast-dependent gain (DeBruyn, Gajewski, & Bonds, 1986), while nicotine has a suppressive effect on gain in cortical layers 2,3 and 5 but an facilitatory effect in layer 4c, where stellate cell bodies are located (Disney, Aoki, & Hawken, 2007). Short-term depression, particularly at the thalamocortical synapse to spiny stellate cells, has also been proposed to play a role (Carandini, Heeger, & Senn, 2002; Chung, Li, & Nelson, 2002); this would fit with the neuronal encoding of precision by neuromodulatory mechanisms.

The mechanisms discussed in relation to encoding precision also seem to have an important role in contrast sensitivity. For example, nicotine increases contrast sensitivity (Disney et al., 2007), especially at low spatial frequencies (Smith & Baker-Short, 1993), while scopolamine, a muscarinic antagonist, universally increases contrast sensitivity (Smith & Baker-Short, 1993), an effect that can be attenuated by increasing the luminance (precision) of the stimulus (Evans, 1975).

Aberrant encoding of precision and uncertainty has been proposed to play a role in a number of neuropsychiatric disorders. In patients with schizophrenia, the mismatch negativity, an evoked potential that is greater in response to deviant or unexpected auditory tones, is decreased in magnitude (Jahshan et al., 2011; Jordanov et al., 2011; Leitman et al., 2010). This may represent a failure to detect statistical regularities and assign higher precision to sensory information, leading to a reduced difference between responses to standard and deviant stimuli (Garrido et al., 2008; Garrido et al., 2009; Strelnikov, 2007). In other words, schizophrenic subjects may never be surprised because they fail to make precise predictions. This explanation for the mismatch negativity speaks to an optimisation of precision or gain associated with sensory prediction errors due to rapid sensory learning and calls upon exactly the same synaptic mechanisms that have been proposed to at mediate attentional gain (Feldman and Friston 2010). In short, sensory surprise depends upon appropriately precise prediction errors and adaptive precision or gain control.

(Corlett, Frith, & Fletcher, 2009) have proposed a hierarchical Bayesian explanation for schizophrenia that rests on the aberrant weighting of top-down and bottom-up information that could lead to both hallucinations and delusions. In predictive coding, this weighting is determined by the precision of prediction errors at different levels in hierarchical generative models. Patients with schizophrenia show a pan-frequency increase in contrast sensitivity threshold (Skottun & Skoyles, 2007; Slaghuis, 1998), which could reflect inadequate increase of synaptic gain at superficial pyramidal cells in response to high-contrast stimuli. These ideas are important, because this study suggests it is possible to measure the neuronal encoding of precision noninvasively using EEG, in a very simple paradigm which would be easy to perform with patients.

In conclusion, I provide evidence that the contrast-dependency of early visual cortical responses is mediated by the gain of superficial pyramidal cells. In computational terms, this gain may encode the precision of prediction errors signalled by these cells. These results suggest that DCM may be useful as an assay of the synaptic (neuromodulatory) mechanisms that underlie perceptual inference.

The work described in this chapter is published as:

Brown H, Friston KJ. (2012). Dynamic causal modelling of precision and synaptic gain in visual perception - an EEG study. Neuroimage. 63:223-31.

Chapter 3

Attention, uncertainty and free-energy

1. Introduction

Attention is a ubiquitous and important construct in cognitive neuroscience. Many accounts of attention fall back on Jamesian formulations, famously articulated as "the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought" (James 1890). More recent and formal accounts appeal to information theory and computational principles (Duncan and Humphreys, 1989; Deco & Rolls 2005; Jaramillo & Pearlmutter 2007; Spratling 2008; Reynolds & Heeger 2009; Bruce & Tsotsos 2009; Spratling 2010), with an increasing emphasis on Bayesian formulations (Rao 2005; Itti and Baldi, 2010; Chikkerur et al., 2010). These attempts to understand attention are pursued in computational terms. This means terms like uncertainty, surprise and precision will be used in a rather formal way. Without exception, these terms refer to properties of probability distributions. Probability distributions are central to modern treatments of perception that cast perception as inference. Inference requires us to represent probability distributions (or densities) over possible causes or explanations for our sensations. These distributions have several important attributes: For example, a broad distribution encodes a high degree of uncertainty about a particular cause. This uncertainty is, mathematically, the average (expected) surprise over all possibilities. A key measure of uncertainty is the width or variance of the distribution, or its inverse, precision. In what follows, I hope to show that attention is more concerned with optimizing the uncertainty or precision of probabilistic

representations, rather than what is being represented. By describing perception in formal terms, one can see almost intuitively where attention fits into the larger picture and how it might be mediated neurobiologically. This is important because a formal framework allows one to link classical psychological constructs to current physiological perspectives on attention (e.g., communication through coherence; Fries 2005). I hope to show that the perspectives afforded by cognitive psychology, neurophysiology and formal (theoretical) treatments are all remarkably consistent.

In this chapter I suggest that perception is the inference about causes of sensory inputs and attention is the inference about the uncertainty (precision) of those causes (Friston 2009). This places attention in the larger context of perceptual inference under uncertainty (Rao 2005; Spratling 2008; Whiteley & Sahani 2008; Chikkerur et al., 2010). In this work,I try to show that attention emerges naturally in a Bayes-optimal scheme used previously to address predictive coding (Friston & Kiebel 2009), perceptual categorisation (Kiebel et al., 2009), learning (Friston 2008) and action (Friston 2010a). In other words, I try to explain some simple attentional phenomena using an established framework that has explanatory power in domains beyond attention. Specifically, I show how attention can be construed as inferring the precision of sensory signals and their causes. The idea is illustrated using computational simulations of neuronal processing that try to establish face-validity in terms of psychophysical and electrophysiological responses. This is done in the context of the Posner paradigm (Posner 1980); a classical paradigm for studying directed spatial attention in vision, using cued targets. This paradigm also makes it possible to address, in a heuristic way, biased competition (Desimone 1996) by presenting validly and invalidly cued targets simultaneously. The hope was to connect psychophysical studies of attention with theories based upon detailed electrophysiological studies in monkeys.

The basic idea pursued in this chapter is that attention entails estimating uncertainty during hierarchical inference about the causes of sensory input. This idea is developed in the context of perception based on Bayesian principles, under the free-energy principle (Friston 2009), as described in the introduction. Formally, this scheme can be regarded as a generalisation of predictive coding (Rao and Ballard 1998) and involves recurrent message passing among hierarchical levels of cortical systems to optimise a probabilistic representation of the world (Mumford 1992; Friston 2009). In these generalised schemes, precision is encoded by the synaptic gain (post-synaptic responsiveness) of units reporting prediction errors (Friston 2008). There are many metaphors for attention that

relate closely to the idea described in this chapter. Perhaps the simplest is that of statistical inference, which treats perception as hypothesis testing (Gregory 1980): Indeed, most modern theories of perception draw on Helmholtz's ideas about the brain as an inference machine (e.g. Gregory 1968; Ballard et al 1983; Dayan et al 1995). These theories regard the brain as inferring how sensory data are generated using generative models (cf, hypotheses) in exactly the same way that we analyse scientific data. The simplest example of this is the Student's t-statistic, where a difference in group means is divided by its standard error to test for group differences. Under the null hypothesis, the observed difference is the prediction error and the standard error is an estimate of its precision (inverse variance). This means that one can regard the t-statistic as a precision-weighted prediction error. Crucially, both the prediction error and its precision have to be estimated, given empirical (sensory) data. The idea here is that attention rests on estimating precision and is therefore an integral part of perception. Things get more interesting if we consider that the precision of sensory signals depend on states of the world. This means that optimising precision entails optimising inferred states of the world that affect the precision or uncertainty about our sensations. It is this generalisation of generative models which is exploited in this chapter. In brief, most generative models (including those used to simulate perception) ignore state-dependent noise or error variance; assuming that it is constant for any (sensory) data channel. In what follows, this assumption is relaxed to consider generative models in which the states of the world (for example the presence of attentional cues) can change the precision of sensory data. A simple example of this would be the direction (state) in which we pointed a searchlight. This determines which part of the sensorium contains precise information; namely visual information reflected by surfaces that are illuminated. Any coupling between the state of the world (content) and the precision of sensory samples (context) should be an inherent part of veridical generative models of sensory input. Under this perspective, searchlight (spotlight) metaphors for attention become a natural way to think about its functional role (Shulman et al., 1979; Crick 1984; Cave and Bichot, 1999; Eckstein et al 2002). Mechanistically, this role is to weight or bias selected sensory channels (Desimone and Duncan, 1995; Maunsell, & Treue 2006; Reynolds & Heeger 2009; Stokes et al., 2009). In statistical terms, this is formally identical to weighted least squares that underlies all optimal (maximum *a posteriori*) estimates of model parameters. Put simply, this involves weighting data in proportion to their estimated precision.

In predictive coding schemes, sensory data are replaced by prediction error, because this is the only sensory information that has yet to be explained. Here, the weighting is implemented by synaptic gain. We therefore return to the central role of precision-

weighted prediction errors in optimal inference. Neurobiologically, this is easy to relate to theories of attentional gain, where the postsynaptic responsiveness of sensory (prediction error) units is modulated by attentional mechanisms (Desimone 1996; Maunsell, & Treue 2006). This chapter will focus on two neurobiological candidates for modulating synaptic gain that have specifically linked to attention: synchronous gain (Chawla et al 1999a) mediated by fast oscillatory or synchronised activity (Womelsdorf & Fries 2006; Fries et al 2008) and classical neuromodulatory (e.g., cholinergic) neurotransmission (Schroeder et al 2001; Hirayama et al 2004):

Electrophysiologically, desynchronisation with increased gamma activity (between 30 and 100 Hz) is seen during attentional tasks in invasive (Steinmetz et al., 2000; Fries et al., 2008; Bichot et al., 2005), and non-invasive EEG and MEG studies (Gruber et al., 1999; Sokolov et al., 1999; Pavlova, Birbaumer and Sokolov, 2006; Vidal et al., 2006). Gamma oscillations induced with subliminal flicker may improve attention-based performance (Bauer et al., 2009). Furthermore, increased gamma is associated with faster reaction times (Womelsdorf *et al.*, 2006; Fründ et al., 2007). Gamma oscillations can control gain by affording synchronised neuronal discharges a greater influence on the firing rate of downstream neurons (Chawla et al 1999a; Salinas and Sejnowski, 2001; Zeitler, Fries and Gielen, 2008). Gamma activity has also been proposed as a solution to the 'binding problem', which I discuss below in relation to attention (Treisman and Schmidt, 1982).

In terms of neurotransmitters, gamma oscillations are profoundly affected by acetylcholine, which is released into sensory cortices from nuclei in the basal forebrain. It acts through both fast ion channel (nicotinic) receptors and slow metabotropic (muscarinic) receptors (Wonnacott, 1997; Zilles et al., 2004; Hasselmo et al., 2006). Disruption of the cholinergic system by drugs or lesions can interfere with attentional processes, including the Posner paradigm (Voytko et al., 1994; Witte et al., 1997; Dalley at al., 2001; Herrero et al 2008; Vossel et al., 2008). Acetylcholine appears to increase synaptic gain directly by, for example, reducing spike-frequency adaptation (McCormick and Prince, 1985; *ibid* 1986). It may also facilitate the induction of gamma oscillations by reducing adaptation in pyramidal cells (Buhl *et al.* 1998; Borgers *et al.* 2005), decreasing activity of inhibitory interneurons (Buia and Tiesinga 2006) or directly inactivating specific interneurons (Xiang *et al.* 1998). However, the time-course of acetylcholine release can be quite protracted (Parikh et al., 2007). This suggests rapid (ten to a hundred milliseconds) attentional mechanisms may rest on an interaction of cholinergic mechanisms with fast activity-dependent modulation of synaptic gain. It is this activity (state) dependent optimisation I pursue in this chapter.

In summary, it may be the case that attention is the process of optimising synaptic gain to represent the precision of sensory information (prediction error) during hierarchical inference. Furthermore, if we allow for state-dependent changes in precision, the neurobiology of attention must involve activity-dependent changes in synaptic gain; assuming that neuronal activity represents the states of the world and synaptic gain represents precision. Given this sort of architecture we can, in principle, simulate attentional processing with established (Bayes-optimal) inversion or recognition schemes, using models with state-dependent noise. What follows is an attempt to do this.

This chapter comprises five sections. The first comprises brief review of attention in psychological and neurobiological terms. This section focuses on directed spatial attention and, in particular, the Posner (cueing) paradigm that emphasises the importance of valid cues in establishing attentional set during target detection (Posner 1980). To complement this psychophysical perspective, biased competition models that are based on careful electrophysiological studies of evoked visual responses using intracranial recordings are considered (Desimone 1996). Biased competition is probably the most established and influential theory that accounts for unit responses in attentional paradigms framed at the level of receptive fields. I also review the concepts of attentional resources and other constructs associated with early and late attentional selection and the feature-integration theory of attention. In the second section, the free-energy principle and form of the generative models described in the introduction are briefly recapitulated. The emphasis here is on generalising previous models to include state-dependent noise and what this means for their neurobiological optimisation or inversion. The resulting inversion scheme corresponds to recognizing the causes of sensory data (that include both states of the world and their precision). Precision is encoded by the synaptic gain of sensory or prediction error units, which pass messages to units representing conditional expectations about the world. In this scheme, optimisation of synaptic gain may correspond to attention. In the third section, simulations of the Posner paradigm are presented using the recognition scheme of the previous section. This allows demonstration some basic characteristics of attention-based inference; including attentional bias, attentional capture and the cue-validity effect. A direct interpretation of the probabilistic representations encoded by simulated neuronal activity is supplemented with simulated psychophysical and electrophysiological data. These simulated responses make some clear predictions about speed-accuracy trade-offs and event-related electrophysiological responses, which I compare against the literature. In the final section, the same simulations are used but both valid and invalidly cued targets are presented

together. This is a rough metaphor for paradigms used to study biased competition and allows us to see if biased competition emerges from the free-energy formulation. This is examined by looking at competition between cues via the effect of an attended cue on the responses evoked by an unattended cue. This chapter concludes with a brief discussion and indications of how the scheme in this chapter could be applied to empirical psychophysical and electroencephalographic observations.

2. Attention, biased competition and the Posner paradigm

In this section, I review some of the key paradigms and theories that have dominated attention research over the past decades. The focus will be the Posner paradigm, which is simulated in later sections, and biased competition, which is one of the most prevalent electrophysiologically grounded theories of attention.

Early cognitive models of attention, although inherently limited by lack of knowledge about the underlying neural processes, elucidated the important difference between early and late selection. Broadbent (1958), working in the auditory domain, suggested that attention operated by selecting stimuli at an early stage of processing, when only basic physical attributes had been encoded. The selected stimulus was then processed by an 'identification system', which could handle only one stimulus at a time; to explain why semantic information about unattended stimuli is unavailable to recall (Broadbent, 1952a,b). However, there are stimuli which violate this principle: Moray (1959) demonstrated that a subject's name, which is salient only after semantic processing, could shift attention to a previously unattended auditory stream. The competing theory, that all stimuli are processed semantically before selection for consciousness recall, was posited by Deutsch and Deutsch (1963), whereas Treisman (1964) suggested that unattended stimuli are attenuated so that attention can be diverted to them, if they become behaviourally salient. Lavie (1995) attempted to reconcile these models by demonstrating that perceptual load plays an important role in attentional selection: Intuitively, early selection occurs with higher attentional load and late selection with lower load. This differential selection rests on the notion of limited capacity. Many of these ideas can be understood in the framework of biased competition theory, which tries to explain some of the phenomena described above using neurobiological mechanisms:

2.1 Biased competition

Biased competition (Desimone and Duncan, 1995) is a model of attention based on electrophysiological studies and earlier behavioural models. Its main contribution was to make the notion of limited capacity or resources more concrete, by suggesting small lower level receptive fields (RFs) compete to drive larger RFs at higher hierarchical levels. Biased competition says that attention is an emergent property of competition between stimuli for attentional resources, which is influenced by the properties of the stimuli and task requirements. Its premise is that, in a crowded visual field, objects must compete for neural representation at some point along the visual processing stream. This can be deduced from the large size of classical receptive fields in higher visual areas, such as monkey area TE, which can cover up to 25^{o} of visual angle (Gross et al., 1972; Desimone and Gross, 1979). Clearly, many objects can fall into such a visual field but the neuron can only represent (report) one thing with its firing. If an object is represented by these higher-order visual neurons, they are unavailable to represent other objects. Thus the object has consumed some finite 'attentional resource'.

This premise leads to a key prediction: if two stimuli are presented within a cell's receptive field, the response to both will be smaller than the sum of the response to the stimuli presented separately (Reynolds et al. 1999). Single-cell electrophysiological studies have confirmed that stimuli interact in this mutually suppressive manner in areas V2 and V4 (Reynolds et al. 1999), IT (Rolls and Tovee, 1995) and MT (Recanzone et al., 1997), but not V1, where receptive fields are so small it is difficult to present competing stimuli (Moran and Desimone, 1985). The average responses of visual cortical areas in fMRI studies in humans mirror the results from electrophysiological studies in animals (Kastner et al. 1998; Beck and Kastner, 2005). An important result is that the maximum spatial separation between stimuli, which induces suppressive interactions, increases at higher levels of visual processing, which is consistent with increasing receptive field size (Kastner et al., 2001).

Large receptive fields thus cause stimuli to compete. The probability with which stimuli are represented by cells is thought to be influenced by a number of top-down and bottom-up biases. Bottom-up biases result from the properties of the stimuli itself, such as visual or emotional salience and novelty. Abrupt-onset stimuli, which have high temporal contrast, and thus salience, can attract attention even if they are task-irrelevant (Yantis and Jonides, 1984). In the visual search paradigm, used to address feature-integration and binding (Treisman & Gelade 1980; Treisman & Schmidt 1982; Treisman 1998), subjects

are required to find a unique object in a display cluttered with distracters. If the unique object is particularly salient, for example if it is brighter than the distracters or has a unique colour, the search time remains constant regardless of the number of distracters. This phenomenon is called 'pop-out'. Salience does not have to be a function of simple visual attributes: distractor faces exhibiting negative emotions slow search times more than neutral faces (Pessoa, Kastner and Ungerleider 2002). Novelty preference, the well-documented tendency for neurons to respond more strongly to a new stimulus than to a familiar one, can also be considered as a bottom-up bias (Desimone, 1996).

Top-down biases reflect the cognitive requirements of the task rather than the stimuli. Top-down biases have been most studied via spatially-directed attention experiments. Electrophysiologically, if attention is directed towards one of two competing stimuli in a receptive field, the mutually suppressive effect disappears and the response of the cell emulates the response to attended stimulus alone (Moran and Desimone, 1985; Chelazzi et al. 1993; Treue and Maunsell, 1996; Desimone, 1998). Even in the absence of visual stimulation, baseline increases in firing rate of 30-40% may be seen, if attention is directed to a location within a cell's receptive field (Luck et al., 1997). Indeed, in fMRI studies, responses are increased in visual areas after attentional cuing but before stimulus onset (Chawla et al 1999b; Kastner et al. 1999; O'Connor et al. 2002; Stokes et al., 2009). In addition, cells respond more strongly to attended than unattended stimuli (Luck et al., 1997). Thus, top-down bias has both additive (baseline shift) and multiplicative (attentional gain) components that may depend on each other (Chawla et al 1999b). In summary, biased competition is a mechanistic framework, which provides a plausible neurobiological account of attention. Later, we will see how biased competition emerges naturally in predictive coding formations of Bayes-optimal perception.

2.2 The Posner paradigm

In later sections optimal perception under the Posner task, a covert attention task, will be simulated. Attending to an object usually involves looking at it; that is placing its image at the fovea (the central area of the retina with highest acuity). However, attention can be directed independently of eye movement (Posner et al., 1978). Under the Posner paradigm, subjects are required to foveate a central spot and respond as quickly as possible to the appearance of a peripheral target. The target is cued with either a central arrow indicating the side it will appear on, or a peripheral box around the target's eventual location. The cue is correct (valid) most (usually 80%) of the time. Posner found that reaction times to validly cued targets were significantly shorter than to invalidly cued

targets, which appeared on the opposite side. This demonstrated that attention could be moved to salient locations in the absence of gaze shift. The cuing seen in the Posner paradigm seems to be separable from the phenomenon of 'alerting', in which a non-directional signal indicates the imminent onset of a target (Fernandez-Duque and Posner, 1997; Posner, 2008). Subjects are quicker to respond to a target if the cue indicates the location of the target than when it only indicates the timing (Davidson and Marrocco, 2000). In addition, a pharmacological double-dissociation exists such that inhibitors of the cholinergic system selectively reduce the benefits of spatial cues, while noradrenergic inhibitors selectively reduce the benefits of alerting cues (Marrocco et al., 1994). Furthermore, dopamine and noradrenalin antagonists can reduce the reaction time cost of invalidly cued targets, while preserving the validly effect (Clark et al., 1989). However, this effect may be due to the role of noradrenalin in task switching (Sara, 1998; Yu and Dayan, 2005).

The two types of cues used in the Posner paradigm – central and peripheral – show the same facilitation effect. However, they may operate by different mechanisms. Peripheral stimuli are labelled as 'exogenous', because the change in attention is triggered by an external event. It is well established that abrupt-onset peripheral stimuli can attract attention via bottom-up mechanisms (Yantis and Jonides, 1984), even when task-irrelevant (Theeuwes, 1991). Central stimuli are 'endogenous' because they do not in themselves indicate target location; attention must be directed to the correct location according to information conveyed by the cue. The most common central cues are inherently directional: an arrow pointing to where the target will appear, or an asterisk just to one side of fixation. Although cues such as this may automatically 'push' attention, even when the subject has been told the cue is invalid (Hommel at al., 2001).

Exogenous and endogenous cuing fit well with biased competition theory: exogenous cuing can be thought of as a bottom-up bias, based on the prior expectation that salient events recur in the same part of the visual field. On the other hand the effect of endogenous cues must be mediated by top-down bias. However, these top-down effects do not necessarily call on semantic or explicit processing: For example, Decaix *et al.* (2002) examined performance on the Posner paradigm when subjects were not informed about the cue-target relationship but subjects still learnt cue-target relationships within 90 trials, and performance was independent of whether the learnt relationship was accessible to verbal report. Bartolomeo et al. (2007) compared performance of informed and non-informed subjects and found no effect of explicit knowledge on reaction time. Finally, Risko and Stolz (2010) demonstrated that knowledge of the proportion of valid trials did

not affect reaction time. In short, the basic phenomena disclosed in the Posner paradigm may not depend on high level cognitive processing. This suggests that a low level simulation of perceptual processing should be able to account for cue validity effects. This is what I attempt to show and demonstrate that cue-validity effects are Bayes-optimal. In the next section, I review the principles that lie behind Bayes-optimal perception and apply these principles to the Posner paradigm in the subsequent section.

3. Mathematical foundations

The hierarchical message passing scheme below was introduced in the introduction as a highly general scheme which allows inversion of hierarchical dynamic models within a predictive coding scheme:

$$\begin{aligned}
\dot{\tilde{\mu}}^{(i,v)} &= \mathrm{D}\tilde{\mu}^{(i,v)} + \chi_{\tilde{v}}^{(i,v)}\xi^{(i,v)} + \chi_{\tilde{v}}^{(i,x)}\xi^{(i,x)} + \lambda_{\tilde{v}}^{(i,v)} + \lambda_{\tilde{v}}^{(i,x)} - \xi^{(i+1,v)} \\
\dot{\tilde{\mu}}^{(i,x)} &= \mathrm{D}\tilde{\mu}^{(i,x)} + \chi_{\tilde{x}}^{(i,v)}\xi^{(i,v)} + \chi_{\tilde{x}}^{(i,x)}\xi^{(i,x)} + \lambda_{\tilde{x}}^{(i,v)} + \lambda_{\tilde{x}}^{(i,x)} - \mathrm{D}^{T}\xi^{(i,x)}
\end{aligned}$$

$$\begin{aligned}
\xi^{(i,v)} &= \tilde{\Pi}^{(i,v)}\tilde{\varepsilon}^{(i,v)} = \tilde{\Pi}^{(i,v)}(\tilde{\mu}^{(i-1,v)} - \tilde{f}^{(i,v)}) \\
\xi^{(i,x)} &= \tilde{\Pi}^{(i,x)}\tilde{\varepsilon}^{(i,x)} = \tilde{\Pi}^{(i,x)}(\mathrm{D}\tilde{\mu}^{(i,x)} - \tilde{f}^{(i,x)})
\end{aligned} \qquad 3.1$$

$$\begin{aligned}
\chi_{w}^{(i,v)} &= \tilde{f}_{w}^{(i,v)T} - \tfrac{1}{2}\tilde{\varepsilon}^{(i,v)T}\tilde{\Omega}_{w}^{(i,v)} \\
\chi_{w}^{(i,x)} &= \tilde{f}_{w}^{(i,x)T} - \tfrac{1}{2}\tilde{\varepsilon}^{(i,x)T}\tilde{\Omega}_{w}^{(i,x)}
\end{aligned}$$

Here, the amplitude of random fluctuations are assumed to be parameterised in terms of log-precisions, where

$$\begin{aligned}
\tilde{\Pi}^{(i,u)} &= R^{(i,u)} \otimes diag(\exp(\pi^{(i,u)})) \\
\tilde{\Omega}^{(i,u)} &= I^{(i,u)} \otimes diag(\pi^{(i,u)})
\end{aligned} \qquad 3.2$$

The vector function $\pi^{(i,u)} := \pi(x, v, \gamma^{(i,u)}) : u \in v, x$ returns state-dependent log-precisions and $R^{(i,u)}$ is the inverse smoothness matrix $V^{(i,u)}$. In what follows the amplitude (variance) of random fluctuations will be quantified in terms of log-precisions, such that the associated variance is $\exp(-\pi^{(i,u)})$. With this particular form for the precisions, the terms $\tilde{\Omega}_{w}^{(i,u)}$ and $\lambda_{w}^{(i,u)} = tr(\tilde{\Omega}_{w}^{(i,u)})$ are constant for states $w \in \tilde{v}, \tilde{x}$ that affect the log-precisions linearly and zero if they have no effect.

In the present context, the key thing about this scheme is that the precisions $\tilde{\Pi}^{(i,u)} := \tilde{\Pi}(v^{(i)}, x^{(i)}, \gamma^{(i,u)})$ depend on the expected hidden causes and states. It is this dependency that I propose mediates attentional processing. Equation 3.1 tells us that the state-dependent precisions modulate the responses of the error-units to their presynaptic inputs. This modulation depends on the conditional expectations about the states and suggests something intuitive; attention is mediated by activity-dependent modulation of the synaptic gain of principal cells that convey sensory information (prediction error) from one cortical level to the next. These are generally thought to be the superficial pyramidal cells responsible for generating EEG signals. More specifically, precision sets the synaptic gain of error-units to their top-down and lateral inputs. In hierarchical models, the gain modulation of error-unit activity $\xi^{(i,u)}$ depends on $\tilde{\Pi}(v^{(i)}, x^{(i)}, \gamma^{(i,u)})$ and therefore depends on the conditional expectations of $x^{(i)}$ in the current level and $v^{(i)}$ in the level above. This translates into a top-down control of synaptic gain in principal (superficial pyramidal) cells elaborating prediction errors and fits comfortably with the modulatory effects of top-down connections in cortical hierarchies that have been associated with attention. Note that the precisions or synaptic gain $\tilde{\Pi}^{(i,u)}$ also depends on the slowly varying parameters $\gamma \subset \varphi$ responsible for learning. It is these parameters I associate with the slower dynamics of classical neuromodulation (e.g., cholinergic neurotransmission; Friston 2008).

4. Simulating the Posner paradigm

In this section, the hierarchical dynamic models described in the introduction are used as a generative model of stimuli used in the Posner paradigm. Inversion of this model, using generalised predictive coding (Equation 16) will be used to simulate neuronal responses. This allows us to explore some of the inferential and empirical aspects of perception the Posner paradigm was designed to study. I first describe the particular model and stimuli used. I then present simulated responses to valid and invalid targets to highlight their differences, in terms of implicit speed-accuracy trade-offs and their electrophysiological correlates.

4.1 The Posner model

I deliberately tried to keep the generative model as simple as possible so that its basic behaviour can be seen clearly. To this end, the model used has two levels, the first

representing visual input and the second representing the causes of that input. The model has the following form, which is unpacked below:

$$
\begin{aligned}
s &= \begin{bmatrix} s_L \\ s_C \\ s_R \end{bmatrix} = v^{(1)} + z^{(1,v)} \\
\dot{x}^{(1)} &= \begin{bmatrix} \dot{x}_L^{(1)} \\ \dot{x}_R^{(1)} \end{bmatrix} = \underbrace{\tfrac{1}{4}\begin{bmatrix} 1 \\ -1 \end{bmatrix} v_L^{(1)} + \tfrac{1}{4}\begin{bmatrix} -1 \\ 1 \end{bmatrix} v_R^{(1)}}_{exogenous} + \underbrace{\tfrac{1}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix} v_C^{(1)}}_{endogenous} - \tfrac{1}{32} x^{(1)} + z^{(1,x)} \\
v^{(1)} &= \begin{bmatrix} v_L^{(1)} \\ v_C^{(1)} \\ v_R^{(1)} \end{bmatrix} = z^{(2,v)}
\end{aligned}
\qquad 3.3
$$

$$
\begin{aligned}
z^{(1,v)} &\sim \mathrm{N}\ (0, diag(\exp(-\pi^{(1,v)}))) \\
z^{(1,x)} &\sim \mathrm{N}\ (0, e^{-8} I) \\
z^{(2,v)} &\sim \mathrm{N}\ (0, I)
\end{aligned}
\qquad
\pi^{(1,v)} = 2 + \gamma \begin{bmatrix} \overline{x}_L^{(1)} \\ 2 \\ \overline{x}_R^{(1)} \end{bmatrix}
$$

This minimal model has all the ingredients needed to demonstrate some complicated but intuitive phenomena. It helps to bear in mind that this is a generative model of how sensory data are caused that is used by the (synthetic) brain; sensory data were actually generated by simply presenting visual cues in various positions. Because this is a model the prior assumptions about the causes of visual input are that they are just random fluctuations about a mean of zero; i.e., $v^{(1)} = z^{(2,v)}$. Perception (model inversion) uses this model to explain sensory input in terms of conditional expectations about what caused that input.

The model is first described in terms of the way that it explains sensory data; in other words, how it maps from causes to consequences. I then reprise the description in terms of its inversion; namely, mapping from consequences (sensory data) to causes (percepts). As a generative model, Equation 3.3 describes how hidden causes generate sensory input. There are three causes, which are just random fluctuations with a mean of zero and a precision of one. Two causes generate targets in the right and left visual fields $v_{L,R}^{(1)}$ respectively and a third $v_C^{(1)}$ that generates a cue. This cue establishes the probabilistic context in which the first two causes are expressed. This context is determined by hidden states $x_{L,R}^{(1)}$, which modulate the log-precision (inverse amplitude) of random fluctuations that are added to the hidden causes to create sensory data. Here, $\overline{x}_{L,R}^{(1)}$ are mean centred

versions and $\gamma \subset \varphi$ is a constant that controls the potency of hidden states. Unless stated otherwise, $\gamma = 2$. Crucially, the hidden causes induce sensory signals directly but also drive increases or decreases in the hidden states (second equality in Equation 20). The two hidden states represent a high precision on the left and a low precision on the right and *vice versa*. In other words, they induce a redistribution of precision to the left and right in a complementary way. The first cause $v_L^{(1)}$ generates a stimulus s_L in the left hemi-field and drives its corresponding hidden state $x_L^{(1)}$ to increase precision on the left; similarly for the right cause. This means that hidden causes not only cause sensory signals but also augment their precision. In other words, they cause precise visual information with spatial specificity.

Note how the log-precision $\pi^{(1,v)}(x^{(1)}, \gamma)$ of sensory noise $z^{(1,v)} \sim \mathrm{N}\,(0, diag(\exp(-\pi^{(1,v)})))$ depends on the hidden states. The motivation for this dependency is simple: high levels of signal are generally associated with lower levels of noise (i.e., high signal to noise). More formally, this represents a prior expectation that sensory input conforms to Weber's law (Formankiewicz & Mollon, 2009): For stimulus intensities with a fixed precision (of sensory noise), under Weber's law (after log-transform) the log-precision scales with the magnitude of the signal. See Appendix 5.

The ensuing increase in local precision can be regarded as analogous to exogenous cuing in the Posner paradigm, in the sense that it co-localises in space and time with its sensory expression. Endogenous effects on precision that do not co-localise correspond to the probabilistic context established by $v_C^{(1)}$ that enables endogenous cuing. This hidden cause drives hidden states to increase precision on the right. One can think of s_C as the corresponding endogenous cue in the centre of the field of view. Note that the hidden states decay slowly. This represents a formal prior that once a cause has been expressed in any part of the visual field, subsequent causes will be expressed in the same vicinity with a high sensory precision. The time constants for the accumulation of hidden causes by hidden states (four and two) and their decay (thirty two) are somewhat arbitrary, because any units of time can be assigned to the dynamics. The important thing is that the decay is slower than the accumulation (by factors of eight and sixteen here).

This model uses two hidden states that are placed in (redundant) opposition to each other. The reason for this is so this model can accommodate situations where hidden states encode a high precision in their circumscribed part of the visual field: This involves

generating data in multiple sensory channels, with a hidden state for each channel or location. These simulations are demonstrated in the section on biased competition below. Note that the 'competition' between these units is a formal prior of the model and can be thought to represent the selection of the single most behaviourally salient stimulus in order to act upon it. One can easily add further hierarchical levels to make the sensory dynamics more realistic (i.e., the causes at the sensory level could excite hidden states in a lower level to produce spatiotemporally structured or moving stimuli; cf, Nobre et al., 2007). However, the basic behaviour I want to illustrate here does not change. Finally, note that there is no hand-crafted gain modulation of sensory signals in the generative model. Attentional boosting of sensory signals is an emergent property of model inversion, which is now considered:

From the perspective of model inversion (mapping from sensory signals to causes) the predictive coding scheme of the previous section implies the following sort of behaviour. When a cue s_C is presented, it induces high-precision prediction errors, which excite the representation of the hidden cause $v_C^{(1)}$ at the higher level. This then drives up the hidden states biasing precision to the valid (right) hemi-field, which remain activated after the cue disappears. If a subsequent (valid) target is presented, it will induce high-precision prediction errors and a consequent representation of its associated cause at the second level $v_R^{(1)}$, with a reasonably high degree of conditional confidence. Conversely, if an invalid target is presented, it faces two challenges. First, the prediction errors it elicits have low precision and will therefore exert less drive on its associated cause $v_L^{(1)}$. Furthermore, this cause has to activate its associated hidden or contextual state $x_L^{(1)}$ from much lower (negative) levels. This means that the invalid target may never actually be perceived or, if it is inferred, then it will take considerably longer before the prediction error increases its own potency (by changing the hidden causes and states). In short, invalid targets will be perceived later and with a lower degree of conditional certainty (cf, Vibell et al., 2007).

Figure 3.1 shows an example of these dynamics. In this simulation, both cue and target stimuli were generated with Gaussian functions presented one quarter and two-thirds of the way during the trial (each trial comprised sixty four 10 ms time bins; i.e., 640 ms). When generating stimuli all random fluctuations were suppressed, using a log-precision of eight. The cue was a simple bump function with duration (standard deviation) of about 45 milliseconds. The target was a (biphasic) time derivative of a Gaussian bump function with duration of about 90 milliseconds. The cue and target stimuli are shown as broken grey

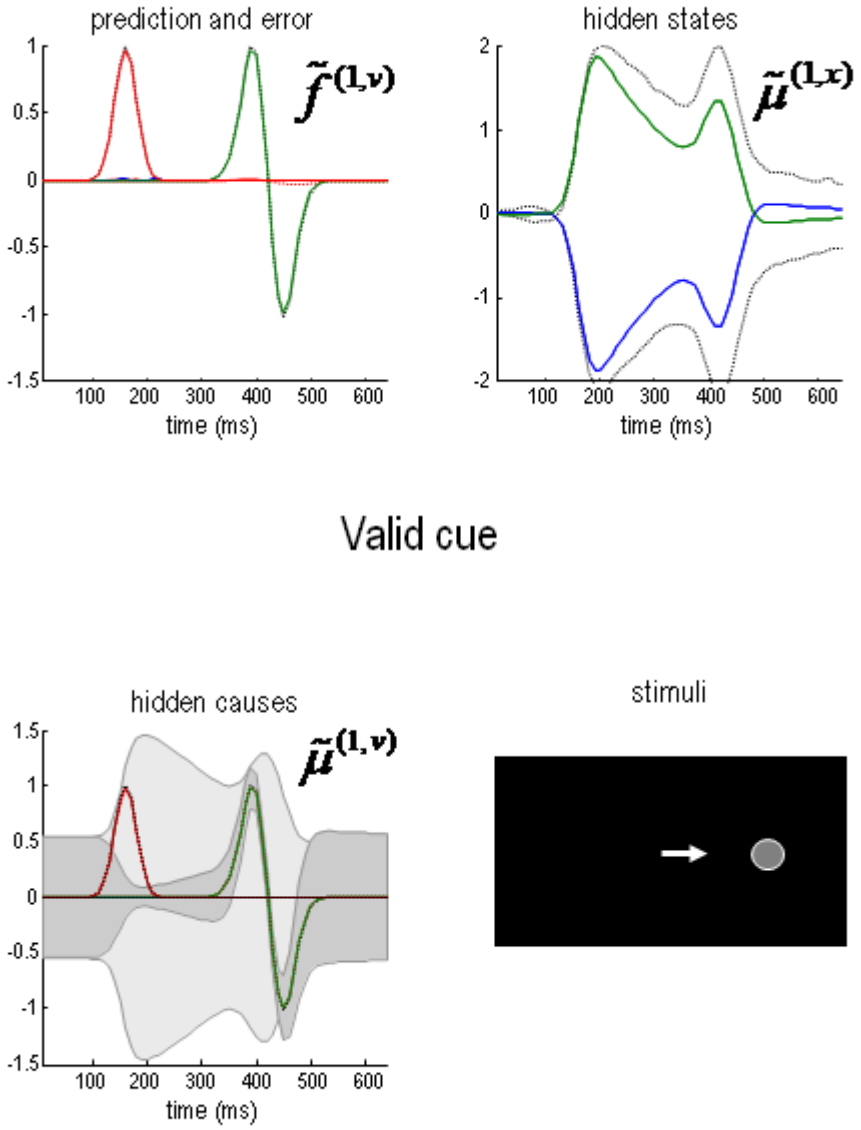

Figure 3.1: Simulation of the Posner task. Here, a validly cued target is simulated. **Upper left panel**: The time-dependent expression of the cue and target stimuli are shown in red s_C and green s_L respectively. **Lower left panel**: The ensuing conditional expectations of the underlying hidden causes $v_R^{(1)}, v_C^{(1)}, v_L^{(1)}$ are shown below. The grey areas correspond to 90% conditional confidence tubes. **Upper right panel**: This confidence reflects the estimated precision of the sensory data, which is encoded by the expectations of the hidden states. The green line corresponds to a precision or attentional bias to the right $x_R^{(1)}$ and the blue line to the left $x_L^{(1)}$. They grey lines are the true precisions. **Lower right panel**: This insert indicates the sort of stimuli that would be generated by these hidden causes.

lines in Figure 3.1. These are nearly underneath the respective predictions in red s_C and green s_L respectively. The dotted red lines show the prediction error and reflect the small amount of noise used in the simulations (a log-precision of eight; see Equation 3.3). The ensuing conditional expectations of the underlying causal states $v_R^{(1)}, v_C^{(1)}, v_L^{(1)}$ are shown below (lower left). The grey areas around the expectations correspond to 90% conditional confidence regions (referred to as tubes). Note that the conditional tube for the cued target (green line $v_R^{(1)}$) is relatively tight because the precision of the prediction errors associated with this location is high. Conversely, the tube for the nontarget $v_L^{(1)}$ is somewhat wider but correctly centred on an expectation of zero. The precisions are determined by the hidden states shown on the upper right. The green line corresponds to a precision or attentional bias to the right $x_R^{(1)}$ and the blue line to the left $x_L^{(1)}$. It can be seen that by the time the target arrives, the log-precision is about four (see Equation 3.3). This is substantially greater than the prior precision on the hidden causes (set this to a log-precision of zero). Therefore, the representation of the hidden cause (target) is driven primarily by sensory input. The insert on the lower level provides a schematic indicating the sort of stimuli that would be generated by these hidden causes. Now, compare these results with the equivalent responses to an identical stimulus but presented in the other hemi-field:

Figure 3.2 uses the same format as Figure 3.1 to show the responses to an invalid target (blue lines) presented on the right. It can be seen here that the predictions on this sensory channel are substantially less than the true value (compare the blue and dotted grey lines) with a consequent and marked expression of prediction error (dotted red line). As anticipated, the conditional confidence regions for the conditional expectation of this invalid target (lower left panel) are now much larger; with the 90% confidence tube always containing the value zero. The reason for this is that this invalid cue has failed to reverse the attentional context and is still operating under low levels of precision. This is reflected by the hidden states. In comparison with the previous figure, the attentional bias (difference between the right and left hidden states) has been subverted by the invalid cue but has not been reversed (the dotted grey lines show the true values of these hidden or contextual states).

The result of this asymmetry between valid and invalid cueing means that responses to valid targets are of higher amplitude and have much tighter confidence tubes, in relation to invalid targets. This is shown on the lower right panel of Figure 3.2, where one can

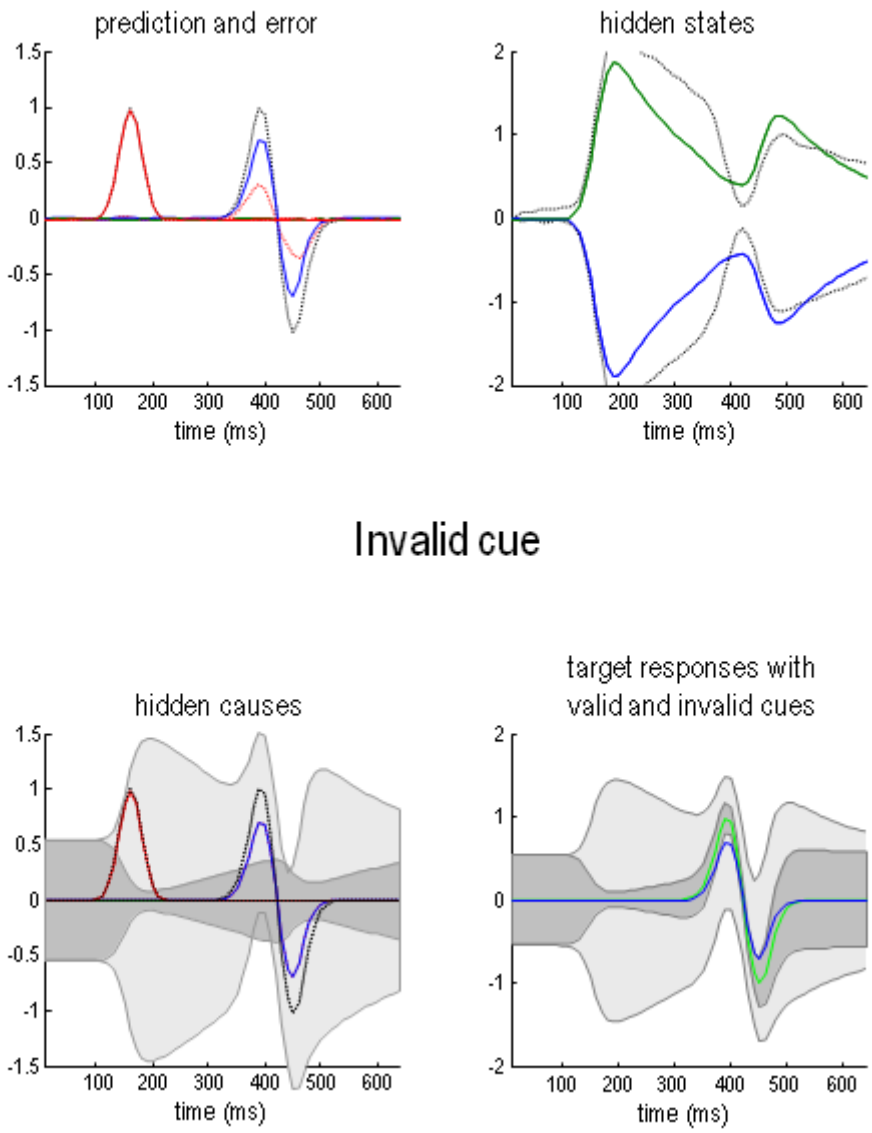

Figure 3.2: Simulation of an invalid cue. This figure uses the same format as Figure 3.1 but shows responses to an invalid target (blue line) presented on the right. The predictions of this sensory channel are substantially less than the true value (compare the blue and dotted grey lines) with a consequent expression of prediction error (dotted red line). The conditional confidence regions for the conditional expectation of this invalid target (**lower left panel**) are now much larger than in the previous figure. This is shown in the **lower right panel**, where one can compare the conditional estimates of the valid (green; see Figure 3.1) and the invalid (blue) hidden cause, with their respective conditional confidences (grey). Note that these responses were elicited using exactly the same stimulus amplitude.

compare the conditional estimates of the valid (green) and the invalid (blue) cause. Note that these profoundly different responses were elicited using exactly the same stimulus amplitude, after the cue had disappeared. This means that the difference is attributable only to the context (hidden states) that is instantiated by the endogenous cue. This is the basic phenomenon that I wanted to demonstrate, namely attentional bias in the ability of stimuli to capture attentional resources, where these resources correspond to the precision of sensory samples encoded by inferred hidden states or context. The reason that precision behaves like a resource is that the generative model contains prior beliefs that log-precision is redistributed over sensory channels in a context-sensitive fashion but is conserved over all channels.

4.2 The psychophysics of the Posner paradigm

The difference in the confidence tubes between valid and invalidly cued targets (Figure 3.2; lower right) can be usefully interpreted in relation to behaviour. At each point in peristimulus time, the conditional density implicit in the conditional mean and precision can be used to compute the conditional probability that the target intensity is present. This provides the posterior probability $p(v_i^{(1)} > 0 \mid \tilde{s}, m) : i \in R, L$ of the presence of a target as a function of peristimulus time shown in Figure 3.3 (left panel). These results can be interpreted in terms of a speed-accuracy trade-off. For example, one can identify the amount of peristimulus time required to accumulate sufficient evidence for a fixed level of accuracy, as determined by the posterior conditional confidence. Note how the conditional probability of the target being present shrinks toward chance (50%) levels, under invalid cueing. In this example, 80% conditional confidence for valid targets (solid line) is attained at about 20 milliseconds before the same accuracy for invalid targets (broken line). This translates into a reaction time advantage for valid targets of about 20 milliseconds.

Figure 3.3 (right panel) shows the time taken to reach 80% conditional confidence after the onset of invalid, neutral and valid cues (simulated with $\gamma = .8$). Neutral cues are modelled by reducing $\gamma = .2$ and removing any spatial bias afforded by the hidden states (by only using valid targets). This produces a temporal facilitation (temporal alerting effect) but without spatial specificity. The reaction time advantage with valid cues and the cost with invalid cues can be seen clearly. The reaction time to neutrally cued stimuli lies between these values. Note the asymmetry between the reaction time benefit of a valid

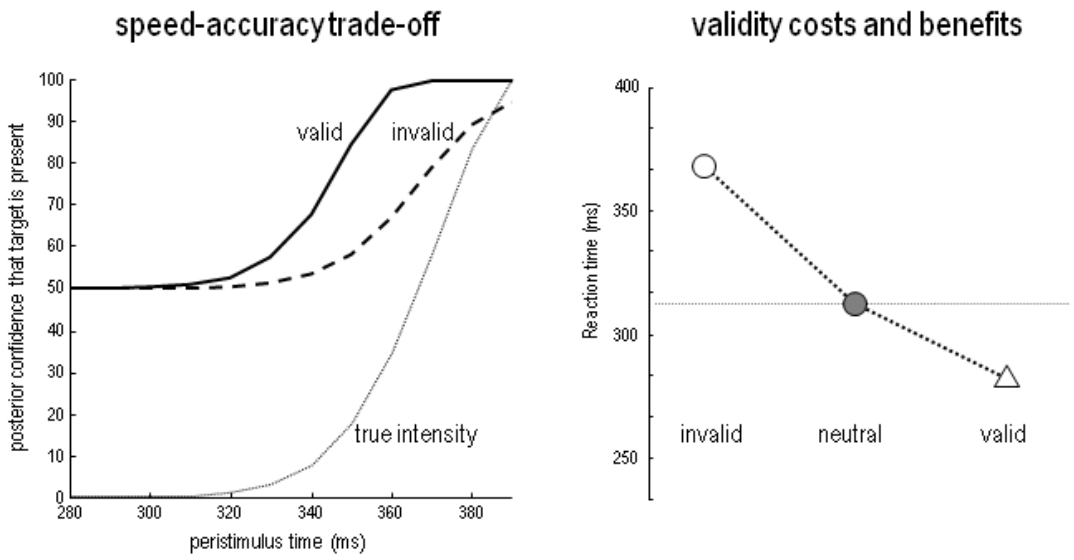

Figure 3.3: Speed-accuracy tradeoffs and replication of psychophysical results. Left panel: The posterior probability of a target being present as a function of peristimulus time, which can be interpreted in terms of a speed-accuracy trade-off. A reaction time can be derived from this data, as the post-stimulus time taken to achieve a fixed level of accuracy, as determined by the posterior conditional confidence. In this example, 80% conditional confidence is attained at about 340 milliseconds for valid targets (solid line). However, for invalid targets (broken line) the same accuracy is only attained after about 360 milliseconds. This translates into a reaction time advantage for valid targets of about 20 milliseconds. **Right panel**: This shows the reaction times for invalid, neural and valid cues (to within a constant), where neutral cues caused a small reduction in precision but with no spatial bias. The reaction times here are shown to within an additive constant, to better reflect empirical data (see Figure 3.4).

cue and the cost of an invalid cue; this asymmetry is evident in behavioural data and is an emergent property of the nonlinearities inherent in this Bayes-optimal scheme.

Recall that the time-course of the Posner effect depends on the slowly-decaying hidden states encoding precision (with a time constant of thirty two in Equation 3.3). This reflects a formal prior that changes in precision show a temporal persistence at any location in visual space. This sort of prior means that attentional biasing will persist but decay monotonically following a cue. This effect manifests in reaction times as a slow decay of benefits and costs with valid and invalid cures respectively. Figure 3.4 (left panel) shows the difference in reaction times following the three types of cue for various asynchronies between cue and stimulus onset (the 'foreperiod'). The small benefit seen for neutral cues is due to a temporal alerting effect and reflects an increase in precision with no spatial bias (i.e., a small increase in precision at both locations). Note that cue-dependent effects emerge over 200 ms, during which time conditional expectations accumulate evidence (see Figure 3.1; upper right panel). The ensuing profiles of reaction times are pleasing similar to empirical observations. The right panel of Figure 3.4 shows the corresponding behavioural results reported in Posner et al. (1978). Note again that the asymmetry in costs and benefits, over different foreperiods, is an emergent property of the scheme used in the simulations.

The speed-accuracy trade-off is a useful psychophysical function, which can also be interpreted in terms of relative accuracies at a fixed reaction time. In this example, at 360 ms after the cue (about 50 ms after the onset of the target), the posterior confidence about the presence of valid targets is about 98%, whereas it is only about 70% for invalid targets (Figure 3.2). The relative position and divergence of the speed-accuracy curves may provide a useful and quantitative link to empirical psychophysical data.

4.3 The electrophysiology of the Posner paradigm

In what follows, I attempt to explain the well characterised electrophysiological correlates of the Posner paradigm using simulated event-related activity evoked by target stimuli. Spatial cueing effects are expressed in the modulation of event-related potentials (ERPs) to valid and invalid cues (Eimer et al., 1993; Mangun and Hillyard, 1991; Perchet et al., 2001). Generally, one sees an increase in P1 and N1 and a decrease in posterior P3 components in validly cued trials with respect to invalid ones. In other words, there is usually a validity-related enhancement of early components and an invalidity-related enhancement of late components. The P1 component is the earliest component showing

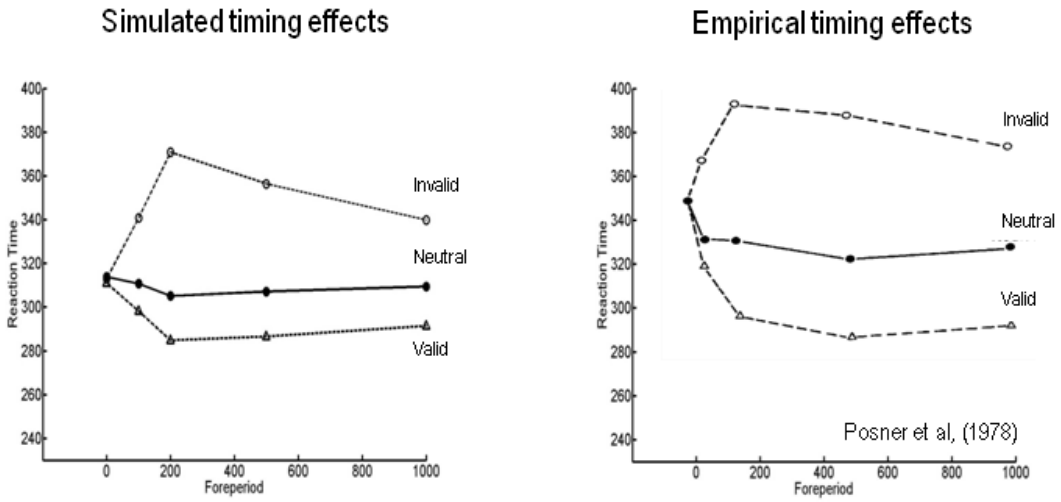

Figure 3.4: Timecourse of the cuing effect. **Left panel**: Simulated reaction times showing the time course of the Posner effect over different delays (foreperiod) between the onset of the cue and the target increases. **Right panel**: Empirical reaction time data, redrawn from Posner et al (1978). In both the simulated and empirical data, reaction time benefit and cost increase swiftly to a maximum and then decay slowly. This reflects the quick rise and slow decay of the inferred hidden states seen in Figures 3.1 and 3.2 (upper right panels). There is a slight reaction time benefit for neutral cues due to a temporal alerting effect. This was modelled by allowing neutral cues to induce a small rise in both the inferred hidden states. The simulated reaction times data were taken as time at which there was 80% confidence that the target was present. The simulated reaction times are shown to within an arbitrary constant (to accommodated unmodelled motor responses). The asymmetric difference between the cost for an invalid cue and the benefit for a valid cue was an emergent property of the simulations.

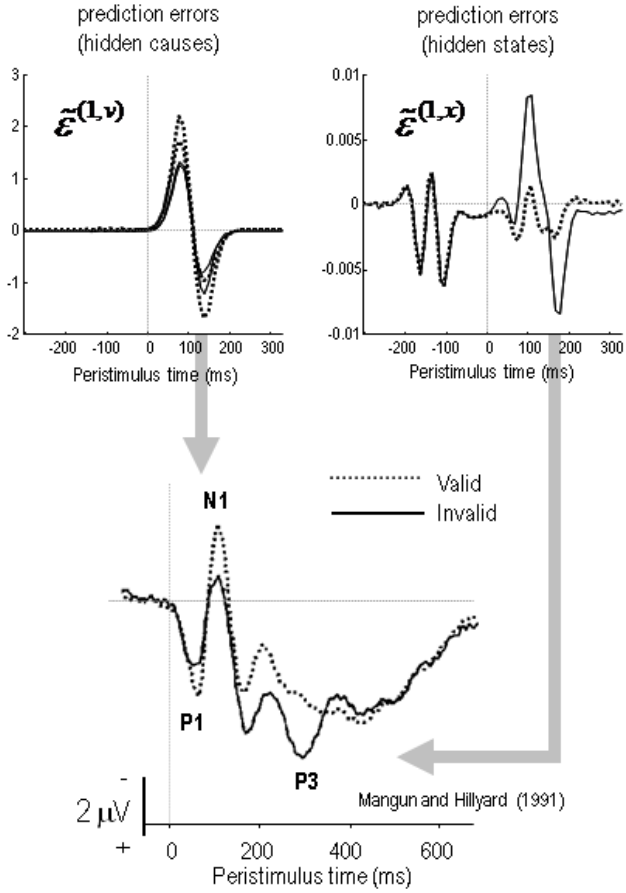

Figure 3.5: Simulation of electrophysiological responses. This figure shows simulated EEG data from our simulations (**upper panels**) and empirical EEG data (**lower panel**) from Mangun and Hillyard (1991). The EEG traces were created from the prediction errors on the hidden causes (**left**) and states (**right**). The empirical data were recorded via EEG from the occipital cortex contralateral to the target (i.e., the cortex processing the target). The simulated data exhibits two important features of empirical studies: Early in peristimulus time, stimulus-driven responses are greater for valid cues (upper left panel) relative to invalid cues. This is often attributed to a validity enhancement of early (e.g., N1) components. Conversely, later in peristimulus time, invalid responses are greater in amplitude. This can be related to novelty (and salience) responses usually associated with late waveform components (e.g., P3). In the simulations, this invalidity effect is explained simply by greater prediction errors on inferred hidden states encoding precision (upper right panel; second bracket in lower panel). It is these prediction errors that report a surprising or novel context, following the failure to predict invalidly cued stimuli in an optimal fashion.

attentional modulation and is considered to reflect attentional gain or the cost of attending to the wrong location (Coull, 1998; Luck et al., 1990; Mangun and Hillyard, 1991). It is well known that the amplitude of the later P3 component is inversely related to the probability of stimulus (Donchin and Coles, 1998). The anterior P3a is generally evoked by stimuli that deviate from expectations. Indeed, novel stimuli generate a higher-amplitude P3a component than deviant but repeated stimuli (Friedman et al., 2001). The P3b is a late positive component with a parietal (posterior) distribution seen in oddball paradigms and is thought to represent a context-updating operation (Donchin and Coles, 1998; Polich, 2007). Increased P3 amplitudes during invalid trials, relative to valid trials, suggest that invalidly cued targets produce a novelty-like effect (P3a) and change the representation of probabilistic contingencies (P3b) or context (Gómez et al., 2008; Vossel et al., 2006). These hypotheses sit very comfortably with the formal scheme in this chapter; in that sensory signals (prediction errors) evoked by valid targets will enjoy a selective gain, leading to enhanced early (P1 and N1) responses. Conversely, initial responses to invalid targets are suppressed until they revise the probabilistic context encoded by inferred hidden states. The prediction errors on the hidden states reflect (and drive) this revision and may contribute the later (P3) ERP components. The prediction errors on the hidden causes and states representing the content and context respectively are shown in Figure 3.5.

Figure 3.5 shows synthetic ERPs based on the simulations in Figures 3.1 and 3.2. Here, I have made the simplifying assumption that electrophysiological signals represent the activity of superficial pyramidal cells (which are presumed to encode prediction error; Friston 2008). This means we can focus on the prediction error as a proxy for electrophysiological responses. The results in the top panels of Figure 3.5 show the prediction errors on the sensory signals ($\tilde{\varepsilon}^{(1,v)}$ - left panel) and hidden states ($\tilde{\varepsilon}^{(1,x)}$ - right panel). The prediction errors for valid trials are shown as dotted lines and invalid trials as solid lines. These simulations show an early suppression of prediction error for an invalidly cued target, as its low precision fails to drive its representation to its veridical level. This violation of predictions causes prediction errors on the hidden states encoding context that are expressed later in peristimulus time and drive the hidden states to revise their conditional expectations (shown in Figures 3.1 and 3.2). This double dissociation between validity effects in early and late peristimulus time is exactly the same as that observed by Mangun and Hillyard (1991). The empirical results of their ERP study are shown in the lower panel of Figure 3.5 and are very similar to the simulations.

4.4 Summary

In summary, this section has applied the Bayes-optimal scheme established in the previous section to a minimal model of the Posner paradigm. This model provides a mechanistic if somewhat simplified explanation for some of the key psychophysical and electrophysiological aspects of the Posner effect, namely, validity effects on reaction times and the time course of these effects as stimulus onset asynchrony increases. Furthermore, the model exhibits an asymmetry in costs and benefits for invalid and valid trials respectively. Electrophysiologically, it suggests early attentional P1 enhancement can be attributed to a boosting or biasing of sensory signals (prediction errors) evoked by a target, while later P3 invalidity (cf, novelty) effects are mediated by prediction errors about the context in which targets appear.

5. Simulating biased competition

In this final section, I revisit the simulations above but from the point of view of biased competition. Although the Posner paradigm considers a much greater spatial and temporal scale than the paradigms normally employed in a monkey electrophysiology, similar phenomena can be emulated by presenting both cued and non-cued targets simultaneously using the Posner model. I hoped to see a competitive interaction between stimuli that favoured the cued target. Furthermore, I hoped to see responses to the unattended (invalid) target changed in the presence of an attended target. This is one of the hallmarks of biased competition and is usually attributed to lateral interactions among competing representations for stimuli, within a cell's receptive field (see Section 2). Although this model is too simple to distinguish between stimuli presented inside and outside the classical receptive field (because the spatial support of sensory channels is not modelled in this chapter), we can assume that targets fall within the extraclassical receptive of field of units representing hidden causes. This is because the response to one target depends on the presence of the other, as we will see next. This competition is explicitly laid out in the model, by forcing cue stimuli to decrease the value of one hidden state as much as they increase the value of the other. The process here is best described as 'competition'; a selection process could be trivially implemented by adding a further hierarchical level exhibiting winner-takes-all behaviour.

Figure 3.6 shows the results of presenting both stimuli simultaneously. Again the cue is in red, the valid target in green and the invalid (unattended) target in blue. It is immediately obvious that biased competition between the targets is profound, such that the response

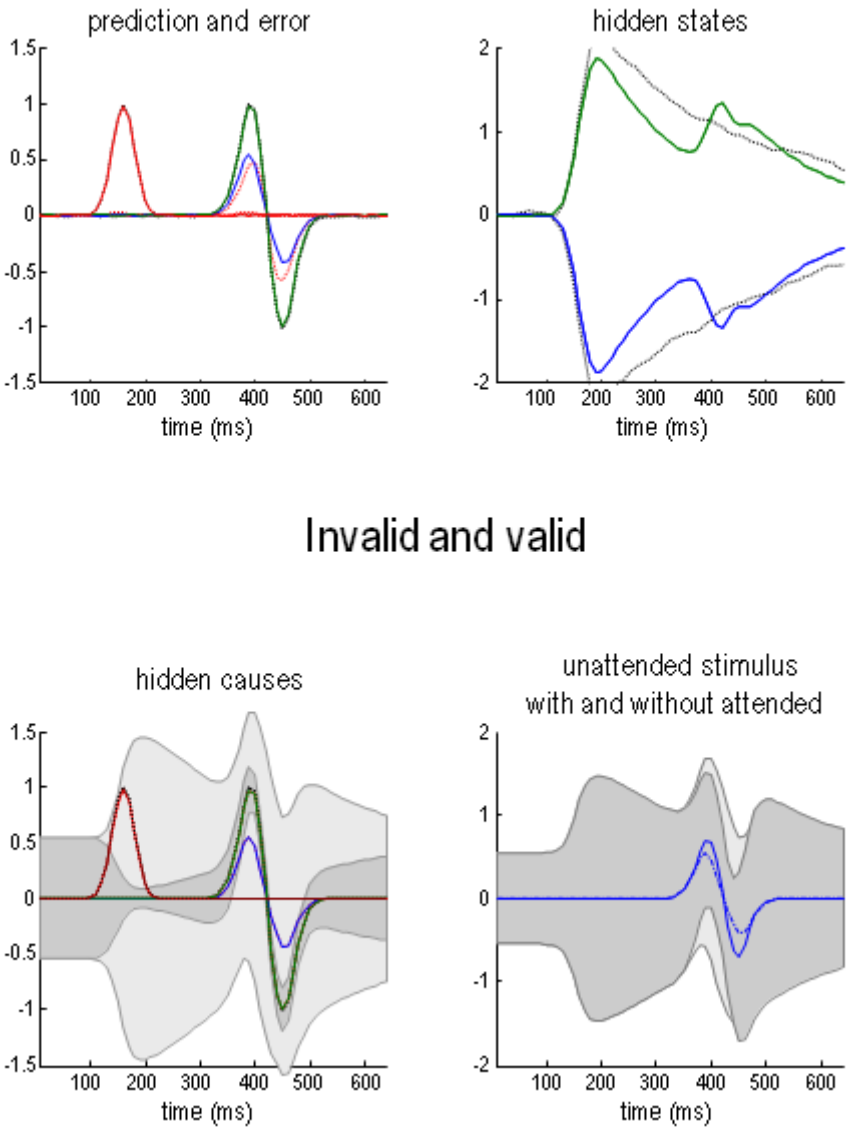

Figure 3.6: Simulation of biased competition. This figure uses the same format as Figures 3.1 and 3.2 but reports the results when both targets are presented simultaneously. The ensuing conditional responses can be compared with the responses in Figure 3.2, when the invalidly cued target was presented alone: When the valid target is also presented, it prevents the invalid target from reversing the precision bias established by the cue; i.e., it fails to capture attention resources. The lower left panel shows the conditional expectation and confidence regions for the invalid target, with and without the valid target, to show how the responses evoked are suppressed; i.e., biased competition.

to the unattended target is about half of the response to the attended target. Furthermore, the conditional confidence about the unattended target is substantially less than that for the attended target (light and dark confidence tubes in the lower left panel). The lower right panel of Figure 3.6 compares the conditional expectations and confidence intervals associated with the unattended (invalid) target presented with and without the attended (valid) target. The latter response is exactly the same as the data presented in the lower left of Figure 3.2 simulating invalid cue responses. One can see that when the same stimulus is presented in conjunction with an attended target, its conditional expectation is attenuated by about 20% and the conditional confidence tubes are much wider (light with an attended distractor and dark without). In other words, the attended target has competed for attentional resources to subvert conditional confidence about the unattended target. This is despite the fact that both unattended targets were identical; they were just presented in a different context.

This context is encoded by the expected hidden states and explains the biased competition for resources: In contrast with the hidden states inferred with the invalid target alone (see the equivalent panel in Figure 3.2) the partial reversal of contextual representations has been precluded by the presence of the valid target. This means that the invalid cue can no longer capture precision and consequently is never able to fully express itself, through precise prediction errors, on the conditional representation of its cause. It is this effect, and only this effect, that is needed to explain biased competition. Note that it is not necessary to model lateral interactions or explicit competition among representations; competition emerges naturally in a Bayes-optimal fashion through the nonlinear effects of precision encoded by the units representing context, where the influence of these units is mediated by top down or lateral projections.

The results in Figure 3.6 are strikingly similar to data obtained from electrophysiological studies. Figure 3.7 (upper panel) shows the conditional expectations about valid (solid line) and invalid (dashed line) targets from Figure 3.6. The lower panel shows peristimulus histograms reported in Luck al. (1997) following simultaneous presentation of two (effective and ineffective) stimuli averaged over V4 neurons that showed a significant attention effect. The solid line reports trials in which attention was directed to the effective stimulus (cf, responses to a valid target) and the dashed line when attention was directed to the ineffective stimulus (cf, responses to an invalid target). The quantitative agreement between these simulated and empirical responses is evident and speaks quantitatively to biased competition among stimuli.

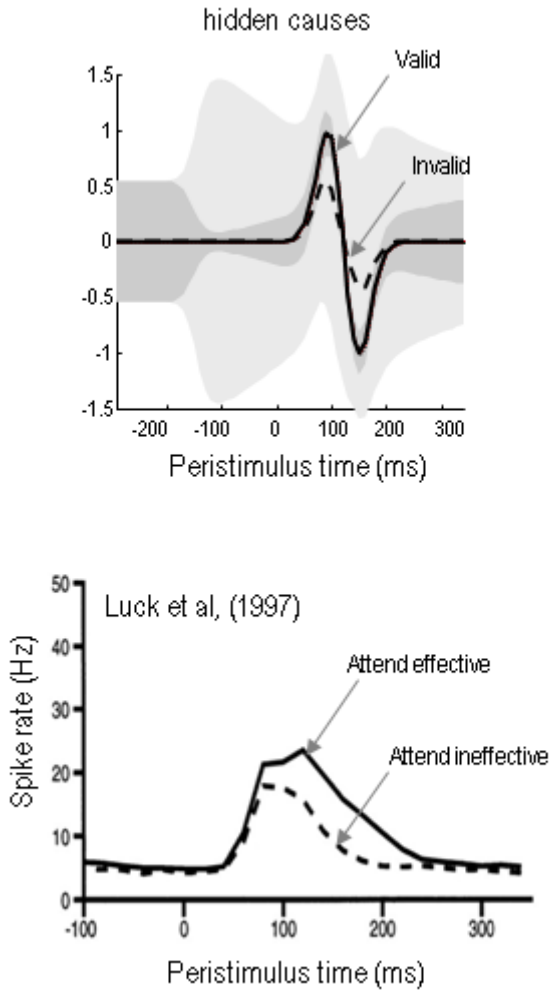

Figure 3.7: This figure demonstrates how generalised predictive coding reproduces some quantitative aspects of biased competition. The simulation (**upper panel**) reproduces the conditional expectations in the previous figure about valid (solid line) and invalid (dashed line) targets, when presented simultaneously. These two responses resemble those reported in Luck et al. (1997): **Lower panel**: Peristimulus histograms (over 20 ms bins) redrawn from Luck at al (1997), following simultaneous presentation of two (effective and ineffective) stimuli averaged over 29 V4 neurons that showed a significant attention effect. The solid line reports trials in which attention was directed to the effective stimulus (cf, responses to a valid target) and the dashed line when attention was directed to the ineffective stimulus (cf, responses to an invalid target). Note that the empirical data are non-negative spike counts, whereas the simulated activity represent firing rate deviations around baseline levels.

5.1 Summary

Biased competition emerges naturally in Bayes-optimal schemes as a simple consequence of the fact that only one context can exist at a time. This unique aspect of context is encoded in the way that the representation of hidden states (context) modulates or distributes precision over sensory channels. Optimising this representation leads to competition among stimuli to make the inferred context more consistent with their existence. This highlights the simplicity and usefulness of appealing to formal (Bayes-optimal) schemes, when trying to understand perception.

6. Discussion

This treatment of attention is one of many accounts that emphasise the role of probabilistic inference in sensory processing; including sensorimotor integration (Körding and Wolpert, 2004; Wolpert et al., 1995), sensory integration (Jacobs, 1999; Knill & Saunders 2003; and Ernst and Banks, 2002; Saunders, 2003; Alais and Burr, 2004), salience and value estimation (Whiteley and Sahani, 2008; Trommershauser et al., 2003b; Seydell et al., 2008) and perception (Langer and Bulthoff, 2001; Adams et al., 2004). There have been some notable Bayesian accounts of attention using formal models (Rao 2005; Spratling 2008, Spratling 2010). Others have tried to define statistical measures of saliency, i.e. that which draws our attention (Duncan and Humphreys, 1989; Bruce & Tsotsos 2009; Itti and Baldi, 2010). These developments are now discussed in the light of the more general free-energy formulation used in this chapter.

The free-energy formulation is a generalisation of information theoretic treatments that subsumes Bayesian schemes by assuming the brain is trying to optimise the evidence for its model of the world. This optimisation involves changing the model to better account for sensory samples or by selectively sampling sensations that can be accounted for by the model (cf, perception and action). Attention can be viewed as a selective sampling of sensory data that have high precision (signal to noise) in relation to the model's predictions. Crucially, the model is also trying to predict precision. It is this (state-dependent) prediction I associate with attention. In short, perception, attention and action are trying to suppress free-energy, which is an upper bound on (Shannon) surprise (or the negative log-evidence for the brain's model of the world). Under some simplifying assumptions, free-energy is just the amount of prediction error, which means free-energy minimisation can be cast as predictive coding. So how does this relate to other formal treatments?

6.1 Attention and surprise

Rao (2005) has introduced a compelling model of visual attention using Bayesian belief propagation. However, although consistent with Bayesian (free-energy) principles, belief propagation schemes rest on (discrete) representations of hidden causes and states, which are not compatible with the dimensionality of states in the real world (Friston 2009). Using a more descriptive approach, Itti and Baldi (2006; 2009) proposed that many factors, which influence visual salience, can be integrated with prior expectations by calculating *Bayesian surprise*. This is (heuristically) related to another measure of saliency, proposed by Bruce and Tsotsos (2009), who suggest that visual searches are attracted to areas of the visual field which maximize the information sampled. Crucially, reducing free-energy or (Shannon) surprise increases Bayesian surprise and increases the changes in the conditional representations afforded by sensory information. This is because Bayesian surprise is the difference (Kullback-Leibler divergence) between the posterior (conditional) and prior densities on hidden causes or states. This difference reports the change in the conditional density after sampling new information. It is also called *complexity* in the Bayesian model comparison literature. Free-energy can be expressed as complexity minus accuracy (Friston 2009). This means that minimising (Shannon) surprise by updating conditional representations to increase accuracy (decrease prediction errors), necessarily entails an increase in complexity (Bayesian surprise). In short, increases in Bayesian surprise are necessarily associated with decreases in free-energy (they are the complexity cost of reducing prediction errors) but Bayesian surprise *per se* is not optimised in Bayes-optimal schemes.

6.2 Biased competition and predictive coding

It is becoming increasingly clear that estimates of the precision play an important role in sensory inference. Whiteley and Sahani (2008) demonstrated very neatly that the brain possesses (and uses) a model of sensory uncertainty (i.e. precision) in decision-making, and that this model is available even under intermittent feedback, showing that is estimated internally rather than learnt. Thinking of attention as optimising representations of uncertainty or precision resolves any potential conflict between biased competition and predictive coding schemes: Spratling (2008) noted the potential difficulty in reconciling these two theories and proposed a variant of predictive coding, in which representations compete via negative feedback. Specifically, he showed that a particular implementation of the biased competition model, in which nodes compete via

inhibition that targets the inputs to a cortical region, is mathematically equivalent to linear predictive coding. This scheme relies on a rather complex neural architecture and employs nonlinear modifications to prevent cells from having a negative firing rate. These modifications are interesting and relate to important theories based on divisive normalisation (Heeger 1993). This form of (divisive) predictive coding can explain a remarkable range of classical and extraclassical receptive field properties in V1 (see Spratling 2010).

The formulation in this chapter reaffirms that there is no tension between biased competition and predictive coding: It demonstrates that the characteristic behaviours of biased competition emerge naturally under predictive coding. They key thing that reconciles these two theories is to realise that predictive coding can be generalised to cover both states and precisions and that (state-dependent) precision is itself optimised. This leads to nonlinear interactions among states implicit in the precision-weighting of prediction errors and provides a simple explanation for attentional gain effects. It will be interesting to relate the ensuing bias or weighting of sensory signals (prediction errors) by precision to the divisive schemes above (e.g., Heeger 1993; Spratling 2010).

6.3 Baseline shifts and precision

In this chapter, I have focussed on reaction time and event-related responses to targets. However, many electrophysiological and neuroimaging studies of attentional paradigms (e.g., Chelazzi et al. 1993; Chawla et al., 1999b; Kastner et al., 1999; Stokes et al., 2009) have demonstrated cue-related increases in the basal firing rate of cells, whose receptive field corresponds to the attended location. A non-invasive electrophysiological correlate of these baseline shifts is called the Contingent Negative Variation component (CNV), which follows a cue that furnishes information about subsequent (imperative) target stimuli (Walter et al., 1964; Rockstroh et al., 1982). Crucially, the cortical sources generating the CNV can include those responsible for processing the stimuli (Gómez et al., 2001). These baseline shifts may be accounted for, in the computational scheme, by the dynamics of expected hidden states, shown in the top left panels of Figures 3.2 and 3.3. These accumulate evidence from cues and represent changes in context that persist over time. It is possible that the activity of these representational units could contribute to the CNV or baseline shift directly. However, it is also possible that they could modulate baseline activity (caused by ambient sensory signals) in the prediction error units they modulate. This would be consistent with baseline shifts seen with fMRI in retinotopically mapped areas of directed attention (e.g., Macaluso et al, 2003), and the reduction in non-attended

areas (Smith et al, 2000). This suggests that baseline (endogenous) activity may be a quantitative proxy for the expected precision of sensory information in the corresponding sensory area (cf., Hesselmann et al, 2008). This hypothesis was tested recently: using fMRI, Hesselmann et al. (2010) linked perceptual estimates of precision with baseline increases in activity; showing that baseline activity before a (subliminal) stimulus was correlated with the accuracy of deciding if the stimulus was present (and not whether the decision was present or absent). This means that baseline activity may reflect the inferred precision of sensory signals. Specifically, they found that neuronal activity in sensory areas (extrastriate visual and early auditory cortex) biases perceptual decisions towards correct inference and not towards a specific percept. They conclude: "In accord with predictive coding models and the free-energy principle, this observation suggests that cortical activity in sensory brain areas reflects the precision of prediction errors and not just the sensory evidence or prediction errors *per se*."

The neurobiological (resp. computational) mechanisms that might underlie these effects tie several strands of evidence together rather neatly: As noted in the introduction the most plausible candidate for modulating activity-dependent (resp. state-dependent) synaptic gain (resp. precision) are fast synchronous interactions associated with attention (Börgers et al 2005; Womelsdorf & Fries 2006; Zeitler, Fries and Gielen, 2008; Fries et al 2008). The associated increase in synchronous gain is necessarily accompanied by increased levels of population activity that are both supported by and support synchrony (Chawla et al 1999a; Salinas and Sejnowski, 2001). These are manifest as high frequency (gamma) activity and elevated fMRI signals seen in attentional paradigms (Gruber et al., 1999; Sokolov et al., 1999; Steinmetz et al., 2000; Bichot et al., 2005; Pavlova, Birbaumer and Sokolov, 2006; Vidal et al., 2006; Fries et al., 2008).

6.4 Attention, gain and learning

In closing, a potentially interesting argument about the specificity of gain mechanisms and attention can now emerge. The idea pursued in this chapter is that attention corresponds to inference about uncertainty or precision and that this inference is encoded by dynamic changes in postsynaptic gain. However, nonlinear (gain) postsynaptic responses are ubiquitous in the brain; so what is special about the nonlinearities associated with attention? I suggest that attention is mediated by gain modulation of prediction error units (forward or bottom up information) in contradistinction to gain modulation of prediction units (backward, lateral or top-down information). The distinction may seem subtle but there is a fundamental difference between inferring the context-dependent contingencies

and causes of sensations (perception) and their precision (attention). For example, in a probabilistic learning task a cue that predicted reward in one particular context would be afforded higher top-down precision (on predictions) than one known to predict reward only in another context, whereas out of two cues predictive of reward, the one which predicted reward with a greater probability would cause higher bottom-up precision (on prediction errors) to be afforded to any subsequent events. In this sense, there is an implicit distinction between inferring what is relevant for a task (as in classical attention tasks like dichotic listening) and the uncertainty about what is relevant. I have side-stepped this issue with the Posner task, because all cues are task relevant, and as such it is only the bottom-up precision on prediction errors which is modulated.

There is a final distinction that may be mechanistically important: This chapter have focussed on activity-dependent optimisation of gain but have not considered the (slower) learning of how and when this optimisation should be deployed. For example, the latency of saccades to a target can be reduced if the target is more likely to appear on one side – and this relationship can be learned in as few as 150 trials (Carpenter and Williams, 1995; Anderson and Carpenter, 2006; Brodersen et al., 2008). This sort of learning corresponds to the optimisation of the precision parameters in Equation 3.2 and may involve modulatory neurotransmitters. This learning may be related to the psychopharmacology of attention and related theories about uncertainty (e.g., Yu & Dayan 2005).

6.5 Conclusion

In this chapter, I have tried to establish the face validity of optimising the precision of sensory signals as an explanation for attention in perceptual inference. We started with an established scheme for perception based upon optimising a free-energy bound on surprise or the log-evidence for a model of the world. Minimising this bound, using gradient descent, furnishes recognition dynamics that are formally equivalent to evidence accumulation schemes. Under some simplifying assumptions, the free-energy reduces to prediction error and the scheme can be regarded as generalised predictive coding. The key thing that I have tried to demonstrate is that all the quantities required for making an inference have to be optimised. This includes the precisions that encode uncertainty or the amplitude of random fluctuations generating sensory information. By casting attention as inferring precision, we can explain several perspectives on attentional processing that fit comfortably with their putative neurobiological mechanisms. Furthermore, by considering how states of the world influence uncertainty, one arrives at a plausible architecture, in which conditional expectations about states modulate their

own precision. This leads naturally to competition and other nonlinear phenomena during perception. I have tried to illustrate these ideas in the context of a classical paradigm (the Posner paradigm) and relate the ensuing behaviour to biased competition evident in electrophysiological responses recorded from awake, behaving monkeys. In chapter 4, I will use the theoretical framework in this chapter to model empirical electrophysiological data and pursue this hypothesis using formal model comparison.

The work described in this chapter is published as:
Feldman H, Friston KJ. (2010). Attention, uncertainty, and free-energy. Front. Hum. Neurosci. 4:215.

Chapter 4

The functional anatomy of attention: a DCM study

1. Introduction

In chapter 3, I suggest that attention can be understood as the selection of processing channels that conveyed precise or salient information within the framework of predictive coding (Feldman & Friston, 2010). The idea is that both the content of visual information and the confidence placed in that information have to be inferred during perception. In predictive coding, top-down predictions of the content are confirmed or disconfirmed by comparison with bottom-up sensory information (Rao & Ballard, 1999; Friston, 2005). However, this comparison rests on estimating the reliability or precision of sensory information – or more exactly the residuals or prediction error that cannot be explained. This precision may be itself context sensitive and has to be updated in exactly the same way as predictions of content, as described in chapter 1 (Brown & Friston, 2012). This leads to view of hierarchical perceptual synthesis in which particular processing channels are selected on the basis of cues that portend spatial locations or featural attributes that are likely to convey precise information. In neuronally plausible implementations of this hierarchical Bayesian inference – namely, generalised Bayesian filtering or predictive coding – expected precision is thought to be encoded by the postsynaptic sensitivity or gain of cells reporting prediction error (Friston & Kiebel, 2009). Given that prediction error is passed forward from sensory cortex to higher cortical areas by ascending or forward connections, the most likely candidates for reporting prediction error are the superficial pyramidal cells that are the source of ascending connections (Bastos et al.,

2012). This means that one can understand attention as the top-down gain control of superficial pyramidal cells passing information that is yet to be explained (i.e., prediction error) deep into the visual hierarchy.

This normative model and its neuronal implementation have been used to simulate and reproduce both the psychophysical and electrophysiological characteristics of the Posner paradigm, as described in chapter 3 (Feldman & Friston, 2010). In brief, predictive cues engage top-down predictions of increased precision in the left or right hemifield that facilitate the rapid processing of (inference about) valid visual targets. However, when an invalid target is presented in the wrong hemifield, the evidence accumulation implicit in predictive coding is slower, because gain or precision acts as a synaptic rate constant, meaning more presynaptic input is needed before firing threshold is reached. This leads to protracted reaction times and an invalidity cost. Simultaneously, the scheme infers that prior beliefs about the target have been violated and prediction errors drive higher levels to update both the deployment of attention (i.e., precision) and target predictions *per se*. This explains the classic electrophysiological correlates of the validity effects in the Posner paradigm – in which invalid targets elicit slightly attenuated P1, N1 and N2 early components and a more pronounced P3b late component (Mangun & Hillyard, 1991; Hugdahl & Nordby, 1994; Talsma et al., 2007). These two electrophysiological characteristics may reflect the initial insensitivity (low precision or gain) of early visual responses and a subsequent *post hoc* revision of top-down precision or gain control, when prediction error cannot be resolved by predictions based upon the (invalid) cue.

In this chapter, I try to verify these explanations for electromagnetic responses to valid and invalid targets in the Posner paradigm using magnetoencephalography (MEG) and dynamic causal modelling of differences in effective connectivity. In particular, I hoped to establish that a sufficient explanation for responses evoked by valid and invalid targets would be provided by a difference in the gain or postsynaptic sensitivity of superficial parietal cells following a cue – and a subsequent top-down modulation of this gain from parietal and higher extrastriate sources. To do this, dynamic causal models based on canonical microcircuits that distinguish between superficial and deep pyramidal cells (Bastos et al., 2012) were used, which explicitly include a top-down modulation of superficial pyramidal cells.

In what follows, I provide a brief description of the dynamic causal models used to address precision or gain control in predictive coding; describe the data and experimental design; and report the results of Bayesian model comparisons that quantify the evidence for

condition-specific differences in superficial pyramidal cell gain. The focus here is on cue-dependent differences in gain prior to the onset of a visual target and subsequent top-down modulation of that gain during target processing. In particular, I ask whether cue-dependent differences in gain, top-down modulation or both were evident in evoked electromagnetic responses – and, whether any differences in gain were restricted to visual sources or extended to the parietal cortex.

2. Methods

2.1 Dynamic causal modelling of predictive coding

In the predictive coding models of inference in the brain described in the introduction (Mumford 1992; Friston 2005; Bastos et al 2012), prediction error ascends to update representations at higher hierarchical levels. Crucially, the excitability of cells reporting prediction error corresponds (mathematically) to the precision of – or confidence in – the information they convey. This precision has been used to explain the psychophysical and electrophysiological correlates of attention and can be regarded as the basis of selective (predictive or attentional) gain – in which sensory processing channels that convey precise information are enabled.

Neurobiological implementations of predictive coding use superficial pyramidal cells to report precision-weighted prediction error: $\xi^{(i)} = \Pi^{(i)} \cdot (\tilde{\mu}^{(i)} - f(\tilde{\mu}^{(i+1)}))$, where $\tilde{\mu}^{(i)}$ corresponds to representations (posterior expectations) of states of the world at level i in a cortical hierarchy and $f(\tilde{\mu}^{(i+1)})$ corresponds to the top-down predictions of these expectations – based upon expectations in the level above. The precision of the ensuing prediction error is modulated by the precision $\Pi^{(i)}$ to weight prediction errors in proportion to their (expected) reliability (c.f., known uncertainty). The encoding of precision – at each level of the hierarchy – can be associated with the strength of inhibitory recurrent connections.

With Dynamic Causal Modelling (Garrido et al., 2008, Bastos et al., 2012), this neurobiological implementation of predictive coding is mapped onto a neural mass model which is capable of simulating MEG data. The depolarisation of the three excitatory cell populations in the model – superficial and deep pyramidal cells, as well as spiny stellate cells, forms the output of the model with the main contribution coming from superficial

pyramidal cells. This activity is transformed by an MEG-specific lead-field which describes the translation from source activity to sensor perturbation.

The four-population neural mass model used here has been described in chapter 2 (Brown et al., 2012). In the neural mass models, $\gamma^{(i)}$, the negative log precision, corresponds to the strength of recurrent inhibitory connections on superficial pyramidal cells. This means that as precision increases, the strength of recurrent inhibition decreases. The strength of intrinsic (recurrent) self-inhibition (on superficial pyramidal cells) is therefore used as a proxy for log precision.

One new feature is introduced in this implementation of the neural mass model. To model top-down modulation of this self-inhibition the following form of (backward) modulatory connectivity is included:

$$\gamma^{(i)} = \gamma_0^{(i)} - 32 \cdot M \cdot (\sigma(V) - \sigma_0)$$

Here, γ_0 is self-inhibition when firing rates are at baseline levels $\sigma_0 = \sigma(0)$. Firing rates $\sigma(V) \in [0,1]$ are a sigmoid function of depolarisation V of afferent neuronal populations (deep pyramidal cells in other sources). The modulatory connection strength matrix M weights the influence of other sources; such that a high value suppresses self-inhibition and (effectively) increases the gain or precision of the superficial pyramidal cells that are targeted. In what follows, condition (valid or invalid) specific effects on γ will be modelled to evaluate the evidence for cue-dependent changes in gain at the onset of target processing and test for condition specific changes in M that mediate target-dependent changes in gain as target is processed. The hope was that evidence will be found for differences in baseline gain and subsequent top-down modulation – and that these would be expressed predominantly in early visual sources.

Specifically, I anticipated that intrinsic self-inhibition would be lower (gain would be higher) in left hemisphere sources after (invalid) cueing of the right hemifield relative to (valid) cueing of the left hemifield, where the target appeared in the left hemifield in both conditions. In other words, I hoped to show differential responses to identical targets could be explained by differences in gain induced by valid and invalid cues. Furthermore, I anticipated differences in descending modulatory effects between valid and invalid trials that would be necessary to reverse the laterality of gain control following an invalid target.

2.2 Participants

14 healthy right-handed subjects participated in the study (8 male; age 20-54). Ethical approval was obtained from the UCL Research Ethics Committee (no. 2715/001). Written informed consent was obtained from all subjects.

2.3 Experimental paradigm

All stimuli were presented using Matlab 7.1 and Cogent. Stimuli were projected onto a screen 70 cm from the subjects. During the task, subjects fixated on a central cross at all times. At the start of each trial, the cross was replaced by an arrow pointing to the bottom left or bottom right corner of the screen, or a double-headed arrow pointing to both (neutral trials). The cues subtended 1.6 degrees of visual angle. After a cue-target interval of 50, 100, 200 or 400ms, a target appeared either where the arrow had indicated (valid) or at the other side (invalid). The target was a white circle subtending 3.1 degrees of visual angle and presented in the lower left or lower right corners of the screen at 14.7 degrees eccentricity. Participants pressed a button with their right hand as soon as the target appeared. 66% of trials were valid, 17% were invalid and 17% uninformative (neutrally cued trials are not considered here). Left and right cues and targets were balanced. Catch trials, in which no target followed the cue, made up 10% of trials. 1800 trials were collected over three sessions on two consecutive days.

2.4 Behavioural data

Reaction times were collected by Cogent and analysed with IBM SPSS 20. A full factorial univariate ANOVA was performed with fixed factors 'side' 'validity' and 'cue-target interval' and random factor 'subject'.

2.5 Data collection and processing

MEG data was obtained using a whole-head 275-channel axial gradiometer MEG system (CTF Systems). The sampling rate was 600Hz and a low-pass filter of 150Hz was applied. Head position was monitored using three localisation coils, placed on the nasion and in front of each ear. An infrared eyetracker (Eyelink 1000) was used to monitor participants' fixation as well as to detect blinks. Stimuli were presented and behavioural data were collected with Cogent.

Data were analysed using SPM12b for EEG/MEG. Data were down-sampled to 200Hz and bandpass-filtered between 2Hz and 32Hz. Baseline-corrected epochs were extracted from the time series starting at 50ms before target onset and ending 400ms after target onset. Trials where the eyetracker detected a blink or saccade were excluded from analysis. Trials were then robustly averaged across cue-target intervals and participants to yield four conditions – left valid cue, right valid cue left invalid cue and right invalid cue. Averaging across participants can reduce the spatial precision of the MEG signal; however, as the hypotheses were not concerned with the spatial location of the signals, data were combined across all participants to increase the signal-to-noise ratio of the waveforms.

2.6 Data feature and source specification

The hypothesis was addressed using condition-specific grand average responses over all subjects. Intuitively, this is like treating each dataset as if it were from the same subject to produce an average ERP. To identify plausible sources a distributed source reconstruction was calculated (using four grand averages: valid right target, invalid right target, valid left target and invalid left target) based on multiple sparse priors (with default settings).
The grand average data were bandpass filtered between 2 and 32 Hz and windowed from 0 to 400 ms of peristimulus time. A lead field based upon the standard MRI template and a boundary element model as implemented in SPM12 (Mattout et al., 2005) was used. After source reconstruction, the power of evoked responses (over all frequencies and peristimulus time) was quantified to produce the maximum intensity projections in Figure 4.1. As one would expect, left targets activate right early visual sources and *vice versa*. Note further, that early visual source responses to valid left targets are greater than the same targets under invalid cues. On the basis of these reconstructions, eight sources were identified corresponding (roughly) to key maxima of source activity. These sources included bilateral early visual sources (V2); bilateral sources near the occipitotemporal-parietal junction (V5); bilateral dorsal (V3) extrastriate sources and bilateral superior parietal sources (PC). The anatomical designation of these sources should not be taken too seriously - they are used largely an aide-memoire for sources at various levels in the visual hierarchy, so that the functional anatomy can be discussed. Clearly, the spatial precision of source localisation does not allow us to associate each source with a specific cytoarchitectonic area – and even if this could be, there is sufficient intersubject variability in cortical architectures to make this association, at best, heuristic.

The distributed network constituting the DCM is shown in Figure 4.2. The parietal sources sent backward connections to the extrastriate (V3 and V5) sources that then sent

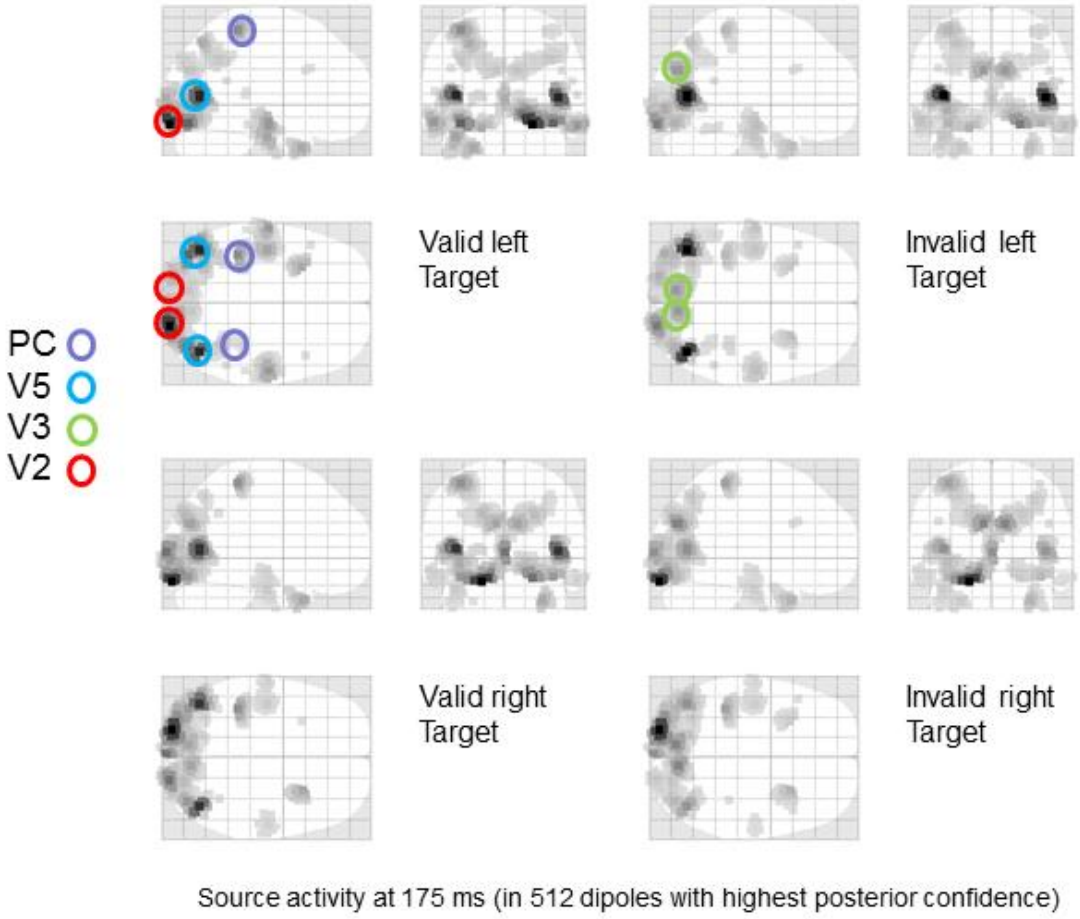

Figure 4.1: Source specification for dynamic causal modelling. A distributed source reconstruction was performed (Mattout et al., 2005) and the power of evoked responses was quantified over the time course of the trial and all frequencies to yield the maximum intensity projections shown. Eight sources corresponding (roughly) to key maxima of source activity were identified: included bilateral early visual sources (V2); bilateral sources near the occipitotemporal-parietal junction (V5); bilateral dorsal (V3) extrastriate sources and bilateral superior parietal sources (PC).

backward connections to the V2 sources. These connections were reciprocated by extrinsic forward connections to produce a simple visual hierarchy with bilateral connections.

2.7 Model space and Bayesian model comparison

The DCM analyses used data from 0 to 400 ms of peristimulus time. To de-noise the data and improve computational efficiency, the first eight canonical modes of the scalp data, given the source locations, were fitted – these can be regarded as the principal components of the data that can be explained by source activity. The sources were modelled as small cortical patches of about 16 mm radius – centred on the source locations in Figure 4.1 – as described in (Daunizeau et al 2006). The vertices of these sources used the same lead fields as in the source reconstruction.

Exogenous (visual target related) input was modelled as a Gaussian function with a prior peak at 120 ms (and a prior standard deviation of 16 ms). This input was delivered to V2 on the appropriate side (left for right target trials and right for left target trials). The ensuing models were optimised to explain sensor responses by adjusting their (neuronal and lead field) parameters in the usual way – this is known as model inversion or fitting. The products of this inversion are posterior estimates of (differences in) intrinsic and extrinsic connectivity and the evidence or marginal likelihood for each model considered.

The hypothesis centred on the gain of superficial pyramidal cells. I therefore estimated a full model in which all intrinsic gains and their extrinsic (backward) modulation could differ between valid and invalid trials. To ensure the same stimuli were used for assessing these differences two sets of analyses were conducted – one for targets presented to the left visual field and another for targets presented on the right. Each DCM estimated all intrinsic, extrinsic and modulatory connection strengths and any differences in intrinsic and modulatory connections due to invalid cuing.

We performed an initial Bayesian model comparison to establish whether validity dependent differences in top-down connections were expressed in their driving effects, modulatory effects or both. To evaluate the ensuing models, we use Bayesian model comparison based upon (a variational free energy) approximation to log evidence (Penny et al 2004; Friston and Penny 2011). The differences in log evidence among the models (shown in Figure 4.3) suggest that we can be nearly 100% confident that validity dependent changes are mediated by top-down modulatory effects (with no changes in top-

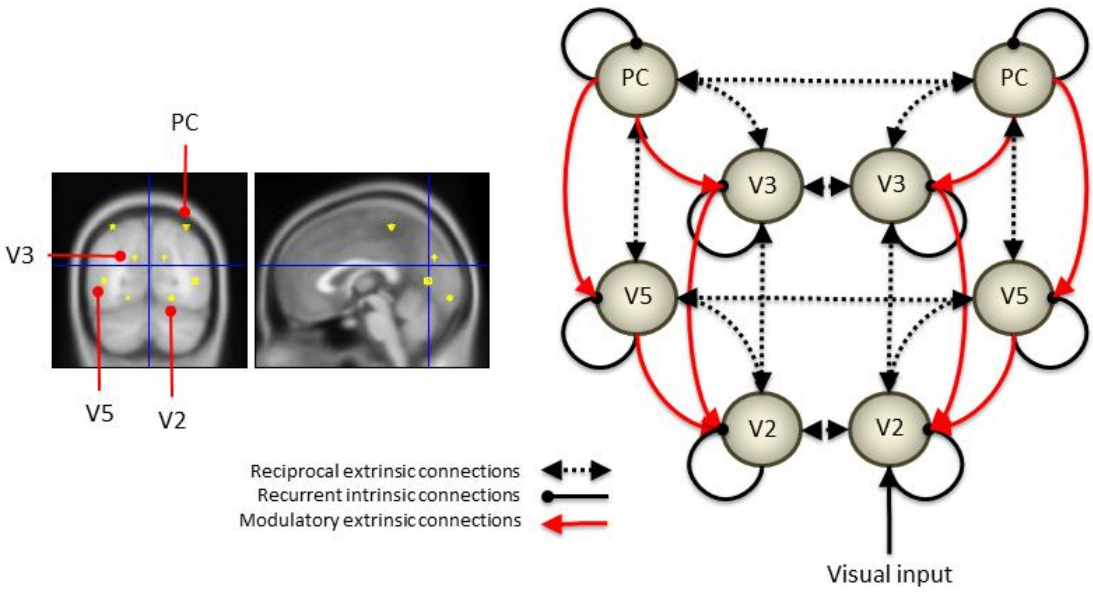

Figure 4.2: DCM network. The location of the eight sources is shown in the panels on the left. To construct the DCM, these sources were connected in the distributed network shown on the right. The parietal sources sent both driving and modulatory backward connections to the extrastriate (V3 and V5) sources that then sent backward connections to the V2 sources. These connections were reciprocated by extrinsic forward connections to produce a simple visual hierarchy with bilateral connections.

down driving effects). We therefore selected the model – in which all top-down modulatory connections could differ between conditions – for further examination.

After inverting the full model, the evidence for reduced versions that constitute alternative hypothesises or models was then evaluated. This model space was created by partitioning connectivity differences into three subsets and considering all eight combinations. These subsets were changes in intrinsic gain in the extrastriate sources (V2, V3 and V5); changes in parietal (PC) gain and changes in extrinsic modulatory connections. This partition was motivated by distinguishing between the effect of the cue on target-related responses – which should be apparent in changes in intrinsic gain in the visual areas – and the effect of the target *per se* – which should be apparent in changes in backward modulation of gain. To evaluate the ensuing models, Bayesian model comparison was used, based upon a (variational free energy) approximation to log evidence. Having identified the model with the greatest evidence, its posterior parameter estimates were examined. This allowed validity effects to be quantitatively characterised and interpreted in computational (predictive coding) terms.

3. Results

3.1 Behavioural data

The ANOVA demonstrated significant main effects of validity, subject and cue-target interval, with significant interactions between cue-target interval*validity, cue-target interval*subject, side*cue-target interval and validity*side*subject. Reaction times to validly cued targets were significantly shorter than to invalidly cued targets (left: mean (SD) 333ms (42ms) vs. 355ms (44ms), $p<0.001$; right: mean (SD) 334ms (42ms) vs. 354ms (44ms)), figure 4.4.

The effects of attention (validity of cueing) on responses to targets presented in the left hemifield are shown – for the first two canonical modes – in Figure 4.5. Although these MEG responses are formally distinct from classic EEG results, they speak to similar effects on early and late responses: the blue lines correspond to valid trials and red lines to invalid trials. The response in the first mode shows the early response (just before 200 ms) has a reduced latency and slightly higher amplitude – consistent with an attenuation of N2 response to invalid targets, as seen in classic EEG studies (Mangun& Hilliard, 1991). In terms of late responses, the second mode shows a protracted and elevated response

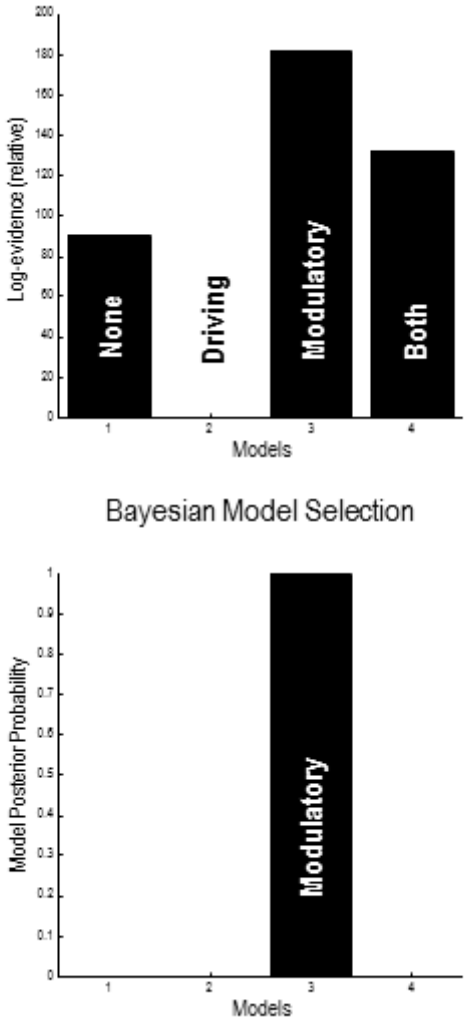

Figure 4.3: Results of provisional Bayesian model selection. The (free energy approximation) to log evidence was assessed for models with and without validity–dependent differences in top-down driving and modulatory connections. The log evidences (upper panel) show that the model with differences in modulatory connections has the greatest posterior probability (lower panel). The log evidences are shown relative to the evidence for a null model with no changes in either driving or modulatory backward connections.

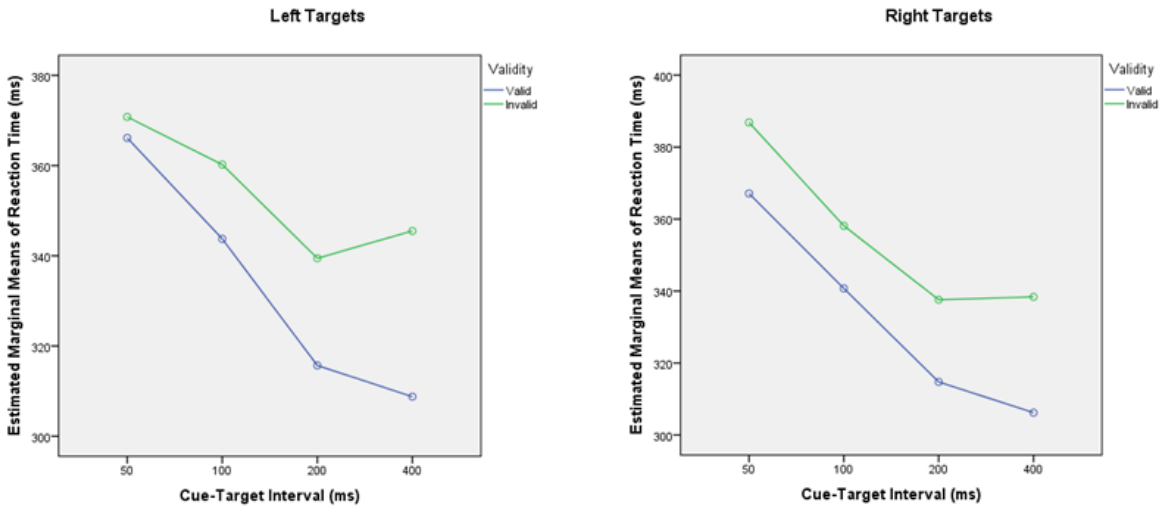

Figure 4.4: Behavioural results. Reaction times to validly and invalidly cued targets at different cue-target intervals for targets appearing on the left (left panel) and right (right panel), averaged across all participants. Reaction times were faster for validly than invalidly cued targets ($p<0.0001$). Reaction times decreased as cue-target interval increase (all $p<0.05$).

around 300 ms that is consistent with a P3b component, when the target location is not attended.

The solid lines report the model predictions of observed responses (broken lines) in sensor space after inversion of the DCM. These illustrate the accuracy of model inversion, capturing both the early and late differences to a considerable level of detail. Examples of the underlying source activity that generates these predictions are shown in the lower panel. These traces represent the depolarisation of three excitatory populations within the left V2 source, contralateral to the visual input modelling the effects of target presentation. The dotted lines correspond to the spiny stellate and deep pyramidal populations, while the solid lines report the superficial pyramidal cells – that are the predominant contributors to sensor data. Note that this level of reconstructed neurophysiological detail rests on having a biologically plausible forward model.

The differential responses to right targets were much less marked (results not shown). Furthermore, model inversion failed to converge for these conditions. Therefore, this analysis is restricted to the left target conditions. The failure to elicit clear validity effects with right targets may relate to the asymmetry of responses – and attentional gain control (see below).

3.2 Bayesian model selection

The comparison of different explanations for the validity effects above focused on differences in the gain of superficial pyramidal cells – either intrinsic to extrastriate or parietal sources, or differences in the modulation of gain, mediated by extrinsic top-down connections. The relative log evidences for all combinations of these condition-specific differences are shown in the upper left panel of Figure 4.6. The labelling of these models indicates the presence or absence of differences in extrastriate gain, parietal gain and gain modulation. It can be seen that the model with the greatest evidence includes differences in extrastriate gain and gain modulation – but not differences in parietal gain. The corresponding posterior probabilities of these models (assuming all were equally plausible *a priori*) are shown in the upper right panel. These suggest that differences in parietal gain cannot be definitively excluded; however, confidence exceeds 80% that parietal effects are not necessary to explain these data, provided validity effects on extrastriate gain and its top-down modulation are allowed for.

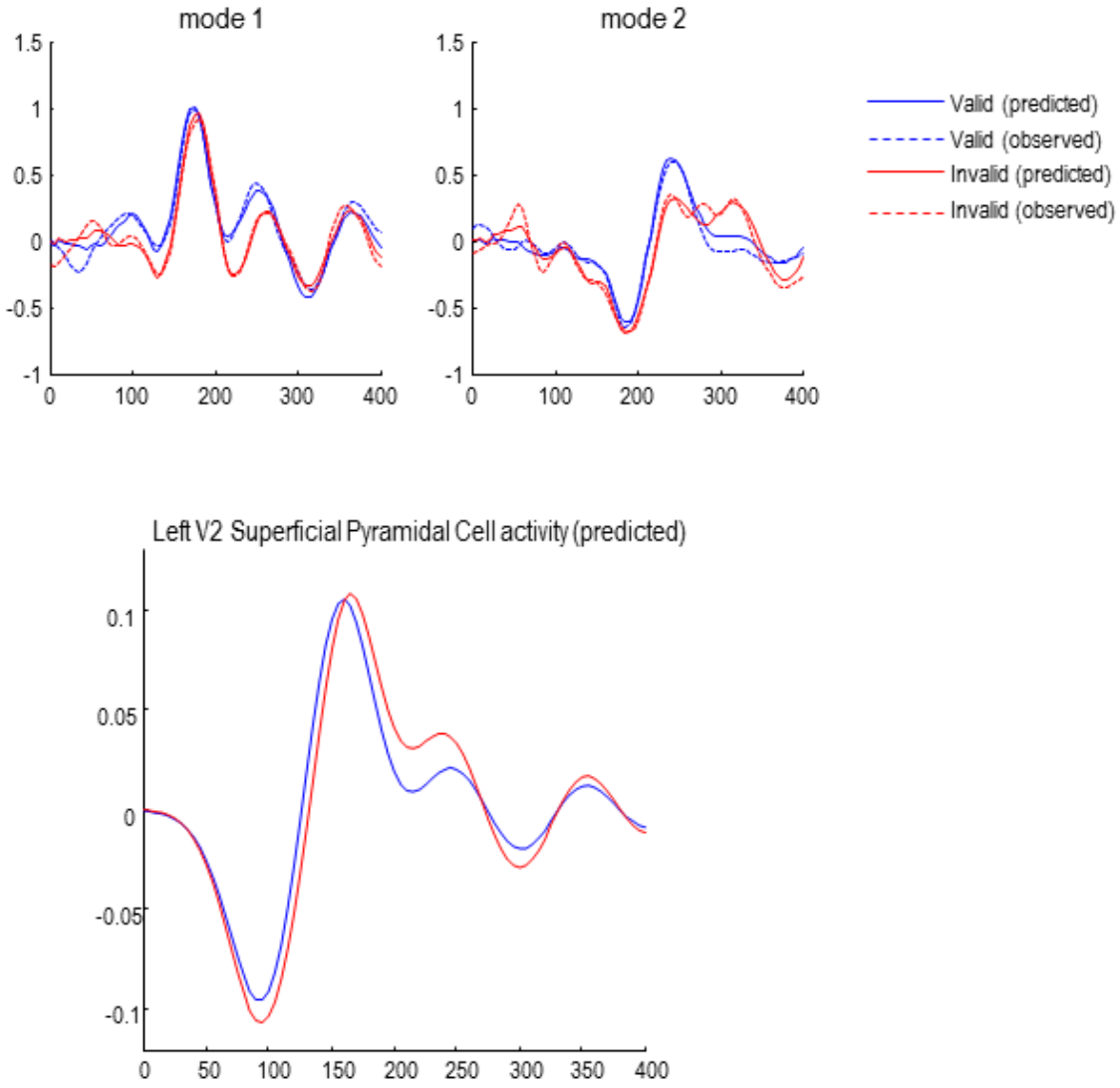

Figure 4.5: **Empirical and predicted responses. Upper panel:** the first two of eight spatial modes (principle components) of the data to which the DCMs were fitted. Observed responses are dashed lines; solid lines show the responses fitted by the winning model (see below), demonstrating a good model fit. **Lower panel:** reconstructed source activity in left V2.

The lower panels show the same log evidences but in image format, to illustrate the relative evidence for gain effects. The image on the right is under extrinsic top-down gain modulation and suggests greater evidence than the corresponding results on the left, where modulatory effects are concluded. In both cases, the model with extrastriate– but

not parietal – gain differences has the greatest evidence. Having identified the best model, the changes in model parameters that explain the validity effect were then quantified.

3.3 Attentional gain effects

Figure 4.7 shows the differences in self-inhibition (top left panels) and backwards modulation of self-inhibition (top right panels) for the model with the highest posterior probability above. The upper panels show the differences as connectivity matrices indicating changes in connection strength. This means that differences in self-inhibition are located along the leading diagonal, while differences in backward connections are restricted to the upper diagonal elements. The middle panels show the same results but in terms of the posterior expectations for differences (in connections that changed) and their 90% confidence intervals.

As anticipated, the recurrent or self-inhibition of early visual sources showed a highly asymmetrical difference when attending to the right hemifield (during invalid trials), compared to attending to the left hemifield (during valid trials). When attending to the right hemifield the left V2 source shows a profound decrease in the self-inhibition of superficial pyramidal cells – consistent with a disinhibition or increase in gain. This is accompanied by a slight decrease in the gain or sensitivity of the left extrastriate V3 source and an increase in the right V5 source. Note that these gain differences are in place before the target is presented and – presumably – are instantiated by the cue. The lower left and right panels in figure 4.7 show the dynamic change in superficial pyramidal cell gain during the trial. When the target arrives, it evokes responses throughout the visual hierarchy that modulate the gain of the lower sources. These effects are mediated by the backward modulatory connections.

With the exception of backward connections from the right parietal source, all the differences in backward modulation between valid and invalid trials are positive, speaking to an increase in gain (or a top-down disinhibition of superficial pyramidal populations). However, it is difficult to predict the changes in gain that are produced by modulatory effects, because this disinhibition could itself be inhibited when top-down afference falls below baseline firing rates. Therefore, the changes in gain in early visual sources as a function of peristimulus time were evaluated for the two conditions. This is possible because of the biologically plausible forward or generative model that allows examination of changes in both neuronal states and connectivity – over peristimulus time – using the posterior parameter estimates.

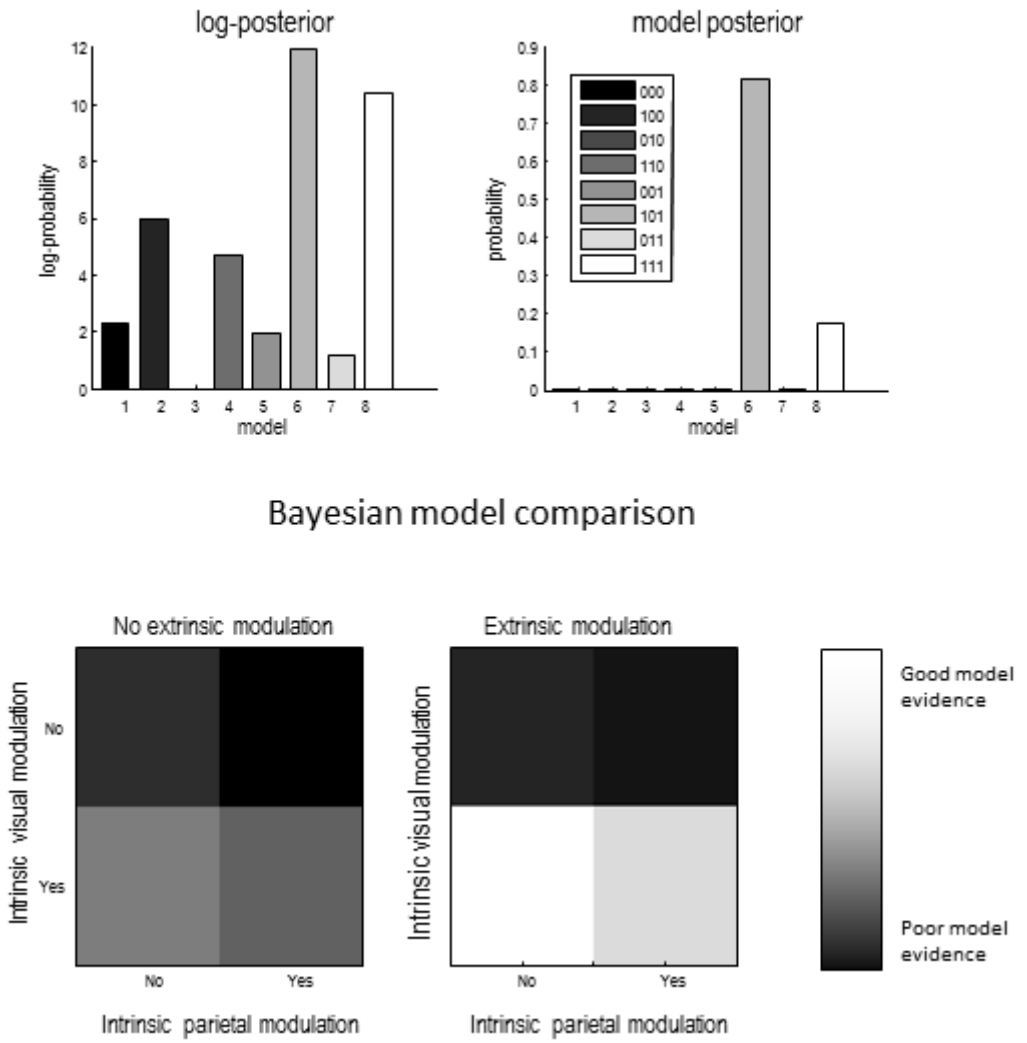

Figure 4.6: Upper left panel: relative log evidence for models which fitted differences between conditions through changes in one of three sets of parameters: superficial pyramidal cell gain in visual areas (1 _ _), superficial pyramidal cell gain in parietal areas (_ 1 _) and strength of backwards modulatory connections (_ _ 1). **Upper right panel:** The winning model had changes in superficial pyramidal cell gain in visual areas and in the strength of backwards modulatory connections, meaning that we can be more than 80% certain that backwards modulatory connections are not necessary explain the electrophysiological signatures of the validity effect. Lower panels show the same data as in the top left panel, but in image format.

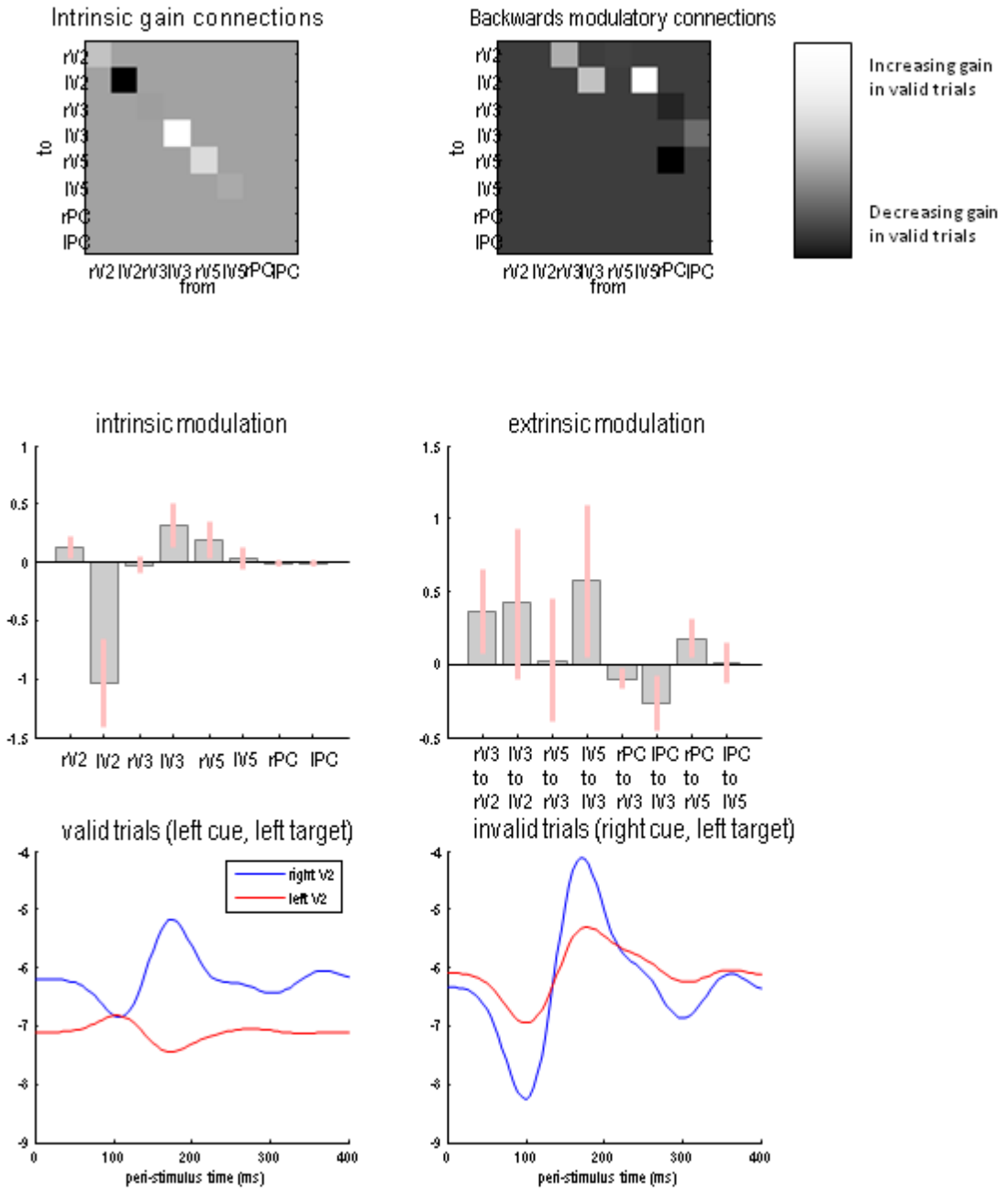

Figure 4.7: Fitted parameter values. Differences in self-inhibition (upper left panels) and backwards modulation of self-inhibition (upper right panels) between valid and invalid trials for the model with the highest posterior probability above. The lower panels show the gain of the superficial pyramidal cells over time in valid and invalid trials.

Figure 4.7 shows the log gain or precision of the early visual sources, following target presentation for valid (lower left panel) and invalid trials (lower right panel). As expected, there is a marked asymmetry in gain modulation during the prestimulus period that is revised or updated after the target is processed – through activity dependent modulatory mechanisms. Specifically, during valid trials the gain is greater in the appropriate (right) early visual source and then reaches a peak shortly before 200 ms. This peak is complemented by a suppression of gain in the unattended (left) visual source. This can be contrasted with the gain modulation during invalid trials. Here, the attended left source starts off with a slightly higher gain. Furthermore, the unattended source is suppressed more acutely with the arrival of the target. However, after about 120 ms its gain increases markedly, to peak just before 200 ms. This redeployment of precision (c.f., reorientation of attention) is the largest gain modulation in both sources and conditions. Interestingly, the gain of the left source also enjoys a slight increase but to a substantially lesser degree. In short, the top-down modulation of gain (through modulatory disinhibition of superficial pyramidal cells) appears to exert a dynamic gain control over peristimulus time and shows marked lateralisation, when attention is switched from one hemifield to another.

4. Discussion

In conclusion, I have used dynamic causal modelling to characterise putative changes in the gain of superficial pyramidal cell populations that might underlie attentional (validity) effects in the Posner paradigm. The focus on gain mechanisms was motivated by theoretical formulations of attention in terms of optimising perceptual inference using the expected precision of particular processing streams, as discussed in chapter 3 (Feldman & Friston, 2010). This formulation rests upon predictive coding schemes that the brain might use to infer the causes of sensory consequences it has to explain (Friston & Kiebel, 2009). The model comparison and quantitative analysis of changes in parameter estimates are remarkably consistent with theoretical predictions.

In brief, the modelling results suggest that, following a cue, sensory channels in the appropriate hemisphere are afforded more precision through the disinhibition of recurrent or self-inhibition of superficial pyramidal cells. These cells are thought to pass sensory information (prediction error) to higher levels to inform perception. When a target appears in an unattended location, the misplaced gain or sensitivity of lower areas is revised or updated by top-down modulatory influences from higher extrastriate and parietal sources. Phenomenologically, this increases the latency and reduces the

amplitude of early responses to invalid targets – because they are processed by channels that have an inappropriately low gain. The resulting prediction error induces an update response that reverses the misattribution of gain, producing differences in late or endogenous response components – such as the P3b. The P3b is known to be sensitive to probabilistic surprise (Mars et al., 2008; Kolossa et al., 2012) as well as to risk (Schuermann et al., 2012). These results suggest that the larger P300 in response to more unexpected events might be a result of exaggerated precision at lower levels incited by the arrival of an unexpected stimulus.

This application of dynamic causal modelling is slightly more focused than normal applications. A large model space was not explored; instead, I focused on particular synaptic mechanisms as sufficient explanations for condition-specific responses. It is more than likely that there are many models of these differential responses that would produce equally good or better explanations. However, I chose to focus on models that were explicitly informed or constrained by computational and biophysical considerations; namely, that the effects have to be mediated by a neurobiologically plausible gain control that is consistent with normative principles of perceptual inference. This allowed the validation of the theoretical proposals empirically, while providing a principled model space within which to test specific hypotheses about the underlying wetware.

Evidence suggests that gain modulation in pyramidal cells is an important mechanism in visual attention. Electrophysiological studies have demonstrated that attention can enhance the response of visual neurons (likely to be pyramidal cells) by a multiplicative factor (McAdams & Maunsell, 1999; Treue & Martinez-Trujillo, 1999). fMRI studies demonstrate increased BOLD response for attended versus unattended stimuli (Kastner et al., 1998), even if these stimuli are predictable (Kok et al., 2012), and early visual ERPs, which are most strongly determined by pyramidal cell firing, are enhanced by attention (Rauss et al., 2011).

Interestingly, although the relative simplicity of neuronal mass models used by dynamic causal modelling almost forces gain control to be modelled using inhibitory self-connections, this particular mechanism makes a lot of sense in relation to current thinking about attention. Convergent evidence implicates local inhibitory processing, mediated by GABAergic neurotransmission, in attention. Drugs working at GABA receptors, such as benzodiazepines, which are positive allosteric modulators of GABA-A receptors, increase the behavioural effect of cues so that reaction time differences to validly and invalidly cued targets become larger, while overall reaction times are slowed (Johnson et al., 1995).

Nicotine (an agonist at nicotinic acetylcholine receptors) also affects reaction times in the Posner paradigm, but it decreases the validity effect while increasing reaction times (Thiel et al., 2005; Meinke et al., 2006), and it is believed that the attentional effects of acetylcholine might be mediated at least partly though depression of inhibitory interneuron activity (Xiang et al., 1998; Buia and Tiesinga, 2006). These contrasting effects suggest that the inhibitory interneurons set the gain of their cortical area to determine reaction times. Increasing their effects increases reaction times due to greater overall inhibition, exaggerating the difference between high- and low-gain cortical areas, and vice versa. This is consistent with the 'biased activation theory' of selective attention (Grabenhorst and Rolls, 2010), which suggests that GABA interneurons mediate competition between stimuli which can be biased through top-down signals (the backwards modulatory connections in this DCM).

In summary, the emerging picture is that attention may be mediated through local intrinsic or recurrent inhibitory mechanisms that form a key part of cortical gain control – and that have characteristic signatures in terms of frequency specific induced responses. This fits comfortably with the theoretical perspective provided by predictive coding – that provides a computational role for recurrent inhibition in encoding the gain or precision of prediction errors in hierarchical processing. The results presented in this chapter provide an initial link between these computational imperatives and plausible mechanisms at the level of synaptic processing and hierarchical neuronal circuits.

The work described in this chapter is published as:

Brown HR, Friston KJ. (2013). The functional anatomy of attention: a DCM study. Front. Hum. Neurosci. 2013 7:784.

Chapter 5

Active inference, attention and motor preparation

1. Introduction

During the preparation and execution of goal-directed movements, processing is biased towards the perceptual attributes of the goal (e.g., Gherri and Eimer 2010; Perfetti et al. 2010; Baldauf and Deubel 2010; Humphreys et al. 2010) and preparation or execution of an action improves perceptual processing in relevant sensory domains (Fagioli et al. 2007). This suggests motor planning and attention are inherently linked, such that "perceptual codes and action plans share a common representational medium, which presumably involves the human premotor cortex" (Fagioli et al. 2007). This relates to the concept of *motor attention* that is specific to the effectors employed (Rushworth et al., 2001) and decision making through attentional selection among motor plans (Goldberg and Segraves 1987). Moreover, the *premotor theory of visual attention* (Rizzolatti et al. 1994) proposes that distinct maps are tuned to different effector representations and become active when a movement is prepared. In short, attention has a fundamental role in the selection and control of action; see Allport (1987) for a review.

The link between action and attention and was first proposed by James (1890) and Woodworth (1899): however, the cognitive and neural mechanisms responsible for this association remain largely unknown (Dalrymple and Kingstone 2010). Greenwald (1970b) provided evidence that attention to a particular sensory modality speeded movements that are detected in that modality: In the oculomotor system, visual

discrimination performance is enhanced at the target location of a prepared saccade (Deubel & Schneider, 1996). Furthermore, stimulation of the superior colliculus can produce both eye movements (Robinson, 1972) and shifts of attention (Müller et al., 2004). Conversely, Craighero et al. (1999) showed that reaction times to visually presented objects are reduced when subjects grasp the objects being presented, illustrating the motor facilitation of sensory processing.

This chapter explores the idea that motor attention uses exactly the same synaptic mechanisms as visual attention. This may sound strange because motor commands are usually considered to be outputs, whereas the visual channels selected by attention are inputs. However, active inference regards action as being driven by proprioceptive prediction errors in exactly the same way that perception is driven by exteroceptive prediction errors, as described in the introduction (Friston et al., 2010). If true, this means that attentional modulation may operate at low levels in the motor system in the same way that it operates in the early visual system. In this chapter, evidence for attentional modulation within the motor system was sought by reproducing a classical visual attention paradigm (Posner 1980) in the motor domain. Furthermore, by cueing attention to different attributes of movements I tried to locate the putative attentional modulation within the motor hierarchy. I hoped to show that attentional effects were expressed in low levels (in an intrinsic frame of reference) in much the same way that directed spatial attention operates in the early visual pathways. This chapter comprises four sections. The first rehearses the theoretical background that motivated a reaction time study described in the second section. The third section presents the results, which are discussed in relation to theoretical considerations in the final section.

1.1 Active inference and motor attention

In this section, motor preparation is considered as attention that is directed towards predicted *proprioceptive* sensations (Galazky et al. 2009), as opposed to the predicted *exteroceptive* consequences of action. This idea is motivated by the work described in chapter 3 (Feldman and Friston 2010). In this model, the effects of orienting cues on reaction times were explained by the Bayes-optimal encoding of precision in a hierarchical message-passing scheme (predictive coding). In this context, precision is the inverse variance or uncertainty associated with particular sensory channels, such that attention can be understood as weighting sensory signals in proportion to their precision (Feldman and Friston 2010; Friston 2010). In these predictive coding schemes, precision is encoded by the gain of units reporting bottom-up sensory information that has yet to be explained

by top-down predictions. This sensory information is called *prediction error* and is generally associated with the activity of superficial pyramidal cells: these cells are the source of forward or bottom-up projections in the brain (Rockland & Pandya, 1979; Mumford 1992; Friston 2010). In these schemes, attention therefore reduces to the optimisation of the postsynaptic gain of superficial pyramidal cells, of the sort associated with gamma-synchronisation (e.g. Womelsdorf et al. 2006) and monoaminergic or cholinergic modulation (e.g. Herrero et al. 2008); both of which have been implicated in attention. Here, I pursue the notion that attention is the optimum weighting of prediction error in the context of action preparation (Mars et al. 2007; Bestmann et al. 2008). In short, attention is considered to boost the gain of proprioceptive channels during motor preparation, in the same way that attention selects particular visual channels when subjects prepare for a visual target. In what follows, predictive coding and active inference are briefly reviewed in the context of this work; for a more complete exposition see the introduction and appendix:

1.2 Predictive coding and active inference

Predictive coding is based on the assumption that the brain makes inferences about the causes of its sensations. These inferences are driven by bottom-up or forward sensory information that is passed to higher brain areas in the form of prediction errors (Rao and Ballard 1999; Friston et al. 2008). Top-down or backward connections convey predictions that try to suppress prediction errors until predictions are optimised and prediction error is minimised. This suppression rests on opposing excitatory and inhibitory effects of top-down predictions and bottom-up inputs on prediction error units (usually considered to be superficial pyramidal cells: Mumford, 1992). Active inference (Friston et al. 2010) generalises this scheme and proposes that exactly the same recursive message passing operates in the motor system. The only difference is that prediction errors at the lowest level (in the cranial nerve nuclei and spinal cord) are also suppressed by movement, through classical reflex arcs. In this view, descending (cortico-spinal) signals are not motor commands *per se* but predictions of proprioceptive signals that the peripheral motor system fulfils (see Friston 2009, 2010 for details). As illustrated in Figure 5.1, a cued movement is not regarded as a simple stimulus-response mapping but is generated by a high level (sensorimotor) percept that predicts a particular pattern of proprioceptive and exteroceptive sensory signals. This percept arises to explain prediction errors caused by a cue in the exteroceptive domain, while motor reflexes suppress the ensuing prediction errors in the proprioceptive domain. This framework has been used to explain several features of the motor system and a series of behaviours, from visual tracking

(Friston et al 2009) to action observation (Friston et al 2010). Active inference formalises much of what is proposed by the ideomotor theory of action (Lotze, 1852; James, 1890). The ideomotor account of motor control posits that moving causes a bidirectional association to be formed between a movement and its perceptual consequences. Learning this association allows the perceptual consequences of a movement to be predicted, and anticipating the sensory consequences of a movement can be used to select an action. At the level of the stretch receptors, the similarity is clear: signalling the predicted sensory consequences of an action (under active inference) causes the action to occur. At higher hierarchical levels, movements can still be initiated in order to change the sensory input in another sensory system; indeed the free-energy principle demands the sampling of predicted information to minimise free energy or, more simply, surprise. See Figure 5.1 for a schematic illustration.

1.3 Attention and active inference

Attention enters this picture through context or state-dependent optimization of the precision of prediction errors. This sort of prediction is about the second-order statistics of sensory signals (i.e., their variability or reliability). In predictive coding, top-down first-order predictions drive (or inhibit) neurons reporting prediction errors; while contextual, second-order predictions optimise their postsynaptic gain. It is this sort of top-down effect that is associated with attention. Neurobiologically, the distinction between first and second order predictions can be related to the distinction between the driving and modulatory effects mediated by AMPA and NMDA receptors. Optimising postsynaptic gain ensures that sensory information (prediction error) is weighted in proportion to its precision. This may sound complicated but is exactly the same procedure used every day in statistics, when weighting a difference in means (prediction error under the null hypothesis) by standard error (inverse precision) to form a *t*-statistic. Precision can thus be regarded as representing the reliability, ambiguity, or uncertainty about sensory signals. In summary, top-down predictions can have a direct (first-order) or a modulatory (second-order) effect on the responses of prediction error units that make the ensuing predictions as efficient as possible. Reaction time (Goodman & Kelso, 1980), corticospinal excitability (Bestmann et al., 2008; Mars et al., 2007) and EEG data (Osman et al., 1995; Mars et al., 2008) all confirm that the motor system is highly sensitive to such second-order effects.

If ascending sensory signals are prediction errors and descending motor commands are predictions, then optimal predictions (and the resulting movements) should depend on

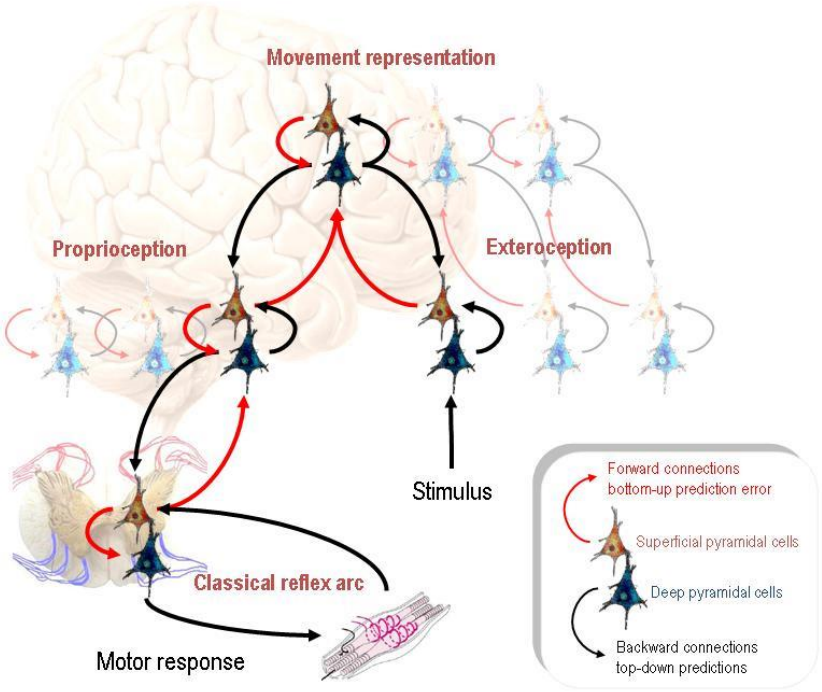

Figure 5.1: Active inference and predictive coding. Active inference is a generalisation of predictive coding that covers motor behaviours and itself is a special instance of the principle of free-energy minimisation. Free-energy is a statistical quantity that bounds the surprise (self-information) associated with sensory signals. This surprise is quantified in relation to a generative model of how those signals were caused. Predictive coding uses prediction error as a proxy for free-energy (cf, surprise) and rests on a hierarchical model, in which prediction errors are passed up the hierarchy (red arrows) to optimise high-level representations that provide top-down predictions (black arrows). In this schematic, prediction-error units are portrayed in red and units encoding the conditional expectations of the hidden causes of sensory input are shown in blue. During perception, the best explanation for sensory input emerges when the top-down predictions can explain as much of the prediction error (at each hierarchical level) as possible. Active inference takes this one step further and notes that certain sensory modalities can use prediction errors to drive motoneurons to eliminate prediction error directly (through classical motor reflex arcs). This is shown schematically on the lower left, using units in the dorsal and ventral horns of the spinal cord. Under active inference, a movement just fulfils the predictions afforded by percepts that predict both exteroceptive (e.g. visual) and interoceptive (e.g., stretch receptor) consequences. This high-level (sensorimotor) percept is activated by an exteroceptive (sensory) cue and the ensuing top-down predictions propagate to both sensory cortex (to suppress exteroceptive prediction error) and the motor system. However, in the motor system, the predictions engender a proprioceptive prediction error that is eliminated by movement. In this schematic, it is assumed that prediction errors are reported by superficial pyramidal cells (Mumford 1992), while conditional representations are encoded by (top-down) projecting deep pyramidal cells. Darker units highlight those activated by the presentation of a target-stimulus.

optimizing precision in exactly the same way as in sensory processing. This suggests that, in the motor domain, cueing has a similar effect to that observed in the sensory domain: Rosenbaum (1980) first demonstrated an effect of movement cueing on reaction time in a way that is analogous to the accelerated detection of visual targets when they are preceded by valid cues in the Posner paradigm (Posner 1980). However the movements cued in Rosenbaum (1980) were button presses, which required either visual or somatosensory attention to guide movement to the target. Thus, these non-proprioceptive aspects of button presses conflate attentional effects in visual, somatosensory and proprioceptive domains. In other words, in previous work movements were planned in relation to an object in extra-personal space. Here, I used a simpler paradigm in which movements (wrist flexion and extension) could be performed using only proprioceptive information. This ensured that any attentional effects could be attributed to proprioception. This motor analogue of the Posner paradigm therefore allows interpretation of the results in relation to visual attention as modelled in chapter 3 (Feldman and Friston, 2010); and to illustrate how active inference provides a framework in which to address questions about the functional anatomy of action preparation and attention.

1.4 Cueing in an extrinsic or intrinsic frame of reference?

In this chapter, 'intrinsic' will refer to a muscle-based, effector-dependent co-ordinate system and 'extrinsic' will refer to an external, effector-independent co-ordinate system. A key question in the functional anatomy of motor attention is where biasing effects are located in the cortical hierarchy: see Grafton and Hamilton (2007) for a review of motor hierarchies. In the sensory domain, attention is usually considered to operate at the lower levels of sensory hierarchies to select among competing sensory processing channels. This is seen in both psychological (e.g., the distinction between object and spatial visual attention: Treisman, 1998; Macaluso et al 2003) and electrophysiological treatments (e.g., biased competition models: Desimone and Duncan 1995). If the functional anatomy of the motor hierarchy recapitulates that of sensory hierarchies, then one might expect to see attentional modulation in lower levels, which will be associated with representations in an intrinsic frame of reference.

Electrophysiological evidence demonstrates that between the ventral premotor cortex and M1 neurons change their response patterns from signalling movements in a visual (*extrinsic*) coordinate system that is independent of starting posture to a motor (*intrinsic*) coordinate system that depends on starting posture (Kakei et al. 1999, 2001, 2003). Thus

in ventral premotor cortex, actions are largely encoded allocentrically, while in M1 they are predominantly encoded in terms of the joint angles and proprioceptive input required to reach the target (Soechting & Flanders 1992). Shipp (2005) suggests that neurons representing movements in an intrinsic frame of reference send descending cortico-spinal predictions from M1. Kakei et al (2003) provide a detailed discussion of movement representations in terms of the coordinate transformations that begin with an "extrinsic coordinate frame representing the spatial location of a target and end with an intrinsic coordinate frame describing muscle activation patterns". It should be noted however, that the segregation of intrinsic and extrinsic representations between motor and premotor cortex may not be complete or unique (Wu and Hatsopoulos 2007)

These observations suggest two possible levels of the motor hierarchy at which attentional cueing effects could operate. Consider movements with two dimensions or attributes that are cued in an extrinsic frame of reference; for example, moving the left or right hand (*where*) inwards or outwards (*what*). If attention operates at high levels of the motor hierarchy, then one might expect cues to move the hand inward will facilitate inward movements, irrespective of which hand is used. This is because the representation of the movement can be primed in extrinsic coordinates, prior to transformation to intrinsic coordinates. Conversely, if attention operates at lower levels, encoding the muscle groups involved in inward movements of the left hand, then attentional priming will only be expressed when the left hand is moved inwards. In short, if attention operates on prediction errors in an intrinsic frame of reference, the effect of the *what* cue will depend upon the *where* cue.

In summary, if sensorimotor constructs mediate attentional biases in an extrinsic frame of reference, cueing effects would be expected on both dimensions independently. Conversely, if these representations instantiate top-down biases at a lower (intrinsic) level of the motor system, then only a particular movement (in an intrinsic frame of reference) will be cued. Figure 5.2 tries to make the different predictions clear in terms of top-down enabling of postsynaptic gain (indicated with blue arrows). Crucially, the profile of speeded responses (under valid and invalid cueing) is different for extrinsic and intrinsic levels of attentional gain. In the intrinsic (motor cortex) model, there should be an interaction between the validity effects of cues over both movement dimensions. Conversely, under the extrinsic (premotor cortex) model, there should be no interaction but two main effects due to the validity of both *what* and *where* aspects of the cue. It was this difference in the profile of validity effects on reaction times that this experiment was designed to reveal.

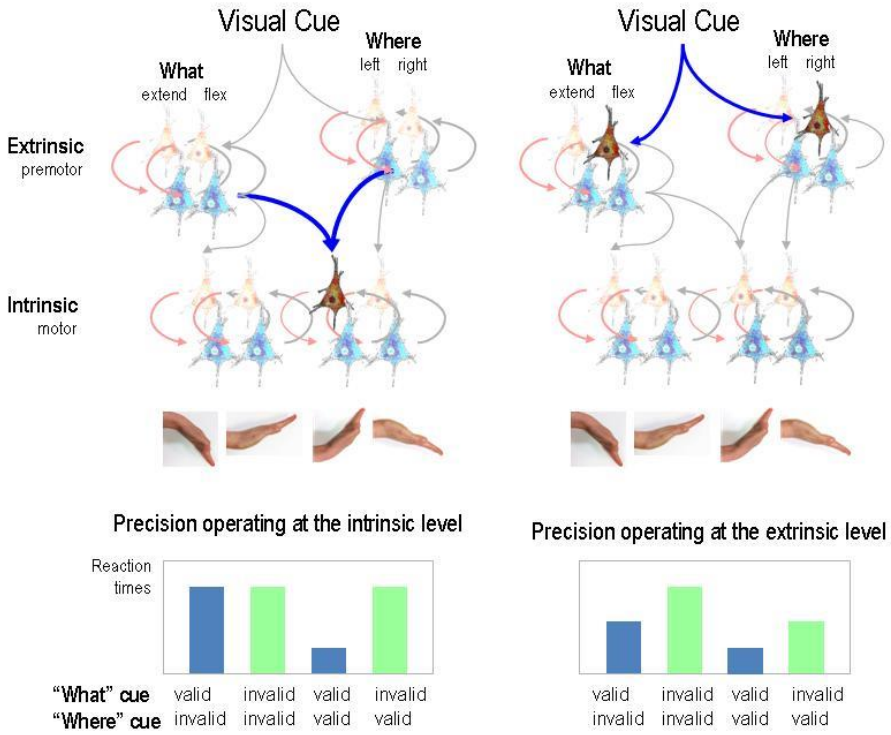

Figure 5.2: Different levels of attentional bias. This schematic illustrates the top-down enabling of postsynaptic gain (blue arrows) at different levels in the motor hierarchy. In the left panel, the predictions of an inward (flexion) movement of the left hand selectively bias the intrinsic prediction-error units that elicit inward movements of the left hand. This means that when a valid target-stimulus appears, these prediction errors will produce a more efficient and speeded movement (be eliciting stronger descending predictions). Conversely, if the attentional bias is mediated at the premotor (extrinsic) level, the prediction errors associated with both *what* and *where* aspects of the movement will facilitate speed responses over both movement dimensions; e.g., all left-hand movements and all inward movements. In this figure, darker units highlight prediction-error units with increased gain. The lower graphs show the predicted profile of reaction times (under valid and invalid cueing) for cueing at extrinsic (right) and intrinsic (left) levels. In the intrinsic (motor cortex) model, there should be an interaction between the validity effects of cues over both movement dimensions. In other words, the benefit using the expected hand will only be seen if the expected movement is required. Conversely, under the extrinsic (ventral premotor cortex) model, there should be no interaction but two main effects due to the validity of *what* and *where* aspects of the movement respectively.

Based on the results of Jentzsch et al. (2004) and the retinotopic frame of reference of attentional effects in the Posner paradigm (Woldorff et al. 1997), I hypothesise that attentional cueing operates in an intrinsic frame of reference. I therefore expected to see an interaction between the validity effects of cueing, with speeded responses when, and only when, both *what* and *where* dimensions were valid.

2. Methods

2.1 Subjects

Eight healthy right-handed volunteers (2 female), aged 19 to 42, participated in this experiment. All subjects provided written and informed consent and the experiments were conducted in compliance with the standards established by the local ethical committee.

2.2 Experimental procedure and EMG recordings

Subjects were seated in a comfortable reclining chair. Their wrists were in a semi-supine position with the palms facing each other and supported by a splint that restricted wrist and hand movement to pure flexion and extension. The hand-splints were mounted on vertical spindles, which allowed rotation in the transverse plane. The hands were positioned such that the wrist joints sat directly above the axes of rotation. Additional support of the forearms further ensured that movements were constrained to the wrists, and reduced fatigue. Stimuli were viewed on a screen placed at eye level. Each trial started with a (150ms) cue-stimulus, followed by a blank screen (see Figure 5.3). 700ms after the appearance of the cue, a target-stimulus appeared for 400ms. A 50ms white-noise mask was presented after the cue and target stimuli to prevent the appearance of visual after-effects. Participants were given 1000ms after the appearance of the target-stimulus to make a response. No feedback was given. At the appearance of the target-stimulus, participants were required to respond as quickly as possible with the movement indicated. Four movements were possible – flexion or extension at the left or right wrist. The cue and target-stimuli had two dimensions – colour (blue, red) and spatial frequency (high, low). For four of the participants, the colour of the stimulus cued the hand (e.g. blue = left, red = right) and the spatial frequency indicated the movement (e.g. high frequency = flexion, low frequency = extension). For the remaining four, the stimulus-response mapping was reversed, so that colour indicated the movement to be made and spatial frequency the hand to be used. The stimuli subtended approximately 35 degrees of visual

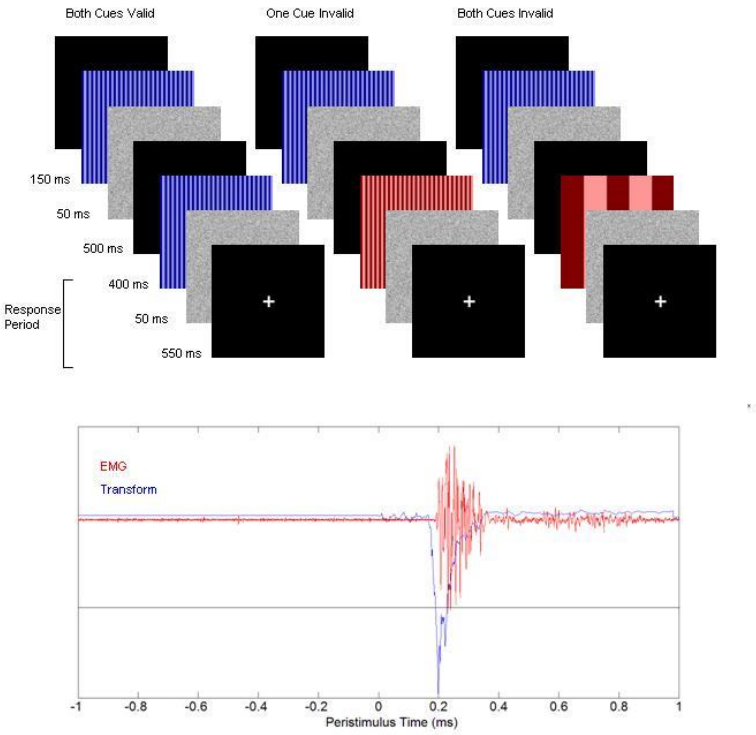

Figure 5.3: Experimental Design. Top panel: Schematic showing the time-line of three experimental trials, which comprised cue stimuli that could be congruent (valid) or incongruent (invalid) over each of their two dimensions (*what*: extension *vs*. flexion; *where*: left *vs*. right hand). **Bottom panel:** Example EMG trace acquired from a single muscle, plotted with the transform used for identifying movement onset. The line shows the *ad-hoc* threshold used to derive reaction times automatically.

angle. High-frequency stimuli were 2.5c/deg, low frequency were 0.25c/deg. The colours had RGB values ([128 0 0] [255 100 100]) and ([0 0 128] [100 100 255]).

Participants were required to relax their hands and lower arms until the appearance of the target-stimulus. This paradigm independently cued which motor and (right or left) would implement one of two movements (wrist flexion or extension). Each cue contained two dimensions – one signalling the hand to be moved and one the movement. For each dimension (colour, spatial frequency), cue-stimuli could be valid (80%) or invalid with regards to the target-stimulus (20%). Since the validity of the cue in each dimension was independent, this gave 64% (0.8x0.8) of trials with a completely valid cue, 32% (0.8x0.2x2) of trials where either the hand or the movement required was invalidly cued and 4% (0.2x0.2) of trials where both the hand and movement were cued invalidly. The experiment comprised one training session and 25 experimental sessions. Each session contained 100 trials, which were balanced for the four types of cue and four movements. The large number of trials was needed to acquire sufficient data from trials with invalid cues in both dimensions. The sessions were conducted over three separate days.

Reaction times were evaluated using surface EMG. Ag/AgCl electrodes were placed on the left and right brachioradialis/extensor carpi radialis longus and flexor carpi ulnaris muscles. Muscle activity was monitored throughout the experiment to ensure the effector muscles were relaxed before the appearance of the target stimulus. Signals were recorded via a CED 1401 laboratory interface (Cambridge Electronic Design Ltd., Cambridge, UK) and stored on a personal computer (for later analysis) at a sample rate of 5 kHz (Signal 2.0, Cambridge Electronic Design Ltd.). Data were bandpass-filtered between 3 Hz and 2.5 kHz.

2.3 Statistical analysis

EMG data were smoothed with a Butterworth low-pass filter with a cutoff frequency of 600Hz to increase signal-to-noise. After full-wave rectification the data were log-transformed to provide normally distributed time series for further analysis. The mean of 100 consecutive data points was compared with the mean of the preceding 5000 data points, using two-sample *t*-tests and a sliding window. Reaction times were defined operationally as the first time at which the absolute value of the *t*-statistic exceeded 50. This *ad hoc* threshold identified the highest number of correctly performed trials. Incorrect trials, where a muscle other than the agonist for the correct movement showed the shortest reaction time, were excluded.

A standard summary statistic method was used for statistical inference, using the log of the mean reaction times (to correct for positive skew) over each of the four conditions, for each subject. Univariate 5-way ANOVA was performed in SPSS, with factors HAND CUE VALIDITY (valid *vs*. invalid), MOVEMENT CUE VALIDITY (valid *vs*. invalid), HAND (left vs. right), MOVEMENT (flexion vs. extension). Factors SUBJECT and STIMULUS-RESPONSE MAPPING were nested and were implemented in two separate ANOVA models.

3. Results

13% of trials (range over subjects 8% - 22%) were discarded. Of these trials, in 2% no movement was made or no movement could be identified. In the remaining 11%, an incorrect movement was made (error trials). Error trial frequency varied significantly by cue type ($p < 0.001$, $\chi^2 > 400$, 1 d.f.), with errors less likely on validly cued trials. The most common error (64% of errors) was making the incorrect movement with the correct hand. The least common error (10% of errors) was making the correct movement with the wrong hand. Among invalidly cued trials, performing the movement specified by the cue stimulus rather than the target stimulus occurred significantly more often ($p < 0.05$, $\chi^2 > 6.01$, 1 d.f.). Since the EMG measured the onset of movement rather than the endpoint, changing the response before the movement was completed resulted in an error trial. This may explain the comparatively high error rate seen here, compared with more traditional button-press paradigms.

The grand average reaction time was 334ms. There was no significant main effect of HAND, MOVEMENT or STIMULUS-RESPONSE MAPPING, so the ANOVA model including SUBJECT as a factor was used for further analysis. There were significant main effects of HAND CUE VALIDITY ($F_{(1,7)} = 90.54$, $P < 0.001$, $\eta^2 = 0.368$), MOVEMENT CUE VALIDITY ($F_{(1,7)} = 171.12$, $P < 0.001$, $\eta^2 = 0.155$) and SUBJECT ($F_{(1,7)} = 9.29$, $P < 0.003$, $\eta^2 = 0.293$). There were two significant two-way interactions – MOVEMENT x MOVEMENT CUE ($F_{(1,7)} = 4.98$,$p = 0.048$, $\eta^2 = 0.001$), and, as anticipated, MOVEMENT CUE VALIDITY x HAND CUE VALIDITY (Fig. 4; $F_{(1,7)} = 233.34$, $P < 0.001$, $\eta^2 = 0.108$). As expected, the fastest mean reaction time was seen when both cues were valid. Figure 5.4 highlights the nature of this interaction with reference to the profiles predicted by high (extrinsic) and low (intrinsic) levels of facilitation in the motor hierarchy. It is clear that this profile is consistent with attentional bias at the (motor cortex) level of representation, in an intrinsic frame of reference. Quantitatively, these results suggests that the validity effect is expressed primarily when both cue dimensions were jointly valid.

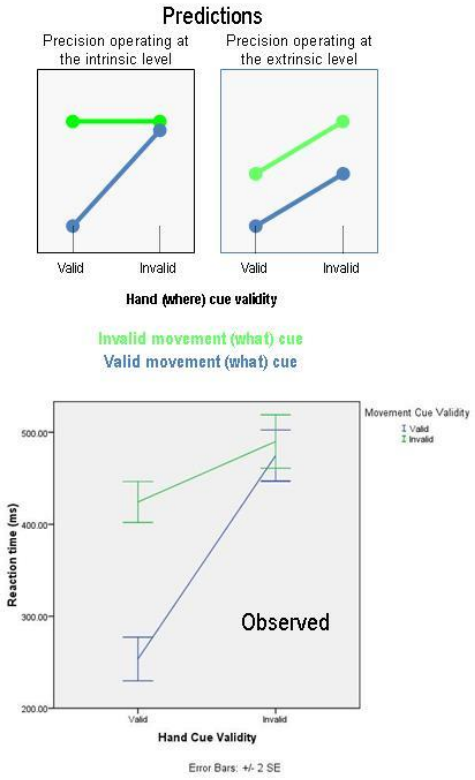

Figure 5.4: Reaction time effects for the four combinations of cue-validity. The top panels show the results predicted by the theoretical architectures of Figure 5.2. The green lines correspond to valid movement cues and the blues lines to invalid movement cues. The empirical results are shown in the lower panel using the same colours. The bars correspond to standard error over subjects. The form of the interaction observed is very close to that predicted under a model where attention biases prediction errors in an intrinsic frame of reference (Figure 5.2).

Paired *t*-tests among the four validity categories confirmed that only one pair failed to show a significant difference (after Bonferroni correction): movement cue valid, hand cue invalid and movement cue invalid, hand cue invalid ($p > 0.2$). All other pairwise differences were highly significant ($p < 0.001$).

4. Discussion

This chapter pursues the idea that attention is an integral part of motor control and expresses itself through biasing the precision afforded to the proprioceptive and somatosensory consequences of an anticipated action (Galazky et al. 2009). This places previous proposals that link motor preparation and attention (cf, Rizzolatti 1994; Rushworth 2001; Humphreys 2010; Allport, 1987; Goldberg and Segraves 1987; see Tipper, 2004 for an overview) in the general framework of active inference and predictive coding. The important perspective provided by active inference is that movements fulfil predictions furnished by percepts with both exteroceptive (e.g. visual) and proprioceptive (e.g., stretch receptor) components.

The work described in chapter 3 demonstrated that the reaction time benefits of cueing can be understood as statistically optimal responses, where the associated optimisation of precision can account for both psychophysical and electrophysiological phenomena fairly accurately (Feldman and Friston, 2010). In this chapter, I asked whether similar reaction time benefits can be seen empirically in the motor domain. To this end, the paradigm developed by Rosenbaum (1980) was adapted, in which two different visual dimensions (colour and spatial frequency) cued the impending movement. As in Rosenbaum & Kornblum (1982), it was predicted and confirmed that cueing effects would occur only when both cue dimensions were valid. These predictions were based on the possible outcomes of attentional bias at different levels in the cortical hierarchy, which I associate with representations in extrinsic (higher) and intrinsic (lower) frames of reference: In an extrinsic model, one would predict that cueing effects enact their influence independently and to a comparable degree. As outlined above, the interaction between the two validity factors argues for an intrinsic model, in which hand and movement are selectively enabled in a way that cannot be separated. In the present case, the observed interaction can be accounted for by a model where precision is increased in proprioceptive channels that represent the confluence of top-down predictions about the nature of a movement and where it will be implemented (see Figure 5.2).

In addition to the interaction above, there was a small reaction time benefit from a valid hand cue, even if the movement cue was invalid. The magnitude of this effect was much smaller compared to the reaction time benefit seen for two valid cues (66ms vs. 237ms). This, and the lack of any benefit for a valid movement cue if the hand cue is invalid, means that a model in which precision operates at the intrinsic level is still the most likely. The small validity effect of a valid hand cue might be explained in the framework of active inference; because the movements performed in this experiment were self-limited, the same muscles were recruited for both flexion and extension movements, to either initiate or terminate the movement. Thus, if the precision of the stretch receptor channels in one forearm were boosted after cuing that side, a small benefit might accrue for the opposite movement.

Rushworth et al. (1997b) also demonstrated a benefit for valid cuing using a similar paradigm. Spatial cues were used, and the motor preparation time was calculated from the difference between two conditions: a simple cuing task in one movement dimension, and a control task where the movement made did not depend on the validity of the cue. A small reaction time benefit was seen for valid cues.

In Rosenbaum (1980), some aspects of the movement were left unspecified until the appearance of the target stimulus. Unlike in this study, Rosenbaum saw separable effects of cuing just the arm, the direction and the extent of the upcoming movement. However, there is a key difference between this paradigm and that of Rosenbaum (1980) that may account for the difference. The button-press responses used in Rosenbaum (1980) entail visuomotor and somatosensory-motor integration. This means that attentional cueing effects in the visual or somatosensory domains cannot be disambiguated from purely proprioceptive attention. This paradigm avoided conflating multiple attention processes by cueing movements that could be performed using only proprioceptive channels (simple, self-terminated flexion and extension movements). This means that one can attribute the cue validity effects to attentional modulation of proprioceptive signals, in accordance with active inference. Furthermore, Rosenbaum's cues were semantic (letters), whereas this paradigm used low-level visual features which were arbitrarily mapped onto flexion and extension movements. The complexity of the semantic cues meant that most of the reaction time advantages seen in Rosenbaum (1980) could be accounted for by validity effects on processing visual targets and their semantic content and not on the movements *per se*. In short, the simplicity of the movements and cues used in this experiment suggests a motoric rather than sensory locus for attentional cueing.

A further study (Rosenbaum & Kornblum 1982), which resembled ours except that only two of four possible movements were possible in each trial, did not find that correctly cuing one response attribute benefitted reaction time. They found the opposite – violating the hand and movement cues increased reaction times relative to violating just the movement cue. Their explanation for this was that both movements were simultaneously prepared, but choosing between two movements on the same hand takes longer because the movements are more 'similar'. The larger number of possible movements in my experiment meant that simultaneously preparing all responses was unlikely (the flexion and extension movements used the same motor plant, while index and middle finger movements were used in Rosenbaum and Kornblum 1982). By contrast, Miller (1982) found a contradictory effect – advance information of which hand to use gave a reaction time advantage, whereas advance information of which finger (on either hand) did not.

How can these discrepancies be resolved? Cui and Deecke (1999) found anatomically congruent movements were performed faster than spatially congruent movements, suggesting that anatomically congruent movements are prepared together in the motor hierarchy, or, alternatively that the mapping from extrinsic to intrinsic coordinates is more efficient. Despite the anatomical distance between [pre]motor cortex in each hemisphere, activity in these areas may be influenced at an early stage during motor preparation. If left and right effectors are competing alternatives for subsequent actions (cf. Cisek and Kalaska, 2010), several (bilateral) representations can in principle occur in an intrinsic frame of reference at the same time. These results suggest that predictions about impending movements are integrated to boost processing in effector-based (intrinsic) coordinates.

Goodman & Kelso (1980) suggested that stimulus-response mapping time is shorter for cued movements. If this were the case, one would expect cues correct in one response dimension to provide some reaction time benefit for the other. The locus of such an effect would likely be before the motor stage; i.e., early in the stimulus-response interval. However, evidence from EEG studies suggests that the effects of cueing occur relatively late, again suggesting an effect in intrinsic coordinates: for example, the lateralised readiness potential (LRP), an EEG potential evoked when one hand is cued, has been suggested to be the halfway point between premotor and motor processing (Osman et al, 1995). This is supported by the finding that it occurs nearer to the movement during trials with informative cues than those without, although the stimulus-LRP latency does not change (Jentzsch et al, 2004). Finally, it can be noted that a locus of the motor attentional effect in intrinsic coordinates provides an interesting parallel with results from the Posner

paradigm. The reaction time benefit associated with cues in most visual paradigms seems to occur in retinotopic (intrinsic) rather than allocentric (extrinsic) frames of reference (Posner & Cohen 1984; Golomb et al., 2008).

4.1 Conclusion

This chapter has explored the idea that motor preparation is an attentional phenomenon that is directed towards proprioceptive sensations (i.e., predicted sensory feedback of the anticipated motor response). This perspective suggests that attention should not be limited to perceptual processing in the exteroceptive (e.g. visual) domain but should also bias interoceptive inference during movement. These predictions were verified by adapting a classical attention (Posner) paradigm for a motor setting. Furthermore, the hierarchical level this attentional bias operates at was established by cueing the movement and effector independently. These behavioural results demonstrate an interaction between the validity of movement and effector cues. This suggests that the bias for the selected action is mediated at a low level in the motor hierarchy, in an intrinsic frame of reference. More generally, the ideas outlined above provide a heuristic framework in which to address questions about the link between motor preparation and attention, and their mechanistic underpinnings.

The work described in this chapter was published as:

Brown H, Friston K, Bestmann S. (2011). Active inference, attention, and motor preparation. Front. Psychol. 2:218.

Chapter 6

Active inference, sensory attenuation and illusions

1. Introduction

Children discover early in life that although they can tickle others, and be tickled by others, it is almost impossible to tickle oneself. The commonplace nature of this observation hides its profundity – two physically identical sensory stimuli can be perceived differently, depending on high-level concepts such as agency or wilfulness. This sort of effect has now been quantified in a number of tasks and has been investigated in numerous neuroimaging studies. However, after more than a decade of research, a simple explanation is still outstanding. In this chapter, I try to provide a principled account of how beliefs about agency depend upon the active sampling of sensory information (active inference) and how this leads naturally to phenomena like sensory attenuation, the force matching illusion and attribution of agency.

1.1 Sensory attenuation and agency

The difference between self-generated and externally-generated tickle has been the focus of many studies (Weiskrantz et al., 1971; Claxton 1975; Blakemore et al., 1999). Self-produced tickle is consistently rated less 'ticklish' than externally produced tickle and its ticklishness can be increased by closing the eyes (Claxton, 1975). Tickle is not the only attribute of sensation affected – self-generated touch stimuli are also perceived as less pleasant and intense (Blakemoreet al., 1999). Indeed, sensory attenuation is not limited to

somatosensation; attenuation of self-generated visual (Hughes et al., 2011; Cardoso-Leite et al., 2010) and auditory sensations have been reported (Martikainen et al., 2005; Weiss et al., 2011a, 2011b; Desantis et al., 2012).

A measure of sensory attenuation is provided by the force-match task (Shergill et al., 2003, 2005). During this task, instead of reporting sensations explicitly, subjects match a reference force, either by pressing directly on themselves, or by using a robot to reproduce the perceived pressure. Higher levels of matched force are produced when the force is self-generated, consistent with self-reports of sensory attenuation.

Sensory attenuation is also evident in neuronal responses. Subcortically, both cerebellar (Blakemore et al., 1998, 1999, 2001) and thalamic (Blakemore et al., 1998) activity is reduced for self-produced vs. externally produced sensations. Early sensory responses are also modulated in auditory paradigms, where these differences can appear as early as 27 ms after stimulus onset (Baess et al., 2008, 2009; Aliu et al., 2009; Martikainen et al., 2005). Repetitive stimulation of M1 (which has a depressive effect on activity) reduces the magnitude of sensory attenuation in the force-match task, as well as in a grip-production task (Therrien et al., 2001; Voss et al., 2007), whereas single-pulse TMS of M1 just before movement onset (which delays the movement) has no effect on the level of sensory attenuation (Voss et al., 2006). This suggests that M1 is involved in determining the level of sensory attenuation but not in mediating it. In visual studies, the only ERP change noted thus far is a late (~150 ms) modulation of frontoparietal potentials (Shaefer & Marcus, 1973; Hughes et al. 2011). Concepts, such as meaning, perception of agency and social factors can influence sensory attenuation. Curio et al. (2000) demonstrated the absence of the late (300 ms) 'oddball' potentials (usually elicited in response to rare stimuli which have 'meaning' or task-relevance) in response to self-generated stimuli, suggesting that they are categorised as distinct from externally generated stimuli at a conceptual level. Sato (2008) observed sensory attenuation both when participants performed a movement resulting in a tone, and when they observed experimenters performing the same movement. Similarly, Weiss et al., (2011) noted greater sensory attenuation when participants triggered the experimenter to produce externally generated tones by tapping them, and vice versa for self-generated tones.

The relationship between sensory attenuation and the experience of agency is complex. An experience of agency over movements that generate sensation seems to be necessary for sensory attenuation (Desantis et al., 2012; Gentsch & Schütz-Bosbach, 2011): sensory attenuation does not occur if movement and sensation are correlated but the relationship

is not perceived as causal. Some authors have suggested that the experience of sensory attenuation is important in labelling movements as self-generated (Blakemore et al., 2002). In support of this idea, Baess et al. (2011) found that sensory attenuation was more pronounced in blocks with mixed self- and externally-produced sensations. In this setting, the attribution of agency is more difficult than during a sequence of sensations that are purely self- or purely externally-generated.

Sensory attenuation is an interesting phenomenon partly because sensory attenuation is reduced in schizophrenia (Blakemore et al., 2000; Shergill et al., 2005), or those at high risk of developing psychosis (Wilquin & Delevoye-Turrel, 2012). In normal subjects, sensory attenuation is (negatively) correlated with the level of delusional beliefs (Teufel et al., 2010). Less sensory attenuation means that the percepts of schizophrenics are more veridical than controls and – in the force match task – they perform more accurately (Shergill et al., 2005). This means that differences between schizophrenics and controls are difficult to attribute to non-specific effects of long-term disease, psychoactive medication or social deprivation, and that investigating this effect might provide clues about the pathogenesis of schizophrenia. A key symptom of schizophrenia is aberrant perception of agency (Frith, 2005), particularly the delusion that one's actions are being controlled by others, suggesting the mechanisms that impair sensory attenuation in schizophrenia are intimately related to the perception of agency.

1.2 Formal theories of sensory attenuation

Previous explanations for the force-match paradigm – that can be applied to sensory attenuation more generally – have come from engineering approaches to motor control (Wolpert and Flanagan, 2001). In the model proposed by Frith, Blakemore & Wolpert (2000), the decision to move initiates a motor command, which is transformed by a forward model into a prediction of the sensations created by that movement. The real sensations produced by the movement are compared to the predictions of the forward model to produce a 'control theory' prediction error, which is used to update predictions and refine the forward model. During self-generated movement, an accurate forward model means there is little prediction error. Under this model, it is suggested that small prediction errors during self-generated movement leads to a percept of a less intense force, relative to the true force.

This model is incomplete in a number of aspects. Firstly, it is unclear why the intensity of a percept is related to the size of prediction error: prediction errors are used to update

predictions but they do not constitute predictions or percepts *per se*. Within predictive coding formulations of perception (Rao & Ballard, 1999; Friston 2005), prediction errors play a crucial role in perception but, again, they are not the percept itself; the percept is a synthesis of prior beliefs and sensory evidence that is conveyed by prediction errors.

Second, this explanation overlooks the multidimensional nature of sensory attributes. In the optimal control explanation, any mismatch between the forward model and sensory input is mapped to a single variable that determines perceived intensity. It is true that parametrically varying the time delay between movement and sensation – or rotating sensory feedback with respect to movement – will alter the force-match illusion (Blakemore et al., 1999). However, the optimal control formulation does not explain how this is caused by the amplitude of prediction error, pooled over all sensory channels. Furthermore, the amplitude of prediction error does not seem to be important in determining the level of sensory attenuation: for example, Baess et al. (2008) show that the predictability of a self-generated sensation does not affect sensory attenuation. Crucially, a self-generated movement that should result in sensation – but does not – can still cause sensory attenuation, despite the implicit production of prediction errors (Bays et al., 2005).

Third, there is a set of results that control theory approach cannot account for. During self-generated movement, sensory attenuation is often noted in response to *externally* generated stimuli (Voss et al., 2008; Rushton et al., 1981; Milne et al., 1988; Chapman et al., 1987). These stimuli are applied by the experimenter so they cannot be predicted by the forward model and therefore cannot be attenuated. Additionally, sensory attenuation has been found for stimuli that occur (up to 400 ms) before the onset of movement (Voss et al., 2008; Bays et al., 2005), when they cannot be predicted from self-generated movement. This attenuation seems to be due to changes in sensitivity (*d*-prime) to external stimuli rather than a change in the response criterion (Juravle and Spence, 2011; Van Hulle et al., 2012). The attenuation of these stimuli – which cannot be predicted from motor commands – suggests that the phenomenon of sensory attenuation is broader than suggested by optimal control formulations.

In this chapter, I put forward an alternative explanation for sensory attenuation based on active inference. Active inference is based on Bayes optimal accounts of behaviour and provides a principled explanation of how sensory attenuation may arise in a Bayes optimal (normative) sense. This is in contrast to previous explanations, which have explained sensory attenuation as a quirk or anomaly of motor control. Instead, I suggest that

sensory attenuation is a necessary consequence of reducing the precision of sensory evidence during movement to allow the expression of proprioceptive predictions that incite movement. This explanation is potentially important because a failure of sensory attenuation may result in false inference about the causes (agency) of self-made acts – a failure that is characteristic of the positive symptoms of schizophrenia. Furthermore, the neuronal mechanisms behind sensory attenuation (and compensatory changes in the precision of beliefs at non-sensory levels) rest on neuromodulatory mechanisms that have been implicated in psychosis.

In the following, active inference and its neurobiological implementation are briefly recapitulated. This is discussed more extensively in the introduction. This implementation is used in later simulations to demonstrate why sensory attenuation is necessary for movement. I then simulate the force match illusion using exactly the same scheme. The chapter concludes by simulating a loss of sensory attenuation and a compensatory increase in non-sensory precision, as might be found in schizophrenia. Crucially, this simulated pathology exposes actors to false beliefs or delusions; interestingly, with a necessarily antagonistic content. These simulations do not model all the aspects of sensory attenuation discussed above (e.g. Sato et al., 2008); however, it is hoped that the principles of active inference – in particular, the optimisation of precision at different levels of a predictive coding hierarchy – may generalise to other settings.

2. Neurobiological implementation of active inference

These simulations use differential equations that minimise the free energy of sensory input using the generalised gradient descent derived in Appendix 1 (Friston, Stephan, Li, & Daunizeau, 2010).

$$\begin{aligned}\dot{\tilde{\mu}}(t) &= \mathcal{D}\tilde{\mu}(t) - \partial_{\tilde{\mu}} F(\tilde{s},\tilde{\mu}) \\ \dot{a}(t) &= -\partial_a F(\tilde{s},\tilde{\mu})\end{aligned} \qquad (6.1)$$

These coupled differential equations describe perception and action respectively and just say that neuronal activity encoding conditional expectations $\tilde{\mu} \equiv (\mu, \mu', \mu'', \ldots)$ and action a change to reduce free energy, where free energy $F(\tilde{s},\tilde{\mu})$ is a function of sensory inputs $\tilde{s} = (s, s', s'', \ldots)$ and conditional expectations encoded by neuronal activity. The first differential equation has the same form as Bayesian (e.g., Kalman-Bucy) filters used in time series analysis. The second differential equation says that action also minimises free

energy. The differential equations above are coupled because sensory input depends upon action, which depends upon perception through the conditional expectations. This circular dependency leads to a sampling of sensory input that is both predicted and predictable, thereby minimising free energy and prediction errors.

To perform neuronal simulations under this scheme, it is only necessary to integrate or solve Equation (6.1) to simulate the neuronal dynamics that encode conditional expectations and the ensuing action. Conditional expectations depend upon the brain's generative model of the world, which is assumed to have the following hierarchical form discussed in the introduction.

In the present context, the key thing about this predictive coding scheme used is that the precisions at each level in the hierarchy depend on the expected hidden causes and states in the level above. It is this dependency that has been proposed to mediate attention or selection in hierarchical inference, as described in chapters 3 and 5 (Feldman & Friston, 2010; Brown et al., 2011; Friston et al., 2012). The mathematics of this scheme (set out in equation 0.3) tells us that the state-dependent precisions modulate the responses of the error-units to their presynaptic inputs. This modulation depends on the conditional expectations about the states and suggests something intuitive – attention is mediated by activity-dependent modulation of the synaptic gain of principal cells that convey sensory information (prediction error) from one cortical level to the next. This translates into a top-down control of synaptic gain in principal (superficial pyramidal) cells elaborating prediction errors and fits comfortably with the modulatory effects of top-down connections in cortical hierarchies that have been associated with attention and action selection.

2.1 Action

In active inference, conditional expectations elicit behaviour by sending top-down predictions down the hierarchy that are unpacked into proprioceptive predictions at the level of the cranial nerve nuclei and spinal-cord. These engage classical reflex arcs to suppress proprioceptive prediction errors and produce the predicted motor trajectory

$$\dot{a} = -\frac{\partial}{\partial a}F = -\frac{\partial \tilde{s}}{\partial a}.\xi_v^{(1)} \qquad (4)$$

The reduction of action to classical reflexes follows because the only way that action can minimise free energy is to change sensory (proprioceptive) prediction errors by changing

sensory signals; cf., the equilibrium point formulation of motor control (Feldman & Levin, 1995). In short, active inference can be regarded as equipping a generalised predictive coding scheme with classical reflex arcs: see (Friston, Daunizeau, Kilner, & Kiebel, 2010; Friston, Daunizeau, & Kiebel, 2009) for details. The actual movements produced clearly depend upon top-down predictions that can have a rich and complex structure, as we will see next.

3. Simulations of sensory attenuation

This section provides a series of simulations – using the active inference scheme of the previous section – to illustrate the basic phenomena of sensory attenuation. In what follows, I describe a minimal model of sensations that can be generated internally or externally. This model is used to illustrate the permissive and necessary role of sensory attenuation in the production of self-made acts. The perceptual consequences of sensory attenuation are then addressed, in terms of detecting externally and internally generated events – that has been the focus of much work in psychology and psychophysics reviewed in the introduction. Using the same model, the force-matching illusion is then reproduced by yoking externally applied forces to the perceived level of self-generated forces. Finally, I demonstrate the disappearance of the illusion and the emergence of false inferences about (antagonistic) external forces, when sensory attenuation (attenuation of sensory precision) is removed.

3.1 The generative process and model

Figure 6.1 describes the generative process and model in terms of equations (that have the same hierarchical form as Equation 0.2) and a schematic showing how the hidden states and causes are interpreted. This model is as simple as it could be, while retaining the key ingredients that are required to demonstrate inference about or attribution of agency. The equations on the left describe the real world (whose states and causes are in boldface), while the equations on the right constitute the subject's generative model. In the real world, there is one hidden state $\mathbf{x}_i$ modelling self-generated force or pressure that is registered by both proprioceptive s_p and somatosensory s_s input. This hidden force increases with action and decays with a time constant of four time bins (where each time bin corresponds to about one hundred milliseconds). Externally generated forces are modelled with $\mathbf{v}_e$ and add to the internally generated forces to provide somatosensory input.

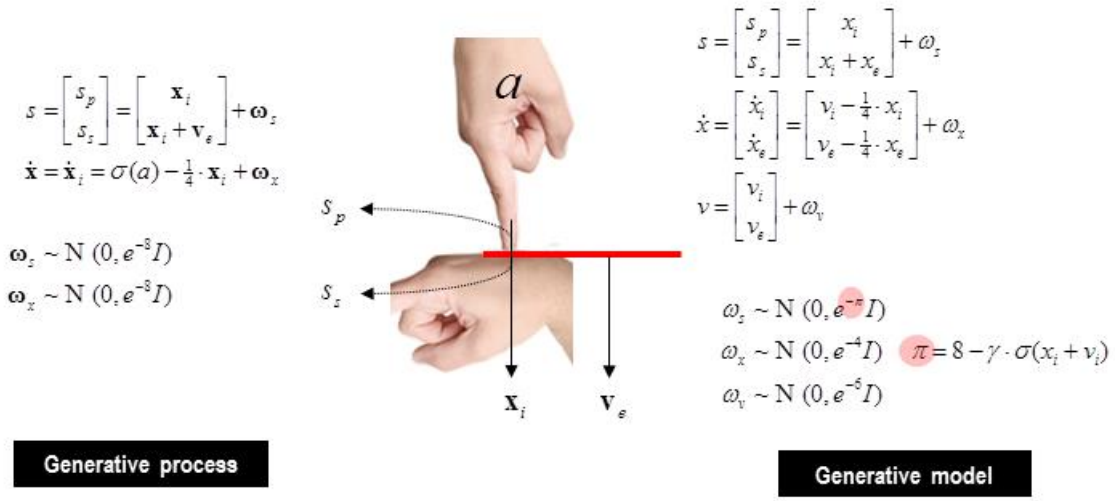

Figure 6.1: Generative model. This figure shows the generative process and model used in these simulations. The generative process (left) models real-world states and causes, while the generative model (right) is used by the subject to make inferences about causes of its sensations. In the real world, the hidden state $\mathbf{x_i}$ models self-generated forces that are sensed by both somatosensory s_s and proprioceptive s_p input channels. External forces are modelled with the hidden cause $\mathbf{v_e}$ and are sensed only by the somatosensory input channel. Action causes the self-generated force to increase and is modified by a sigmoid squashing function σ (a hyperbolic tangent function). The hidden state decays slowly over four time bins. In the generative model, causes of sensory data are divided into internal causes v_i and external causes v_e. The hidden cause excites dynamics in hidden states x_i and x_e, which decay slowly over time as above. Internal force is perceived by both proprioceptive and somatosensory receptors, while external force is perceived only by somatosensory receptors. Crucially, the precision of the sensory prediction error π is influenced by the level of internal force, again modulated by a squashing function, and controlled by a parameter γ which governs the level of attenuation of precision. The pink circles highlight this state dependent precision, which effectively controls the influence of sensory prediction errors during active inference.

The key thing about this model is that somatosensory sensations are caused ambiguously, by either internally or externally generated forces. The only way that the underlying cause of the sensations can be resolved is by reference to proprioceptive input – which, here, we use to refer to as the subset or pattern of inputs generated only internally. This is a very simple model, where the somatosensory input is being used metaphorically to stand in for the sensory consequences of events that could either be caused by self or others, while proprioceptive input represents those sensory signals that can only be caused by self-made acts. Active inference now compels the subject to infer the causes of its sensations.

The generative model used for this inference is shown on the right. In this model, internally and externally generated forces (x_i, x_e) are modelled symmetrically, where changes in both are attributed to internal and external hidden causes (v_i, v_e), with the same restorative dynamics associated with action above. The hidden causes trigger the dynamics associated with the hidden states, much like a push which sets a swing in motion. This means that proprioceptive and somatosensory inputs are explained in terms of hidden causes, where proprioceptive sensations are caused by internally generated forces and somatosensory consequences report a mixture of internal and external forces. Crucially, the precision of the sensory prediction errors depends upon the magnitude of the internally generated force (and its hidden cause). This dependency is controlled by a parameter γ that mediates the attenuation of sensory precision: as internally generated forces rise, sensory precision falls, thereby attenuating the amplitude of (precision weighted) sensory prediction errors. These context or state-dependent changes in precision enable the agent to attend to sensory input, or not – depending upon the relative precision of prediction errors at the sensory and higher levels. This context sensitive sensory precision is shown in Figure 6.1 as π.

Notice that, from the point of view of the subject, there is no real difference between hidden causes of internal and external forces – other than that the internal forces affect both proprioceptive and somatosensory inputs, while external forces only produce somatosensory sensations. Although action can fulfil proprioceptive predictions, the subject does not need to know this. In other words, it is not aware of its reflexes; it simply attributes particular sensations to particular hidden causes, which is labelled as self-generated.

3.2 Precision and the psychophysics of sensory attenuation

In the simulations which follow, I try to reconcile the literature on stimulus detection and ratings of intensity by associating the reported intensity of a stimulus with its 90% lower posterior confidence bound. This means that the detectability and subjective intensity are functions of both the conditional expectation and confidence or precision – such that stimulus intensity is reported to be greater when the confidence that it exceeds some threshold is larger. This is an important assumption because it implicates the subject's confidence in the estimation of intensity and therefore speaks to a role for precision in subjective reports of sensory attenuation. Invoking a (signal detection or decision theoretic) notion of a threshold rests on the fact that sensory attenuation is only observed for stimulus attributes that can be above a threshold; for example, loudness, pressure, unpleasantness and so on. Stimulus attributes that do not have an intensity threshold could not be treated in this fashion and – I would suggest – could not show sensory attenuation. For example, although one can attenuate the loudness of an auditory tone, one cannot attenuate its frequency (which can only change by going up or down). Put simply, sensory attenuation can only be expressed in sensory modalities that have the attribute of intensity.

The relationship between physical stimulus intensity and perceived stimulus intensity is not linear. In many domains, the relationship is approximated by a power law: that is, perceived intensity is proportional to physical intensity raised to the power of an exponent (Stevens, 1967). In the case of somatosensory pressure, this exponent is less than one (Xiong et al., 2012), meaning that – at higher levels of pressure – the same increase in physical pressure produces a smaller increase in perceived pressure. A clue as to why this might be is found in Weber's law (Weber, 1846), which states that the just-noticeable difference between figure luminance and background luminance increases as background luminance increases. Higher background light levels increase the amplitude of random fluctuations in the stimulus, making discrimination more difficult. It could be that this 'diminishing returns' effect seen in pressure perception results from higher levels of noise attenuating the perception of the stimulus.

As noted above, attentional processing can also be cast in terms of state-dependent precision. In chapter 3 (Feldman & Friston, 2010), I suggest that attention is the process of optimising precision in neural hierarchies, such that attended locations or objects are afforded high precision. This process is exactly opposite to the process of sensory attenuation described above: during sensory attenuation, attention is withdrawn from the

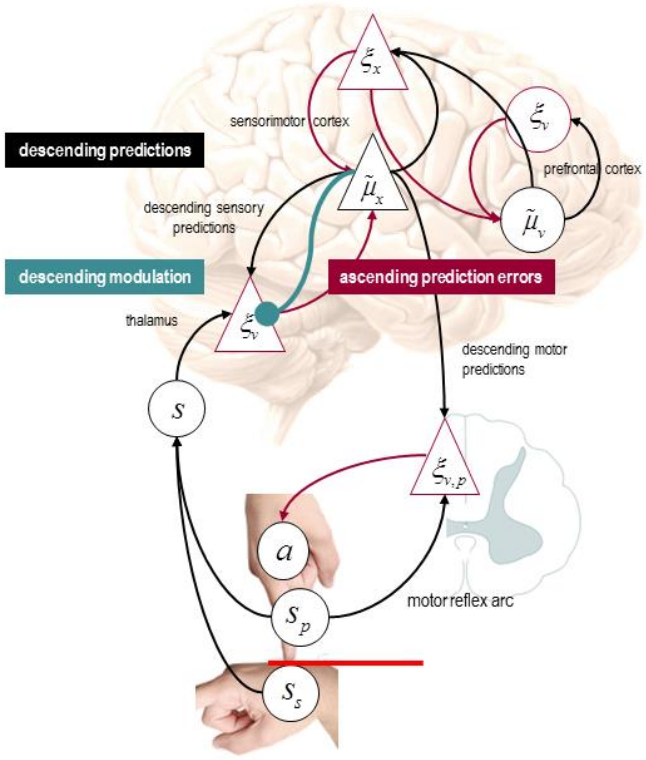

Figure 6.2: Functional anatomy. Speculative mapping of Equation (3) onto neuroanatomy. Somatosensory and proprioceptive prediction errors are generated by the thalamus, while conditional expectations and prediction errors about hidden states (circles) (the forces) are placed in sensorimotor cortex. The expectations and prediction errors about the hidden causes of forces (triangles) have been placed in the prefrontal cortex. In active inference, proprioceptive predictions descend to the spinal-cord and elicit output from alpha motor neurons (playing the role of proprioceptive prediction error units) via a classical reflex arc. Red connections originate from prediction error units (ξ cells) and can be regarded as intrinsic connections or ascending (forward) extrinsic connections from superficial principal cells. Conversely, the black connections represent intrinsic connections and descending (backward) efferents from (deep) principal cells encoding conditional expectations ($\tilde{\mu}$ cells). The cyan connections denote descending neuromodulatory effects that mediate sensory attenuation. The crucial point to take from this schematic is that conditional expectations of sensory states (encoded in the pyramidal cell $\tilde{\mu}_x$) can either be fulfilled by descending proprioceptive predictions (that recruit classical reflex arcs), or they can be corrected by ascending sensory prediction errors. In order for descending motor efferents to prevail, the precision of the sensory prediction errors must be attenuated.

consequences of movement so that movement can occur. Directing attention to a stimulus can increase its perceived intensity: in the visual domain, this has been demonstrated in the cases of contrast (Liu et al., 2012; Carrasco et al., 2002; Treue et al., 2004), colour saturation (Fuller et al., 2006), speed (Turatto et al., 2006), flicker rate (Montagna & Carrasco, 2006) and spatial frequency (Gobell & Carrasco, 2005; Abrams et al., 2010). Given that judgements of stimulus intensity are necessarily subjective, the corollary – that withdrawing attention should decrease intensity – is entirely sensible. There is little empirical work directly addressing the effect of stimulus uncertainty (sensory precision) on perceived intensity. However, it has been demonstrated that in the auditory domain, loudness is attenuated by the addition of a noise mask (Richards, 1968; Lochner and Burger, 1961; Stevens, 1966, 1967).

3.3 Functional anatomy

If this model is placed in the predictive coding scheme described in the introduction, one obtains a simple architecture that is shown schematically in Figure 6.2. The precise anatomy illustrated in the figure should not be taken too seriously but illustrates how a generative model can be transcribed into a plausible neuronal architecture for predictive coding and active inference. In this particular example, sensory prediction errors are assigned to the thalamus, while corresponding expectations and prediction errors about hidden states (forces) are associated with the sensorimotor cortex. The expectations and prediction errors about the hidden causes of forces have been placed – somewhat agnostically – in the prefrontal cortex. Notice how proprioceptive predictions descend to the spinal cord to elicit output from alpha motor neurons (playing the role of proprioceptive prediction error units) to elicit movements through a classical reflex arc. Red connections originate from prediction error units and can be regarded as intrinsic connections or ascending (forward) extrinsic connections from superficial principal cells. Conversely, the black connections represent intrinsic connections and descending (backward) efferents from (deep) principal cells mediating conditional predictions. The cyan connections denote descending neuromodulatory effects that mediate attenuation of sensory precision. The ensuing hierarchy conforms to the functional form of the predictive coding scheme in Equation 0.3. In this architecture, predictions based on expected states of the world $\tilde{\mu}_v$, can either be fulfilled by reflex arcs or they can be corrected by ascending sensory prediction errors. Which of these alternatives occur depends on the relative precisions along each pathway – that are set by the descending modulatory connection to sensory prediction errors. In the following sections I use this model to demonstrate some key points:

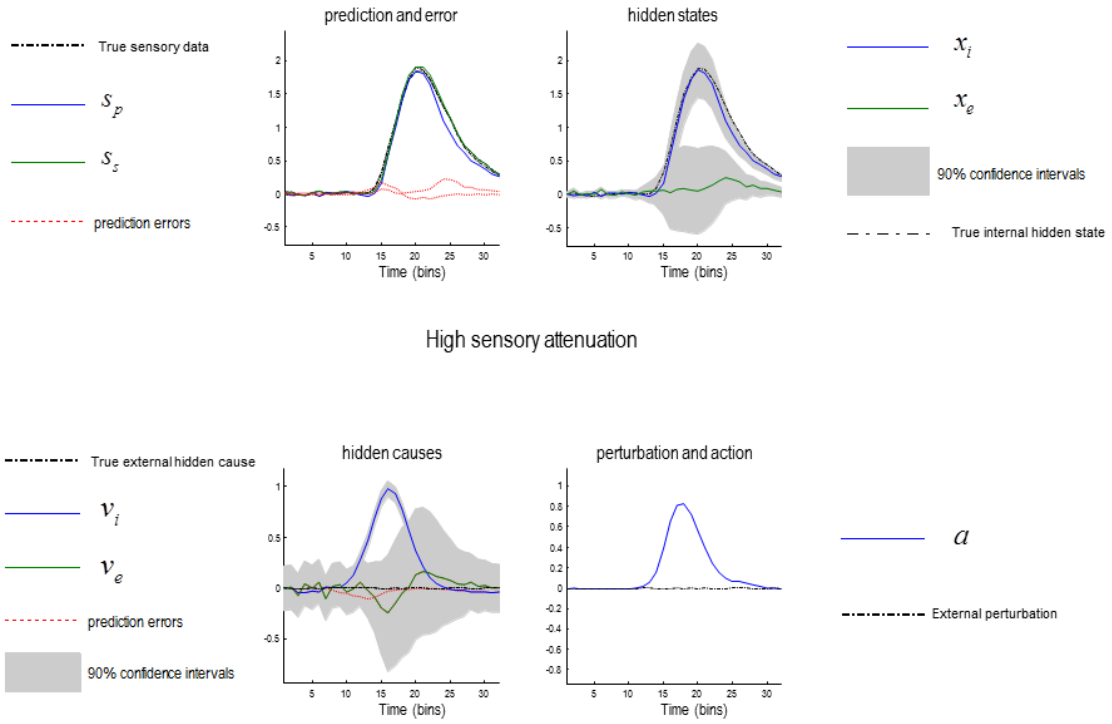

Figure 6.3: Sensory attenuation and action. Simulation results illustrating the permissive effect of sensory attenuation in movement. The model was supplied with a prior belief about the hidden cause of internally generated movement, while sensory attention was high ($\gamma = 6$). This prior expectation was a simple Gaussian function of time (blue line in the lower left panel) and engenders beliefs about forces (upper right panel), which produce proprioceptive predictions (upper left panel). Action is enslaved to fulfil these predictions (lower right panel). Note the confidence interval around the external cause temporarily inflates during action (lower left panel), reflecting the attenuation of sensory precision.

3.4 The permissive role of sensory attenuation in action

In the first simulations, I illustrate the necessary role of state-dependent changes in sensory precision (sensory attention) in permitting self-generated behaviour. To produce internally generated movements, I supplied the subject with prior beliefs that the internal hidden cause increased transiently to a value of one, with high sensory attenuation $\gamma = 6$. Figure 6.3 shows the results of this simulation. The lower left panel shows the internal hidden cause (blue line) with relatively tight 90% confidence intervals (grey areas), reflecting the relatively high log-precision on this hidden cause of six. Log-precisions are a convenient way of quantifying confidence or certainty about prediction errors and correspond to the logarithm of the associated precision or confidence. Prior beliefs about this hidden cause excite posterior beliefs about internally generated forces, while at the same time attenuating the precision of sensory prediction errors. This is reflected by the rise in the conditional expectation of the internal force (blue line in the upper right panel) and the transient increase in the confidence interval about this expectation, due to the attenuation of sensory precision. The resulting proprioceptive predictions are fulfilled by action, and they are sensed very accurately (shown in the upper left panel). Note that proprioceptive prediction (blue line) corresponds to somatosensory prediction (green line) and that both are close to the real values (broken black line). This simulation shows normal self-generated movement under permissive sensory attenuation.

Compare these results with the equivalent simulation when sensory attenuation was reduced from six to two (Figure 6.4). Here, the sensory attenuation leaves the sensory precision higher than the precision of the prior beliefs about internal hidden causes. This means that bottom-up sensory prediction errors predominate over top-down projections and the expected internal hidden force is profoundly suppressed – and inferred with a high degree of confidence. Because there are no predictions about proprioceptive changes, there is a consequent hypokinesia and failure of movement.

These simulations demonstrate the inherent trade-off between sensation and action that is necessary in free-energy minimisation models of the brain. There are two possible ways of minimising free-energy within these schemes – to change internal models of the world to represent current sensory input, or to predict future sensory input which will be achieved by moving. The simulations here demonstrate how these strategies are mediated by allocation of precision. High precision on prediction error prioritises sensation and the updating of internal models of the world, at the expense of movement, while high

precision on predictions of movement prioritises the fulfilling of those predictions through action, at the expense of sensation.

There is an interesting link between this simulation and a body of clinical, behavioural and experimental evidence regarding the impairment of movement by self-focussed attention; i.e., attending to the actual process of moving. Attention towards movement has been recognised as a major factor in the phenomenon of "choking" under pressure in professional sportspeople; where they are sometimes rendered unable to produce over-learned movements in a performance situation (Beilock & Carr, 2001). Less extreme versions of this phenomenon are part of normal experience: most of us can probably recall an incident when our movement has been impaired when we focus on it too much. This phenomenon has been described as "re-investment" in movement, and has been shown to impair performance and motor learning in a number of behavioural simulations (Maxwell et al., 2006; Chell et al., 2003; Zhu et al., 2011; Malhotra et al., 2012). Experimentally, asking healthy subjects to attend the production of an over-learned sequence of key presses impairs performance and elicits activation in prefrontal and anterior cingulate cortex, which is not activated during natural (unattended) sequence production (Jueptner et al., 1997). The suggestion, in light of this model, is that attending to the sensory consequences of movement increases the precision of sensory evidence, so that descending predictions of the intended proprioceptive state are foreshadowed by precise sensory prediction error – and movement is precluded. In other words, movement is imperceptible, for both the subject and any observer.

Figure 6.5 (solid line) shows the results of simulations repeated over a range of sensory attenuations, where γ was decreased from 6 to -4 and the internally generated force was recorded. As the prior precision increases in relation to sensory precision, prior beliefs are gradually able to incite more confident movement, with movement being around half its maximum amplitude when prior and sensory precision are in balance ($\gamma = 2$, vertical line). In short, this simple demonstration shows that sensory attenuation is necessary if prior beliefs are to supervene over sensory evidence, during self-generated behaviour. However, there is a price to be paid for the sensory attenuation, which is considered next:

3.5 Sensory attenuation and perception

Clearly, reducing the precision of sensory prediction errors reduces the posterior confidence in beliefs about their causes. Figure 6.3 shows an inflation of the posterior uncertainty (90% grey confidence intervals) due to sensory attenuation. The consequence

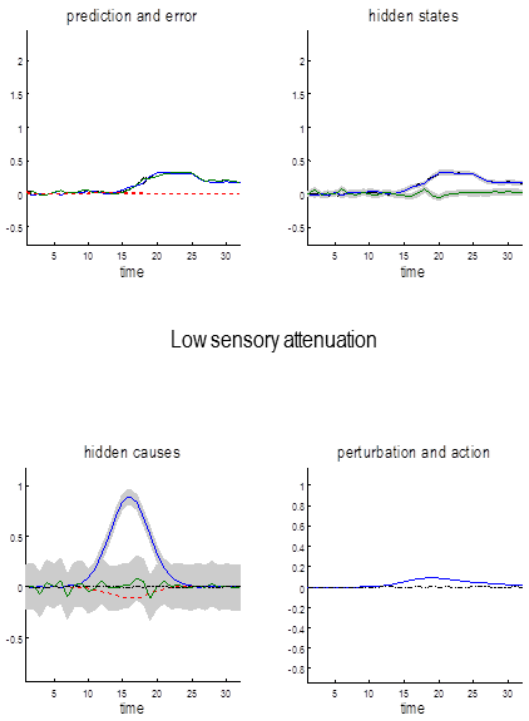

Figure 6.4: A simulation of akinesia. This figure uses the same format as previous figure but reports the results of simulations when sensory attenuation is much lower ($\gamma = 2$). In this case, bottom-up prediction errors retain a higher precision than descending predictions during movement. Conditional expectations that are updated by ascending prediction errors (upper right panel) overwhelm prediction errors based upon top-down predictions, and consequently infer that there is no change in the state of the world. This means that proprioceptive prediction errors are not produced (upper left panel) and action is profoundly suppressed (lower right panel).

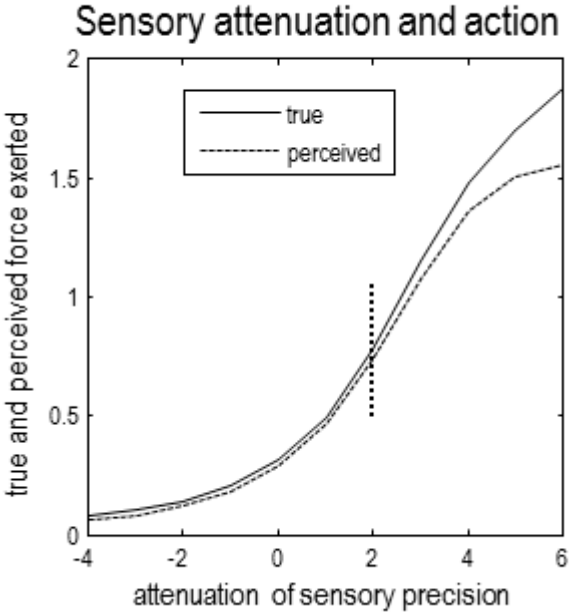

Figure 6.5: Movement and precision. True internally generated force $\mathbf{x_i}$ and perceived internally generated force (lower 90% confidence interval of x_i) simulated over a range of sensory attenuations, where $\gamma = \{6, \ldots -4\}$. Confident movement gradually emerges as the prior precision increases in relation to sensory precision, with movement being around half its maximum amplitude when prior and sensory precisions are balanced ($\gamma = 2$, vertical line). Force on the y axis is measured in arbitrary units.

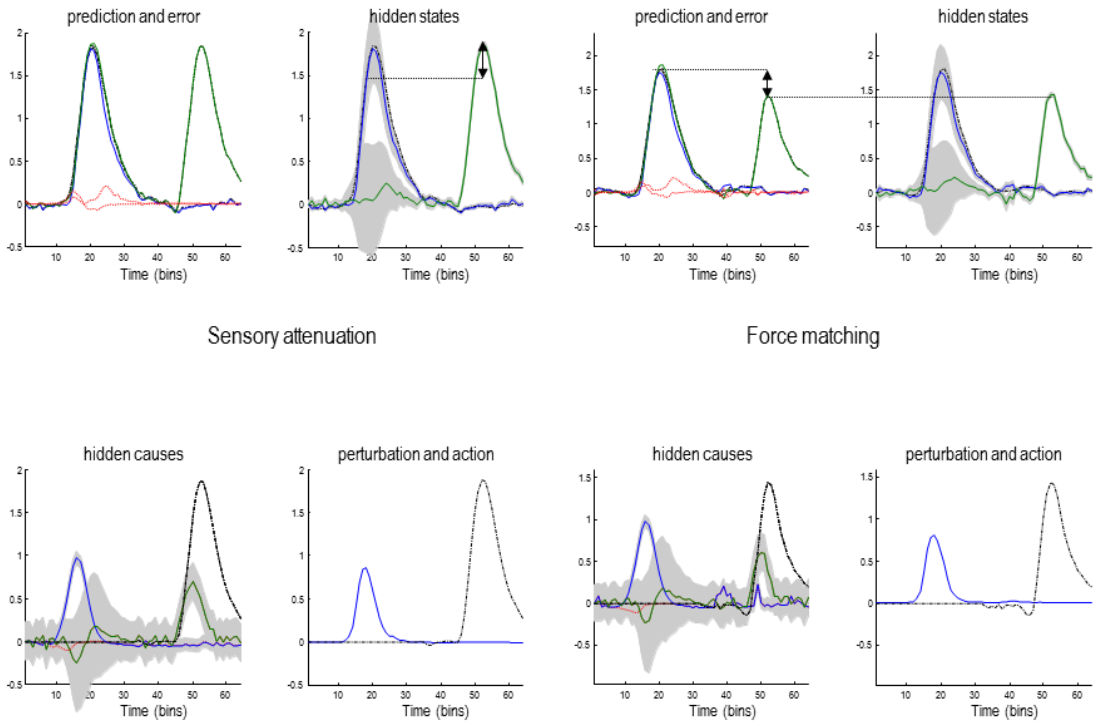

Figure 6.6: Simulation of the Force Match task. In the first part of this simulation (left hand panels), an internal force is generated (from a prior belief about the hidden cause v_i), followed by the presentation of an external force. The estimates of the hidden states (upper right panel) are similar, but the confidence interval around the force for the internally generated state is much broader. If perceptual inference is associated with the lower 90% confidence bound of the estimate of the hidden state, it will be lower when the force is self-generated (double-headed arrow, upper right panel). This is demonstrated in the right-hand panels. This is a simulation the force match paradigm where the external force is matched to the lower bound of the 90% confidence interval of the internal force. This means that internally generated force is now greater than the externally applied force (double-headed arrow, upper left panel).

of this transient uncertainty – due to a temporary suspension of attention to sensory input – provides a simple explanation for sensory attenuation in terms of psychophysical detection. This can be demonstrated fairly easily by presenting the forces generated by the subject exogenously and comparing the posterior beliefs about internal and external hidden states (forces). The left panels of figure 6.6 show the results of this simulation, in which there has been a veridical inference about the succession of internal and external hidden causes (blue and green lines in the lower left panels), with a reasonable degree of confidence. Furthermore, the predictions about internally and externally generated sensations are accurate and subtended by veridical conditional expectations. However, the confidence interval around the estimate of the internal hidden state is much greater than for the external hidden state. This means that if we asked the subject to reports somatosensory sensations at 90% confidence, the externally generated sensations would be detected much more readily than the internally generated sensations. This is the essence of sensory attenuation in psychophysical studies and – in this simulation – rests upon the inflation of the confidence interval associated with internally generated consequences. In other words, one would expect a reduction in *d*-prime for events that were self generated, relative to exactly the same events that were generated externally – as demonstrated experimentally (Cardoso-Leite et al., 2010). As this reduction in precision is applied to the entire sensory channel for the duration of the movement, a reduction in *d*-prime will also been seen for external stimuli produced during voluntary movement. This result has also been demonstrated experimentally (Juravle and Spence, 2011; Van Hulle et al., 2012). This attenuation is shown by the double headed arrow in Figure 6.6. Exactly the same interpretation can be applied to the force matching paradigm:

3.6 Sensory attenuation and the force matching illusion

The right-hand panels of figure 6.6 show exactly the same results as in the left hand panels; however here, I have yoked the exogenous force to the self generated force perceived at 90% confidence, (as opposed to the true force exerted by the subject). In other words, the external force corresponds to the force that would be reported by the subject to match the perceived force at 90% confidence. Crucially, the internally generated force is now much greater than the matched external force. This is the key finding in the force matching illusion and is entirely consistent with the sensory attenuation literature mentioned above. In this setting, the loss of confidence in posterior estimates of hidden states that are self generated translates into an illusory decrease in the intensity of percept

and hence an increase in the force applied, relative to the equivalent force in the absence of sensory attenuation.

To simulate the force-match paradigm, these simulations were repeated under different levels of self-generated forces by modulating the prior beliefs about the internal hidden cause (from a half to twice the normal amplitude). The results are shown in Figure 6.7 (blue line) by plotting the self-generated force against the yoked or matched external force with a corresponding 90% confidence interval. These results are remarkably similar to those obtained empirically (Shergill et al., 2003, 2005) and reveal sensory attenuation through an illusory increase in the self generated force, relative to matched forces over a wide range of forces. In the final simulations, I ask what would happen if subjects compensated for a failure in sensory attenuation by increasing the precision of their prior beliefs.

3.7 False inference and precision

To simulate pathology of sensory attenuation, sensory attenuation was reduced and – to compensate –precision of prediction errors at higher levels in the hierarchy was increased (by reducing sensory attenuation and increasing the log-precision of prediction errors on hidden states and causes by four log units). In the absence of sensory attenuation movement can only be elicited when there is a compensatory increase in the precision of proprioceptive predictions. In other words, beliefs about an intended movement have to be held with undue conviction (precision) to render them immune from contradictory sensory evidence that has not been attenuated.

These changes to precision mean that sensory attenuation is abolished, as indicated by the red line in Figure 6.7 (left panel). This reports the results of repeating the above force matching simulations over a range of internally generated forces but with a compensated loss in sensory attenuation. The resulting behaviour is very reminiscent of empirical results found in schizophrenia (right panel – Shergill et al 2003; 2005). One might ask why optimal subjects do not simply adopt this strategy and use very precise prior beliefs about hidden causes?

The answer is evident in Figure 6.8, which shows the results of a simulation with low sensory attenuation and augmented precisions at non-sensory levels of the generative model. Here, there is an almost perfect and precise inference about internally and externally generated sensations. However, there is a failure of inference about their

hidden causes. This can be seen on the lower left, where the subject has falsely inferred an antagonistic external hidden cause that mirrors the internal hidden causes: i.e., it believes that when it presses its finger on its hand, something also pushes its hand against its finger. Note that this false inference does not occur during normal sensory attenuation (see previous figures), where the true external hidden cause always lies within the 90% confidence intervals. The reason for this false inference or delusion is relatively simple: action is driven by proprioceptive prediction errors that always report less force than that predicted (if they did not, the reflex would not be engaged). However, when sensory precision increases, somatosensory prediction errors become very precise and need to be explained – and can only be explained by falsely inferring an opposing exogenous force. In more general terms, to reconcile a mismatch between the predicted consequences of action and the state of the world that precedes action, external forces are falsely invoked. This only occurs when both the predictions and their consequences are deemed to be very precise. This false inference could be interpreted as a delusion in the same sense that the sensory attenuation is an illusion. Having said this, it should be noted that – from the point of view of the subject – its inferences are Bayes-optimal. It is only our attribution of the inference as false that gives it an illusory or delusionary aspect. In the context of these simulations, the only difference between an illusion and a delusion is the level of the supposed failure of inference. Here, false inference at the perceptual level of hidden states is associated with illusions and false inference at the conceptual level of hidden causes with delusions.

3.8 Precision and psychopathology

Associating false inference at a conceptual level with delusions has some face validity in relation to empirical studies of the force matching illusion. This illusion is attenuated in normal subjects that score highly on ratings of delusional beliefs. Furthermore, subjects with schizophrenia – who are prone to positive symptoms like delusions – are less susceptible to the force matching illusion. In other words, there may be a trade-off between illusions at a perceptual level and delusions at a conceptual level that is mediated by a (failure of) sensory attenuation. A mechanistic contribution of the treatment in this chapter is to link sensory attenuation with putative neurobiological mechanisms that involve neuromodulatory changes in the gain of principal cells reporting prediction error. One important candidate for this modulation is the dopaminergic system, a classical ascending neuromodulatory transmitter system (Howes & Kapur, 2009).

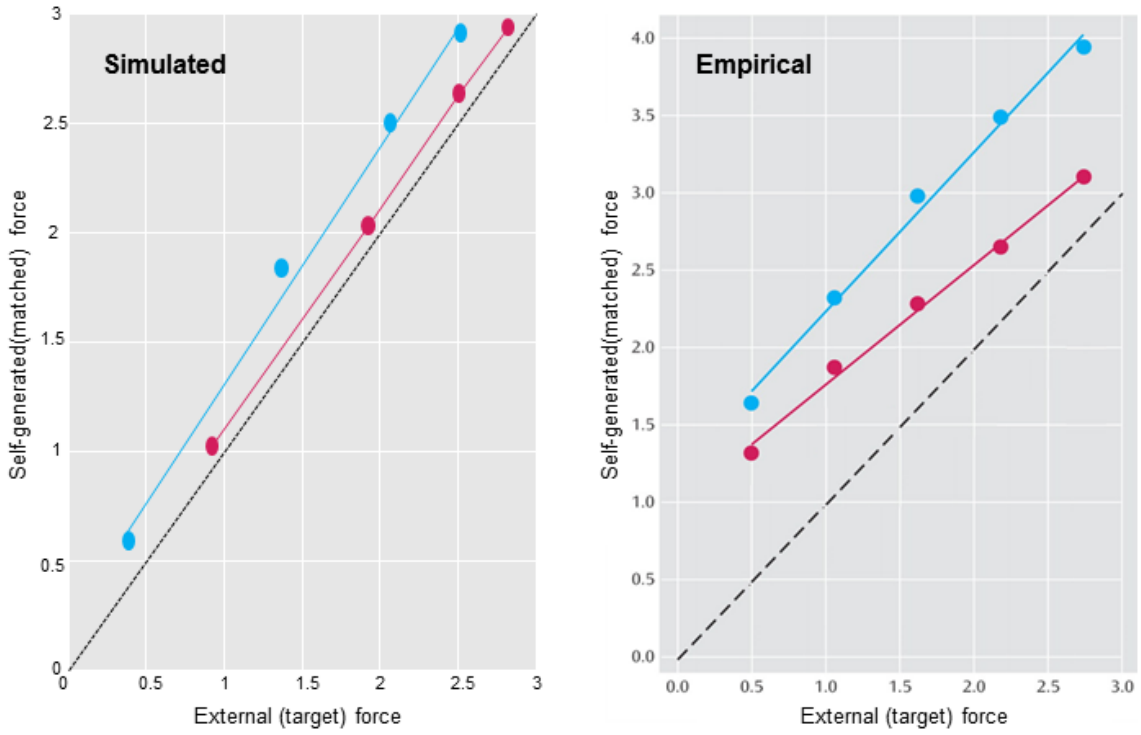

Figure 6.7: Sensory attenuation in schizophrenia. Left panel: results of the force-match simulation repeated under different levels of self-generated force. For normal levels of sensory attenuation (blue circles), internally produced force is higher than externally generated force at all levels of force, consistent with published data. Force matching typical of schizophrenia (red circles) was simulated by reducing sensory attenuation and increasing the precision of prediction errors at higher levels of the hierarchy. This resulted in a more veridical perception of internally generated force (small circles). **Right panel**: empirical results using the same format adapted (with permission) from (Shergill et al., 2003, 2005).

Loosely speaking, these simulation results are entirely consistent with two known pathologies in schizophrenia: the loss of sensory attenuation and a hyper-dopaminergic drive to the striatum in acute psychosis. I demonstrated earlier that an uncompensated loss of sensory attenuation results in an inability to move (Figure 6.4). This state is very reminiscent of the psychomotor poverty (catatonic) symptoms of schizophrenia (and other psychotic disorders) such as immobility, mutism, catalepsy and waxy flexibility. With waxy flexibility, patients may maintain a fixed posture for a long time, even though their limbs can be moved easily by an observer. Increased dopaminergic transmission in the striatum could increase the gain – i.e. precision – of prior beliefs about the causes of internally generated behaviour and may reflect a compensation for the loss of sensory attenuation (as in ther simulations above). A hyper-dopaminergic drive in schizophrenia could then lead to false inferences about external forces attributed to exogenous causes (such as in delusions of somatic passivity) or others in the acute psychotic state. Although it is overstretching the argument, it is tempting to equate the antagonistic aspect of falsely inferred hidden causes to the paranoid content of delusions that are typically seen in schizophrenia.

These simulations have several important similarities with some recent simulations of schizophrenic motor symptoms (Yamashita & Tani, 2012). In this work, the authors used a hierarchical predictive coding network to control a humanoid robot, and observed the effects of network lesions on both neural processing and behaviour. They showed that increasing the noise (i.e. decreasing the precision) in connections from higher to lower hierarchical areas could lead to catatonic motor symptoms, such as disorganised, stereotyped or loss of movements. Exactly the same effects are seen in similar models when reducing the precision of empirical priors in simulations of motor behaviour (Figure 13 in Friston et al., 2010a).

Finally, one might also speculate that the hypo-dopaminergic states seen in Parkinson's disease would produce similar symptoms, for slightly different reasons; here, sensory attenuation might be intact, but hypokinesia may reflect prior beliefs about self-generated movement that are held with insufficient precision and are overwhelmed by sensory evidence that the patient is not moving.

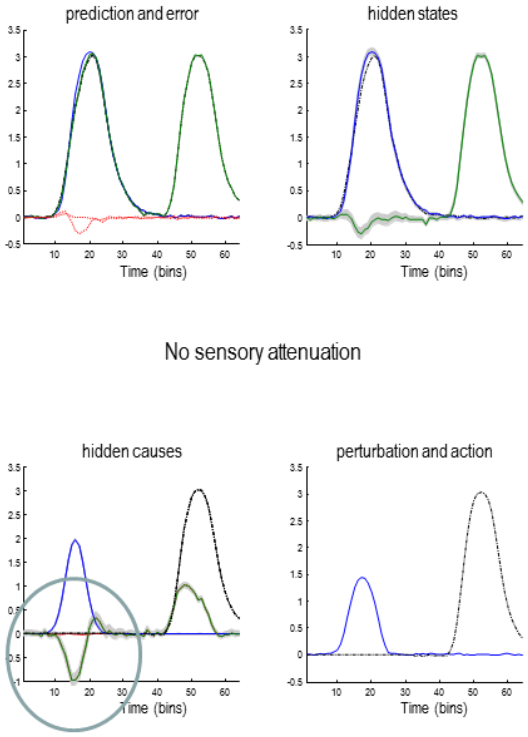

Figure 6.8: Pathology of sensory attenuation. To simulate the force-matching results seen in schizophrenia, sensory attenuation was reduced and precision at non-sensory levels of the hierarchy increased to allow movement. This results in a precise and accurate perception of internally and externally generated sensations (upper left panel). However, the causes of sensory data are not accurately inferred: an illusory cause (circled response in the lower left panel) is perceived during internally generated movement that is antagonistic to the movement. This is because the proprioceptive prediction errors driving action are rendered overly precise, meaning higher levels of the hierarchy must be harnessed to explain them, resulting in a 'delusion'.

4. Discussion

The ideas presented in this chapter suggest that attribution of agency – in an ambiguous situation – can be resolved by attenuating the precision of sensory evidence during movement: in other words, attending away from the sensations caused by self-made acts. When implemented in the context of active inference, this context–dependent attenuation provides a Bayes-optimal explanation for sensory attenuation in terms of perceptual psychophysics. Furthermore, it explains the force-matching illusion and reproduces quantitative results. Finally, if attenuation is withdrawn, the force-matching illusion disappears and false (delusional) inferences about agency emerge. This is important, given the negative correlation between sensory attenuation and predisposition to delusional beliefs in normal subjects and the resistance to the force matching illusion in schizophrenia. Active inference therefore links the neuromodulatory optimisation of precision to sensory attenuation and illusory phenomena during the attribution of agency in normal subjects. It also provides a functional account of deficits in syndromes characterised by false inference and impaired movement that are associated with abnormal neuromodulation.

This interplay between precision, attention, hierarchical inference and neuromodulation may also have important implications for functional movement disorders. It has previously been suggested that functional motor symptoms can be thought of as a pathological attention to predictions about movement that is mediated by abnormally high levels of precision in the motor hierarchy (Edwards at al., 2012). The results of these simulations make the strong prediction that patients with functional movement disorders should resemble people with schizophrenia and show no force matching illusion, which does seem to be the case (Parees et al., in preparation).

The work described here was published as:

Brown H, Adams RA, Parees I, Edwards M, Friston K. (2013). Active inference, sensory attenuation and illusions. Cogn. Process. 14:411-27.

Conclusion

This thesis demonstrates how the free-energy principle can be used to link levels of explanation in the brain. Marr (1982) suggested that complex systems in the brain should be understood on three levels of analysis – computational, algorithmic and physical implementation. The free energy principle is almost unique in offering linked explanations of complex behaviour at each of these levels. At the computational level, the free energy principle suggests that organisms attempt to reduce their entropy, or long-term average of surprise, by minimising their free energy. This suggests that hierarchical Bayesian modelling and predictive coding would be efficient strategies to use at the algorithmic level. In chapters 1, 3, 5 and 6 I demonstrate how models which minimise free energy using hierarchical Bayesian models and predictive coding, used either formally or informally, can model human behaviour, explaining how counterintuitive perceptual phenomena can be adaptive. Implementation of predictive coding in the cortex, constrained by cortical anatomy, suggests certain computational roles for forward and backwards connections and their cells of origin (Bastos, 2012). A test of these theories is provided in chapters 2 and 4.

Chapter 1 revisits the proposal that visual illusions are Bayes-optimal from the perspective of the free-energy principle. Using the Cornsweet illusion, I illustrate that this percept represents the most likely explanation for ambiguous sensory input, based on simple and plausible prior expectations about the spatial deployment of illuminance and reflectance. I also show that the contrast-dependence of the Cornsweet and Mach bands illusions emerges naturally within this simulation when precision, or signal-to-noise ratio, is used as a proxy for contrast. These simulations show how Bayesian modelling can be used to explain visual illusions based on simple and universal principles rather than complex and specific heuristics.

Chapter 2 tests a prediction of the model described in chapter 1 – that increasing contrast is encoded algorithmically as increasing precision on sensory input and prediction error. On the physical implementation level, this suggests that increasing contrast should increase the gain in superficial pyramidal cells, and this hypothesis was confirmed by dynamic causal modelling of EEG data. This validates the free-energy principle and predictive coding models of cortical function, as well as showing how principled computational modelling and dynamic causal modelling might be used to assay physical parameters in the brain such as cortical gain.

Chapter 3 demonstrates how a dynamic, nonlinear implementation of predictive coding can be used to give a principled explanation of directed attention phenomena. It also bridges the gap between behavioural and electrophysiological data by demonstrating how predictive coding schemes can give rise to biased completion-like behaviour. All the quantities required for making an inference have to be optimised – that is, both the conditional expectations and the precision with which the brain holds those expectations must be derived. These precisions encode the uncertainty or the amplitude of random fluctuations generating sensory information. By considering how states of the world influence uncertainty, it is clear that conditional expectations about states modulate their own precision. This leads naturally to competition and other nonlinear phenomena during perception.

The MEG study in chapter 4 demonstrates that the cortical responses seen during cued attention tasks can be modelled using just backwards modulation of intrinsic gain, giving face validity to the model. These computational parameters in the model correspond to the local intrinsic or recurrent inhibitory mechanisms that mediate cortical gain control. The results presented in this chapter provide an initial link between the computational imperatives of predictive coding, and plausible neuronal mechanisms at the level of synapses and cortical columns.

Chapter 5 extends these models to the motor system and describes how attention processes might operate in a motor cuing paradigm, by recasting motor preparation as attention directed towards proprioceptive sensations. This perspective suggests that attention should only optimise the precision of exteroceptive (e.g. visual) perceptual signals, but should also bias interoceptive inference during movement. The results presented in this chapter suggest that the bias for the selected action is mediated at a low level in the motor hierarchy, in an intrinsic (muscle-based) frame of reference.

Chapter 6 extends the formulation of attention as precision, and the reimagining of illusions under Bayes-optimal perception, to provide an account of sensory attenuation. The simulations also explore how movement, sensory attenuation and the attribution of agency might be dynamically linked concepts rather than having a clear order in which they are inferred by the brain. As well as providing a simulation of the force-match illusion, this model highlights the tension between reducing surprise by moving to alter the environment (action) or by changing models of the world to accommodate present sensations (perception). We extend these ideas to the clinical domain to discuss Parkinson's disease and schizophrenia as abnormalities of precision modulation.

In conclusion, the work presented in this thesis demonstrates that the free-energy principle is a powerful and flexible approach to explain brain function in a principled, evolutionarily plausible way. I demonstrate that reducing free energy by inferring the values of causes and parameters in hierarchical dynamic models using predictive coding is a plausible algorithmic task for the brain to be performing. I also show how these computations might be implemented in the brain, and how this can be tested empirically. I show how computational deficits in precision modulation might be an appropriate level of explanation for diseases such as schizophrenia. My future work will focus on testing the hypotheses about cortical function that follow from predictive coding schemes, specifically with respect to frequency responses in magnetoencephalography recordings.

Appendices

Appendix 1: Recognition from basic principles

The objective, given a model (brain), *m,* is to minimise the average uncertainty (entropy) about some generalised sensory states, $\tilde{s} = s \oplus s' \oplus s'' \oplus \ldots \in S$ it experiences ($\oplus$ means concatenation). Generalised states comprise the state itself, its velocity, acceleration, jerk, *etc*. This average uncertainty is

$$H(S \mid m) = -\int p(\tilde{s} \mid m) \ln p(\tilde{s} \mid m) d\tilde{s} \qquad \text{A.1}$$

Under ergodic assumptions, this is proportional to the long-term average of surprise, also known as negative log-evidence $-\ln p(\tilde{s}(t) \mid m)$

$$H(S \mid m) \propto -\int_0^T dt \ln p(\tilde{s}(t) \mid m) \qquad \text{A.2}$$

Minimising sensory entropy therefore corresponds to maximising the accumulated log-evidence for a model of the world. Although sensory entropy cannot be minimised directly, we can create an upper bound $\mathrm{S}(\tilde{s}, q) \geq H(S)$. This bound is induced with a recognition density $q(t) := q(\vartheta)$ on the causes (i.e., environmental states and parameters) of sensory signals. We will see later that these causes comprise time-varying states $u(t) \subset \vartheta$ and slowly varying parameters $\varphi(t) \subset \vartheta$. The recognition density is sometimes

called a proposal density and becomes the conditional density over causes, when it minimises the bound. The bound itself is the path-integral of free-energy $\mathrm{F}(t)$, which is created simply by adding a non-negative function of the recognition density to surprise:

$$\mathrm{S} = \int dt \mathrm{F}(t)$$

$$\begin{aligned} \mathrm{F}(t) &= -\ln p(\tilde{s}(t) \mid m) + D_{KL} \\ D_{KL} &= \left\langle \ln q(\vartheta) - \ln p(\vartheta \mid \tilde{s}, m) \right\rangle_q \end{aligned} \qquad \text{A.3}$$

This function is a Kullback-Leibler divergence D_{KL} and is greater than zero, with equality when $q(\vartheta) = p(\vartheta \mid \tilde{s}, m)$ is the true conditional density. This means that minimising free-energy, by changing $q(\vartheta)$, makes the recognition density an approximate conditional density on sensory causes. This is Bayes-optimal recognition. The free-energy can be evaluated easily because it is a function of the recognition density and a generative model $\mathrm{L}(t)$ entailed by m

$$\mathrm{F}(t) = \left\langle \mathrm{L}(t) \right\rangle_q - \mathrm{H}(t)$$

$$\begin{aligned} \mathrm{L}(t) &= -\ln p(\tilde{s}(t), \vartheta \mid m) \\ \mathrm{H}(t) &= -\left\langle \ln q(t) \right\rangle_q \end{aligned} \qquad \text{A.4}$$

The free-energy has been expressed here in terms of $\mathrm{H}(t)$, the negentropy of $q(t)$ and an energy $\mathrm{L}(t)$ expected under $q(t)$. The energy $\mathrm{L}(t)$ reports the surprise about sensations and their causes under a generative model. If we assume that recognition density $q(\vartheta) = \mathrm{N}(\mu, \mathrm{C})$ is Gaussian (known as the Laplace assumption), we can express free-energy in terms of the mean and covariance of the recognition density

$$\mathrm{F} = \mathrm{L}(\mu) + \tfrac{1}{2} tr(\mathrm{C}\mathrm{L}_{\mu\mu}) - \tfrac{1}{2} \ln |\mathrm{C}| - \tfrac{n}{2} \ln 2\pi e \qquad \text{A.5}$$

Where $n = \dim(\mu)$ and subscripts denote derivatives. We can now minimise free-energy with respect to the conditional precision (inverse covariance). The free-energy is minimised when $\mathrm{P} = \mathrm{C}^{-1} = \mathrm{L}_{\mu\mu} \Rightarrow \mathrm{F}_{\mathrm{C}} = 0 \Rightarrow \delta_{\mathrm{C}}\mathrm{S} = 0$ and allows us to eliminate C from Equation A.5

$$\mathrm{F} = \mathrm{L}(\mu) + \tfrac{1}{2}\ln|\mathrm{L}_{\mu\mu}| - \tfrac{n}{2}\ln 2\pi \qquad \text{A.6}$$

Crucially, this means the free-energy is only a function of the conditional mean or expectation. The expectations that minimise free-energy are the solutions to the following differential equations. For the generalised states $\tilde{u}(t) \subset \vartheta$

$$\begin{aligned}
\dot{\tilde{\mu}}^{(u)} &= \mathrm{D}\tilde{\mu}^{(u)} - \mathrm{F}_{\tilde{u}} \\
&\Leftrightarrow \\
\dot{\mu}^{(u)} &= \mu'^{(u)} - \mathrm{F}_{u} \\
\dot{\mu}'^{(u)} &= \mu''^{(u)} - \mathrm{F}_{u'} \\
\dot{\mu}''^{(u)} &= \mu'''^{(u)} - \mathrm{F}_{u''} \\
&\vdots
\end{aligned} \qquad \text{A.7}$$

Where D is a derivative matrix operator with identity matrices above the leading diagonal, such that $\mathrm{D}\tilde{u} = u' \oplus u'' \oplus \ldots$. Here and throughout, we assume all gradients are evaluated at the mean; here $\tilde{u} = \tilde{\mu}^{(u)}$. The stationary solution of Equation A.7 minimises free-energy and its path integral: $\dot{\tilde{\mu}}^{(u)} - \mathrm{D}\tilde{\mu}^{(u)} = 0 \Rightarrow \mathrm{F}_{\tilde{u}} = 0 \Rightarrow \delta_{\tilde{u}}\mathrm{S} = 0$. This ensures that when free-energy is minimised the mean of the motion is the motion of the mean; that is $\mathrm{F}_{\tilde{u}} = 0 \Rightarrow \dot{\tilde{\mu}}^{(u)} = \mathrm{D}\tilde{\mu}^{(u)}$.

For slowly varying parameters $\varphi(t) \subset \vartheta$ this motion disappears and we can use the following scheme

$$\dot{\mu}^{(\varphi)} = \mu'^{(\varphi)}$$
$$\dot{\mu}'^{(\varphi)} = -\mathrm{F}_{\varphi} - \kappa\mu'^{(\varphi)} \qquad \text{A.8}$$

Here, the solution $\dot{\tilde{\mu}}^{(\varphi)} = 0$ minimises free-energy, under constraint that the motion of the expected parameters is small: $\dot{\mu}^{(\varphi)} = \dot{\mu}'^{(\varphi)} = 0 \Rightarrow \mathrm{F}_{\varphi} = 0 \Rightarrow \delta_{\varphi}\mathrm{S} = 0$. The last equality $\delta_{\varphi}\mathrm{S} = 0$ just means that variations in the parameters do change the path integral of free-energy (cf, keeping to the floor of a valley to minimise the average height of ones path).

Equations A.7 and A.8 prescribe recognition dynamics for the expected states and parameters of the world respectively. The dynamics for states can be thought of as a gradient descent in a frame of reference that moves with the expected motion of the world (cf, surfing a wave). Conversely, the dynamics for the parameters can be thought of as a gradient descent that resists transient fluctuations with a damping term ($-\kappa\mu'^{(\varphi)}$), which instantiates our prior belief that the fluctuations in the parameters are small. We use $\kappa = N$, where N is the number of sensory samples.

In summary, we have derived recognition dynamics for expected states (in generalised coordinates of motion) and parameters, which cause sensory samples. The solutions to these equations minimise free-energy and therefore minimise a bound on surprise or (negative) log-evidence. Optimization of the expected states and parameters corresponds to perceptual inference and learning respectively. The precise form of the recognition depends on the energy $\mathrm{L}(t) = -\ln p(\tilde{s}(t), \vartheta \mid m)$ associated with a particular generative model. In what follows, we examine dynamic models of the world.

Appendix 2: Hierarchical Dynamic Models

We next introduce a very general model based on the hierarchal dynamic model discussed in Friston (2008). We will assume that any sensory data can be modelled with a special case of this model

$$
\begin{aligned}
s &= f^{(v)}(x, v, \theta) + z^{(v)} : z^{(v)} \sim \mathrm{N}\,(0, \Sigma^{(v)}(x, v, \gamma)) \\
\dot{x} &= f^{(x)}(x, v, \theta) + z^{(x)} : z^{(x)} \sim \mathrm{N}\,(0, \Sigma^{(x)}(x, v, \gamma))
\end{aligned}
\qquad \text{A.9}
$$

The nonlinear functions $f^{(u)} : u \in v, x$ represent a sensory mapping and equations of motion respectively and are parameterised by $\theta \subset \varphi$. The variables $v \subset u$ are referred to as hidden causes, while hidden states $x \subset u$ meditate the influence of the causes on sensory data and endow the system with memory. We assume the random fluctuations $z^{(u)}$ are analytic, such that the covariance of $\tilde{z}^{(u)}$ is well defined. Unlike our previous treatments (Friston 2008), this model allows for state-dependent changes in the amplitude of random fluctuations. It is this generalisation that furnishes a model of attention and introduces the key distinction between the effect of states on first and second-order sensory dynamics. These effects are meditated by the vector and matrix functions $f^{(u)} \in \Re^{dim(u)}$ and $\Sigma^{(u)} \in \Re^{dim(u) \times dim(u)}$ respectively, which are parameterised by first and second-order parameters $\{\theta, \gamma\} \subset \varphi$.

Under local linearity assumptions, the generalised motion of the sensory response and hidden states can be expressed compactly as

$$
\begin{aligned}
\tilde{s} &= \tilde{f}^{(v)} + \tilde{z}^{(v)} \\
\mathrm{D}\tilde{x} &= \tilde{f}^{(x)} + \tilde{z}^{(x)}
\end{aligned}
\qquad \text{A.10}
$$

Where the generalised predictions are

$$
\tilde{f}^{(u)} = \begin{bmatrix}
f^{(u)} = f^{(u)} \\
f'^{(u)} = f_x^{(u)} x' + f_v^{(u)} v' \\
f''^{(u)} = f_x^{(u)} x'' + f_v^{(u)} v'' \\
\vdots
\end{bmatrix}
\qquad \text{A.11}
$$

Equation A.10 means that Gaussian assumptions about the random fluctuations specify a generative model in terms of a likelihood and empirical priors on the motion of hidden states

$$\begin{aligned} p\left(\tilde{s} \mid \tilde{x}, \tilde{v}, \theta, m\right) &= \mathrm{N}\left(\tilde{f}^{(v)}, \tilde{\Sigma}^{(v)}\right) \\ p\left(D\tilde{x} \mid x, \tilde{v}, \theta, m\right) &= \mathrm{N}\left(\tilde{f}^{(x)}, \tilde{\Sigma}^{(x)}\right) \end{aligned} \qquad \text{A.12}$$

These probability densities are encoded by their covariances $\tilde{\Sigma}^{(u)}$ or precisions $\tilde{\Pi}^{(u)} := \tilde{\Pi}(x, v, \gamma^{(u)})$ with precision parameters $\gamma \subset \varphi$ that control the amplitude and smoothness of the random fluctuations. Generally, the covariances factorise; $\tilde{\Sigma}^{(u)} = V^{(u)} \otimes \Sigma^{(u)}$ into a covariance proper and a matrix of correlations $V^{(u)}$ among generalised fluctuations that encodes their smoothness.

Given this generative model we can now write down the energy as a function of the conditional means, which has a simple quadratic form (ignoring constants)

$$\mathrm{L} = \mathrm{L}^{(v)} + \mathrm{L}^{(x)} + \mathrm{L}^{(\varphi)}$$

$$\begin{aligned} \mathrm{L}^{(v)} &= \tfrac{1}{2}\tilde{\varepsilon}^{(v)T}\tilde{\Pi}^{(v)}\tilde{\varepsilon}^{(v)} - \tfrac{1}{2}\ln|\tilde{\Pi}^{(v)}| \\ \mathrm{L}^{(x)} &= \tfrac{1}{2}\tilde{\varepsilon}^{(x)T}\tilde{\Pi}^{(x)}\tilde{\varepsilon}^{(x)} - \tfrac{1}{2}\ln|\tilde{\Pi}^{(x)}| \\ \mathrm{L}^{(\varphi)} &= \tfrac{1}{2}\tilde{\varepsilon}^{(\varphi)T}\tilde{\Pi}^{(\varphi)}\tilde{\varepsilon}^{(\varphi)} - \tfrac{1}{2}\ln|\tilde{\Pi}^{(\varphi)}| \end{aligned} \qquad \text{A.13}$$

$$\begin{aligned} \tilde{\varepsilon}^{(v)} &= \tilde{s} - \tilde{f}^{(v)} \\ \tilde{\varepsilon}^{(x)} &= \mathrm{D}\tilde{\mu}^{(x)} - \tilde{f}^{(x)} \\ \tilde{\varepsilon}^{(\varphi)} &= \tilde{\mu}^{(\varphi)} - \tilde{\eta}^{(\varphi)} \end{aligned}$$

Here, the auxiliary variables $\tilde{\varepsilon}^{(j)} : j \in v, x, \varphi$ are prediction errors for sensory data, the motion of hidden states and parameters respectively. The predictions for the states are $\tilde{f}^{(u)}(\mu) : u \in v, x$ and the predictions for the parameters are the prior expectations $\tilde{\eta}^{(\varphi)}$. Equation A.13 assumes flat priors on the states and that priors $p(\varphi \mid m) = \mathrm{N}\left(\tilde{\eta}^{(\varphi)}, \tilde{\Sigma}^{(\varphi)}\right)$

on the parameters are Gaussian. We next consider hierarchical forms of this model. These are just special cases of Equation A.9, in which we make certain conditional independencies explicit. Although they may look more complicated, they are simpler than the general form above. They are useful because they provide an empirical Bayesian perspective on inference and learning that may be exploited by the brain. Hierarchical dynamic models have the following form

$$
\begin{aligned}
s &= f^{(v)}(x^{(1)}, v^{(1)}, \theta) + z^{(1,v)} \\
\dot{x}^{(1)} &= f^{(x)}(x^{(1)}, v^{(1)}, \theta) + z^{(1,x)} \\
&\vdots \\
v^{(i-1)} &= f^{(v)}(x^{(i)}, v^{(i)}, \theta) + z^{(i,v)} \\
\dot{x}^{(i)} &= f^{(x)}(x^{(i)}, v^{(i)}, \theta) + z^{(i,x)} \\
&\vdots \\
v^{(h-1)} &= \eta^{(v)} + z^{(h,v)}
\end{aligned}
\qquad \text{A.14}
$$

Again, $f^{(i,u)} := f^{(u)}(x^{(i)}, v^{(i)}, \theta) : u \in v, x$ are continuous nonlinear functions and $\eta^{(v)}(t)$ is a prior mean on the hidden causes at the highest level. The random terms $z^{(i,u)} \sim \mathrm{N}(0, \Sigma(x^{(i)}, v^{(i)}, \gamma^{(i,u)}))$ are conditionally independent and enter each level of the hierarchy. They play the role of observation error or noise at the first level and induce random fluctuations in the states at higher levels. The causes $v = v^{(1)} \oplus v^{(2)} \oplus \ldots$ link levels, whereas the hidden states $x = x^{(1)} \oplus x^{(2)} \oplus \ldots$ link dynamics over time. In hierarchical form, the output of one level acts as an input to the next. This input can enter nonlinearly to produce quite complicated generalised convolutions with deep (hierarchical) structure. This structure appears in the energy as empirical priors $\mathrm{L}^{(i,u)} : u \in x, v$ where, ignoring constants

$$\mathrm{L} = \sum_i \mathrm{L}^{(i,v)} + \sum_i \mathrm{L}^{(i,x)} + \mathrm{L}^{(\varphi)}$$

$$\begin{aligned}
\mathrm{L}^{(i,v)} &= \tfrac{1}{2}\tilde{\varepsilon}^{(i,v)T}\tilde{\Pi}^{(i,v)}\tilde{\varepsilon}^{(i,v)} - \tfrac{1}{2}\ln|\tilde{\Pi}^{(i,v)}| \\
\mathrm{L}^{(i,x)} &= \tfrac{1}{2}\tilde{\varepsilon}^{(i,x)T}\tilde{\Pi}^{(i,x)}\tilde{\varepsilon}^{(i,x)} - \tfrac{1}{2}\ln|\tilde{\Pi}^{(i,x)}|
\end{aligned} \qquad \text{A.15}$$

$$\begin{aligned}
\tilde{\varepsilon}^{(i,v)} &= \tilde{v}^{(i-1)} - \tilde{f}^{(i,v)} \\
\tilde{\varepsilon}^{(i,x)} &= D\tilde{x}^{(i)} - \tilde{f}^{(i,x)}
\end{aligned}$$

Note that the data enter the prediction errors at the lowest level; $\tilde{\varepsilon}^{(1,v)} = \tilde{s} - f^{(1,v)}$. At intermediate levels the prediction errors mediate empirical priors on the causes.

In summary, these models are as general as one could imagine; they comprise hidden causes and states, whose dynamics can be coupled with arbitrary (analytic) nonlinear functions. Furthermore, these states can be subject to random fluctuations with state-dependent changes in amplitude and arbitrary (analytic) autocorrelation functions. A key aspect is their hierarchical form, which induces empirical priors on the causes. In the next section, we look at the recognition dynamics entailed by this form of generative model, with a particular focus on how recognition might be implemented in the brain.

Appendix 3: Perception and attention

If we now write down the recognition dynamics (Equation A.7) using precision-weighted prediction errors $\xi^{(i,u)} = \tilde{\Pi}^{(i,u)}\tilde{\varepsilon}^{(i,u)}$ from Equation A.15, one can see the hierarchical message passing this scheme entails (ignoring the derivatives of the energy curvature):

$$\dot{\tilde{\mu}}^{(i,v)} = \mathrm{D}\tilde{\mu}^{(i,v)} + \chi_{\tilde{v}}^{(i,v)}\xi^{(i,v)} + \chi_{\tilde{v}}^{(i,x)}\xi^{(i,x)} + \lambda_{\tilde{v}}^{(i,v)} + \lambda_{\tilde{v}}^{(i,x)} - \xi^{(i+1,v)}$$
$$\dot{\tilde{\mu}}^{(i,x)} = \mathrm{D}\tilde{\mu}^{(i,x)} + \chi_{\tilde{x}}^{(i,v)}\xi^{(i,v)} + \chi_{\tilde{x}}^{(i,x)}\xi^{(i,x)} + \lambda_{\tilde{x}}^{(i,v)} + \lambda_{\tilde{x}}^{(i,x)} - \mathrm{D}^{T}\xi^{(i,x)}$$

$$\begin{aligned}\xi^{(i,v)} &= \tilde{\Pi}^{(i,v)}\tilde{\varepsilon}^{(i,v)} = \tilde{\Pi}^{(i,v)}(\tilde{\mu}^{(i-1,v)} - \tilde{f}^{(i,v)}) \\ \xi^{(i,x)} &= \tilde{\Pi}^{(i,x)}\tilde{\varepsilon}^{(i,x)} = \tilde{\Pi}^{(i,x)}(\mathrm{D}\tilde{\mu}^{(i,x)} - \tilde{f}^{(i,x)})\end{aligned} \qquad \text{A.16}$$

$$\begin{aligned}\chi_{w}^{(i,v)} &= \tilde{f}_{w}^{(i,v)T} - \tfrac{1}{2}\tilde{\varepsilon}^{(i,v)T}\tilde{\Omega}_{w}^{(i,v)} \\ \chi_{w}^{(i,x)} &= \tilde{f}_{w}^{(i,x)T} - \tfrac{1}{2}\tilde{\varepsilon}^{(i,x)T}\tilde{\Omega}_{w}^{(i,x)}\end{aligned}$$

Please note that this is a different form of Equation 0.3. Here, we have assumed the amplitude of random fluctuations is parameterised in terms of log-precisions, where

$$\begin{aligned}\tilde{\Pi}^{(i,u)} &= R^{(i,u)} \otimes diag(\exp(\pi^{(i,u)})) \\ \tilde{\Omega}^{(i,u)} &= I^{(i,u)} \otimes diag(\pi^{(i,u)})\end{aligned} \qquad \text{A.17}$$

The vector function $\pi^{(i,u)} := \pi(x, v, \gamma^{(i,u)}) : u \in v, x$ returns state-dependent log-precisions and $R^{(i,u)}$ is the inverse smoothness matrix $V^{(i,u)}$. In what follows we will quantify the amplitude (variance) of random fluctuations in terms of log-precisions, such that the associated variance is $\exp(-\pi^{(i,u)})$. With this particular form for the precisions, the terms $\tilde{\Omega}_{w}^{(i,u)}$ and $\lambda_{w}^{(i,u)} = tr(\tilde{\Omega}_{w}^{(i,u)})$ are constant for states $w \in \tilde{v}, \tilde{x}$ that affect the log-precisions linearly and zero if they have no effect.

Appendix 4: Integrating the recognition dynamics (generalised filtering)

Generalised filtering (Friston 2010b) involves integrating the ordinary differential Equations A.7 and A.8 to optimise the conditional means. We can simplify the numerics for hierarchical dynamic models by first collapsing over the hierarchy, then over generalised motion and finally over hidden causes and states:

$$\mu^{(v)} = \begin{bmatrix} \mu^{(1,v)} \\ \mu^{(2,v)} \\ \vdots \end{bmatrix} \quad f^{(v)} = \begin{bmatrix} f^{(1,v)} \\ f^{(2,v)} \\ \vdots \end{bmatrix} \quad \tilde{\varepsilon}^{(v)} = \begin{bmatrix} \tilde{s} \\ \tilde{\mu}^{(v)} \end{bmatrix} - \begin{bmatrix} \tilde{f}^{(v)} \\ \tilde{\eta}^{(v)} \end{bmatrix}$$

$$\mu^{(x)} = \begin{bmatrix} \mu^{(1,x)} \\ \mu^{(2,v)} \\ \vdots \end{bmatrix} \quad f^{(x)} = \begin{bmatrix} f^{(1,x)} \\ f^{(2,x)} \\ \vdots \end{bmatrix} \quad \tilde{\varepsilon}^{(x)} = \mathrm{D}\tilde{\mu}^{(x)} - \tilde{f}^{(u)}$$

$$\tilde{\varepsilon}^{(u)} = \begin{bmatrix} \tilde{\varepsilon}^{(v)} \\ \tilde{\varepsilon}^{(x)} \end{bmatrix} \quad \tilde{\Pi}^{(u)} = \begin{bmatrix} R^{(x)} \otimes diag(\exp(\pi^{(x)})) & \\ & R^{(x)} \otimes diag(\exp(\pi^{(x)})) \end{bmatrix} \quad \tilde{u} = \begin{bmatrix} \tilde{v} \\ \tilde{x} \end{bmatrix}$$

A.18

This gives a simple form for the (Gibbs) energy that comprises a log-likelihood and prior

$$\begin{aligned} \mathrm{L} &= \mathrm{L}^{(u)} + \mathrm{L}^{(\varphi)} \\ \mathrm{L}^{(u)} &= \tfrac{1}{2}\tilde{\varepsilon}^{(u)T}\tilde{\Pi}^{(u)}\tilde{\varepsilon}^{(u)} - \tfrac{1}{2}\ln|\tilde{\Pi}^{(u)}| \\ \mathrm{L}^{(\varphi)} &= \tfrac{1}{2}\tilde{\varepsilon}^{(\varphi)T}\tilde{\Pi}^{(\varphi)}\tilde{\varepsilon}^{(\varphi)} - \tfrac{1}{2}\ln|\tilde{\Pi}^{(\varphi)}| \end{aligned} \qquad \text{A.19}$$

with the following integration scheme

$$\dot{y} = \begin{bmatrix} \dot{\tilde{s}} \\ \dot{\tilde{\mu}} \end{bmatrix} = \begin{bmatrix} \dot{\tilde{s}} \\ \dot{\tilde{\mu}}^{(u)} \\ \dot{\mu}^{(\theta)} \\ \dot{\mu}^{(\gamma)} \\ \dot{\mu}'^{(\theta)} \\ \dot{\mu}'^{(\gamma)} \end{bmatrix} = \begin{bmatrix} \mathrm{D}\tilde{s} \\ \mathrm{D}\tilde{\mu}^{(u)} - \mathrm{F}_{\tilde{u}} \\ \mu'^{(\theta)} \\ \mu'^{(\gamma)} \\ -\mathrm{F}_{\theta} - \kappa\mu'^{(\theta)} \\ -\mathrm{F}_{\gamma} - \kappa\mu'^{(\gamma)} \end{bmatrix} \quad \Im = \frac{\partial \dot{y}}{\partial y} = \begin{bmatrix} \mathrm{D} & & & & & \\ -\mathrm{F}_{\tilde{u}\tilde{s}} & \mathrm{D} - \mathrm{F}_{\tilde{u}\tilde{u}} & & & & \\ & & & & I & \\ & & & & & I \\ -\mathrm{F}_{\theta\tilde{s}} & -\mathrm{F}_{\theta\tilde{u}} & -\mathrm{F}_{\theta\theta} & -\mathrm{F}_{\theta\gamma} & -\kappa & \\ -\mathrm{F}_{\gamma\tilde{s}} & -\mathrm{F}_{\gamma\tilde{u}} & -\mathrm{F}_{\gamma\theta} & -\mathrm{F}_{\gamma\gamma} & & -\kappa \end{bmatrix}$$

A.20

This system can be solved (integrated) using a local linearisation (Ozaki 1992) with updates $\Delta y = (\exp(\Delta t \Im) - I)\Im(t)^{-1}\dot{y}$ over time steps Δt, where $\Im(t)$ the filter's Jacobian. Note that we have omitted terms that mediate changes in the motion of state estimates due to changes in parameter estimates. This is because changes in parameter estimates are negligible at the time scale of changes in states. The requisite gradients (evaluated at the conditional expectation) are, with a slight abuse of notion when dealing with derivatives with respect to vectors

$$
\begin{aligned}
\mathrm{F}_{\tilde{u}} &= \tfrac{1}{2}\tilde{\varepsilon}^{(u)T}\tilde{\Pi}^{(u)}_{\tilde{u}}\tilde{\varepsilon}^{(u)} + \tilde{\varepsilon}^{(u)T}_{\tilde{u}}\tilde{\Pi}^{(u)}\tilde{\varepsilon}^{(u)} - \tfrac{1}{2}tr(\tilde{\Pi}^{(u)}_{\tilde{u}}\tilde{\Sigma}^{(u)}) + \tfrac{1}{2}tr(\mathrm{P}_{\tilde{u}}\mathrm{C}) \\
\mathrm{F}_{\gamma} &= \tfrac{1}{2}\tilde{\varepsilon}^{(u)T}\tilde{\Pi}^{(u)}_{\gamma}\tilde{\varepsilon}^{(u)} + \Pi^{(\gamma)}\mu^{(\gamma)} - \tfrac{1}{2}tr(\tilde{\Pi}^{(u)}_{\gamma}\tilde{\Sigma}^{(u)}) + \tfrac{1}{2}tr(\mathrm{P}_{\gamma}\mathrm{C}) \\
\mathrm{F}_{\theta} &= \tilde{\varepsilon}^{(u)T}_{\theta}\tilde{\Pi}^{(u)}\tilde{\varepsilon}^{(u)} + \Pi^{(\theta)}\mu^{(\theta)} + \tfrac{1}{2}tr(\mathrm{P}_{\theta}\mathrm{C})
\end{aligned}
\qquad \text{A.21}
$$

The corresponding curvatures are (neglecting second-order terms involving states and parameters and second-order derivatives of the conditional entropy)

$$
\begin{aligned}
\mathrm{F}_{\tilde{u}\tilde{s}} &\approx \mathrm{L}_{\tilde{u}\tilde{s}} = \tilde{\varepsilon}^{(u)T}_{\tilde{u}}\tilde{\Pi}^{(u)}\tilde{\varepsilon}^{(u)}_{\tilde{s}} \\
\mathrm{F}_{\gamma\tilde{s}} &\approx \mathrm{L}_{\gamma\tilde{s}} = \tilde{\varepsilon}^{(u)T}\tilde{\Pi}^{(u)}_{\gamma}\tilde{\varepsilon}^{(u)}_{\tilde{s}} \\
\mathrm{F}_{\theta\tilde{s}} &\approx \mathrm{L}_{\theta\tilde{s}} = \tilde{\varepsilon}^{(u)T}_{\theta}\tilde{\Pi}^{(u)}\tilde{\varepsilon}^{(u)}_{\tilde{s}} \\
\mathrm{F}_{\tilde{u}\tilde{u}} &\approx \mathrm{L}_{\tilde{u}\tilde{u}} = \tilde{\varepsilon}^{(u)T}_{\tilde{u}}\tilde{\Pi}^{(u)}\tilde{\varepsilon}^{(u)}_{\tilde{u}} + \tilde{\varepsilon}^{(u)T}_{\tilde{u}}\tilde{\Pi}^{(u)}_{\tilde{u}}\tilde{\varepsilon}^{(u)} + \tilde{\varepsilon}^{(u)}\tilde{\Pi}^{(u)}_{\tilde{u}}\tilde{\varepsilon}^{(u)}_{\tilde{u}} \\
\mathrm{F}_{\tilde{u}\theta} &\approx \mathrm{L}_{\tilde{u}\theta} = \tilde{\varepsilon}^{(u)T}_{\tilde{u}}\tilde{\Pi}^{(u)}\tilde{\varepsilon}^{(u)}_{\theta} + \tilde{\varepsilon}^{(u)T}\tilde{\Pi}^{(u)}_{\tilde{u}}\tilde{\varepsilon}^{(u)}_{\theta} \\
\mathrm{F}_{\theta\theta} &\approx \mathrm{L}_{\theta\theta} = \tilde{\varepsilon}^{(u)T}_{\theta}\tilde{\Pi}^{(u)}\tilde{\varepsilon}^{(u)}_{\theta} + \Pi^{(\theta)} \\
\mathrm{F}_{\gamma\theta} &\approx \mathrm{L}_{\gamma\theta} = \tilde{\varepsilon}^{(u)T}\tilde{\Pi}^{(u)}_{\gamma}\tilde{\varepsilon}^{(u)}_{\theta} \approx \mathrm{F}^{\ T}_{\theta\gamma} \\
\mathrm{F}_{\tilde{u}\gamma} &\approx \mathrm{L}_{\tilde{u}\gamma} = \tilde{\varepsilon}^{(u)T}_{\tilde{u}}\tilde{\Pi}^{(u)}_{\gamma}\tilde{\varepsilon}^{(u)} \\
\mathrm{F}_{\gamma\gamma} &\approx \mathrm{L}_{\gamma\gamma} = \tfrac{1}{2}\tilde{\varepsilon}^{(u)T}\tilde{\Pi}^{(u)}_{\gamma\gamma}\tilde{\varepsilon}^{(u)} + \Pi^{(\gamma)}
\end{aligned}
\qquad \text{A.22}
$$

$$
\begin{aligned}
\tilde{\Pi}^{(v)}_{w} &= R^{(v)} \otimes diag(\pi^{(v)}_{w} \times \exp(\pi^{(v)})) \\
\tilde{\Pi}^{(x)}_{w} &= R^{(x)} \otimes diag(\pi^{(x)}_{w} \times \exp(\pi^{(x)}))
\end{aligned}
$$

Finally, the conditional precision and its derivatives are given by the curvature of the (Gibbs) energy

$$C^{-1} = P = L_{\mu\mu} \approx \begin{bmatrix} L_{\tilde{u}\tilde{u}} & L_{\tilde{u}\theta} & L_{\tilde{u}\gamma} \\ L_{\theta\tilde{u}} & L_{\theta\theta} & L_{\theta\gamma} \\ L_{\gamma\tilde{u}} & L_{\gamma\theta} & L_{\gamma\gamma} \end{bmatrix}$$

$$P_{\tilde{u}} \approx \begin{bmatrix} & & \\ \tilde{\varepsilon}_{\theta\tilde{u}}^{(u)T}\tilde{\Pi}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} & 2\tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}^{(u)}\tilde{\varepsilon}_{\theta\tilde{u}}^{(u)} & \\ & & 0 \end{bmatrix} + \begin{bmatrix} 3\tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\tilde{u}}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} & 2\tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\tilde{u}}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} \\ 2\tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\tilde{u}}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} & \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\tilde{u}}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} \\ \tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} & \tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}^{(u)T}\tilde{\Pi}_{\gamma\gamma}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} \end{bmatrix}$$

$$P_{\theta} \approx \begin{bmatrix} 2\tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}^{(u)}\tilde{\varepsilon}_{\tilde{u}\theta}^{(u)} & \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}^{(u)}\tilde{\varepsilon}_{\tilde{u}\theta}^{(u)} & \\ \tilde{\varepsilon}_{\tilde{u}\theta}^{(u)T}\tilde{\Pi}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & & \\ & & 0 \end{bmatrix} + \begin{bmatrix} 2\tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\tilde{u}}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\tilde{u}}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} \\ \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\tilde{u}}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & & \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} \\ \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} & \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}^{(u)T}\tilde{\Pi}_{\gamma\gamma}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} \end{bmatrix}$$

$$P_{\gamma} \approx \begin{bmatrix} \tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} & \tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}_{\tilde{u}}^{(u)T}\tilde{\Pi}_{\gamma\gamma}^{(u)}\tilde{\varepsilon}^{(u)} \\ \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} & \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\gamma}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}_{\theta}^{(u)T}\tilde{\Pi}_{\gamma\gamma}^{(u)}\tilde{\varepsilon}^{(u)} \\ \tilde{\varepsilon}^{(u)T}\tilde{\Pi}_{\gamma\gamma}^{(u)}\tilde{\varepsilon}_{\tilde{u}}^{(u)} & \tilde{\varepsilon}^{(u)}\tilde{\Pi}_{\gamma\gamma}^{(u)}\tilde{\varepsilon}_{\theta}^{(u)} & \tilde{\varepsilon}^{(u)T}\tilde{\Pi}_{\gamma\gamma\gamma}^{(u)}\tilde{\varepsilon}^{(u)} \end{bmatrix}$$

A.23

Note that we have simplified the numerics here by neglecting conditional dependencies between the precisions and the states or parameters. These equations may look complicated but can be evaluated automatically using numerical derivatives. All the simulations in this paper used just one routine - **spm_LAP.m**. Demonstrations of this scheme are available as part of the SPM software (http://www.fil.ion.ion.ucl.ac.uk/spm; **DEM_demo.m**) and reproduce the examples in the figures.

Appendix 5: State-dependent noise and Weber's Law

Sensory signals are invariably registered as non-negative quantities (e.g., firing rates of photoreceptors). If we assume the sensory signals $s \approx \ln \varsigma$ are an approximate log-transform of some non-negative variables $\varsigma \in \Re^{+}$ sampled from a Poisson distribution with rate λ, we have from Equation 9 (and using a first-order Taylor expansion):

$$
\begin{aligned}
& s = f^{(v)} + z^{(v)} : s \sim \mathrm{N}\ (f^{(v)}, \Sigma^{(v)}) \\
& s = \ln \lambda + \frac{(\varsigma - \lambda)}{\lambda} \approx \ln \varsigma : \varsigma \sim Pois(\lambda) \Rightarrow s \sim \mathrm{N}\ (\ln \lambda, \lambda^{-1}) \Rightarrow \\
& f^{(v)} = \ln \lambda \\
& \Sigma^{(v)} = \lambda^{-1}
\end{aligned}
\qquad \text{A.25}
$$

This means that as the expected amplitude of the sensory input increases, $f^{(v)} = \ln \lambda$, so does its precision $\Pi^{(v)} = \lambda = \exp(f^{(v)})$.

Bibliography

Abrams J, Barbot A, Carrasco M. (2010). Voluntary attention increases perceived spatial frequency. Atten. Percept. Psychophys. 72:1510-21.

Adams RA, Shipp S, Friston KJ. (2012). Predictions not commands: active inference in the motor system. Brain Struct. Funct. 218:611-43.

Adams WJ, Graf, EW & Ernst MO. (2004). Experience can change the 'light-from-above' prior. Nat. Neurosci. 7:1057–1058.

Alais D & Burr D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Curr. Biol. 14:257–62.

Albrecht DG, Farrar SB, Hamilton DB. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. J Physiol. 347:713 -739.

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L. (2010). Stimulus predictability reduces responses in primary visual cortex. J Neurosci. 30:2960-6.

Aliu SO, Houde JF, Nagarajan SS. (2009). Motor-induced suppression of the auditory cortex. J Cogn. Neurosci. 21:791-802.

Allport DA. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In Heuer H and Sanders AF (eds.) *Perspectives on perception and action*, Hillsdale, NJ: Erlbaum. pp 395-419.

Anderson AJ, Carpenter RHS. (2006). Changes in expectation consequent on experience, modeled by a simple, forgetful neural circuit. J Vis. 6:822–835.

Anderson EJ, Dakin SC, Rees G. (2009). Monocular signals in human lateral geniculate nucleus reflect the Craik-Cornsweet-O'Brien effect. J Vis. 9:14-18.

Baess P, Jacobsen T, Schröger E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. Int. J Psychophysiol. 70:137-43.

Baess P, Widmann A, Roye A, Schröger E, Jacobsen T. (2009). Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. European J of Neurosci. 29:1514–1521.

Baess P, Horváth J, Jacobsen T, Schröger E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. Psychophysiology. 48:1276-83.

Baldauf D, Deubel H. (2010). Attentional landscapes in reaching and grasping. Vision Res. 50:999-1013.

Bays PM, Flanagan JR, Wolpert DM. (2006). Attenuation of Self-Generated Tactile Sensations Is Predictive, not Postdictive. PLoS Biol. 4: e28.

Bays PM, Wolpert DM, Flanagan JR. (2005). Perception of the consequences of self-action is temporally tuned and event driven. Curr. Biol. 15:1125-8.

Baldi P, Itti L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. Neural Netw. 23:649-66.

Ballard DH, Hinton GE, Sejnowski TJ. (1983). Parallel visual computation. Nature 306:21-6.

Barlow HB. (1974). Inductive inference, coding, perception, and language. Perception 3:123-34.

Bartolomeo P, Caroline Decaix C, Siéroff E. (2007). The phenomenology of endogenous orienting. Conscious. Cogn. 16:144–61.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. (2012). Canonical microcircuits for predictive coding. Neuron, 76:695-711.

Bauer F, Cheadle SW, Parton A, Müller HJ, Usher M. (2009). Gamma flicker triggers attentional selection without awareness. Proc. Natl. Acad. Sci. USA. 106:1666-71.

Beck DM, Kastner S. (2005). Stimulus context modulates competition in human extrastriate cortex. Nat. Neurosci. 8:1110–16.

Beilock SL, Carr TH. (2001). On the fragility of skilled performance: what governs choking under pressure? J Exp. Psychol. Gen 130:701-725.
Bestmann S, Harrison LM, Blankenburg F, Mars RB, Haggard P, Friston KJ, Rothwell JC. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. Curr. Biol. 18:775-80.
Bichot NP, Rossi AF, Desimone R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. Science 308:529-34.
Blakemore SJ, Wolpert DM, Frith CD. (1999). The Cerebellum Contributes to Somatosensory Cortical Activity during Self-Produced Tactile Stimulation. NeuroImage 10:448-459
Blakemore SJ, Wolpert DM, Frith CD (1998). Central cancellation of self-produced tickle sensation. Nat. Neurosci. 1:635-640.
Blakemore SJ, Smith J, Steel R, Johnstone CE, Frith CD. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. Psychol. Med. 30:1131-9.
Blakemore SJ, Frith CD, Wolpert DM. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. J Cogn. Neurosci. 11:551-9.
Blakemore SJ, Frith CD, Wolpert DM. (2001). The cerebellum is involved in predicting the sensory consequences of action. Neuroreport. 12:1879-84.
Blakemore SJ, Wolpert DM, Frith CD. (2002). Abnormalities in the awareness of action. Trends Cogn. Sci. 6:237-242.
Börgers C, Epstein S, Kopell NJ. (2005). Background gamma rhythmicity and attention in cortical local circuits: a computational study. Proc. Natl. Acad. Sci. USA 102:7002-7.
Brady N, Field DJ. (2000). Local contrast in natural images: normalisation and coding efficiency. Perception 29:1041-1055.
Broadbent DE. (1952). Listening to one of two synchronous messages. J Exp. Psychol. 44:51-5.
Broadbent DE. (1952). Failures of attention in selective listening. J Exp. Psychol. 44:428-33.
Broadbent DE. *Perception and communication.* New York: Pergamon Press, 1958.
Brodersen KH, Penny WD, Harrison LM, Daunizeau J, Ruff CC, Duzel E, Friston KJ, Stephan KE. (2008). Integrated Bayesian models of learning and decision making for saccadic eye movements. Neural Netw. 21:1247-60.
Brown H, Friston K, Bestmann S. (2011). Active inference, attention, and motor preparation. Front. Psychol. 2:218.
Brown H, Friston KJ. (2012). Free-energy and illusions: the cornsweet effect. Front. Psychol. 3:43.
Brown H, Friston KJ. (2012). Dynamic causal modelling of precision and synaptic gain in visual perception - an EEG study. Neuroimage. 63:223-31.
Brown HR, Friston KJ. (2013). The functional anatomy of attention: a DCM study. Front. Hum. Neurosci. 2013 7:784.
Brown H, Adams RA, Parees I, Edwards M, Friston K. (2013). Active inference, sensory attenuation and illusions. Cogn. Process. 14:411-27.
Bruce ND, Tsotsos JK. (2009). Saliency, attention, and visual search: an information theoretic approach. J Vis. 9:1-24.
Buchsbaum G Gottschalk A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. Proc. R. Soc. Lond. B Biol. Sci. 220:89-113.
Buhl EH, Tamás G, Fisahn A. (1998). Cholinergic activation and tonic excitation induce persistent gamma oscillations in mouse somatosensory cortex in vitro. J Physiol. 513:117-26.
Buia C, Tiesinga P. (2006). Attentional modulation of firing rate and synchrony in a model cortical network. J. Comput. Neurosci. 20:247–264.
Burton GJ, Moorhead IR. (1987). Color and spatial structure in natural scenes. Appl. Opt. 26:157-70.
Carandini M, Heeger DJ. (1994). Summation and division by neurons in primate visual cortex. Science 264:1333-6.
Carandini M, Heeger DJ, Senn W. (2002). A Synaptic Explanation of Suppression in Visual Cortex. J Neurosci. 22:10053-10065.

Cardoso-Leite P, Mamassian P, Schütz-Bosbach S, Waszak F. (2010). A new look at sensory attenuation. Action-effect anticipation affects sensitivity, not response bias. Psychol. Sci. 21:1740-5.
Carpenter RHS & Williams ML. (1995). Neural computation of log likelihood in control of saccadic eye movements. Nature. 377:59-62.
Carrasco M, Ling S, Read S. (2004). Attention alters appearance. Nat. Neurosci. 7:308-13.
Cave CR & Bichot NP. (1999). Visuospatial attention: Beyond a spotlight model. Psychon. B. Rev. 6:204-223.
Chapman CE, Bushnell MC, Miron D, Duncan GH, Lund JP. (1987). Sensory perception during movement in man. Exp. Brain Res. 68:516-24.
Chawla D, Lumer ED, Friston KJ. (1999). The relationship between synchronization among neuronal populations and their mean activity levels. Neural Comput. 11:1389-411.
Chawla D, Rees G, Friston KJ. (1999). The physiological basis of attentional modulation in extrastriate visual areas. Nat. Neurosci. 2:671-6.
Chelazzi L, Miller EK, Duncan J, Desimone R. (1993). A neural basis for visual search in inferior temporal cortex. Nature 363:345-7.
Chell BJ, Graydon JK, Crowley PL, Child M. (2003). Manipulated stress and dispositional reinvestment in a wall-volley task: an investigation into controlled processing. Percept. Motor Skills 97:435-48.
Chikkerur S, Serre T, Tan C, Poggio T. (2010). What and where: A Bayesian inference theory of attention. Vision Res. 50:2233-47.
Chung S, Li X, Nelson SB. (2002). Short-Term Depression at Thalamocortical Synapses Contributes to Rapid Adaptation of Cortical Sensory Responses In Vivo. Neuron, 34: 437-446.
Cisek P, Kalaska JF. (2010). Neural mechanisms for interacting with a world full of action choices. Annu. Rev Neurosci. 33:269-98.
Clark CR, Geffen GM, Geffen LB. (1989). Catecholamines and the covert orientation of attention in humans. Neuropsychologia 27:131-139.
Claxton G (1975). Why can't we tickle ourselves? Percept. Motor Skills 41:335-338.
Corlett PR, Honey GD, Krystal JH, Fletcher PC. (2010). Glutamatergic Model Psychoses: Prediction Error, Learning, and Inference. Neuropsychopharmacology 36:294-315.
Corlett PR, Frith CD, Fletcher PC. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacology, 206:515-530.
Cornsweet T. (1970). *Visual Perception*. New York: Academic Press.
Coull JT. (1998). Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. Prog. Neurobiol. 55:343-361.
Craighero L, Fadiga L, Rizzolatti G, Umilta C. (1999). Action for perception: a motor-visual attentional effect. J Exp. Psychol. Hum. Percept. Perform. 25:1673–1692.
Craik KJ. (1938). The effect of adaptation on differential brightness discrimination. J Physiol. 92:406-21.
Craik KJ. (1966). *The nature of psychology.* (Sherwood SL, ed.). Cambridge, UK: Cambridge UP.
Crick F. (1984). Function of the thalamic reticular complex: the searchlight hypothesis. Proc Natl Acad Sci USA 81:4586–4590.
Cui R-Q, Deecke L. (1999). High Resolution DC EEG of the Bereitschaftspotential Preceding Anatomically Congruent Versus Spatially Congruent Bimanual Finger Movements. Brain Topog. 12:117-27.
Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. Hum. Brain Mapp. 9: 183–191.
Dalley JW, McGaughy J, O'Connell MT, Cardinal RN, Levita L, Robbins TW. (2001). Distinct changes in cortical acetylcholine and noradrenaline efflux during contingent and noncontingent performance of a visual attentional task. J Neurosci. 21:4908-14.
Dalrymple KA, Kingstone A. (2010). Time to act and attend to the real mechanisms of action and attention. Br. J Psychol. 101:213-6.

Daunizeau J, Mattout J, Clonda D, Goulard B, Benali H, Lina JM. (2006). Bayesian spatio-temporal approach for EEG source reconstruction: conciliating ECD and distributed models. IEEE Trans. Biomed. Eng. 53:503-16.
Daunizeau J, David O, Stephan KE. (2009). Dynamic causal modelling: A critical review of the biophysical and statistical foundations. Neuroimage 58:312-22
Davidson MC, Marrocco RT. (2000). Local infusion of scopolamine into intraparietal cortex slows covert orienting in rhesus monkeys. J Neurophysiol. 83:1536-49.
Dayan P, Hinton GE, Neal RM. (1995). The Helmholtz machine. Neural Comput. 7:889-904.
DeBruyn EJ, Gajewski YA, Bonds AB. (1986). Anticholinesterase agents affect contrast gain of the cat cortical visual evoked potential. Neurosci. Lett. 71:311-316.
Decaix C, Siéroff E, Bartolomeo P. (2002). How voluntary is 'voluntary' orienting of attention? Cortex 38:841-5.
Deco G, Rolls ET. (2005). Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. J Neurophysiol. 94:295-313.
den Ouden HE, Daunizeau J, Roiser J, Friston KJ, Stephan KE. (2010). Striatal prediction error modulates cortical coupling. J Neurosci. 30:3210-9.
Desantis A, Weiss C, Schütz-Bosbach S, Waszak F. (2012). Believing and Perceiving: Authorship Belief Modulates Sensory Attenuation. PLoS One 7: e37959.
Donchin E, Coles MGH. (1988). Is the P300 component a manifestation of context updating? Behav Brain Sci. 11:355–72.
Desimone R, Gross CG. (1979). Visual areas in the temporal cortex of the macaque. Brain Res. 178:363-80.
Desimone R, Duncan J. (1995). Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18:193-222.
Desimone R. (1996). Neural mechanisms for visual memory and their role in attention. Proc. Natl. Acad. Sci. USA. 93:13494-9.
Desimone R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex Phil. Trans. R. Soc. Lond. B 353:1245-55.
Deubel H, Schneider WX. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. Vision Res. 36:1827-37.
Deutsch JA, Deutsch D. (1963). Attention: Some theoretical considerations. Psychol. Rev. 70:80-90.
Disney AA, Aoki C, Hawken MJ. (2007). Gain modulation by nicotine in macaque v1. Neuron 56:701-713.
Duncan J, Humphreys GW. (1989). Visual search and stimulus similarity. Psychol. Rev. 96:433-458.
Eagleman DM. (2001). Visual Illusions and Neurobiology. Nat. Rev. Neurosci. 2:920-6.
Edwards MJ, Adams RA, Brown H, Parees I, Friston KJ. (2012). A Bayesian Approach to "Hysteria". 135:3495-512.
Eckstein MP, Shimozaki SS, Abbey CK. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. J Vis. 2:25-45.
Eimer M. (1993). Spatial cueing, sensory gating and selective response preparation: an ERP study on visuo-spatial orienting. Electroencephalogr. Clin. Neurophysiol. 88:408–420.
Ernst MO, Banks MS. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.
Evans HL. (1975). Scopolamine effects on visual discrimination: modifications related to stimulus control. J Pharmacol. Exp. Therapeut. 195: 105-113.
Fagioli S, Hommel B, Schubotz RI. (2007). Intentional control of attention: action planning primes action-related stimulus dimensions. Psychol. Res. 71:22-9.
Fechner GT. (1860). Specielles zur Methode richtiger und falscher Fälle, in Anwendung auf die Gewichtsversuche. In *Elemente der Psychophysik* Leipzig: Breitkopf & Härtel.
Feldman AG, Levin MF. (1995). The origin and use of positional frames of reference in motor control. Behav. Brain Sci. 18:723-806.
Feldman H, Friston KJ. (2010). Attention, uncertainty, and free-energy. Front. Hum. Neurosci. 4:215.

Felleman DJ, Van Essen DC. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1:1-47.
Fernandez-Duque D & Posner MI. (1997). Relating the mechanisms of orienting and alerting. Neuropsychologia. 35:477-86.
Field DJ. (1987). Relations between the statistics of natural images and the response properties of cortical cells. J Opt. Soc. Am. A 12:2379-2394.
Formankiewicz MA, Mollon JD. (2009). The psychophysics of detecting binocular discrepancies of luminance. Vision Res. 2009 Jul;49(15):1929-38.
Friedman, D., Cycowicz, Y.M., Gaeta, H., 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neurosci Biobehav Rev. 25:355–373.
Fries P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci. Oct;9(10):474-80.
Fries P, Womelsdorf T, Oostenveld R, Desimone R. (2008). The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4. J Neurosci. Apr 30;28(18):4823-35.
Friston K. (2005). A theory of cortical responses. Philos. Trans. R Soc. Lond. B Biol. Sci. 360:815-36.
Friston K, Kilner J, Harrison L. (2006). A free energy principle for the brain. J Physiol. Paris 100:70-87.
Friston K. (2008). Hierarchical models in the brain. PLoS Comput. Biol. e1000211.
Friston K. (2009). The free-energy principle: a rough guide to the brain? Trends Cogn. Sci. :293-301.
Friston K, Kiebel S. (2009). Predictive coding under the free-energy principle. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364:1211-21.
Friston K, Kiebel S. (2009). Cortical circuits for perceptual inference. Neural Netw. 22:1093-104.
Friston KJ, Daunizeau J, Kiebel SJ. (2009). Active inference or reinforcement learning? PLoS One. 4:e6421.
Friston KJ, Daunizeau J, Kilner J, Kiebel SJ. (2010). Action and behavior: a free-energy formulation. Biol. Cybern. 102:227-60.
Friston K, Stephan KE, Li B, Daunizeau J. (2010). Generalised filtering. Math. Probl. Eng. 621670.
Friston K, Mattout J, Kilner J. (2011). Action understanding and active inference. Biol. Cybern. 104:137–160.
Friston K, Penny W. (2011). Post hoc Bayesian model selection. Neuroimage. 56:2089-99.
Friston KJ, Fitzgerald T, Galea JM., Adams R, Brown H, Dolan RJ, et al. (2012). Dopamine, affordance and active inference. PLoS Comput. Biol. 8:e1002327.
Frith C. (2005). The self in action: lessons from delusions of control. Conscious Cogn. 14:752-70.
Fründ I, Busch NA, Schadow J, Körner U, Herrmann CS. (2007). From perception to action: phase-locked gamma oscillations correlate with reaction times in a speeded response task. BMC Neurosci. 8:27.
Fuller S, Carrasco M. (2006). Exogenous attention and color perception: performance and appearance of saturation and hue. Vision Res. 46:4032-47.
Galazky I, Schütze H, Noesselt T, Hopf JM, Heinze HJ, Schoenfeld MA. (2009). Attention to somatosensory events is directly linked to the preparation for action. Journal of Neurological Science 279:93-8.
Geisler WS, Albrecht DG. (1992). Cortical neurons: isolation of contrast gain control. Vision Res. 32:1409-10.
Gentsch A, Schütz-Bosbach S. (2011). I did it: unconscious expectation of sensory consequences modulates the experience of self-agency and its functional signature. J Cogn. Neurosci. 23:3817-28.
Gherri E, Eimer M. (2010). Manual response preparation disrupts spatial attention: an electrophysiological investigation of links between action and attention. Neuropsychologia 48:961-9.
Gilbert CD, Wiesel TN. (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. Nature 280:120-125.
Ginzburg VL, Landau LD. (1950). On the theory of superconductivity. Zh. Eksp. Teor. Fiz. , 20, 1064.

Gobell J, Carrasco M. (2005). Attention alters the appearance of spatial frequency and gap size. Psychol. Sci. 16:644-51.
Goldberg ME, Segraves MA. (1987). Visuospatial and motor attention in the monkey. Neuropsychologia 25:107–18.
Golomb JD, Chun MM, Mazer JA. (2008). The native coordinate system of spatial attention is retinotopic. J Neurosci. 28:10654–10662.
Gómez CM, Delinte A, Vaquero E, Cardoso MJ, Vazquez M, Crommelynck M, Roucoux A. (2001). Current source density analysis of CNV during temporal gap paradigm. Brain Topogr. 13:149-159.
Gómez CM, Flores A, Digiacomo MR, Ledesma A, González-Rosa J, (2008). P3a and P3b components associated to the neurocognitive evaluation of invalidly cued targets. Neurosci. Lett. 430:181-185.
Goodman D, Kelso JAS. (1980). Are movements prepared in parts? Not under compatible (naturalized) conditions. J Exp. Psychol. Gen. 109:475–495.
Gordon IE. (1967). Stimulus probability and simple reaction time. Nature. 215:895-6.
Grabenhorst F, Rolls ET. (2010). Attentional modulation of affective versus sensory processing: functional connectivity and a top-down biased activation theory of selective attention. J Neurophysiol. 104, 1649-60.
Grafton ST, Hamilton AF. (2007). Evidence for a distributed hierarchy of action representation in the brain. Hum. Movem. Sci. 26:590-616.
Greenwald A. (1970). A double stimulation test of ideomotor theory with implications for selective attention. J Exp. Psych. 84:352-398.
Gregory RL. (1968). Perceptual illusions and brain models. Proc. R. Soc. Lond. B 171:179 - 196.
Gregory, RL. (1980). Perceptions as hypotheses. Phil. Trans. R. Soc. Lond. B 290:181-197.
Gross CG, Rocha-Miranda CE, Bender DB. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. J. Neurophysiol. 35:96-111.
Grossberg S, Hong S. (2006). A neural model of surface perception: lightness, anchoring, and filling-in. Spat Vis. 19:263-321.
Grossberg S. & Versace M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. Brain Res. 1218:278-312.
Gruber T, Müller MM, Keil A, Elbert T. (1999). Selective visual-spatial attention alters induced gamma band responses in the human EEG. Clin. Neurophysiol. 110:2074-85.
Haken H. (1983). Synergetics: An introduction. In: *Non-equilibrium phase transition and self-selforganisation in physics, chemistry and biology* (3rd ed.). Berlin: Springer Verlag.
Hart MA, Reeve TG. (2007). Equivalency of reaction times for simple and primed tasks. Acta Psychologica 125:291-300.
Hasselmo ME, Giocomo LM. (2006). Cholinergic modulation of cortical function. J Mol. Neurosci. 30:133-5.
Heeger DJ. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. J Neurophysiol. 70:1885-98.
Herrero JL, Roberts MJ, Delicato LS, Gieselmann MA, Dayan P, Thiele A. (2008). Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. Nature 454:1110–4.
Hesselmann G, Kell CA, Kleinschmidt A. (2008). Ongoing activity fluctuations in hMT+ bias the perception of coherent visual motion. J Neurosci. 28:14481-5.
Hesselmann G, Sadaghiani S, Friston KJ, Kleinschmidt A. (2010). Predictive coding or evidence accumulation? False inference and neuronal fluctuations. PLoS One. 5:e9926.
Hirayama J, Yoshimoto J, Ishii S. (2004). Bayesian representation learning in the cortex regulated by acetylcholine. Neural Netw. 17:1391-400.
Hommel B, Pratt J, Colzato L, Godijn R. (2001). Symbolic control of visual attention. Psychol. Sci. 12:360-5.
Howes OD, Kapur S (2009) The dopamine hypothesis of schizophrenia: version III–the final common pathway. Schizophr. Bull. 35: 549–562.

Hugdahl K, Nordby H. (1994). Electrophysiological correlates to cued attentional shifts in the visual and auditory modalities. Behav. Neural Biol., 62:21-32.
Hughes G, Waszak F. (2011). ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. Neuroimage 56:1632-40.
Humphreys GW, Yoon EY, Kumar S, Lestou V, Kitadono K, Roberts KL, Riddoch JM. (2010). Attention and its coupling to action. Br. J. Psychol. 101:217-9.
Itti L, Baldi P. (2006). Bayesian surprise attracts human attention. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (pp. 1–8). Cambridge, MA: MIT press.
Itti L & Baldi P. (2009). Bayesian surprise attracts human attention. Vision Res. 49:1295-306.
Jacobs RA. (1999). Optimal integration of texture and motion cues to depth. Vision Res. 39:3621-29.
James W. (1890). *The Principles of Psychology*. New York: Henry Holt, Vol. 1, pp. 403-404.
Jansen BH, Rit VG. (1995) Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern 73:357-366.
Jaramillo S, Pearlmutter BA. (2007). Optimal coding predicts attentional modulation of activity in neural systems. Neural Comput. 19:1295-312.
Jentzsch I, Leuthold H, Ridderinkhof KR. (2004). Beneficial effects of ambiguous precues: parallel motor preparation or reduced premotoric processing time? Psychophysiology 41:231-44.
Johnson DN, Weingartner HJ, Andreason P, George DT. (1995). An effect of triazolam on visual attention and information processing. Psychopharmacology, 121:145-149.
Jueptner M, Stephan KM, Frith CD, Brooks DJ, Frackowiak RS, Passingham RE. (1997). Anatomy of motor learning. I. Frontal cortex and attention to action. J Neurophysiol. 77: 1313-1324.
Juravle G, Spence C. (2011) Juggling reveals a decisional component to tactile suppression. Exp. Brain Res., 213: 87-97.
Kakei S, Hoffman DS, Strick PL. (1999). Muscle and movement representations in the primary motor cortex. Science 285:2136-9.
Kakei S, Hoffman DS, Strick PL. (2001). Direction of action is represented in the ventral premotor cortex. Nat. Neurosci. 4:1020-5.
Kakei S, Hoffman DS, Strick PL. (2003). Sensorimotor transformations in cortical motor areas. Neurosci. Res. 46:1-10.
Kastner S, De Weerd P, Desimone R, Ungerleider LG. (1998). Mechanisms of Directed Attention in the Human Extrastriate Cortex as Revealed by Functional MRI. Science 282:108-11.
Kastner S, De Weerd P, Pinsk MA, Elizondo MI, Desimone R & Ungerleider LG. (2001). Modulation of sensory suppression: Implications for receptive fields sizes in the human visual cortex. J Neurophysiol. 86:1398–1411.
Kastner S, Pinsk MA, De Weerd P & Desimone R, and Leslie G. Ungerleider (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751–61.
Kiebel SJ, Garrido MI, Moran RJ, Friston KJ. (2008). Dynamic causal modelling for EEG and MEG. Cogn. Neurodynamics 2: 121-136.
Kiebel SJ, von Kriegstein K, Daunizeau J, Friston KJ. (2009). Recognizing sequences of sequences. PLoS Comput. Biol. 5:e1000464.
Kok P, Rahnev D, Jehee JFM, Lau HC, de Lange FP. (2011). Attention Reverses the Effect of Prediction in Silencing Sensory Signals. Cereb. Cortex 22:2197-206.
Kolossa A, Fingscheidt T, Wessel K, Kopp B. (2012). A model-based approach to trial-by-trial p300 amplitude fluctuations. Front. Hum. Neurosci. 6:359.
Knill DC, Pouget A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. 27:712-9.
Knill DC, Saunders JA. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? Vision Res. 43:2539–2558.
Körding KP, Wolpert DM. (2004). Bayesian integration in sensorimotor learning. Nature. 427:244-7.

Langer MS, Bulthoff HH. (2001). A prior for global convexity in local shape-from-shading. Perception. 30:403-10.
Lavie N. (1995). Perceptual load as a necessary condition for selective attention. J Exp. Psychol. Hum. Percept. Perform. 21:451-68.
Liu T, Abrams J, Carrasco M. (2009). Voluntary attention enhances contrast appearance. Psychol. Sci. 20:354-62.
Lochner, JPA, Burger, JF. (1961). Form of the loudness function in the presence of masking noise. J. Acoust. Soc. Am. 33:1705–1707.
Lotto RB, Williams SM, Purves D (1999). Mach bands as empirically derived associations. Proc. Nat. Acad. Sci. 96:5245–50.
Lotto RB, Purves D. (2001). An empirical explanation of the Chubb illusion. J Cogn. Neurosci. 13:547-55.
Lotze RH. (1852). *Medizinische Psychologie oder die Physiologie der Seele* [Medical psychology or physiology of the soul] Weidmann'sche Buchhandlung; Leipzig.
Luck SJ, Chelazzi L, Hillyard SA, Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol. 77:24-42.
Macaluso E, Eimer M, Frith CD, Driver J. (2003). Preparatory states in crossmodal spatial attention: spatial specificity and possible control mechanisms. Exp Brain Res. 149(1):62-74.
Mach E. (1865). Über die Wirkung der raumlichen Verteilung des Lichtreizes auf die Netzhaut. Sitzungsberichte der Akademie der Wissenschaften in Wien, Mathematisch-naturwissenschaftliche Klasse 52:303-322.
Malhotra N, Poolton JM, Wilson MR, Ngo K, Masters RS. (2012). Conscious monitoring and control (reinvestment) in surgical performance under pressure. Surg. Endosc. 2012 26(9):2423-9.
Mangun GR, Hillyard SA. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. J Exp. Psychol. Hum. Percept. Perform. 17:1057-74.
Marr DC. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* New York: Freeman.
Marrocco RT, Witte EA, Davidson MC. (1994). Arousal systems. Curr. Opin. Neurobiol. 4:166-70.
Mars RB, Bestmann S, Rothwell JC, Haggard P. (2007). Effects of motor preparation and spatial attention on corticospinal excitability in a delayed-response paradigm. Exp. Brain Res. 182:125-9.
Mars RB, Debener S, Gladwin TE, Harrison LM, Haggard P, Rothwell JC, Bestmann S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. J Neurosci. 28:12539-45.
Martikainen MH, Kaneko K, Hari R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. Cereb Cortex. 15:299-302.
Mattout J, Phillips C, Penny WD, Rugg MD, Friston KJ. (2006). MEG source localization under multiple constraints: an extended Bayesian framework. Neuroimage, 30, 753-67.
Maunsell JH, van Essen DC. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. J Neurosci. 3:2563-2586.
Maunsell JH, Treue S. (2006). Feature-based attention in visual cortex. Trends Neurosci. 29:317-322.
Maxwell JP, Masters RS, Poolton JM. (2006). Performance breakdown in sport: the roles of reinvestment and verbal knowledge. Res. Q Exerc. Sport. 77(2):271-6.
Mazurek ME, Roitman JD, Ditterich J, Shadlen MN. (2003). A role for neural integrators in perceptual decision making. Cereb Cortex. 13:1257-69.
McAdams CJ, Maunsell JH. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J Neurosci. 19, 431-41.
McCormick DA, Prince DA. (1985). Two types of muscarinic response to acetylcholine in mammalian cortical neurons. Proc. Natl. Acad. Sci. USA 82:6344-8.
McCormick DA, Prince DA. (1986). Mechanisms of action of acetylcholine in the guinea-pig cerebral cortex in vitro. J Physiol. 375:169-94.

Meinke A, Thiel CM, Fink GR (2006). Effects of nicotine on visuo-spatial selective attention as indexed by event-related potentials. Neuroscience 141:201-212.
Miller J. (1982). Discrete versus continuous stage models of human information processing: in search of partial output. J Exp. Psychol. Hum. Percept. Perform. 8:273-96.
Milne RJ, Aniss AM, Kay NE, Gandevia SC. (1988). Reduction in perceived intensity of cutaneous stimuli during movement: a quantitative study. Exp. Brain Res. 70:569-76.
Montagna B, Carrasco M. (2006). Transient covert attention and the perceived rate of flicker. J Vis. 6:955-65.
Moray N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. Quart. J Exp. Psychol. 11:56-60.
Müller JR, Philiastides MG, Newsome WT. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. Proc. Natl. Acad. Sci. USA. 102:524-9.
Mumford D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol. Cybern. 66:241-51.
Nobre A, Correa A, Coull J. (2007). The hazards of time. Curr Opin Neurobiol. 17:465-70.
O'Brien V. (1959). Contrast by contour-enhancement. Am. J Psychol. 72:299–300.
O'Connor DH, Fukui MM, Pinsk MA, Kastner S. (2002). Attention modulates responses in the human lateral geniculate nucleus. Nat. Neurosci. 5:1203-9.
Ohzawa I, Sclar G, Freeman RD. (1982). Contrast gain control in the cat visual cortex. Nature 298:266-268.
Osman A, Moore C, Ulrich R. (1995). Bisecting RT with lateralized readiness potentials: Precue effects after LRP onset. Acta Psychol. 90:111–127.
Ozaki T (1992). A bridge between nonlinear time-series models and nonlinear stochastic dynamical systems: A local linearization approach. Statistica Sin. 2:113-135.
Parikh V, Kozak R, Martinez V, Sarter M. (2007). Prefrontal acetylcholine release controls cue detection on multiple timescales. Neuron 56:141-54.
Pavlova M, Birbaumer N, Sokolov A. (2006). Attentional modulation of cortical neuromagnetic gamma response to biological movement. Cereb. Cortex. 16:321-7.
Penny WD, Stephan KE, Mechelli A, Friston KJ. (2004). Comparing dynamic causal models. Neuroimage 22:1157-72.
Perchet C, Revol O, Fourneret P, Mauguière F, Garcia-Larrea L, (2001). Attention shifts and anticipatory mechanisms in hyperactive children: an ERP study using the Posner paradigm. Biol. Psychiatry. 50:44-57.
Perfetti B, Moisello C, Lanzafame S, Varanese S, Landsness EC, Onofrj M, et al. (2010). Attention modulation regulates both motor and non-motor performance: a high-density EEG study in Parkinson's disease. Arch. Ital. Biol. 148:279-88.
Pessoa L, Kastner S, Ungerleider LG. (2002). Attentional control of the processing of neutral and emotional stimuli. Cog. Brain Res. 15:31-45.
Peterson BS, Skudlarski P, Gatenby JC, Zhang H, Anderson AW, Gore JC. (1999). An fMRI study of Stroop word-color interference: evidence for cingulate subregions subserving multiple distributed attentional systems. Biol. Psych. 45:1237-1258.
Podzebenko K, Egan GF, Watson JDG (2005). Real and imaginary rotary motion processing: functional parcellation of the human parietal lobe revealed by fMRI. JCogn. Neurosci. 17:24-36.
Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. Clin. Neurophysiol. 118:2128–2148.
Posner MI, Nissen MJ, Ogden WC. (1978). Attended and unattended processing modes: The role of set for spatial location. In Pick HL, Saltzman NJ. (Eds), *Modes of Perceiving and Processing Information.* Hillsdale, NJ: Lawrence Erlbaum Associates
Posner MI. (1980). Orienting of attention. Quart. J Exp. Psychol. 32:3-25.
Posner MI, Snyder CR, Davidson BJ. (1980). Attention and the Detection of Signals. J Exp. Psychol: General 109:160-74.
Posner MI, Cohen Y (1984) Components of visual orienting. In *Attention and Performance Vol. X* (Bouma H, Bouwhuis D, eds.) Erlbaum.

Posner MI. (2008). Measuring alertness. Ann. N Y Acad. Sci. 1129:193-9.
Purves D, Shimpi A, Lotto RB. (1999). An empirical explanation of the Cornsweet effect. J Neurosci. 19:8542-51.
Purves D, Williams SM, Nundy S, Lotto RB. (2004). Perceiving the intensity of light. Psychol. Rev. 111:142-158.
Rafal RD, Calabresi PA, Brennan CW, Sciolto TK. (1989). Saccade Preparation Inhibits Reorienting to Recently Attended Locations. J Exp. Psychol: Hum. Percept. Perform. 15:673-85.
Rahnev D, Lau H, de Lange FP. (2011). Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. J Neurosci. 31:10741-10748.
Rahnev D, Maniscalco B, Graves T, Huang E, de Lange FP, Lau H. (2011). Attention induces conservative subjective biases in visual perception. Nat. Neurosci. 14:1513-1515.
Rao RP and Ballard DH. (1998). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive field effects. Nature Neuroscience 2:79-87.
Rao RP. (2005). Bayesian inference and attentional modulation in the visual cortex. Neuroreport. Nov 7;16(16):1843-8.
Rauss K, Schwartz S, Pourtois G. (2011). Top-down effects on early visual processing in humans: A predictive coding framework. Neurosci. Biobehav. Rev., 35: 1237-1253.
Ratliff F. (1965). *Mach Bands: Quantitative Studies on Neural Networks in the Retina.* Holden-Day, San Francisco.
Recanzone GH, Wurtz RH & Schwarz U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. J Neurophysiol. 78:2904-15.
Reeve TG, Proctor RW. (1984). On the advance preparation of discrete finger responses. J Exp. Psych. Hum. Percep. Perfor. 10:541–53.
Reynolds JH, Chelazzi L & Desimone R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci. 19:1736-53.
Reynolds JH, Heeger DJ. (2009). The normalization model of attention. Neuron. Jan 29;61(2):168-85.
Richards AM. (1968). Monaural loudness functions under masking. J Acoust. Soc. Am. 44:599-605.
Risko EF & Stolz JA. (2010) The proportion valid effect in covert orienting: Strategic control or implicit learning? Conscious. Cogn. 19:432-42.
Rizzolatti G. (1994). Nonconscious motor images. Behavioral and Brain Sciences 17:187-245.
Robinson DA. (1972). Eye movements evoked by collicular stimulation in the alert monkey. Vision Res. 12:1795-808.
Rockland KS, Pandya DN. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res. 179:3–20.
Rockstroh, B., Elbert, T., Birbaumer, N., Lutzenberger, W. (1982). *Slow brain potentials and behavior*: Urban & Schwarzenberg, Baltimore-Munich.
Rolls ET & Tovee MJ. (1995). The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. Exp Brain Res. 103:409-20.
Rosenbaum DA. (1980). Human movement initiation: specification of arm, direction, and extent. J Exp. Psychol. Gen. 109:444-74.
Rosenbaum DA, Kornblum S. (1982). A priming method for investigating the selection of motor responses. Acta psychol. 51:223-243.
Ruderman DL, Bialek W. (1994) Statistics of natural images: Scaling in the woods. 73:814-817.
Ruderman DL. (1997) Origins of scaling in natural images. Vision Res 37:3385-98.
Rushton DN, Rothwell JC, Craggs MD. (1981). Gating of somatosensory evoked potentials during different kinds of movement in man. Brain 104:465-91.
Rushworth MF, Nixon PD, Renowden S, Wade DT, Passingham RE. (1997). The left parietal and motor attention. Neuropsychologia 35:1261–1273.
Rushworth MF, Ellison A, Walsh V. (2001). Complementary localization and lateralizatio of orienting and motor attention. Nat. Neurosci. 4:656-61.

Salinas E, Sejnowski TJ. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. Neuroscientist. Oct;7(5):430-40.
Sara SJ. (1998). Learning by neurones: role of attention, reinforcement and behaviour. C R Acad Sci III. 321:193-8.
Sato A. (2008). Action observation modulates auditory perception of the consequence of others' actions. Conscious Cogn. 17:1219-27.
Schafer EW, Marcus MM. (1973). Self-stimulation alters human sensory brain responses. Science 181:175-7.
Schuermann B, Endrass T, Kathmann N. (2012). Neural correlates of feedback processing in decision-making under risk. Front. Hum. Neurosci. 6:204.
Schroeder CE, Mehta AD, and Foxe JJ. (2001). Determinants and mechanisms of attentional modulation of neural processing. Front Biosci. 6:D672-84.
Schrödinger E. (1992). *What is life?: With mind and matter and autobiographical sketches.* Cambridge University Press.
Seydell A, McCann BC, Trommershauser J & Knill DC. (2008). Learning Stochastic Reward Distributions in a Speeded Pointing Task. J Neurosci. 28:4356-67.
Shams L, Ma WJ, Beierholm U. (2005) Sound-induced flash illusion as an optimal percept. Neuroreport 16:1923-7.
Shergill SS, Samson G, Bays PM, Frith CD, Wolpert DM. (2005). Evidence for sensory prediction deficits in schizophrenia. Am. J Psychiatry 162:2384-6.
Shergill SS, Bays PM, Frith CD, Wolpert DM. (2003) Two eyes for an eye: the neuroscience of force escalation. Science 301(5630):187.
Shikata E, Hamzei F, Glauche V, Koch M, Weiller C, Binkofski F, Büchel C. (2003). Functional properties and interaction of the anterior and posterior intraparietal areas in humans. Eur. J Neurosci. 17: 1105-1110.
Shipp S. (2005). The importance of being agranular: a comparative account of visual and motor cortex. Philos. Trans. R Soc. Lond. B 360:797-814.
Shulman GL, Remington RW& McLean JP. (1979). Moving Attention through Visual Space. J Exp. Psychol: Hum. Percept. Perform. 3:522-526.
Slaghuis WL. (1998). Contrast sensitivity for stationary and drifting spatial frequency gratings in positive- and negative-symptom schizophrenia. J Abn. Psychol. 107:49-62.
Smith AT, Baker-Short, CM. (1993). Pharmacological separation of mechanisms contributing to human contrast sensitivity. Visual Neurosci. 10: 1073-1079.
Smith AT, Singh KD, Greenlee MW. (2000). Attentional suppression of activity in the human visual cortex. Neuroreport. Feb 7;11(2):271-7.
Sokolov A, Lutzenberger W, Pavlova M, Preissl H, Braun C, Birbaumer N. (1999). Gamma-band MEG activity to coherent motion depends on task-driven attention. Neuroreport. Jul 13;10(10):1997-2000.
Spratling MW (2008). Predictive-coding as a model of biased competition in visual attention. Vision Research 48:1391-1408.
Spratling MW. (2010). Predictive coding as a model of response properties in cortical area V1. J Neurosci. Mar 3;30(9):3531-43.
Srinivasan MV, Laughlin SB, Dubs A. M(1982). Predictive coding: A fresh view of inhibition in the retina. Proc. R. Soc. Lond. B Biol. Sci. 216:427-459.
Steinmetz PN, Roy A, Fitzgerald PJ, Hsiao SS, Johnson KO, Niebur E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. Nature. Mar 9;404(6774):187-90.
Stevens SS. (1966). Power transformations under glare, masking recruitment. J. Acoust. Soc. Am. 39:725–735.
Stevens SS, Guirao M. (1967). Loudness functions under inhibition. Percept. Psychophys. 2:459–465.
Stevens SS. (1967). Intensity functions in sensory systems. Int. J Neurol. 6:202-9.

Stokes M, Thompson R, Nobre AC, Duncan J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. Proc Natl Acad Sci U S A. Nov 17;106(46):19569-74.
Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nat. Neurosci, 11:1004-1006.
Talsma D, Mulckhuyse M, Slagter HA, Theeuwes J. (2007) Faster, more intense! The relation between electrophysiological reflections of attentional orienting, sensory gain control, and speed of responding. Brain Res. 1178:92-105.
Teufel C, Kingdon A, Ingram JN, Wolpert DM, Fletcher PC. (2010). Deficits in sensory prediction are related to delusional ideation in healthy individuals. Neuropsychologia. 48:4169-72.
Theeuwes J. (1991). Exogenous and endogenous control of attention: the effect of visual onsets and offsets. Percept Psychophys. 49:83-90.
Therrien AS, Richardson BA, Balasubramaniam R. (2011). Continuous theta-burst stimulation to primary motor cortex reduces the overproduction of forces following removal of visual feedback. Neuropsychologia. 49:2941-6.
Thiel CM, Zilles K, Fink GR (2005). Nicotine modulates reorienting of visuospatial attention and neural activity in human parietal cortex. Neuropsychopharmacology 30:810-820.
Tipper SP. (2004). Attention and Action. The Cognitive Neurosciences, 3rd Edition, pp619-630. MIT Press
Tolhurst DJ, Tadmor Y, Chao T. (1992) Amplitude spectra of natural images. Ophthalmic Physiol. 12:229-32.
Treisman A. (1964). Verbal cues, language and meaning in selective attention. Am. Jour. Psychol. 77:206-9.
Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12:97-136.
Treisman A & Schmidt H. (1982). Illusory conjunctions in the perception of objects. Cogn. Psychol. 14:107-41.
Treisman, A. (1998). Feature binding, attention and object perception. Philosophical Transactions of the Royal Society of London Series B: 353(1373):1295-1306.
Treue S. & Maunsell JHR. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. Nature 382:539-41.
Treue S, Martínez Trujillo JC. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. Nature. 399:575-9.
Treue S. (2004). Perceptual enhancement of contrast by attention. Trends Cogn. Sci. 8:435-7.
Trommershauser J, Maloney LT & Landy MS (2003b) Statistical decision theory and the selection of rapid, goal-directed movements. J Opt. Soc. Am. A 20:1419–1433.
Turatto M, Vescovi M, Valsecchi M. (2007). Attention makes moving objects be perceived to move faster. Vision Res. 47:166-78.
Van Hulle L, Juravle G, Spence C, Crombez G, Damme S. (2013) Attention modulates sensory suppression during back movements. Conscious. Cogn. 22: 420-429.
Vibell J, Klinge C, Zampini M, Spence C, Nobre AC. (2007). Temporal order is coded temporally in the brain: early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. J Cogn Neurosci. Jan;19(1):109-20.
Vidal JR, Chaumon M, O'Regan JK, Tallon-Baudry C. Visual grouping and the focusing of attention induce gamma-band oscillations at different frequencies in human magnetoencephalogram signals. J Cogn Neurosci. 2006 Nov;18(11):1850-62.
Voss M, Ingram JN, Haggard P, Wolpert DM. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. Nat. Neurosci. 9:26-7.
Voss M, Bays PM, Rothwell JC, Wolpert DM. (2007). An improvement in perception of self-generated tactile stimuli following theta-burst stimulation of primary motor cortex. Neuropsychologia. 45:2712-7.
Voss M, Ingram JN, Wolpert DM, Haggard P. (2008). Mere expectation to move causes attenuation of sensory signals. PLoS One. 2008 3:e2866.

Vossel, S., Thiel, C.M., Fink, G.R. (2006). Cue validity modulates the neural correlates of covert endogenous orienting of attention in parietal and frontal cortex. Neuroimage. 32:1257-1264.
Vossel S, Thiel CM, Fink GR. (2008). Behavioral and neural effects of nicotine on visuospatial attentional reorienting in non-smoking subjects. Neuropsychopharmacology. Mar;33(4):731-8.
Voytko ML, Olton DS, Richardson RT, Gorman LK, Tobin JR & Price DL. (1994). Basal forebrain lesions in monkeys disrupt attention but not learning and memory. J Neurosci. 14:167-86.
Walter, W.G., Cooper, R., Aldridge, W.J., McCallum, W.C. (1964). Contingent negative variation: An electrophysiological sign of sensoriomotor association and expectancy in the human brain. Nature. 203:380-384.
Watson AB, Pelli DG. (1983). QUEST: a Bayesian adaptive psychometric method. Percept Psychophys. 33:113-20.
Weber E. (1846) Der Tastsinn and das Gemeingefuhl. In *Handwörterbuch der Physiologie.* (Wagner R – ed.) Leipzig: Wilhelm Engelmann
Weiskrantz L, Elliott J, Darlington C (1971). Preliminary Observations on Tickling Oneself. Nature 230:598-599.
Weiss C, Herwig A, Schütz-Bosbach S. (2011). The self in action effects: selective attenuation of self-generated sounds. Cognition. 121:207-18.
Weiss C, Herwig A, Schütz-Bosbach S. (2011) The Self in Social Interactions: Sensory Attenuation of Auditory Action Effects Is Stronger in Interactions with Others. PLoS ONE 6:e22723.
Weiss Y, Simoncelli EP, Adelson EH. (2002) Motion illusions as optimal percepts. Nat Neurosci. 5:598-604.
Whiteley L, Sahani M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. J Vis. Mar 6;8(3):2.1-15.
Wilquin H, Delevoye-Turrell Y. (2012). Motor agency: a new and highly sensitive measure to reveal agency disturbances in early psychosis. PLoS One. 7:e30449.
Witte EA, Davidson MC & Marrocco RT. (1997). Effects of altering brain cholinergic activity on covert orienting of attention: comparison of monkey and human performance. Psychopharmacology (Berl). 132:324-34.
Woldorff M, Fox P, Matzke M, Lancaster J, Veeraswamy J, Zamarripa F, Seabolt M, Glass T, Gao J, Martin C, Jerabek P. (1997). Retinotopic organization of the early visual spatial attention effects as revealed by PET and ERPs. Hum Brain Map. 5:280–6.
Wolpert, DM, Ghahramani Z, & Jordan MI. (1995). An internal model for sensorimotor integration. Science 269:1880-2.
Wolpert DM, Flanagan JR. (2001). Motor prediction. Curr. Biol. 11:R729-32.
Womelsdorf T and Fries P. (2006). Neuronal coherence during selective attentional processing and sensory-motor integration. J Physiol Paris. Oct;100(4):182-93.
Womelsdorf T, Fries P, Mitra PP & Desimone R. (2006) Gamma-band synchronization in visual cortex predicts speed of change detection. Nature 439:733-6.
Wonnacott S. (1997). Presynaptic nicotinic ACh receptors. Trends Neurosci. Feb;20(2):92-8.
Woodworth RS. (1899). The accuracy of voluntary movement. Psychol Rev. 3:1-119.
Wu W, Hatsopoulos NG. (2007). Coordinate system representations of movement direction in the premotor cortex. Exp. Brain Res. 176:652-7..
Wyart V, Nobre AC, Summerfield C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. Proc. Natl. Acad. Sci. USA. 109:3593-8
Xiang Z, Huguenard JR, Prince DA. (1998). Cholinergic switching within neocortical inhibitory networks. Science. 28:985-8.
Xiong S, Goonetilleke RS, Rodrigo WD, Zhao J. (2013). A model for the perception of surface pressure on human foot. Appl. Ergon. 44:1-10.
Yamashita Y, Tani J. (2012). Spontaneous prediction error generation in schizophrenia. PloS One: e37843.
Yantis S, Jonides J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. J Exp. Psychol: Hum. Percept. Perform. 10:601-21.
Yu AJ and Dayan P. (2005). Uncertainty, neuromodulation and attention. Neuron. 46:681-692.

Yuille A, Geiger D, Bulthoff H. (1991). Stereo integration, mean field theory and psychophysics. Network 2:423–442.
Zeitler M, Fries P, Gielen S. (2008). Biased competition through variations in amplitude of gamma-oscillations. J Comput Neurosci. Aug;25(1):89-107.
Zhu FF, Poolton JM, Wilson MR, Maxwell JP, Masters RS. (2011). Neural co-activation as a yardstick of implicit motor learning and the propensity for conscious control of movement. Biol. Psychol. 87:66-73.
Zilles K, Palomero-Gallagher N, Schleicher A. (2004). Transmitter receptors and functional anatomy of the cerebral cortex. J Anat. Dec;205(6):417-32.