Computational Modelling of Information Gathering

Muammer Berk Mirza

University College London

A thesis submitted for the degree of
Doctor of Philosophy

February 2019

I Muammer Berk Mirza 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.

Thesis materials

The second and the third chapters have been drawn and derived from (Mirza, Adams, Mathys, & Friston, 2016).

Mirza, M. B., Adams, R. A., Mathys, C. D., & Friston, K. J. (2016). Scene Construction, Visual Foraging, and Active Inference. Front Comput Neurosci, 10, 56. doi:10.3389/fncom.2016.00056

The fourth chapter has been drawn and derived from *(Mirza, Adams, Mathys, & Friston, 2018).*

Mirza MB, Adams RA, Mathys C, Friston KJ. Human visual exploration reduces uncertainty about the sensed world. PloS one. 2018 Jan 5;13(1):e0190429.

The fifth chapter has been drawn and derived from Mirza et al., in submission

Mirza, M. B., Adams, R. A., Parr, T., & Friston, K. J. Selective attention and active inference

The sixth chapter has been drawn and derived from (Mirza, Adams, Parr, & Friston, 2019)

Mirza, M. B., Adams, R. A., Parr, T., & Friston, K. (2019). Impulsivity and Active Inference. *Journal of cognitive neuroscience*, 31(2), 202-220. https://www.mitpressjournals.org/doi/abs/10.1162/jocn_a_01352

Abstract

This thesis describes computational modelling of information gathering behaviour under active inference – a framework for describing Bayes optimal behaviour. Under active inference perception, attention and action all serve for same purpose: minimising variational free energy. Variational free energy is an upper bound on surprise and minimising it maximises an agent's evidence for its survival. An agent achieves this by acquiring information (resolving uncertainty) about the hidden states of the world and uses the acquired information to act on the outcomes it prefers. In this work I placed special emphasis on the resolution of uncertainty about the states of the world. I first created a visual search task called scene construction task. In this task one needs to accumulate evidence for competing hypotheses (different visual scenes) through sequential sampling of a visual scene and categorising it once there is sufficient evidence. I showed that a computational agent attends to the most salient (epistemically valuable) locations in this task. In the next, this task was performed by healthy humans. Healthy people's exploration strategies provided evidence for uncertainty driven exploration. I also showed how different exploratory behaviours can be characterised using canonical correlation analysis. In the next study I showed how exploration of a visual scene under different instructions could be explained by appealing to the computational mechanisms that may correspond to attention. This entailed manipulating the precision of task irrelevant cues and their hidden causes as a function of instructions. In the final work, I was interested in characterising impulsive behaviour using a patch leaving paradigm. By varying the parameters of the MDP model, I showed that there could be at least three distinct causes of impulsive behaviour, namely a lower depth of planning, a lower capacity to maintain and process information, and an increased perceived value of immediate rewards.

Impact statement

The work presented in this thesis uses a number of discrete space models of active inference to describe how events evolve through time, giving rise to categorical outcomes at each discrete time point. The behaviour under these models depends upon prior beliefs and different combinations of these prior beliefs can account for diverse human behaviours. This makes this framework suitable for studying the differences between individuals in terms of their prior beliefs. Given that a number of clinical disorders (including mental disorders) are linked with abnormal prior beliefs, this framework can be used to study psychopathologies as well. Further advances in these models may allow them to be used as diagnostic tools in the future.

The models presented here are not implausible in the sense that the hidden state space is factorised in a fashion that bears a resemblance to functional segregation in the brain. An example of this is given in the models of visual search presented in this thesis. These models make a distinction between ‘what’ and ‘where’ attributes of a visual scene. This is similar to the functionally segregated ventral (what) and dorsal (where) streams in the brain. Using the same principles used in these models one can create neurobiologically plausible models.

Active inference describes a neuronal process theory by minimisation of variational free energy. This framework proposes that neuronal responses can be described in terms of belief propagation which rests upon a gradient descent on variational free energy. Under neurobiologically plausible models one can use these belief updates to predict the electrophysiological responses one might expect in empirical studies. Moreover, one can fit models to the electrophysiological responses from empirical studies (in a similar fashion as fitting models to behavioural responses) to estimate subject (or group) specific prior beliefs, which would allow for studying individual and/or group level differences.

In summary, the work presented here creates an opportunity to answer many other research questions by using model simulations and model fitting. Ultimately, these methods have the prospect of being used as a diagnostic tool, distinguishing between different clinical conditions, providing better and more customised clinical therapies.

Acknowledgements

I am grateful to my supervisors Rick Adams and Karl Friston for their guidance and support throughout my PhD. I also would like to thank my colleague Thomas Parr for his friendship and help. I would like express my gratitude to Chris Mathys for his help in the first stages of my PhD.

Among many other people at the Wellcome Centre for Human Neuroimaging, I would like to thank Kamlyn, Ramkissoon for always helping me and being a safety net, Shiv Jhala, Monica Bumbury and Maddy Scott for their friendship, Chris Freemantle, Liam Reilly and Ric Davis for providing help with the technical equipment and recovering impossible to recover files from my desktop, and David Bradbury for his help with experimental designs and his positive approach to everything.

I would like to thank my friends Diego Lorca, Andrea Gajardo and Anne-Catherine Huys for always being there. I would also like to thank Philipp Ludersdorfer, Rebecca Lawson, Cornelia McCormick, Daniel Barry, Sophie Roberts, Jochen Michely, Iris Vilares, Yael Balbastre, Mikael Brudfors and Micah Allen for their friendship.

I could not have accomplished this work without the love and support of my family. I would like to thank my brother Cenk Mirza and my parents Meral Mirza and Ertugrul Mirza. I am grateful to my girlfriend Gabi Leleikaite for her encouragement, patience and love.

Last but not least, I would like to express my gratitude to PACE Network and the members of this network for their support, which made this PhD possible.

Contents

1. Introduction

Perception of an environment is crucial to take appropriate actions. This process requires accumulation of evidence sampled through our sensory organs piece by piece. Here, what I refer to as evidence is observations (under the same or different modalities) made at each unit time. Then our job as living agents is to find (or make) the observations that acquire the most information about our environment and use this information to take the actions that achieve our goals. But what is information? Although there are a number of different accounts of what information is, the account I use in this thesis is based on the concept of Bayesian surprise (Itti & Baldi, 2009) which suggests that the informative observations cause big shifts in our beliefs about the world. As an example one might think that he/she has a large amount of money in his/her wallet judging by the wallet's thickness. Finding nothing but blank banknotes in the wallet and not real money would suddenly shift the beliefs about having a large amount of money to having no money at all. But how do we sample the informative observations? In order to understand this, we first need to understand the computational mechanisms underlying our behaviour. One can then ask the question 'How these computations are carried out physiologically?'

In this thesis, my objective is to introduce computational models of information gathering. I use models based upon a Markov decision process variation of active inference throughout the thesis. Active inference is a Bayesian model of decision making which expresses behaviour in terms of a *generative model* of the task at hand. A *generative model* is a joint probability distribution over the outcomes and their hidden causes. A *generative model* can be thought of as an agent's internal model that describe how the real-world dynamics generate observations (which is expressed by the *generative process*). These models express how the observed outcomes might have been generated by specifying a likelihood mapping from the hidden causes (hidden states and model parameters) to the observed outcomes and they depend upon prior beliefs. The prior beliefs correspond to the parameters or the hyperparameters of probability distributions that define the model.

The use of generative models affords an opportunity to compare competing hypotheses that may underlie information gathering behaviour in terms of model

evidence obtained under each model. This would allow for interpreting the computational mechanisms underlying people's behaviour in terms of the parameters of the generative model (with the greatest log evidence). This is important as a number of clinical conditions (e.g., schizophrenia, autism, anxiety, etc.) have been associated with atypical exploratory (or information gathering) behaviour and yet the computational mechanisms underlying the behaviour under these conditions are poorly understood. The exploratory behaviour in schizophrenia, for example have been described in terms of fewer fixations, shorter mean scan-paths (Tsunoda et al., 1992) and on tasks which require information gathering, the scan-paths are much more restricted compared to that of healthy human subjects (Kojima et al., 1992). According to aberrant salience hypothesis of schizophrenia (Kapur, 2003) there is an abnormal attribution of salience to objects. People with a diagnosis of anxiety spectrum disorder attend less to the core facial features of people's faces (Pelphrey et al., 2002) and superior to controls on visual search tasks that involve visual illusions (Happé, 1996). It has been suggested that these behaviours might be due to diminished use of prior beliefs when contextualising sensory data (Frith, 2003). Threat of shock induced anxiety studies show that a threat of shock during a facial expression identification task enhances people's perception of fearful faces (stimuli that are congruent with anxiety or threat condition) while leaving the perception of happy faces unchanged (Robinson, Charney, Overstreet, Vytal, & Grillon, 2012). This is a finding that suggests that anxiety might have adaptive value. These results suggest that there might be distinct computational mechanisms underlying information gathering behaviour. The atypical exploratory behaviours mentioned in schizophrenia, autism and anxiety might be due to a number of things such as making false inferences about context which leads to abnormal attribution of salience to task irrelevant objects, context insensitive behaviour which might be due to inferring each context to be equally likely as the others, increased sensory perception for context relevant stimuli, respectively. In addition to these, one might ask whether people who suffer from clinical conditions have a bad model of the task at hand or whether they do not know the mapping from hidden states to outcomes.

In psychiatry, the symptoms under the same psychiatric diagnosis might differ substantially (Fernandes et al., 2017). The generative models can be used for subtyping under the same psychiatric condition (Stephan & Mathys, 2014) and

classifying people who suffer from these conditions as mainly deficient in certain computational mechanisms. Associating these aberrant computational mechanisms with the physiological processes underlying them may lead to more personalised and improved therapy choices in the future.

This thesis comprises seven chapters. The first is an introduction about what is described in the subsequent chapters. The final chapter is a general discussion of results and how these relate to the other works in the literature. The following paragraphs summarise the contents of the chapters of this thesis.

In the second chapter, I explain active inference and its underlying formalism. Active or embodied inference is a corollary of the free energy principle that tries to explain the function of nervous systems in terms of the minimisation of variational free energy. Variational free energy is an upper bound on surprise or similarly a lower bound on Bayesian model evidence. This means that minimising free energy corresponds to maximising model evidence or equivalently avoiding surprises. The embodied or situated aspect of active inference acknowledges the fact that we are the authors of the sensory evidence we garner. This means that the consequences of sampling or action must themselves be inferred. This means that the problem that is tackled is not choosing the optimal action but to make optimal inference about actions. In turn, this implies that we have (prior) beliefs about our behaviour. Active inference assumes that the only self-consistent prior belief is that our actions will minimise free energy; in other words, we (believe we) will behave to avoid surprises or resolve uncertainty through active sampling of the world.

In the third chapter I consider the (difficult) problem of categorising a scene. In the ‘scene construction’ task I devised, scene categories are defined by the spatial relationships among visual objects. The objects are sampled through a sequence of saccadic eye movements, and the subject must decide on the category of the scene. In this task the evidence for competing hypotheses (categories) about the scene has to be accumulated sequentially, calling upon both prediction (planning) and postdiction (memory). Here, postdiction refers to the integration of information that is already acquired, whereas prediction corresponds to choosing the best location to look at next to resolve uncertainty about the scene category and categorise it once sufficient

evidence has been accumulated. In this task the resolution of uncertainty about hidden states of the world is crucial for efficient categorisation of the scene.

In the fourth chapter, I present the results of an empirical study. The scene construction task was performed by healthy human subjects. I use active inference to explain the visual searches of these subjects; enabling me to answer some key questions about visual foraging and salience attribution. First, I asked whether there is any evidence that subjects' exploratory behaviour resolves uncertainty about hidden states (e.g. scene category). In brief, I used Bayesian model comparison to compare Markov decision process (MDP) models of scan-paths that did – and did not – contain the uncertainty-resolving imperatives for action selection. In the course of this model comparison, I discovered that it was necessary to include heuristic policies to explain some observed behaviour (e.g., a reading-like strategy that involved scanning from left to right). Despite this use of heuristic policies, model comparison showed that there is substantial evidence for uncertainty driven visual exploration in even simple scenes like the ones I used. Second, I compared MDP models that did – and did not – allow for changes in prior expectations over successive blocks of the visual search paradigm. I found that implicit prior beliefs about the speed and accuracy of visual searches changed systematically with experience. Finally, I characterised intersubject variability in terms of subject-specific prior beliefs. Specifically, I used canonical correlation analysis to see if there were any mixtures of prior expectations that could predict between-subject differences in performance; thereby establishing a quantitative link between different behavioural phenotypes and Bayesian belief updating. I demonstrated that better scene categorisation performance is consistently associated with lower reliance on heuristics; i.e., a greater use of a generative model of the scene to direct its exploration.

In the fifth chapter, I engage with the problem of how exploration strategies can be affected by a global context in a visual search task similar to the scene construction task. In active inference uncertainty resolving behaviour favours actions whose (sensory) consequences acquire information about the hidden states of the world (i.e. are salient). In other words, the actions that would cause the greatest shift in beliefs about the hidden states that define an environment. However, not all the information about the hidden states of an environment is relevant to the task at hand. While it may

be crucial to acquire information about one hidden state of an environment under one context, this information may completely be irrelevant under another context. In this chapter, I introduce a contextual epistemic search scheme, based upon active inference. I show that the salience of a visual cue depends upon contextual beliefs. Such beliefs ensure a new observation only causes a large shift in beliefs about the hidden aspects of an environment if it is *task relevant* (context dependent).

The sixth chapter is about characterising impulsive behaviour using a patch leaving paradigm and active inference. This paradigm comprises different environments (patches) with limited resources that decline over time at different rates. Looking at this task from the perspective of the exploration and exploitation trade-off, it is highly relevant to computational modelling of information gathering as the challenge is to decide when to stop exploiting the current patch and explore other patches, to maximise reward. I chose this task because it offers an operational characterisation of impulsive behaviour; namely, maximising proximal reward at the expense of future gain. I use a Markov Decision Process (MDP) formulation of active inference to simulate behavioural and electrophysiological responses under different models and prior beliefs. My main finding is that there are at least three distinct causes of impulsive behaviour, which I demonstrate by manipulating three different components of the MDP model. These components comprise i) the depth of planning, ii) the capacity to maintain and process information, and iii) the perceived value of immediate (relative to delayed) rewards. I show how these manipulations change beliefs and subsequent choices, through variational message passing. Furthermore I appeal to the process theories associated with this message passing to simulate neuronal correlates.

2. Active inference

Active inference is based upon the premise that every living thing minimises variational free energy. This single premise leads to some surprisingly simple update rules for action, perception, policy selection, and the encoding of salience or precision. In principle, the active inference scheme described below can be applied to any paradigm or choice behaviour. Indeed, earlier versions have already been used to model waiting games (K. Friston et al., 2013), two-step maze tasks (K. Friston et al., 2015), the urn task and evidence accumulation (FitzGerald, Schwartenbeck, Moutoussis, Dolan, & Friston, 2015), trust games from behavioural economics (Moutoussis, Trujillo-Barreto, El-Deredy, Dolan, & Friston, 2014) and addictive behaviour (Schwartenbeck et al., 2015). It has also been used in the setting of computational fMRI (P. Schwartenbeck, T. H. FitzGerald, C. Mathys, R. Dolan, & K. Friston, 2014).

2.1. Variational free energy

A characteristic attribute of biological systems is their adaptive exchange with changing environments (K. Friston, 2010). This adaptive exchange requires i) the change in the environment to be recognised (perception), and ii) action to be taken, in order to keep a biological system in the environmental states conducive to existence (K. Friston, Kilner, & Harrison, 2006), e.g. living creatures can only exist in a narrow range of all possible temperatures. Another way to put this is that they must maintain a low entropy, or *surprise* (averaged over time). Active inference describes how an agent's adaptive exchange with its environment can occur in a Bayes optimal fashion by minimising variational free energy.

An agent's perception of its environment and the actions it takes both suppress variational free energy. For an agent to infer (perceive) the state of its environment, it requires a *generative model* that describes how observed outcomes are generated by the environment (K. Friston, 2010; Thomas Parr, Rees, & Friston, 2018). Variational free energy F is a functional of two things: the *generative model* $P(\tilde{o}, x \mid m)$, and an approximate posterior distribution over the hidden causes $Q(x)$ (Eq 1).

Variational free energy is an upper bound on *surprise* $-\ln P(\tilde{o} \mid m)$, and minimising it minimises *surprise*. One can show this using Jensen's inequality for concave functions:

$$
\begin{aligned}
-\ln P(\tilde{o} \mid m) &= -\ln \sum_x P(\tilde{o}, x \mid m) \\
&= -\ln \sum_x Q(x) \frac{P(\tilde{o}, x \mid m)}{Q(x)} \leq \underbrace{-\sum_x Q(x) \ln \frac{P(\tilde{o}, x \mid m)}{Q(x)}}_{\text{Variational free energy } F} \\
F &= E_{Q(x)}\left[-\ln P(\tilde{o}, x \mid m)\right] - H\left[Q(x)\right] \qquad (1)
\end{aligned}
$$

Here F is variational free energy, $\tilde{o}$ represents the sequence of observations over time $\tilde{o} = [o_1, o_2, ..., o_T]^T$, x represents the hidden causes and $Q(x)$ is the approximate posterior distribution over the hidden causes and m represents the model under which *surprise* is evaluated.

Rearranging Eq 1 reveals that the variational free energy is an upper bound on *surprise* because the KL divergence in Eq 2 can never be less than zero.

$$
\begin{aligned}
F &= \underbrace{D_{KL}\left[Q(x) \| P(x \mid \tilde{o})\right]}_{\text{Relative entropy}} - \underbrace{\ln P(\tilde{o} \mid m)}_{\text{log evidence}} \qquad (2) \\
&= \underbrace{D_{KL}\left[Q(x) \| P(x)\right]}_{\textit{complexity}} - \underbrace{E_{Q(x)}\left[\ln P(\tilde{o} \mid x)\right]}_{\textit{accuracy}} \qquad (3)
\end{aligned}
$$

Rearranging free energy reveals that it can be expressed as a mixture of different components. Eq. 3 shows that free energy is expressed in terms of complexity and accuracy. This means that minimising free energy is equivalent to minimising the complexity of accurate of explanations for observed outcomes.

In variational Bayesian inference, model inversion entails minimising variational free energy with respect to the sufficient statistics (i.e., parameters) of the posterior beliefs (see Table 1 for a glossary of expressions). Minimising the variational free energy shows that the Jensen's inequality is a good way to upper bound the surprise (or negative log evidence) as minimising the variational free energy minimises the distance between the approximate and true posterior distributions over the hidden causes, making this distribution an approximation of the true distribution. The

approximate posterior distribution becomes exact when the KL divergence in Eq. 2 is equal to zero and the variational free energy becomes the negative log-evidence obtained under the generative model.

$$Q(x) = \arg\min_{Q(x)} F \approx P(x \mid \tilde{o}) \tag{4}$$

2.2. Markov decision processes

The Markov decision process (MDP) models use a discrete state space to describe how events evolve through time, giving rise to categorical outcomes at each (discrete) time point (T. Parr & Friston, 2018). Below is a graphical representation of this process.

The graphical representation of Markov decision process is shown in Fig 2.1. This graph shows that the outcomes at each time step o_t depend only on the hidden states in the same time step s_t and nothing else. Hidden states in the current time step s_t depend on the hidden states in the previous time step s_{t-1} and the policies π. A policy prescribes the sequence of actions one takes over time. Crucially, in active inference the agent has some control over parts of the environment. This means that the agent can control the transitions between some of the hidden states, through actions sampled from beliefs about policies π. This means that the observed sensory input depends on the hidden states that can be controlled through actions.

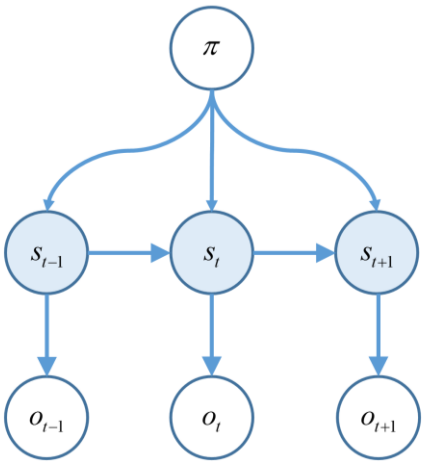


Figure 2. 1 Markov decision process

This graph shows a graphical representation of the Markov decision process and the conditional dependencies among the terms in the model.

Expressing variational free energy in terms of policies π, hidden states s and outcomes o reveals that it consists of three main terms.

$$
\begin{aligned}
F &= \mathrm{E}_{Q(\tilde{s})}[-\ln P(\tilde{o},\tilde{s}) + \ln Q(\tilde{s})] \\
&= \mathrm{E}_{Q(\tilde{s},\pi)}[-\ln P(\tilde{o},\tilde{s},\pi) + \ln Q(\tilde{s},\pi)] \\
&= \mathrm{E}_{Q(\pi)}[\ln Q(\pi) - \ln P(\pi)] + \mathrm{E}_{Q(\tilde{s}|\pi)Q(\pi)}[-\ln P(\tilde{o},\tilde{s}\mid\pi) + \ln Q(\tilde{s}\mid\pi)] \\
&= D[Q(\pi)\,||\,P(\pi)] + \mathrm{E}_{Q(\pi)}[F(\pi)] \\
&= \boldsymbol{\pi}.(\ln \boldsymbol{\pi} + G(\pi) + F(\pi))
\end{aligned} \tag{5}
$$

The final equation above shows that variational free energy is formulated in terms of the posterior beliefs about the policies $\mathrm{Q}(\pi)$, expected free energy under a policy $\mathrm{G}(\pi) = -\ln \mathrm{P}(\pi)$ and free energy of hidden states under a policy $\mathrm{F}(\pi)$, where $\mathrm{Q}(\pi)$ is represented by $\boldsymbol{\pi}$, and a particular policy is represented by π. Minimising free energy ensures expectations encode posterior beliefs, given observed outcomes. However, beliefs about policies rest on future outcomes. This means that policies should, *a priori*, minimise the free energy of beliefs about the future. This can be formalised by making the log probability of a policy proportional to the free energy expected in the future (K. Friston et al., 2015). The free energy of hidden states and the expected free energy are given below:

$$
\begin{aligned}
\mathrm{F}(\pi) &= \textstyle\sum_{\tau} \mathrm{F}(\pi,\tau) \\
\mathrm{F}(\pi,\tau) &= \mathrm{E}_{\mathrm{Q}(\mathrm{s}_\tau,\mathrm{s}_{\tau-1}|\pi)}\left[\ln \mathrm{Q}(\mathrm{s}_\tau\mid\pi) - \ln \mathrm{P}(\mathrm{o}_\tau,\mathrm{s}_\tau\mid\mathrm{s}_{\tau-1},\pi)\right] \\
&= \underbrace{\mathrm{E}_{\mathrm{Q}(\mathrm{s}_{\tau-1}|\pi)}\left[\mathrm{D}\left[\mathrm{Q}(\mathrm{s}_\tau\mid\pi)\,||\,\mathrm{P}(\mathrm{s}_\tau\mid\mathrm{s}_{\tau-1},\pi)\right]\right]}_{complexity} - \underbrace{\mathrm{E}_{\mathrm{Q}(\mathrm{s}_\tau|\pi)}\left[\ln \mathrm{P}(\mathrm{o}_\tau\mid\mathrm{s}_\tau)\right]}_{accuracy}
\end{aligned} \tag{6}
$$

$$
G(\pi) = \textstyle\sum_{\tau} G(\pi,\tau)
$$

$$
\begin{aligned}
G(\pi,\tau) &= \mathrm{E}_{\tilde{Q}}[-\ln P(o_\tau,s_\tau\mid\pi) + \ln Q(s_\tau\mid\pi)] \\
&\approx -\underbrace{E_{\tilde{Q}}[\ln Q(s_\tau\mid o_\tau,\pi) - \ln Q(s_\tau\mid\pi)]}_{epistemic\ value} - \underbrace{E_{\tilde{Q}}[\ln P(o_\tau)]}_{extrinsic\ value} \\
&= \underbrace{D[Q(o_\tau\mid\pi)\,||\,P(o_\tau)]}_{Risk} + \underbrace{E_{\tilde{Q}}[H[P(o_\tau\mid s_\tau)]]}_{Ambiguity}
\end{aligned} \tag{7}
$$

where $\tilde{Q} = Q(o_\tau, s_\tau \mid \pi) = P(o_\tau \mid s_\tau)Q(s_\tau \mid \pi) \approx P(o_\tau, s_\tau \mid \tilde{o}, \pi)$ and $Q(o_\tau \mid s_\tau, \pi) = P(o_\tau \mid s_\tau)$ for $\tau > t$.

Expected free energy can be interpreted in several different ways. The second line in Eq. 7 shows that the expected free energy can be expressed in terms of epistemic and extrinsic value. Epistemic value is the expected information gain about the hidden states. It expresses how much the beliefs about the hidden states would be shifted by taking into account the expected outcomes in the future. Essentially, informative – epistemically valuable – observations are the ones that cause the biggest shifts in prior to posterior beliefs about hidden states. Epistemic value is Bayesian surprise (Itti & Baldi, 2009) expected under (expected) outcomes in the future. Extrinsic value expresses how much utility is expected to be acquired by pursuing different policies. Rearranging the expected free energy shows that it can be written in terms of risk and ambiguity (see Eq. 7 third line). Risk is the expected divergence from preferred outcomes, whereas ambiguity is the expected uncertainty in the mapping from hidden states to observations. Policies that minimise both risk and ambiguity are more likely to be chosen.

2.3. Markovian generative model

Active inference rests upon a generative model of observed outcomes. This model is used to infer the most likely causes of outcomes in terms of expectations about states of the world. These states are called *hidden states* because they can only be inferred indirectly through, possibly limited, sensory observations. In terms of active inference, making inferences about the hidden states corresponds to perception. Crucially, observations depend upon action, which requires the generative model to entertain expectations under different policies or action sequences. This means that under any policy the agent will form beliefs about the hidden states and evaluate the policies in terms of their free energy. Because the model generates the consequences of sequential action, it has explicit representations of the past and future; in other words, it is equipped with a working memory and expectations about future (counterfactual) states of the world under competing policies. While the working memory allows the generative model to accumulate information for competing hypothesis, the expectations about the future enables the generative model to have a free energy map

over policies, given the previously observed outcomes. These expectations are optimised by minimising variational free energy, which renders them (approximately) the most likely (posterior) expectations about states of the world, given the current observations.

One crucial aspect of the free energy principle is that the free energy itself is a function of policies. Under each policy the agent entertains a free energy. This allows the generative model to evaluate the quality of each policy in terms of its free energy. A policy is more likely to be chosen if it returns the least free energy (or greatest model evidence). As I showed above, this expected free energy can be expressed in terms of epistemic and extrinsic value, where epistemic value scores the information gain or reduction in uncertainty about states of the world – and extrinsic value depends upon prior beliefs about future outcomes. These prior preferences play the role of utility in economics and reinforcement learning. A softmax function of free energies of all policies returns the relative probability of each policy with respect to each other. This softmax function is almost identical to the Boltzmann or Gibbs distribution in thermodynamics. In active inference, one can think of different policies as different models. Because expected free energy is a function of policies, one can evaluate the evidence for each policy in terms of the expected free energy obtained under different policies. Note that expected free energy is expressed in terms of log probabilities and therefore it is necessary to exponentiate it to convert to a probabilistic quantity. To ensure this quantity is a true probability distribution (i.e. sums to one), one needs to normalise it. Applying a softmax function (normalised exponential operator) to the (negative) expected free energies under all policies returns a probability distribution over policies. A policy with a high probability is more likely to fulfil an agent's goals expressed in terms of extrinsic value (pragmatically driven behaviour) and epistemic value (uncertainty driven behaviour).

Having evaluated the relative probability of different policies, expectations under each policy can then be averaged in proportion to their (posterior) probability. In statistics, this is known as *Bayesian model averaging*, given that each policy is considered as a model.

$$\mathrm{p}(\mathrm{s}\,|\,\mathrm{o}) = \int \mathrm{p}(\mathrm{s}\,|\,\mathrm{o},\pi)\,\mathrm{p}(\pi\,|\,\mathrm{o})\,\mathrm{d}\pi \tag{8}$$

where s is the hidden states, o is the observed outcomes and π refers to the models. The results of this averaging specify the next most likely hidden states and through that the most likely outcomes, which determines the next action. Once an action has been sampled, the generative process produces a new outcome and the (perception and action) cycle starts again. The resulting behaviour is a principled interrogation and sampling of sensory cues that has both epistemic and pragmatic aspects. Generally, behaviour in an ambiguous and uncertain context is dominated by epistemic drives until there is no further uncertainty to resolve. Once the level of uncertainty is reduced substantially extrinsic value takes over and drives the behaviour. At this point, explorative behaviour gives way to exploitative behaviour. In this thesis, I am primarily interested in the epistemic behaviour, and only use extrinsic value to encourage the agent to report its decision, when it is sufficiently confident.

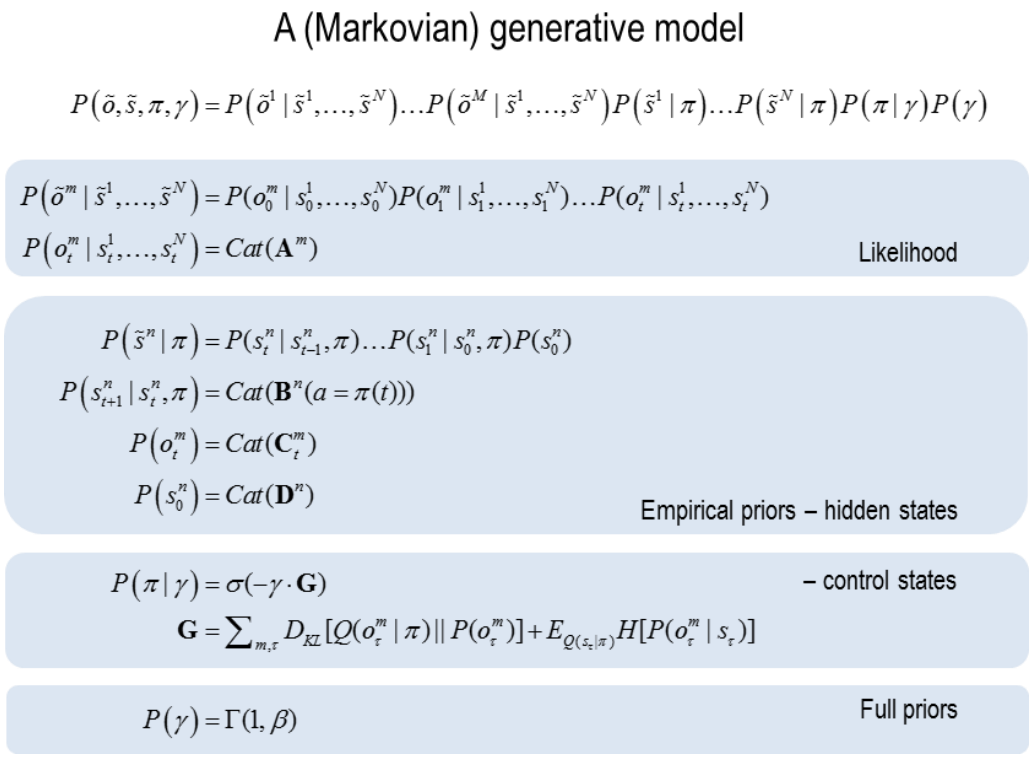


Figure 2. 2 Formal specification of the generative model

These equations specify the form of the (Markovian) generative model used in this thesis. A generative model is essentially a specification of the joint probability of outcomes or consequences and their (latent or hidden) causes. Usually, this model is expressed in terms of a *likelihood* (the probability of consequences given causes) and priors over the causes. When a prior depends upon a random variable it is called an *empirical prior*. Here, the generative model specifies the mapping between hidden states and observable outcomes in terms of the likelihood. The priors in this instance pertain to transitions

among hidden states that depend upon action, where actions are determined probabilistically in terms of policies (sequences of actions). The key aspect of this generative model is that, *a priori*, policies are more probable if they minimise the (path integral of) expected free energy **G**. Bayesian model inversion refers to the inverse mapping from consequences to causes; i.e., estimating the hidden states and other variables that cause outcomes.

In more detail: expectations about hidden states (and the precision of beliefs about competing policies) are updated to minimise variational free energy under a generative model. The generative model considered here is fairly generic (see Fig 2.2). Outcomes at any particular time depend upon hidden states, while hidden states depend upon action and the hidden states in the previous time step (as the Markov property mandates), and the policies depend on the inverse temperature or precision γ. Formally, this model is specified by two sets of matrices (strictly speaking these are multidimensional arrays). The first is a likelihood matrix, $\mathbf{A}^m$, that probabilistically maps from hidden states to the *m*-th outcome. The *m*-th sort of outcome here can be considered the *m*-th modality; for example, exteroceptive or proprioceptive observations. The second set of matrices $\mathbf{B}^n\left(a\right)$, encode the transitions among the *n*-th dimension of hidden states, given an action, *a*. The *n*-th sort of hidden state can correspond to different factors or attributes of the world; for example, the location of an object and its identity. These different hidden state dimensions can account for different characteristics of the world, such as what object is located where in the world. The remaining parameters encode prior beliefs about the initial states $\mathbf{D}^n$, the precision of beliefs about policies γ, where a policy returns an action at a particular time $a = \pi(t)$ and prior preferences $\mathbf{C}^m$ (see Fig 2.2).

The form of the generative model in Fig 2.2 means that outcomes are generated in the following way: first, a policy is sampled from the beliefs about the repertoire of policies using a softmax function of expected free energy for each policy, where the inverse temperature or precision is selected from a prior (exponential) density. The sampled policy and the hidden states of different dimensions in the previous time step then determines the hidden states of different dimensions in the current time step. The hidden states in the current time step then generates outcomes under each modality. These new outcomes are available for the agent to observe and thus the perception and actions cycle starts again. Perception or inference corresponds to inverting or

fitting this generative model, given a sequence of outcomes. This corresponds to optimising the expected hidden states, policies and precision with respect to variational free energy. These (posterior) estimates constitute posterior beliefs, usually denoted by the probability distribution $Q(x)$, where $x = \tilde{s}, \pi, \gamma$ are the hidden or unknown variables.

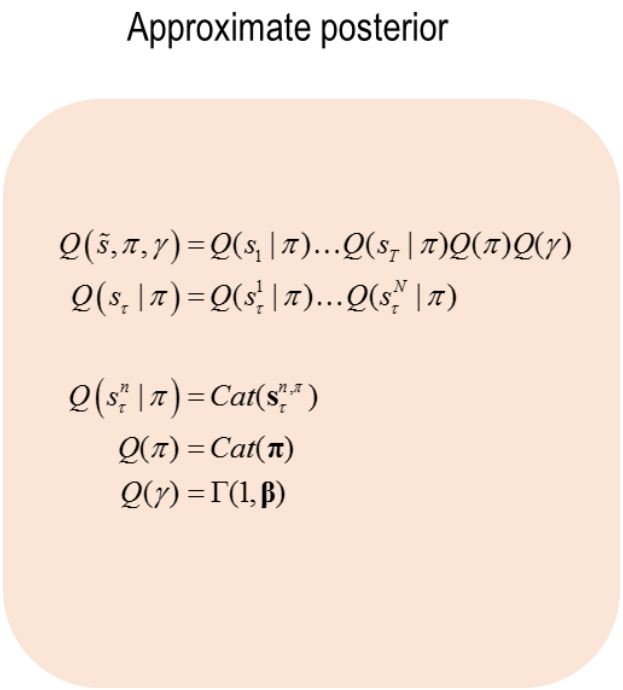


Figure 2. 3 Approximate posterior distribution

In variational Bayesian inversion, one has to specify the form of an approximate posterior distribution. This particular form uses a mean field approximation, in which posterior beliefs are approximated by the product of marginals or factors. Here, a mean field approximation is applied both to posterior beliefs at different points in time and different sorts of hidden states. See the main text and Table 1 for more detailed explanation of the variables.

I now turn to belief updating that is based on minimising free energy under the generative model described above.

2.4. Belief updates and message passing

Under active inference, perception arises as a result of minimising variational free energy with respect to beliefs about hidden variables. Mathematically, this is implemented by using a gradient descent on the variational free energy for each hidden variable. It is easy to see that the updates minimise variational free energy

because they converge when the free energy gradients are zero: i.e., $\nabla F = 0$. The resulting belief update equations show how message passing occurs under this scheme (see Fig 2.4).

Taking the gradient of variational free energy with respect to the hidden states after observing a new outcome gives the optimal solution to state estimation shown with s^* (state estimation: first equation). The difference between s^* and the current beliefs about the hidden states s_τ^π generates a state prediction error ε_τ^π (second equation). A gradient descent on state prediction errors is used to infer the most likely hidden states (third equation). Here τ represents the time steps from 1 to $t+1$. When $\tau \leq t$, the term in brackets (first equation) returns 1 and otherwise it returns 0. This means that the inference about the hidden states at the current time step t depends on the observed outcomes from time step 1 to t, which allows for evidence accumulation over time. When $\tau > t$ beliefs about the hidden states do not depend on the outcomes as the outcomes have not been observed yet. This means that beliefs about the hidden states at $\tau > t$ depend only upon beliefs about hidden states in the previous $s_{\tau-1}^\pi$ and next $s_{\tau+1}^\pi$ time steps.

Beliefs about the inferred hidden states are projected into the future to form expectations about the most likely observations in the future under different policies. These expectations are used to compute the probability distribution over policies π (policy evaluation: first equation) such that the most likely policy is that with the smallest free energy $\mathbf{F}$ (second equation) and expected free energy $\mathbf{G}$ (third equation). $\mathbf{F}$ and $\mathbf{G}$ are vectors with elements corresponding to each policy. The free energy F_π under a policy is a function of the state prediction errors under that policy and the beliefs about states under that policy. The expected free energy $\mathbf{G}_\pi$ is expressed in terms of risk and ambiguity. Risk is the expected divergence from preferred outcomes and expressed as the (expected) difference between (log) expected outcomes $\ln \mathbf{o}_\tau^\pi$ and preferred (log) outcomes C_τ expected under beliefs about future outcomes $\mathbf{o}_\tau^\pi \cdot \left(\ln \mathbf{o}_\tau^\pi - \mathbf{C}_\tau \right)$. Ambiguity is the expected uncertainty in the mapping from hidden states to observations expected under beliefs about the hidden

states $\mathrm{H} \cdot s_\tau^\pi$, where H is the entropy of outcomes under all possible combinations of hidden states.

The gradient of the variational free energy with respect to the temperature parameter (inverse precision) gives the optimal solution to this parameter β^* (precision: first equation). The second equation shows that the prior beliefs about policies is expressed in terms of a softmax function of precision and expected free energy. The difference between β^* and the current temperature parameter β generates a precision prediction error (third equation). A gradient descent on precision prediction error is used to update the inverse temperature parameter whose inverse is equal to the precision term γ (fourth equation).

In the action selection phase an action a_t is sampled from the most likely policies (action selection: first equation), where $\boldsymbol{\pi}$ corresponds to the beliefs about the policies. The expected states and outcomes are acquired by taking Bayesian model averages of the states s_τ (second equation) and outcomes o_τ (third equation) expected under each policy. Once an action is selected the environment will generate a new outcome that can be fed back to the generative model and thus the perception and action cycle begins again.

These updates are remarkably plausible in terms of neurobiological schemes – as discussed elsewhere (K. Friston et al., 2014). For example, expectations about hidden states are a softmax function (c.f., neuronal activation function) of two terms. The first is a decay term, because the log of a probability is always zero or less. The second is the free energy gradient, which is just a linear mixture of (spiking) activity from other representations (expectations). Similarly, the precision updates are a softmax function of free energy and its expected value in the future, weighted by precision or inverse temperature. The expected precision is driven by the difference between the dot product of expected free energy with posterior beliefs about policies and the dot product of expected free energy with prior beliefs about policies. While the posterior beliefs about the policies $\boldsymbol{\pi}$ depend upon both the free energy based upon previous observations $\mathbf{F}$ and the expected free energy $\mathbf{G}$ (weighted by its precision γ), the

prior beliefs about the policies depend only upon the expected free energy $\mathbf{G}$ (weighted by its precision γ). If the dot product of the posterior beliefs about policies (which is based upon previous observations and future expected outcomes) and the expected free energy is smaller than the dot product of the prior beliefs (which is based upon only the future expected outcomes) about policies and the expected free energy, the temperature parameter β decreases and precision γ increases (see Fig 2.4). This is very much like how dopamine is driven by the difference in expected and observed rewards (Schultz, Dayan, & Montague, 1997). See (K. Friston et al., 2015) for further discussion.

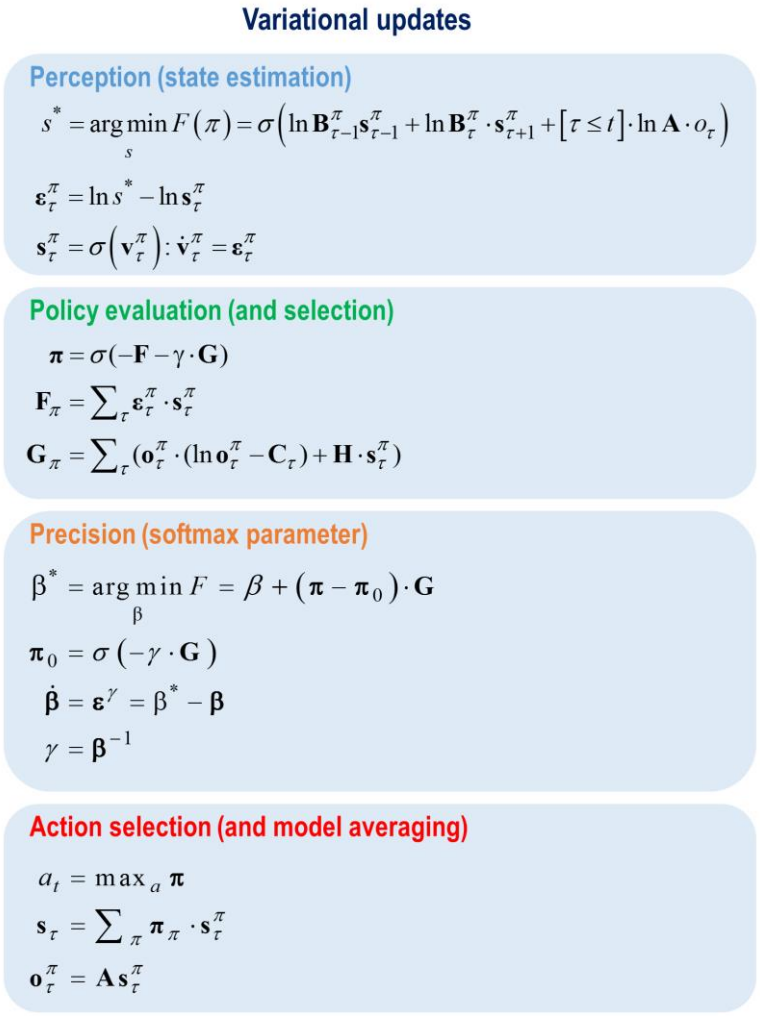


Figure 2. 4 Belief update equations

The belief update equations shown in this panel summarise the variational message passing. Belief updates occur through three phases, namely perception, policy evaluation and action selection. See the main text for details.

The key thing about these updates is that they provide a process theory that implements the normative theory offered by active inference. In other words, they constitute specific processes that make predictions about neuronal dynamics and responses. Although the focus of this paper is on behaviour and large-scale functional anatomy, I will illustrate the simulated neuronal responses associated with active inference in later sections.

3. Scene construction, visual foraging and active inference

We have a remarkable capacity to sample our visual world in an efficient fashion, resolving uncertainty about the causes of our sensations so that we can act accordingly. This capacity calls on the ability to optimise not just beliefs about the world that is 'out there' but also the way in which we sample information (Andreopoulos & Tsotsos, 2013; Howard, 1966; Pezzulo, Rigoli, & Chersi, 2013; Shen, Valero, Day, & Paré, 2011; Wurtz, McAlonan, Cavanaugh, & Berman, 2011). This is particularly evident in active vision, where discrete and restricted (foveal) visual data is solicited every few hundred milliseconds, through saccadic eye movements (S Grossberg, K Roberts, M Aguilar, & D Bullock, 1997; K Srihasam, D Bullock, & S Grossberg, 2009). In this chapter, I consider the principles that underlie this visual foraging – and how it is underwritten by resolving uncertainty about the visual scene that is being explored. I approach this problem from the point of view of active inference; namely, the assumption that action and perception serve to minimise surprise or uncertainty under prior beliefs about how sensations are caused.

I consider the problem of categorising a scene based upon the sequential sampling of local visual cues to construct a picture or hypothesis about how visual input is generated. This is essentially the problem of scene construction (Hassabis & Maguire, 2007; Zeidman, Lutti, & Maguire, 2015), where each scene corresponds to a hypothesis. The main point that emerges from this perspective is that the scene exists only in the eye of the beholder: it is represented in a distributed fashion through recurrent message passing or belief propagation among functionally segregated representations of where (we are currently sampling) and what (is sampled). This application of active inference emphasises the epistemic value of free energy minimising behaviour – as opposed to the pragmatic (utilitarian) value of searching for preferred outcomes. Although the exploratory behaviour is driven by epistemic value, ensuing behaviour, namely, reporting beliefs about the hypothesis (or picture) requires pragmatic value to play a role. However, having said this, the theory (resp. simulations) uses exactly the same mathematics (resp. software routines) that is previously used to illustrate foraging behaviour in the context of reward seeking (K. Friston et al., 2015).

My aim is to introduce a model of epistemic foraging that can be applied to empirical saccadic eye movements and, ultimately, be used to phenotype individual subjects in terms of their prior beliefs: namely, the prior precision of beliefs about competing epistemic policies and the precision of prior preferences (c.f., 'incentive epistemic' and motivational salience). This may be particularly interesting when looking at schizophrenia and other clinical phenotypes that show characteristic abnormalities during visual (saccadic) exploration. For example, schizophrenia has been associated with 'aberrant salience', in which subjects attend to – and hence saccade to – inconsequential features of the environment (Beedie, Clair, & Benson, 2011; Kapur, 2003). It is unclear, however, whether 'aberrant' salience is epistemic or motivational, or both; put simply, do subjects with schizophrenia fail to gather information, and/or fulfil their goals?

This chapter comprises four sections. The first section describes a scene construction task and how this task is modelled using the active inference formalism. In brief, this paradigm requires agents to categorise a scene based upon sampling discrete (visual) cues. These visual cues are located in four quadrants in the scene and the agent starts engaging with this task from a central fixation location. The category of a scene is defined in terms of the relative locations of these visual cues, which means that the absolute locations of these cues (on what quadrant the cue is) by themselves do not speak to the category of a scene. To get the agent to report its beliefs about the category of a scene, equipped the agent's generative model is equipped with preferences for correct feedback and a dislike of incorrect feedback. The second section describes mean field assumption and functional segregation. In the third section, agent's exploratory behaviour has been described using simulated saccadic patterns. The variational belief updating that has occurred in the course of a trial has been used to emulate the electrophysiological responses observed in empirical studies. Simulated behavioural and electrophysiological responses have been shown on a sequence of trials as well. The fourth section shows how exploratory behaviour and the ensuing categorization change on trials under different levels of prior precision and preferences (for avoiding incorrect feedback). The results are characterised in terms of simulated saccadic intervals, and the usual behavioural measures of speed and accuracy. I conclude with a discussion of how this model might be used in an empirical (computational psychiatry) setting.

3.1. Active inference and visual foraging

This section uses active inference for Markov decision processes to illustrate epistemic foraging in the setting of visual searches. Here, the agent has to categorise a scene on the basis of the relative position of various visual objects – that are initially unknown. The agent samples one location (thus one cue) at a time and accumulates evidence for how the scene was generated. When the agent is sufficiently confident about its perceptual categorisation, it makes a saccade to a choice location – to obtain feedback (*'right'* or *'wrong'*). *A priori*, the agent expects to obtain 'correct' feedback and not 'incorrect' feedback. I first illustrate a single trial in terms of behaviour and underlying electrophysiological responses. The next section then considers sequences of trials and how average behaviour (accuracy, number of saccades and saccadic intervals) depends upon prior preferences and precision.

This demonstration uses a mean field approximation to the posterior over different hidden states (*context*, *sampling location*, and *spatial transformations*). In addition, two outcome modalities (exteroceptive or *'what'* and proprioceptive or *'where'*) are considered. In this example, the agent has to categorise a scene that comprises cues at four peripheral locations, starting from a central fixation point. This involves a form of scene construction, in which the category of the scene is determined by the spatial relationship between various cues. The scene always contains a bird and seed, or a bird and a cat. If the bird is next to the seed or the cat, then the scene is categorised as *'feed'* or *'flee'* respectively. Conversely, if the seed is diagonally opposite the bird, the category is *'wait'*. As long as the relative locations of the cues remain the same, the absolute locations of the cues in the scene do not change the context of the scene, e.g. if the bird is next to seed then the particular locations of these objects (whether they are in the bottom or top quadrants) in the two by two grid wouldn't matter in terms of the context, this scene would always be of the context '*feed*'. This means hidden states have to include spatial mappings that induce invariances to spatial transformations. These are reflections around the horizontal and vertical axes.

Fig 3.1B shows the hidden states and outcomes in more detail: there are two outcome modalities (*what* and *where*), encoding one of six cues (*distractor*, *seed*, *bird*, *cat* and *right* or *wrong* feedback) and one of eight sampled locations (central *fixation*, four quadrants and three *choice locations* that provide feedback about the respective

decision). There are four dimensions of hidden states; corresponding to context (*feed*, *flee* and *wait*), the currently sampled location (one of the eight locations above) and two further dimensions modelling the absolute locations of the cues in a scene (i.e., with and without reflections about the vertical and horizontal axes). The three scenes under each context (flee, feed and wait) in the top panel of Fig 3.1B are referred to as base scenes. The context or category defines the objects (distractor, seed, bird and cat) and their relative locations. The hidden states mediating (vertical and horizontal) transformations define the absolute locations and are implemented with respect to the base scenes. For example, in the case of a flee scene, the bird and cat may exchange locations under a vertical transformation. Since the absolute and relative positions of the objects (and the objects themselves) are hidden causes of the scene's appearance, they are not affected by the agent's actions.

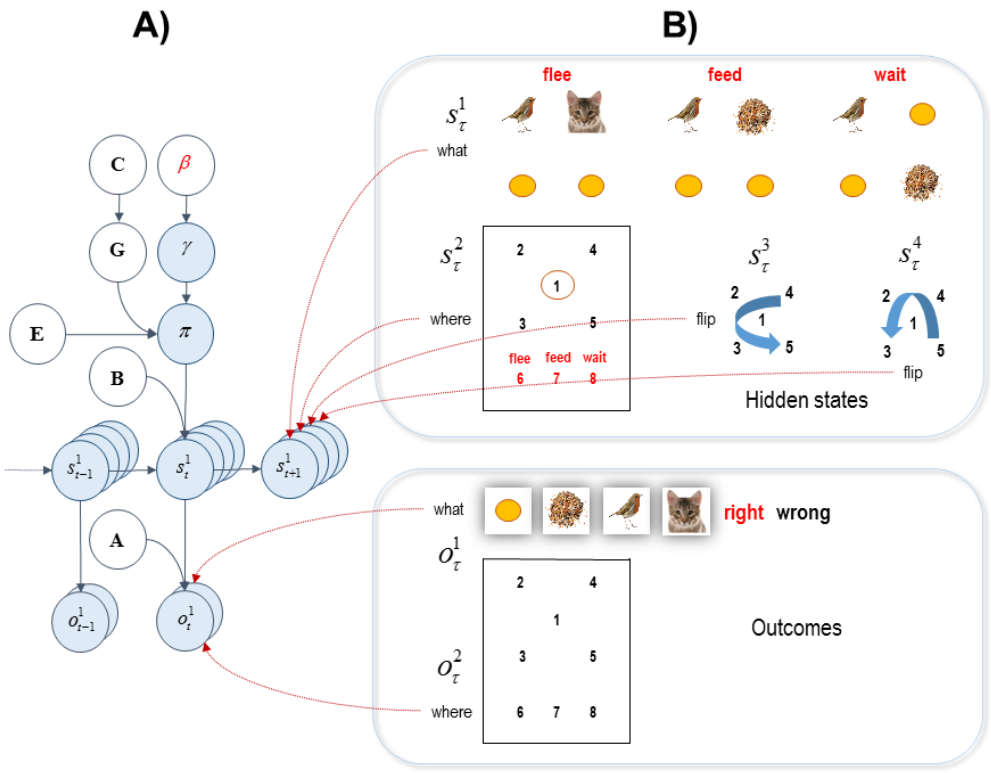


Figure 3. 1 Graphical model corresponding to the generative model

A) The structure of the environment is expressed in terms of the transition and likelihood matrices. The likelihood matrix ($\mathbf{A}$) is a mapping from the hidden states (s_t) to the outcomes (o_t). The state transitions are mediated by the transition matrix ($\mathbf{B}$) which expresses how likely the current state (s_t) is given the previous state (s_{t-1}). Crucially, the transition matrix is a function of action which can be

sampled from the beliefs about the policies. The beliefs about the policies (π) depend on the expected free energy ($\mathbf{G}$) and the precision of policy selection (γ). The expected free energy comprises extrinsic and epistemic values. Extrinsic value is a function of the prior preference matrix ($\mathbf{C}$) which encodes how much one outcome is expected relative to another. Precision of policy selection (γ) is a function of the temperature term (β). The smaller the temperature the more deterministic the policy selection becomes. **B)** The right panels show an example of different hidden states and outcomes modalities. This particular example will be used later to model perceptual categorisation in terms of three scenarios or scenes (*flee*, *feed* or *wait*). The two outcome modalities effectively report what is seen and where it is seen. See the main text for a more detailed explanation.

Given that every context is equally likely, the relative locations of the cues render certain locations more informative than others based on the first seen cue. This means that certain locations become more attractive or salient because of the information they hold. Crucially, there are better ways of exploring a scene. Given that the first seen cue is bird, the quadrant adjacent to where the bird is, is more informative. Fig 3.2 shows certain ways of exploring a scene given the first seen cue in order to categorize the scene, taking minimum number of actions. Obviously the exploratory behaviour is belief based and is dependent on the agent's perception of the scene. There are certain parameter combinations that enable the agent to explore the scene that is described as in Fig 3.2.

Heuristically, the model in Fig 3.1 generates outcomes in the following way. First, one of the three canonical scenes or contexts is selected. This scene can be flipped vertically or horizontally (or both) depending upon the spatial transformation states. Finally, the sampled location specifies the exteroceptive visual cue and the proprioceptive outcome signalling the location. This model can be specified in a straightforward way by specifying the two outcomes for every combination of hidden states in $\mathbf{A}^1 \in \mathbb{R}^{6\times(3\times8\times2\times2)}$ and $\mathbf{A}^1 \in \mathbb{R}^{6\times(3\times8\times2\times2)}$. The arrays in these two matrices just contain a one for each outcome when the combination of hidden states generates the appropriate outcome, and zeros elsewhere. These two matrices encode the observation likelihoods in the two outcome modalities *what* and *where*. Here, $\mathbf{A}^1$ defines the identity (*what*) of objects that are likely to be sampled (i.e., observed), under all possible combinations of hidden states, while $\mathbf{A}^2$ defines the likely locations (*where*) of the objects. The transition matrices are similarly simple: since the only hidden state that can change is the sample locations, the other states that define the

scene (through context and spatial formations) map onto themselves (identity matrices). Action dependent transition matrix encoding the transitions among locations is given by:

$$\mathbf{B}_{ij}^{2}(k) = \mathbb{R}^{(8\times 8)\times 8} = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases} \tag{9}$$

where $k \in \{1,2,...,8\}$. Prior beliefs about the initial states $\mathbf{D}^{n}$ (context and projections) were uniform distributions; apart from the sampled location, which always started at the central fixation $\mathbf{D}^{2} = [1,0,\ldots,0]$ where $\mathbf{D}^{2}$ is a one by eight vector. Here, n indicates the dimension of the hidden states with $n \in \{1,2,3,4\}$. There are four dimensions of hidden states, namely context or category of the scene, sampling location (one of eight locations), and the two spatial transformations (horizontal and vertical flips).

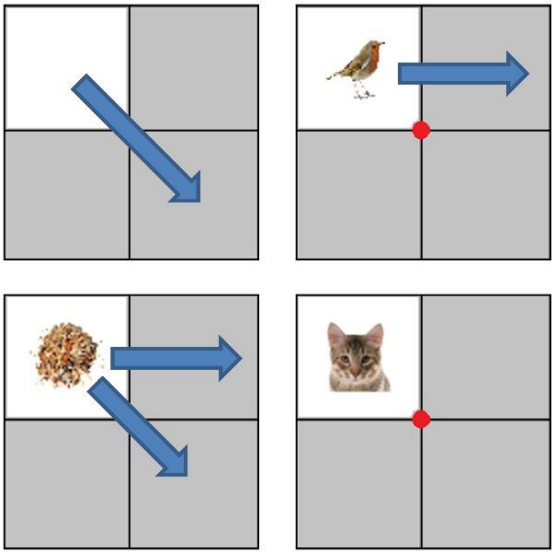


Figure 3. 2 Minimum number action policy

There are certain ways of exploring the scene making minimum number of saccades before categorizing the scene, given the first seen cue. If the first seen cue is a null content then a good way to explore the scene is to look at the quadrant that is on the same diagonal as the 'null'. If the first seen cue is the 'bird' then looking at the quadrant adjacent to where the bird is ensures taking the minimum number of actions. Given the first seen cue is the 'seed' there are two possible good ways to explore the scene, these are either exploring the juxtaposed quadrant or the diagonal. Finally if the cat is seen first then there is no reason to explore the scene, since the cat only exists in the scenes of flee context.

After the generative model has been specified, all its parameters are specified through minimisation of free energy. This means there are only two parameters that can be

adjusted; namely, prior preferences about outcomes, C and prior precision or confidence in beliefs about policies, γ. In this case, the agent has no prior preference (i.e., flat priors) about locations but believes that it will correctly categorize a scene after it has accumulated sufficient evidence. Prior preferences over the outcome modalities were therefore used to encourage the agent to choose policies that elicited correct feedback $C^1 = [0, \dots, 0, c, -2c]: c = 2$, with no preference for sampled locations $C^2 = [0, \dots, 0]$. Here, c is the utility of making a correct categorization, $-2c$ is the utility of being wrong. These preferences mean that the agent expects to obtain correct feedback $\exp(c)$ times more than visual cues – and believes it will solicit incorrect feedback very rarely. Although the main role of these utilities is to encourage the agent to categorize the scene correctly, the relative utility of the correct and incorrect feedback can change the way the agent explores the scene. If the balance between these two utilities is off, the agent may get too eager to categorize the scene, neglecting the importance of information accumulation, e.g. the utility of the correct feedback is much greater than incorrect feedback or it may get apathetic to categorize the scene, e.g. the utility of correct feedback is much smaller than the incorrect feedback. The prior precision of beliefs about behaviours (policies or future actions) $\gamma = \frac{1}{\beta}$ plays the role of an inverse temperature parameter. As the precision increases, the sampling of the next action tends to be more deterministic; favouring the policy with the lowest expected free energy. Conversely as the precision of beliefs decreases the distribution of beliefs over the policies becomes more uniform; i.e., the agent becomes more uncertain about the policy it is pursuing.

With these preferences, the agent should act to maximise epistemic value or resolve uncertainty about the unknown context (the scene and its spatial transformations), until the uncertainty about the scene is reduced to a minimum. At this point, agent's beliefs about the context of the scene allows it to maximise extrinsic value by sampling the choice location it believes will provide feedback that endorses its beliefs. This speaks to the trade-off between exploration and exploitation. Essentially, actions associated with exploration of the scene (one of four quadrants) have no extrinsic value - they are purely epistemic. In contrast, actions associated with the choice locations (locations that are used to report the scene's category) have extrinsic value, because the agent is equipped with prior preferences about the consequences of these actions. The

contributions of epistemic and extrinsic value to policy (and subsequent action) selection are determined by their contributions to expected free energy (see Eq 7). In other words, there is only one imperative (to minimise free energy); however, free energy can always be expressed as a mixture of epistemic and extrinsic value. The relative contribution is determined by the precision of prior preferences, in relation to the epistemic part. The exploration and exploitation dilemma is resolved such that when the extrinsic value of the policies associated with a choice is greater than the epistemic value, the agent terminates the exploration and exploits one of the choice locations (i.e. declares its decision). This reflects a general behavioural pattern during active inference; namely, uncertainty is resolved via minimising a free energy that is initially dominated by epistemic value – until extrinsic value or prior preferences dominate and exploitation supervenes. Notice that pragmatic behaviour (choice behaviour) is driven by preferences in one modality (exteroceptive outcomes), while action is driven by predictions in another (proprioceptive sampling location). Despite this, action brings about preferred outcomes. This rests upon the recurrent belief updating that links the *'what'* and *'where'* streams in Fig 3.3. In this graphic, I have assumed that proprioceptive information has been passed from the trigeminal nucleus, via the superior colliculus to visual cortex (Donaldson, 2000).

Notice that in this work, the agent cannot resolve uncertainty about the sampling location afforded by the observations since there is no posterior uncertainty about the sampled locations. The resolution of uncertainty in terms of epistemic value gains information about the rest of the hidden state dimensions, namely, the context and the two spatial transformation dimensions.

3.2. Functional segregation and the mean field approximation

An important aspect of the belief updating in Fig 2.4 is that it is formulated for a particular form of posterior density. This form rests upon something called a *mean field approximation*, which is a ubiquitous device in Bayesian statistics (and statistical physics) (Jaakkola & Jordan, 1998). Fig 2.3 expresses the posterior as a product of independent marginal distributions over different sorts of hidden states (i.e., factors) at different time points, distributions over the policies and precision. Instead of encoding expectations of a full joint distribution over several factors (e.g., where an object is and what an object is), one just needs to represent both attributes in terms of

their marginal distributions. Similarly, instead of representing the entire trajectory of hidden states over time, one can approximate the trajectory by encoding expectations at each time point separately. This leads to an enormous simplification of the numerics and belief updating. However, there is a price to pay: because the posterior beliefs are conditionally independent, dependencies among the factors are ignored. Generally, this leads to overconfidence, when inferring hidden states – there is an example of this in simulated responses (see the next section).

From a neurobiological perspective, the mean field approximation corresponds to the principle of *functional segregation*, in which representations are anatomically segregated in the cortical mantle (Zeki & Shipp, 1988). A nice example of this is the segregation of ventral and dorsal visual processing streams that deal with *'what'* and *'where'* attributes of a visual scene respectively (Ungerleider & Mishkin, 1982). In the absence of a mean field approximation, there would be neuronal representations of every possible object in every location. It is this aspect of approximate (variational) Bayesian inference I emphasise in this chapter, by sketching the implications for large-scale functional anatomy. The segregation or factorisation into *'what'* and *'where'* attributes is particularly prescient for the oculomotor control of saccadic eye movements. This is because action is only specified by the states or attributes of the world that it can change. Clearly, saccadic eye movements only change where one is looking but not any other aspect of the world (e.g. what is sampled). This means that only one factor or posterior marginal is sufficient to prescribe action.

In visual search paradigms one can use the saccadic choices of the subjects as a surrogate for the brain responses as the cognitive processes give rise to vision. In theory one can fit neurobiologically plausible models (such as the one introduced in this chapter) to the choice behaviour of the subjects and estimate subject specific model parameters. Crucially, these model parameters are hypothesised to be related to distinct neural circuits. Estimating these model parameters and validating their associations with distinct neural components may allow for bridging the gap between the behaviour under psychiatric conditions and their pathophysiologies.

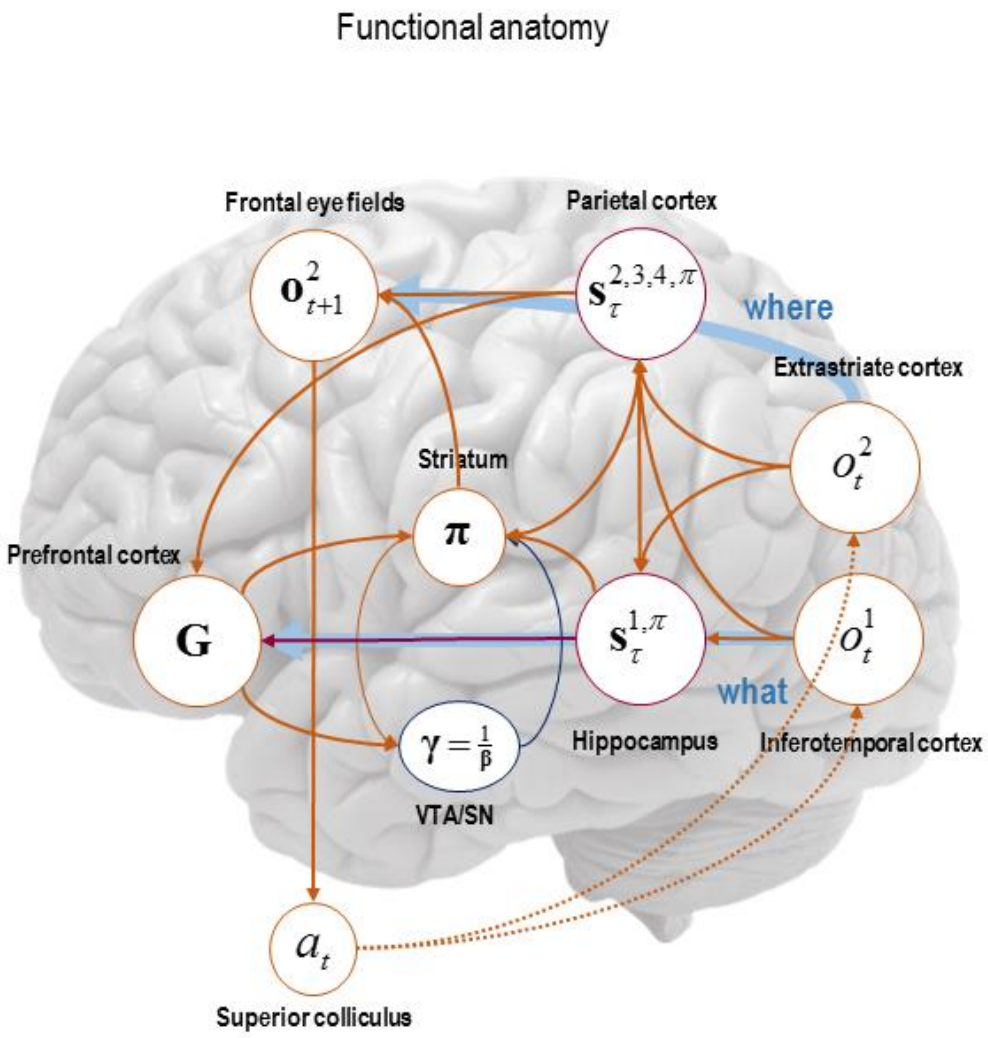


Figure 3. 3 Schematic overview of the belief updates describing active inference

In this figure the quantities that are updated (sufficient statistics or expectations) are assigned to various brain areas. The implicit attribution should not be taken too seriously but serves to illustrate the functional anatomy implied by the form of the belief updates. Here, observed outcomes have been assigned to visual representations in the occipital cortex; with exteroceptive (*what*) modalities entering a ventral stream and proprioceptive (*where*) modalities originating a dorsal stream. Hidden states encoding context have been associated with the hippocampal formation, while the remaining states encoding sampling location and spatial invariance have been assigned to the parietal cortex. The evaluation of policies, in terms of their (expected) free energy, has been placed in the ventral prefrontal cortex. Expectations about policies *per se* and the precision of these beliefs have been associated with striatal and ventral tegmental areas to indicate a putative role for dopamine in encoding precision. Finally, beliefs about policies are used to create Bayesian model averages of future outcomes (in the frontal eye fields) – that are fulfilled by action, via the deep layers of the superior colliculus. The arrows denote message passing among the sufficient statistics of each factor or marginal. Please see the text and Table 1 for an explanation of the equations and variables.

For illustrative purposes, Fig 3.3 shows how the variables and model parameters in this scheme could be encoded in the brain. The encoding of object identity is assigned to inferotemporal cortex (Seeck et al., 1995). The representation of location is associated with (dorsal) extrastriate cortex (Haxby et al., 1994). Beliefs about sampling locations and spatial invariances are attributed to parietal cortex, which anticipates the retinal location of stimuli in the future and updates the locations of stimuli sampled in the past (Duhamel, Colby, & Goldberg, 1992). Inference about scene identity (based on the spatial relationships among objects) is attributed to the hippocampus (Rudy, 2009). Beliefs about policies are assigned to the striatum (Frank, 2011), which receives inputs from prefrontal cortex, ventral tegmental area and hippocampus to coordinate planning (in prefrontal cortex, see (Tanji & Hoshi, 2001)) and execution (VTA/SN), given a particular context. In active inference, action selection depends upon the precision of beliefs about policies (future behaviour), encoded by dopaminergic projections from VTA/SN to the striatum (Philipp Schwartenbeck, Thomas HB FitzGerald, Christoph Mathys, Ray Dolan, & Karl Friston, 2014). Frontal eye fields are involved in saccade planning (Krishna Srihasam, Daniel Bullock, & Stephen Grossberg, 2009) and the superior colliculus mediates eye movement control (Stephen Grossberg, Karen Roberts, Mario Aguilar, & Daniel Bullock, 1997) – by fulfilling expectations about action that are conveyed from frontal eye fields.

Summary

By assuming a generic (Markovian) form for the generative model, it is fairly simple to derive Bayesian updates that clarify the interrelationships between perception, policy selection, precision and action. In brief, the agent first infers the hidden states under each model or policy based on the outcomes it observes. It then evaluates the evidence for each policy in terms of the free energy of hidden states, expected free energy and precision. Beliefs about policies are used to form a Bayesian model average of the next states and thus outcomes, which is realised through action. The anatomy of the implicit message passing is not inconsistent with functional anatomy in the brain: see (K. Friston et al., 2014) and Fig 3.3. Fig 3.3 shows the functional anatomy implied by the mean field approximation and belief updating in Fig 2.3 and Fig 2.4, respectively. Here, I have assumed two input modalities (*what* and *where*) and

four sets of hidden states; first one encodes the context of the scene, second one encodes the location that the agent sampled, and the last two encode spatial transformations that define the absolute locations of the cues.. The anatomical designation in Fig 3.3 should not be taken too seriously – the purpose of this illustration is to highlight the recurrent message passing among the expectations that constitute beliefs about segregated or factorised states of the world. Here, I emphasise the segregation between what and where streams – and how the dorsal *where* stream supplies predicted outcomes (to frontal eye fields) that action can realise (via the superior colliculus). The precision of beliefs about policies has been assigned to dopaminergic projections to the striatum. I will use this particular architecture in the next section to illustrate the behavioural (and electrophysiological) responses that emerge under this scheme.

The following simulations can be reproduced (and modified) by downloading the DEM Toolbox and invoking **DEM_demo_MDP_search.m**. This annotated code can also be edited and executed via a graphical user interface; by typing >> DEM and selecting the **Visual foraging** demo. This demo can be compared with the equivalent variational filtering scheme (for continuous state-space models) in the **Visual search** demo, described in (K. Friston, Adams, Perrinet, & Breakspear, 2012).

3.3. Simulating saccadic searches

Fig 3.4 shows the results of updating the equations in Fig 2.4, using 16 belief updates between each of five saccades. Beliefs about hidden states are updated using a gradient descent on variational free energy. This gradient descent usually converges to a minimum within about 16 iterations. I therefore fixed the number of iterations to 16 for simplicity. This imposes a temporal scheduling on belief updates and ensures that the majority (here, more than 80%) of epochs attain convergence (this convergence can be seen in later figures, in terms of simulated electrophysiological responses). The belief updates are shown in terms of posterior beliefs about hidden states (upper left panels), posterior beliefs about action (upper centre panel) and the ensuing behaviour (upper right panel). Here, the agent constructed policies on-the-fly by adding all possible actions (saccadic movement to the eight possible locations) to previous actions. This means that the agent only looks one move ahead – and yet manages to make a correct categorisation in the minimum number of saccadic eye

movements: in this trial, the agent first looks to the lower right quadrant and finds a distractor (omitted in the figures for clarity). It then samples the upper quadrants to resolve uncertainty about the context, before soliciting feedback by choosing the (correct) choice location. The progressive resolution of uncertainty over the three initial saccades is shown in more detail in the lower panels.

Here, posterior beliefs about the state of the world (the nature of the canonical scene and spatial transformations) are illustrated graphically by weighting the predicted visual cue – under each state – in proportion to the posterior beliefs about that state. Each successive image reports the posterior beliefs after the first three saccades to the peripheral locations, while the insert in the centre is the visual outcome after each saccade. Initially, all four peripheral cue locations could contain any of the visual objects; however, after the first saccade to the lower right quadrant, the agent believes that the objects (bird and seed or cat) are in the upper quadrants. It then confirms this belief and resolves uncertainty about vertical reflection by sampling the upper right quadrant to disclose a bird. Finally, to make a definitive decision about the underlying scene, it has to sample the juxtaposed location to resolve its uncertainty about whether this contains seed or cat. Having observed cat, it can then make the correct choice and fulfil its prior beliefs or preferences by reporting its beliefs about the category of the scene.

This particular example is interesting because it illustrates the overconfidence associated with a mean field approximation. Note that after the first saccade the agent assumes that the scene must be either a *feed* or *flee* category, believing the scene is flipped horizontally. This is reflected in the fact that the lower quadrants are perceived as empty and the belief that the bird cannot exist in the top left quadrant under its posterior beliefs. If the agent was performing exact Bayesian inference it would allow for the possibility of a *wait* scenario, with the bird and seed on the diagonal locations. In this instance, it would know that there must have been either a vertical or horizontal reflection (but not both). However, this belief cannot be entertained under the mean field approximation, because inferring a vertical or horizontal reflection depends on whether or not the scene is a *wait* category, and the beliefs about the hidden state dimensions that encode the spatial transformations and the context are conditionally independent. These conditional dependencies are precluded by the mean field

approximation; in other words, posterior beliefs about one hidden states (e.g., reflection) cannot depend upon posterior beliefs about another (e.g., scene category). The agent therefore finds the most plausible explanation for the current sensory evidence, in the absence of conditional dependencies; namely, there has been no vertical reflection and the scene is not *'wait'*. If the brain does indeed use mean field approximations – as suggested by the overwhelming evidence for functional segregation – one might anticipate similar perceptual synthesis and saccadic eye movements in human subjects. In principle, one could compare predictions of empirical eye movements under active inference schemes with and without mean field approximations – and test the hypothesis that the brain uses marginal representations of the sort assumed here (see discussion).

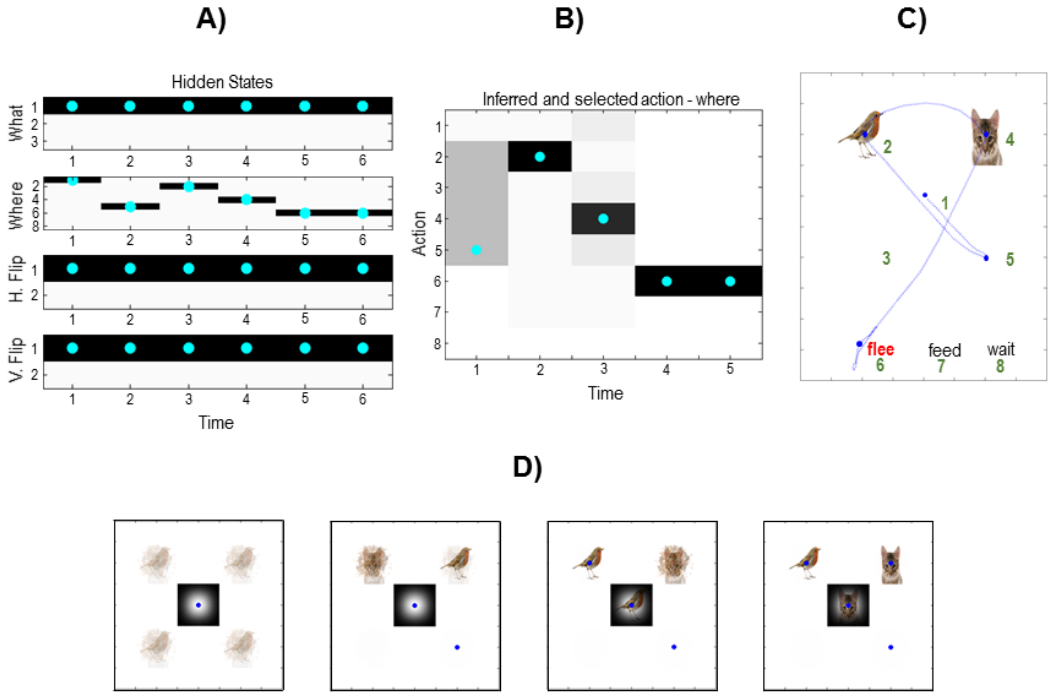


Figure 3. 4 Simulated visual search

A) This panel shows the expectations about hidden states and the expectations of actions are shown in **B)** (upper middle), producing the search trajectory in **C)** – after completion of the last saccadic movement. Expectations are shown in image format with black representing 100% probability. For the hidden states each of the four factors or marginals are shown separately, with the true states indicated by cyan dots. Here, there are five saccades and the agent represents hidden states generating six outcomes (the initial state and five subsequent outcomes). The results are shown after completion of the last saccadic, which means that, retrospectively, the agent beliefs it started in a *flee* context, with no horizontal or vertical reflection. The sequence of sampling locations indicates that the agent first interrogated the lower right quadrant and then emitted saccades to the upper locations to correctly infer

the scene – and make the correct choice (indicated by the red label). The lower panel **D)** illustrates the beliefs about context during the first four saccades. Initially, the agent is very uncertain about the constituents of each peripheral location; however, this uncertainty is progressively resolved through epistemic foraging, based upon the cues that are elicited by saccades (shown in the central location). The blue dots indicate the sampling location after each saccade.

3.3.1. Electrophysiological correlates of variational belief updating

Fig 3.5 shows the belief updating during the above visual search to emulate electrophysiological responses measured in empirical studies. The upper left panel shows simulated neuronal activity (firing rates) for units encoding the first (scene category) hidden state using an image (or raster) format. Each row of blocks shows the beliefs about the category of the scene over six time steps. In each row of blocks the top, middle and bottom lanes correspond to the beliefs about the scene categories flee, feed and wait respectively (the change in beliefs about the context are shown with colours grey, white and black). Units here correspond to the expectations (posterior probabilities) about hidden states of the world. Each block (square) of the raster encodes the activity over 16 time bins (belief updates) between one saccade and the next, with one hidden state in each row. The beliefs about the context (*flee*, *feed* or *wait*) hidden state dimension are shown over 6 different time steps. Crucially, there are two sorts of time shown in these responses. Each row of blocks reports expectations about one of the three categories at different times in the future (or past) – here, beliefs about the context following each of the six saccades. Each column of blocks shows the expectations (about the past and future) at a particular point during the trial. These expectations in the columns show the agent's beliefs about the context projected into the future and the past. This effectively shows the beliefs about the hidden states in the past and the future. For example, the second row of blocks summarises belief updates about the second epoch over subsequent saccades; i.e., expectations about the context in the second saccade are updated in the following saccades (blocks to the right), while the first column of blocks encodes beliefs about future states prior to emission of the first saccade; i.e., expectations about the context in the second time step is projected into the past (one block above) and into the future (one block below). This means beliefs about the current state occupy blocks along the leading diagonal (highlighted in red), while expectations about states in the past and future are above and below the diagonal respectively. For example, the colour density

in the first row denotes the posterior probability of the context being ‘flee’ during the first epoch at different saccades: this expectation about context prior to the first saccade only becomes definitive at around 0.9 s (during the fourth saccade). Conversely, row 12 denotes the posterior probability of ‘wait’ during the fourth epoch: note that this converges to zero before the fourth saccade has occurred.

This format illustrates the encoding of states over time, emphasising the implicit representation of the past and future. To interpret these responses in relation to empirical results, one can assume that outcomes are sampled every 250 ms (K Srihasam et al., 2009). Note the changes in activity after each new outcome is observed. For example, the two units encoding the first two hidden states start off with uniform expectations over the three scenes that switches after the second and fourth saccade to eventually encode the expectation that the first (*flee*) scene is being sampled. Crucially, by the end of the visual search, these expectations pertain to the past; namely, the context at the start of the trial. In other words, these memories are based upon postdiction. Postdiction allows two things here. Firstly, it enables evidence accumulation under each policy (one of eight) by projecting the beliefs about the hidden states into the past as new observations become available. This means that the beliefs about the hidden states become approximately equal. Secondly, since the evidence under each policy (one of eight) is approximately equal, the agent needs to consider only one of eight actions when looking one step ahead while having the same accumulated evidence under each policy in the past time steps. Postdiction allows evidence accumulation and reduces the computation burden of having a large policy tree. Although not illustrated here, this can be very useful when updating beliefs between trials (when the context does not change).

The upper right panel plots the same information (expectations about the hidden states) to highlight saltatory evidence accumulation, in which expectations diverge as the search progresses. This belief updating is very similar to evidence accumulation described by drift diffusion or race-to-bound models (de Lafuente, Jazayeri, & Shadlen, 2015; Kira, Yang, & Shadlen, 2015). Drift diffusion models describe evidence accumulation for different alternatives over time and this process is subject to noise. Under these models a response is initiated once a decision boundary is reached as a result of evidence accumulation. This is similar to how evidence accumulation occurs

under active inference where the evidence for different alternatives is accumulated as new observations are made. One can define a decision boundary – a criterion to terminate belief updates over the hidden states – in terms of the entropy over the posterior probability distribution over the hidden states as well.

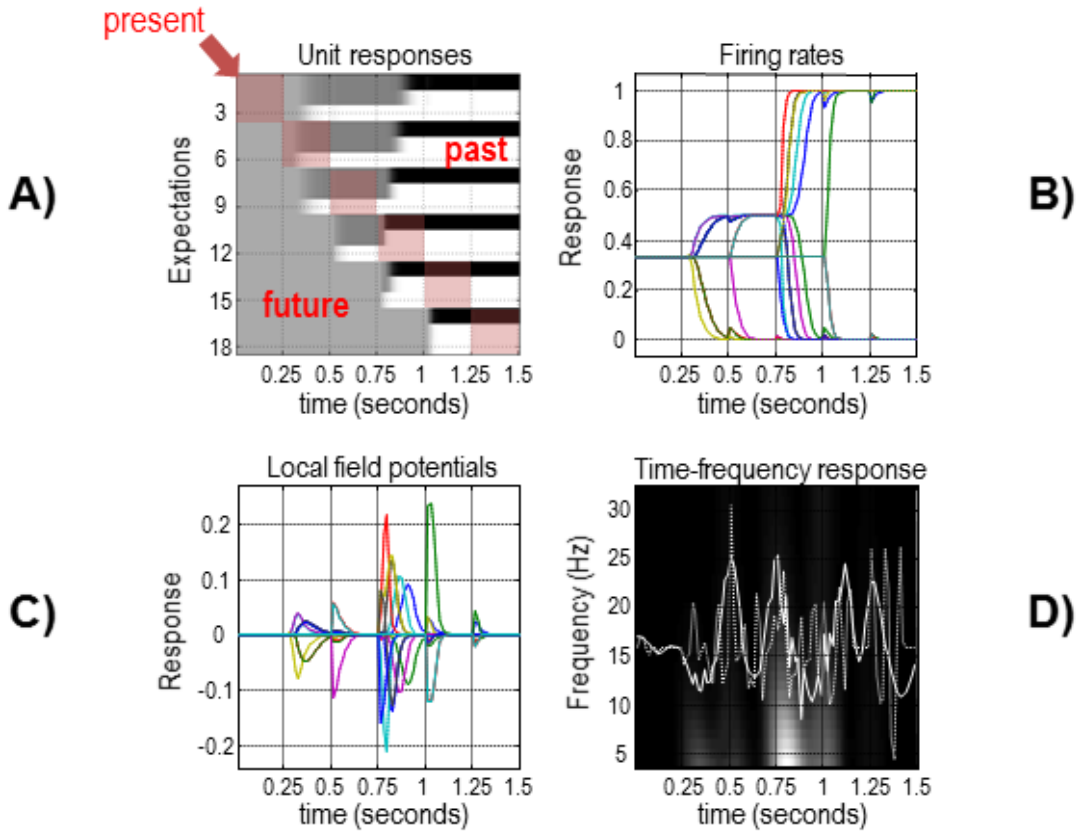


Figure 3. 5 Simulated electrophysiological responses

This figure reports the belief updating behind the behaviour shown in the previous figure. **A)** The upper left panel shows the activity (firing rate) of units encoding the context or scene in image (raster) format, over the six intervals between saccades. These responses are organised such that the upper rows encode the probability of alternative states in the first epoch, with subsequent epochs in lower rows. **B)** The upper right panel plots the same information to illustrate the evidence accumulation and the resulting disambiguation of context. **C)** The simulated local field potentials for these units (i.e. the rate of change of neuronal firing) are shown in the middle left panel. **D)** The middle right panel shows average local field potential over all units before (dotted line) and after (solid line) bandpass filtering at 4 Hz, superimposed upon its time frequency decomposition.

Furthermore, the separation of timescales implicit in variational updating reproduces the stepping dynamics seen in lateral intraparietal responses during decision-making in macaque studies (Latimer, Yates, Meister, Huk, & Pillow, 2015). The middle left

panel shows the associated local field potentials, which are simply the rate of change of neuronal firing (beliefs about contexts) shown on the upper right panel. The bottom right panel of Fig 3.5 shows the simulated local field potential averaged over all units before (dotted line) and after (solid line) bandpass filtering at 4 Hz. These responses are superimposed on its time frequency decomposition. The key observation here is that depolarisation in the theta range coincides with induced responses – a theme that is pursued elsewhere in terms of theta-gamma coupling in the brain (Canolty et al., 2006; K. Friston et al., 2014; J. Lisman & Redish, 2009). Theta-gamma coupling has been observed in hippocampal place cell activity in rodent studies (J. E. Lisman & Jensen, 2013). Although not shown here, the same variational message passing scheme can produce phase precession effect seen in hippocampal place cells (K. Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017). Moreover, studies on rodents suggest that theta-gamma coupling in the hippocampus might be necessary for memory retrieval (Shirvalkar, Rapp, & Shapiro, 2010).

Collectively, these simulated electrophysiological responses are not dissimilar to the sorts of responses recorded in empirical studies; however, in this chapter, I am primarily interested in modelling (epistemic) behaviour. Fig 3.5 shows some of the expectations that are updated using the equations presented in Fig 2.4. These simulated electrophysiological responses can be associated with activity in the various brain regions in Fig 3.3; e.g.., expectations about hidden states encoding context $s_\tau^{1,\pi}$ with the hippocampus. In the final section, I consider multiple trials and how performance depends upon prior preferences and precision.

3.3.2. Sequences of simulations

Fig 3.6 summarises the (simulated) behavioural and physiological responses over 32 successive trials in which the context (scene and spatial transformations) was selected at random. Each trial comprises six saccades following an initial fixation. The first panel shows two sorts of things. Firstly, the coloured circles from one to four on the y-axis show the initial states on each trial. These are context, sampling location on the first time step (always central fixation first), horizontal and vertical flips, respectively. The initial states (except for the sampling location) determines how the scene is going to look like in terms of the absolute and relative locations of the visual cues. Secondly, the numbers one to eight on the same y-axis show the subsequent policy selection (in

image format) over the eight actions (i.e., locations) considered on the final time step**.** Here, the actions 1 to 5 correspond to visiting the central fixation point and quadrants with cues (locations 2 to 5), whereas actions 6-8 select the locations reporting the choice (flee, feed and wait). Choice locations are just there to enable the agent to report its beliefs about the scene category. Since the agent usually arrives at a decision in a few saccades, the selected action on the final time step is usually one of the choice locations, with the exception of the third trial where the agent fails to categorize the scene even after the 6-th saccade and chooses the sampling location two (the top left quadrant). The second panel shows three sorts of quantities; namely, agent's decision about the category of the scene and whether this categorization is correct (encoded by coloured circles) and performance in terms of expected utility and reaction time. The first row of coloured dots shows the sampled choice location (*flee*, *feed* or *wait*) on the last saccade, whereas the second row of dots shows whether the agent's beliefs about the category of the scene is correct. Only on the third trial agent is unable to correctly categorize a scene, here the brown dot shows that the sampled cue on the last saccade is *seed* and not the cue associated with the correct feedback. Second quantity, expected utility (black bars), is the utility of the observed outcomes averaged over time and defined with the following formula:

$$U = \sum_{g \in \{what, where\}} \sum_{t=1}^{T=6} \frac{\ln\left(C^g(o_t^g)\right)}{T} \tag{10}$$

Where C^g is the preference matrix over the outcome modality g and o_t^g designates the entry in the preference matrix on the *o-th* row and *t-th* column under the modality *g*. The utility of an outcome is defined by the prior preference. Note that because preferences are log probabilities they are always negative – and the best outcome is zero. The performance measure differs across trials because the number of saccades the agent employs before categorizing a scene differs from trial to trial. The third quantity, reaction times or saccadic intervals (cyan dots), here are based upon the actual processing time in the simulations and are shown after normalisation to a mean of zero and standard deviation of one. Reaction time is defined as the actual processing time (using Matlab *tic-toc* facility) in the simulations. This definition is based upon the assumption that belief updates in the brain – via neuronal message passing

– follow a similar scheduling to the exchange of sufficient statistics described in Fig 3.3.

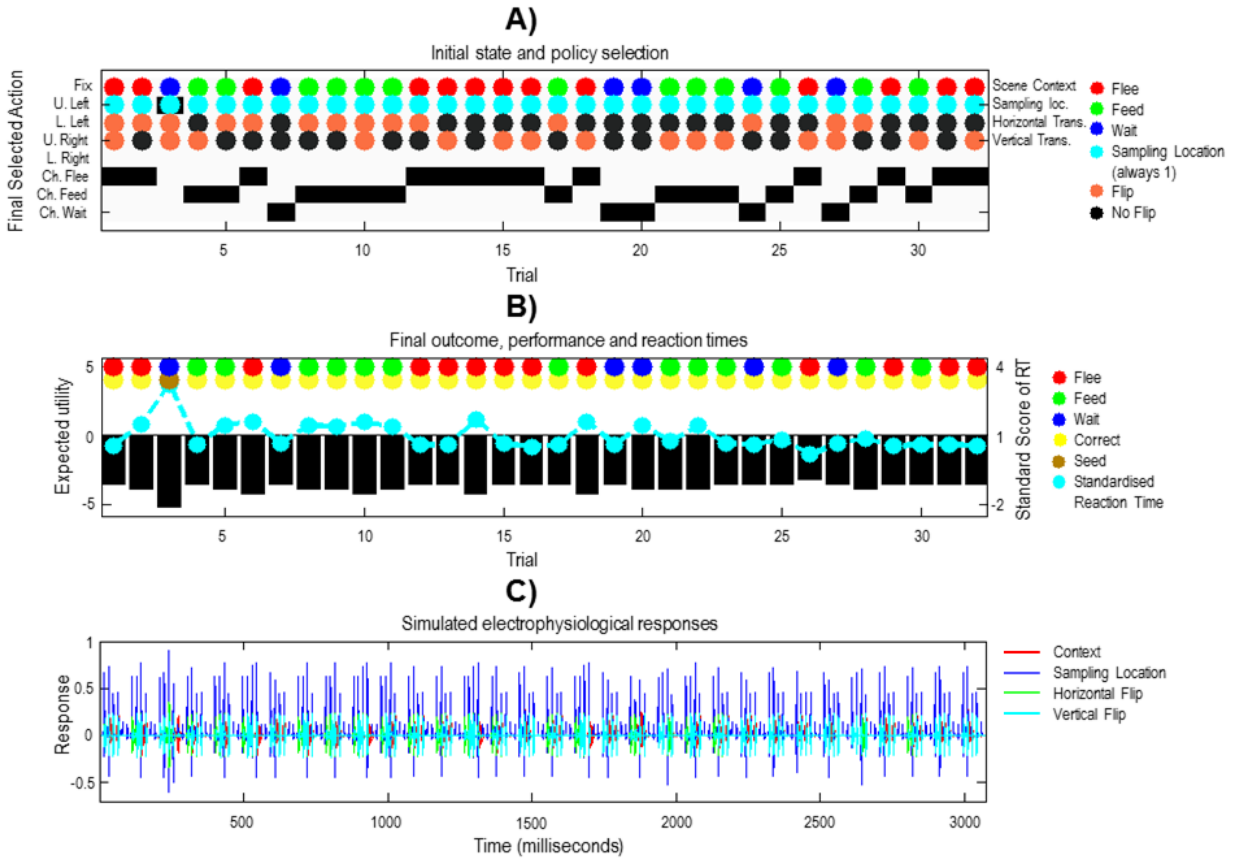


Figure 3. 6 Simulated responses over 32 trials

This figure reports the behavioural and (simulated) physiological responses during 32 successive trials. The scenes in these 32 trials were specified via randomly selected hidden states of the world. **A)** The first panel shows the hidden states of the scene (as coloured circles) and the selected action (i.e. the sampled location) on the last saccade. The y-axis on this panel shows two quantities. The selected action is shown using black bars. The agent can saccade to locations 1 to 8, where the locations 6 to 8 correspond to the choice locations the agent uses to report the scene category. The true hidden states are shown with coloured circles. These specify the objects in the scene and their locations (in terms of the context and spatial transformations). The second row of cyan dots indicates that the agent always starts exploring a scene from the central fixation point. Individual rows in the y-axis indicate the sampled locations according to the following: Fix (Fixation), U. Left (Upper left), L. Left (Lower Left), U. Right (Upper Right), L. Right (Lower Right) and Ch. Flee (Choose Flee), Ch. Feed (Choose Feed) and Ch. Wait (Choose Wait). **B)** The second panel reports the final outcomes (encoded by coloured circles) and performance measures in terms of preferred outcomes (utility of observed outcomes), summed over time (black bars) and standardized reaction times (cyan dots). The final outcomes are shown for the sample location (upper row of dots) and outcome (lower role of dots): yellow means the agent made a correct choice. **C)** The third panel shows a succession of simulated event related potentials following

each outcome. These are taken to be the rate of change of neuronal activity, encoding the expected probability of hidden states encoding context (i.e., simulated hippocampal activity).

These simulations show that, with the exception of the third trial, the agent makes veridical decisions on every occasion. Interestingly, the third (incorrect) trial is associated with the greatest reaction time. Reaction time here varies because the minimisation of free energy converges on different number of iterations under a certain tolerance (here, the variational updates terminate when the decrease in free energy falls below 1/128). The lower panel shows the simulated electrophysiological responses using the same format as in the previous figure. Here bursts of high-frequency activity is seen every hundred milliseconds or so; in other words, a nesting of gamma activity in the alpha range.

The associated behaviour, over the first nine trials is depicted in Fig 3.7. Again, with the exception of the third trial, optimal search behaviour is seen, with a correct choice after the minimum number of saccades. For example, on the first trial, the first saccade samples a bird, which just requires a second saccade to the adjacent location in order to completely disambiguate the context (see Fig 3.2). A detailed analysis of the belief updating for the failed trial suggested that this was an unlucky failure of the mean field approximation; particularly the factorisation over time – and a partial failure of convergence due to the use of a fixed number (i.e., 16) of iterations. These sorts of failures highlight the distinction between exact Bayesian inference and approximate Bayesian inference that may underlie bounded rationality in real agents. With these simulated responses is at hand, I can now assess the effects of changing prior preference and priors over the precision of beliefs about action or policies.

Clearly, there are many model parameters (and hyperparameters) one could consider, in terms of their effects on simulated behaviour. Here, I focused on the precision of preferences and policies because these correspond intuitively to the different aspects of salience. Motivational salience can be associated with the preferences that incentivise choice behaviour. Conversely, the precision of beliefs about policies speaks to the visual salience associated with information gain and epistemic value, and confidence in beliefs about policies. Heuristically, one might expect different patterns of behaviour depending upon whether subjects have imprecise preferences (i.e., are not confident about what to do), as opposed to imprecise beliefs about the

consequences of their actions (i.e., not confident about how to do it). In what follows, I address this heuristic using simulated behaviour.

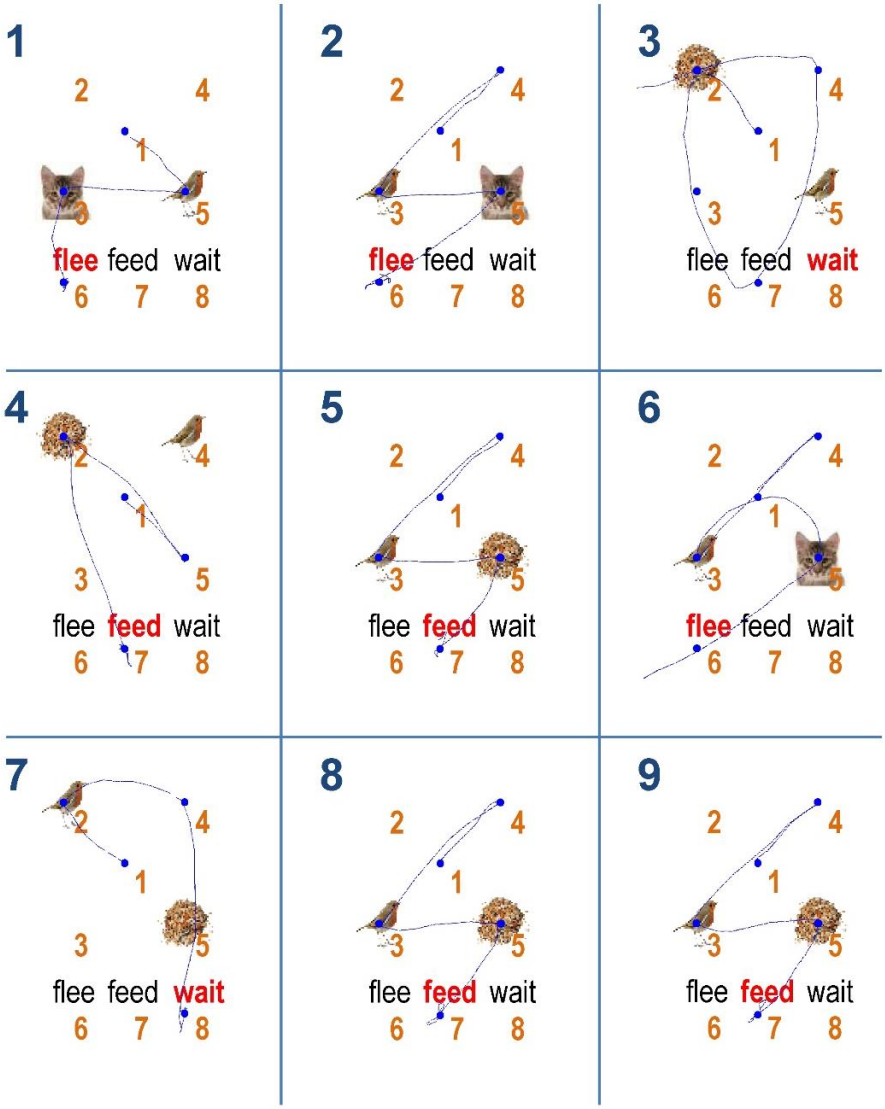


Figure 3. 7 Sequences of saccades

This figure illustrates the behaviour for the first nine trials shown in the previous figure using the same format as Fig 3.4C. The numbers on the top left in each cell show the trial number. With the exception of the third trial, the agent is able to recognise or categorise the scene after a small number of epistemically efficient saccades.

3.4. The effects of priors

Finally, Fig 3.8 reports the performance during presentations of 300 trials, where hidden states of the world were selected randomly – and the agent was allowed to make up to 8 saccades. I measured the performance over these trials in terms of

percent accuracy (a correct choice in the absence of an incorrect choice), decision time or number of saccades until any (correct or incorrect) choice and reaction time or saccadic interval (measured in seconds). Here, I repeated the 300 trial paradigm over all combinations of nine levels of prior preference and precision. To manipulate the precision of preferences, I increased the parameter *c* – specifying the prior preference for not making an incorrect decision from no preferences to very precise preferences (from zero to four), while keeping the prior preference for making a correct decision fixed.

The left panel in Fig 3.8 (Accuracy) shows that accurate categorisation requires both precise preferences and a high precision. Interestingly, precise prior preferences degrade accuracy when the prior precision is very low, as it can be seen on the first column. With greater prior preference, the agent does not want to make mistakes. However a low prior precision precludes a resolution of uncertainty about the scene. The combination of these two priors discourages the agent from making a choice, resulting in an incorrect categorization. The trials where agent doesn't attempt to categorize the scene are considered an incorrect categorization. When prior preferences are less precise, the agent is less afraid of making an incorrect choice thus allowing the agent to make some categorizations which turn out to be correct (notice that the accuracy is still below the chance level). Similarly, greater prior precision does not improve accuracy when prior preference is low. In short, the agent only respond accurately when prior preference and precision are high, as seen on the upper right portion of the image.

The centre panel (Decision Time) shows decision time in terms of number of saccades before choosing a choice location. When prior preferences are high and prior precision is very low (first column), it takes 7 or 8 saccades for the agent to make a decision. Comparing this figure with the accuracy results, it can be seen that accuracy is low even though the agent is making more saccades; i.e. taking its time. This is because the agent accumulates information against the competing hypothesis but in the absence of precise preferences, it is unable to fulfil its goal of categorizing the scene. When prior precision is high but prior preference is very low, the agent rushes to make a decision – but in the absence of precise prior preferences it makes mistakes (see

left panel). In short, the agent successfully categorizes a scene when it deploys 3-4 saccades (upper right quadrants), under precise preferences and high precision.

The right panel (Reaction Time) shows the reaction time in terms of actual processing time of the simulations. Although, quantitatively, reaction times only vary between about 800 and 900 ms, there seems to be a systematic effect of prior precision, with an increase in reaction time at very low levels.

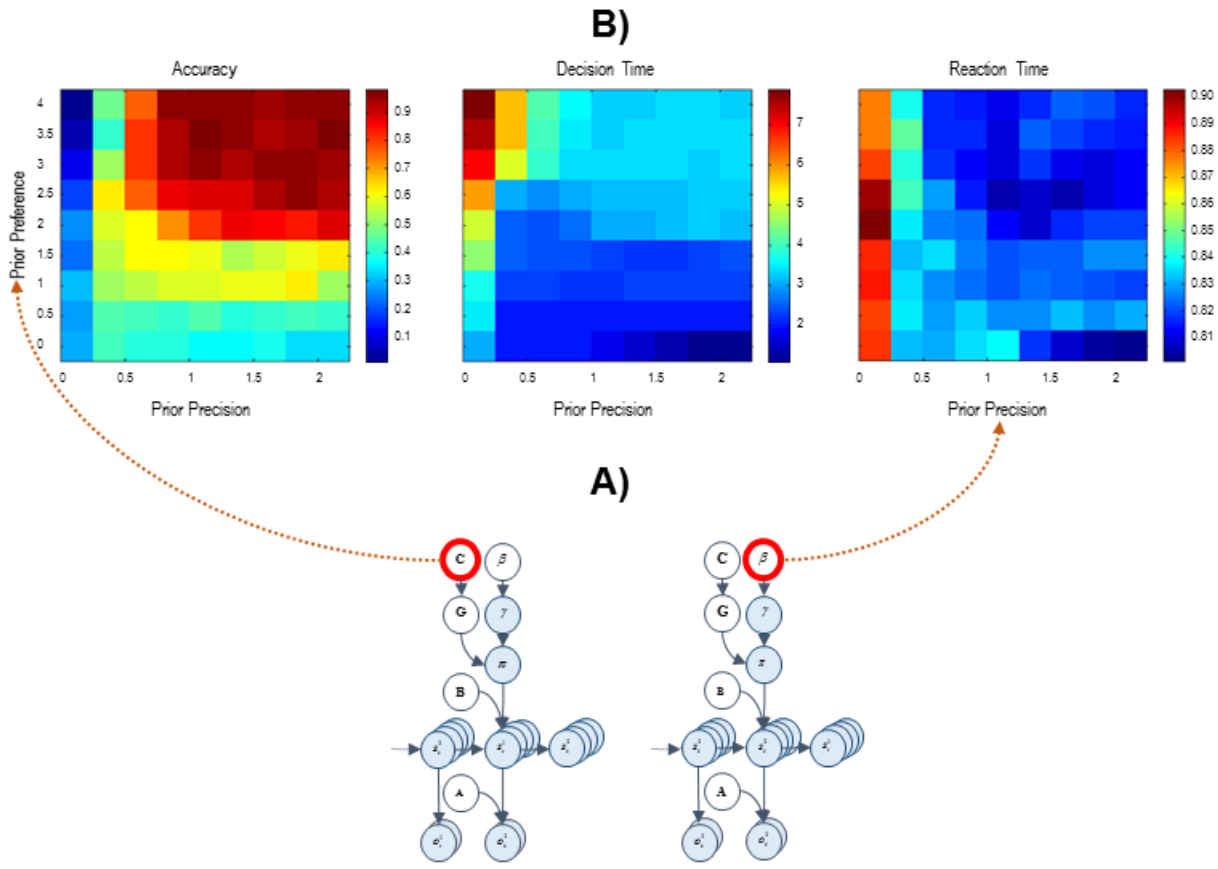


Figure 3. 8 Performance and priors

This figure illustrates the average performance over 300 trials. **A)** The insert (lower panel) shows the prior parameters that were varied; namely, prior preference and precision. These parameters are varied over nine levels. **B)** For each combination, the accuracy, decision and reaction time were evaluated using simulations (upper row). The accuracy is expressed as the percentage of correct trials (defined as a correct choice in the absence of a proceeding or subsequent incorrect choice). Decision time is defined in terms of the number of saccades until a (correct or incorrect) decision. Reaction time or the interval between saccades is measured in seconds and corresponds to the actual computation time during the simulations.

Crucially, results demonstrate a distinct dependency of accuracy and decision time on prior preference and prior precision. This speaks to the possibility of distinct behavioural phenotypes that are characterised by different combinations of prior preference and precision. For example, agents who do not expect themselves to make mistakes may choose more assiduously, inducing a classical speed accuracy trade-off. Conversely, subjects with more precise beliefs about their choices may behave in a more purposeful and deliberate fashion, taking less time to obtain preferred outcomes. I pursue this theme in the discussion.

3.5. Discussion

In this chapter, I have presented an active inference formulation of epistemic foraging that provides a framework for understanding the functional anatomy of visual search entailed by sequences of saccadic eye movements. This formulation provides an elementary solution to the problem of scene construction in the context of active sensing and sequential policy optimisation, while incidentally furnishing a model of spatial invariance in vision.

Although the problem considered in this chapter is relatively simple, it would confound most existing approaches. For example, reinforcement learning and optimal control theories are not applicable because the problem is quintessentially epistemic (belief-based) in nature. This means that the optimal action depends on beliefs or uncertainty about hidden states. This context sensitivity precludes any state-action policy and implicitly any scheme based on the Bellman optimality principle (Bellman, 1952). This is because the optimal action from any state depends upon beliefs about that state and all others. Although, in principle, a belief-state (partially observed) Markov decision process could be entertained (Bonet & Geffner, 2014), the combinatorics of formulating beliefs states over $3\times 8\times 2\times 2 = 96$ hidden states are daunting and computationally burdensome. Furthermore, given the problem calls for sequential policy optimisation – and that five moves are necessary to guarantee a correct categorisation – one would have to evaluate $8^5 = 32768$ policies. Use of the postdiction in this application of active inference eliminates the issue of evaluating large number of policies.

The active inference solution offered here is based upon minimising the path-integral of (expected) free energy under a mean field approximation. The exciting thing about this approach is that, computationally, it operates (nearly) in real-time. For example, the reaction times in Fig 3.8 are based on the actual computation time using a standard desktop personal computer. This computational efficiency may be useful for neurorobotic applications. Having said this, the primary motivation for developing this scheme was to characterise empirical (human) visual searches given observed performance, eye movement and electrophysiological responses.

The example in this chapter has some limitations: for example, all potential spatial combinations of objects can be obtained using just two transformations (e.g. the cat can never be below the bird), and scenes in larger grid worlds may not be describable in terms of simple transformations from a small number of contexts. Clearly, the brain does not use the mean field approximation used to illustrate the scheme – but questions about different forms of meaningful approximations can, in principle, be answered empirically using Bayesian model comparison of such approximations when explaining behavioural or neuroimaging data.

This toy example shows how a scene comprising 2x2 quadrants can be explored using the resolution of uncertainty. A scene of this small size could be explored systematically, if inefficiently, (e.g., in a clockwise manner) or by just visiting all locations randomly. However, more complex scenes – which I hope to use in future work – could not be categorised efficiently in such a fashion. I used this paradigm to characterise different behavioural phenotypes in terms of the free parameters of this model (see chapter 4).

Although the accuracy, number of saccades and saccadic intervals (Fig 3.8) provide a degree of validation for active inference in this setting, it is unlikely that these responses will provide an efficient estimate of subject-specific priors, such as prior preferences and precision. However, it is relatively easy to fit the individual saccadic eye movements by evaluating the probability of *each saccade* in relation to posterior beliefs about action, using the history of action and outcomes in the model above. This allows estimating things like prior preference and precision efficiently, given the sequence of eye movements from any subject. I used the active inference scheme described in this chapter to explain empirical eye movements in terms of subject-

specific priors (see chapter 4). This enables one to simulate or model electrophysiological responses or identify the regional correlates of belief updating, using functional magnetic resonance imaging.

In chapter 4, I implement the scene construction task described above for visual searches on human subjects. I investigate the behavioural measures and the saccadic choices of the subjects that are registered using an eye-tracking device. Using a model inversion scheme formulated in terms of active inference (Schwartenbeck & Friston, 2016), I estimate model parameters from the MDP model (e.g. prior precision) by fitting a model to the saccadic choices of the subjects.

4. Characterising salience attribution under active inference

This chapter is about salience attribution in visual searches. In other words, how do we identify salient targets during saccadic (visual) searches of our visual scenes – and what sorts of policies and prior beliefs underwrite this attribution and subsequent epistemic foraging. To address this question, I applied the model described in the previous chapter in normal subjects and tried to explain their eye movements in terms of Bayes optimal epistemic sampling. In this chapter, I consider the evidence that normal subjects conform to normative (i.e., Bayesian) principles and how this can be used to characterise individual differences.

Visual exploration entails seeking out relevant information, given a context. But what is information? Shannon's definition of information (Shannon, 1948) implies that an outcome that is less predictable contains more information. Shannon entropy is the average or expected information. Shannon entropy is highest when all outcomes are equally likely; i.e., when the outcome is most unpredictable. However, Itti and Baldi (2009) demonstrated that whilst human visual attention is attracted to areas of high Shannon information, it is attracted most strongly to areas that cause the greatest shifts in our beliefs about the world. This notion is formalised as 'Bayesian surprise' (Itti and Baldi 2009): the KL divergence between prior and posterior beliefs about how our sensory data are generated. In the active inference framework, stimuli of greater Bayesian surprise have more epistemic value and this is the formal basis of the current work.

In the previous chapter I introduced an active inference scheme for visual searches using a scene construction task. I showed how a scene currently explored optimally, when a synthetic subject engages in epistemic foraging. In this chapter, I ask whether human subjects perform the same task in an epistemic fashion; i.e., resolving uncertainty about the hidden states of the world. To test this hypothesis I fitted an active inference model to the saccadic choices of my subjects and evaluated the evidence for epistemic foraging.

This chapter comprises three sections. In the first section I reiterate the MDP formation for the scene construction task and preview the analyses of the saccadic scan-paths of subjects performing the scene construction task. The second section describes the

empirical methods for the gaze-contingent protocol I used in the eye-tracking study, the subjects that performed the task and scan-path recording methods. In the third section, I report behavioural results that characterise task performance. This is followed by the analysis of the scan-path choices, using the model of the first section to estimate their prior beliefs. I then report the canonical correlations between prior beliefs and the behavioural measures to understand overt behaviour in terms of characteristic subject 'types'. In the discussion, I discuss the results in terms of active inference and their implications for computational phenotyping of individual subjects.

4.1. Active inference and visual search

4.1.1. The MDP formation

The Fig 3.1A shows the generic form of the MDP model and the conditional dependencies in the generative models. This panel shows how the outcomes are generated from the hidden states in terms of probabilistic transitions that depend on policies. To accommodate the fact that subjects may have preferred heuristic strategies for visual search, the model of their policies included a fixed-form (i.e., heuristic) policy that was estimated on a subject by subject basis. This fixed-form policy corresponds, technically, to a state-action policy. In other words, it is a policy that prescribes (in a probabilistic way) the next target location given the current location. This can be encoded as a single policy in terms of a probability transition matrix among different saccade locations. The heuristic policy was estimated using the empirical transition frequencies for every sequence of saccadic eye movements analysed – and included in the repertoire of policies for each subject. Although this policy is subject-specific, it plays exactly the same role in every instance; namely, a state-action policy that has no uncertainty reducing or epistemic aspects. This can be seen easily because the most probable next action or saccade does not depend upon posterior beliefs that would otherwise contextualise an epistemic saccade.

Operationally, I parameterised the propensity of a subject to engage in a fixed-form (heuristic) policy with a single log probability (an $\mathbf{E}_{\mathbf{heurisitc}}$ coefficient). This value was specified relative to a value of zero for the remaining (eight) policies that could be deployed in an epistemic fashion. Including the heuristic policy allowed the evaluation of how likely different subjects were to engage in epistemic versus non-epistemic

searches – and whether this propensity changed with exposure to the task. Interestingly, I found that the fixed-form policies in several subjects resembled a reading-like strategy, while in others there was a tendency to proceed clockwise around the quadrants. See the rightmost side of Fig 4.1B for an exemplar empirical probability transition matrix encoding these subject-specific policies (in this case the reading-like strategy). To include the propensity of a subject to engage in a fixed-form (heuristic) policy in the MDP scheme, the priors are added to the variational free energy scoring the evidence for each policy based upon past outcomes (**F**) and expected free energy in the future (**G**) weighted by their inverse precision (**β**). This leads to a posterior belief over policies:

$$\boldsymbol{\pi} = \sigma\left(\mathbf{E} - \mathbf{F} - \gamma \cdot \mathbf{G}\right) \tag{11}$$

where $\mathbf{E} = \ln \mathrm{E}$ and E corresponds to the prior preferences over the policies. See (Friston et al. 2017) for the details. All the remaining update equations remained the same as in the previous chapter.

The generative model used to generate stimuli in our experimental paradigm (see the next section) is illustrated in Fig 4.1. The likelihood matrices mapping from hidden states to outcomes are shown in the upper panel. For illustrative purposes, the likelihood matrices are provided for the sampling of the second location (the top left quadrant) under a vertical transformation of the scene. The middle panel shows the action-dependent transition matrices that generate transitions among hidden states following each action. Crucially, the first eight action-dependent transition matrices – that encode the transition probabilities between sampled locations – map deterministically onto the same location as the action; whereas the transition matrix for the ninth action corresponds to a fixed-form (heuristic) state-action policy. This prescribes the next location given the current location in a deterministic way. The transition matrices for the other hidden state dimensions; namely, the *what* or scene context and spatial transformations are identity matrices (because these do not change within each trial). The rules of the game, in terms of scoring points, have been modelled in terms of the prior preferences over outcomes in the **C** matrices. These rules are explained in detail in the next section. As noted above, prior preferences over the policies correspond to **E**. Finally, the generative model assumes uniform beliefs

about the hidden states of the world, apart from the initial sampling location; namely, central fixation.

4.1.2. Characterising empirical behaviour in terms of active inference

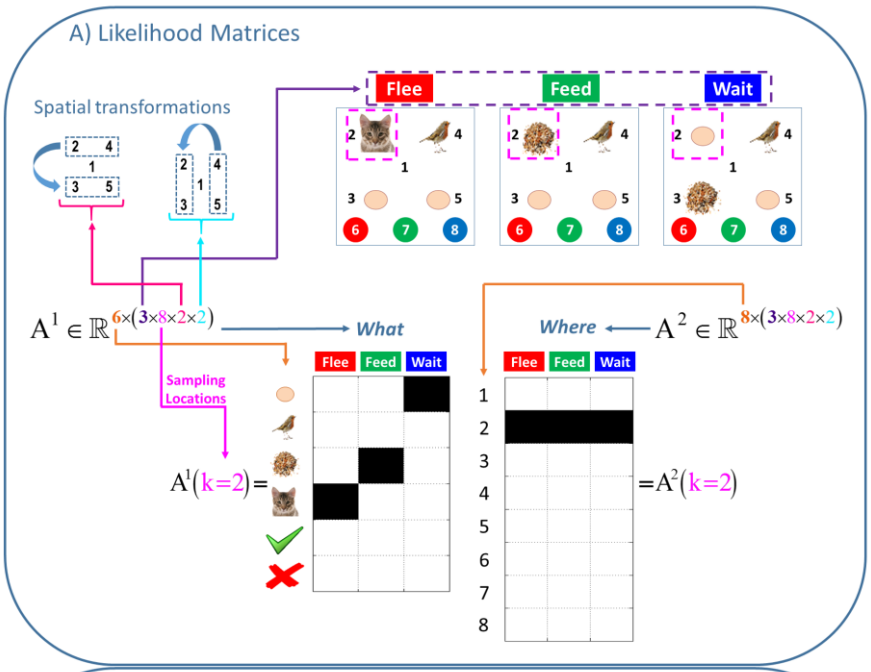
I summarised the hundred trials within each of five blocks in terms of scan-paths (sequences of saccadic locations). Using the stimuli that were disclosed during each epoch of every trial, I was then able to optimise the parameters of an MDP scheme that best explained each subject's behaviour. The (free) parameters of the MDP scheme included a hyperprior on the (inverse) precision of beliefs about policies (**B**eta), prior preferences for outcomes (**C**ost) and a prior expectation (**E**xpectations) or bias towards non-epistemic (heuristic) policies. These (**B**, **C**, **E**) parameters encode prior beliefs (about behaviour, preferences and prior policies respectively). This enabled me to optimise the model of each subject's responses, while accommodating subject-specific preferences. As a prelude to analysis of empirical data, I ensured that fitting the MDP model to observed behaviour has face validity. To do this, I estimated the model parameters using the saccadic choices of the first subject (on the third testing block) and used them to simulate a block of 100 trials. Using the simulated data, I then estimated the MDP parameters to ensure that I could recover the same values used to generate the data. The results of an exemplar analysis are shown in Fig 4.2. One can see that the scheme was able to recover the parameters used to generate the data, with a reasonable degree of confidence (the pink bars correspond to 90% Bayesian confidence intervals). This inversion scheme was applied to the empirical data to address the hypotheses about whether subjects evidence epistemic behaviour and whether this behaviour increases with exposure to the paradigm described here. The subsequent analysis of the empirical behavioural data comprised three components.

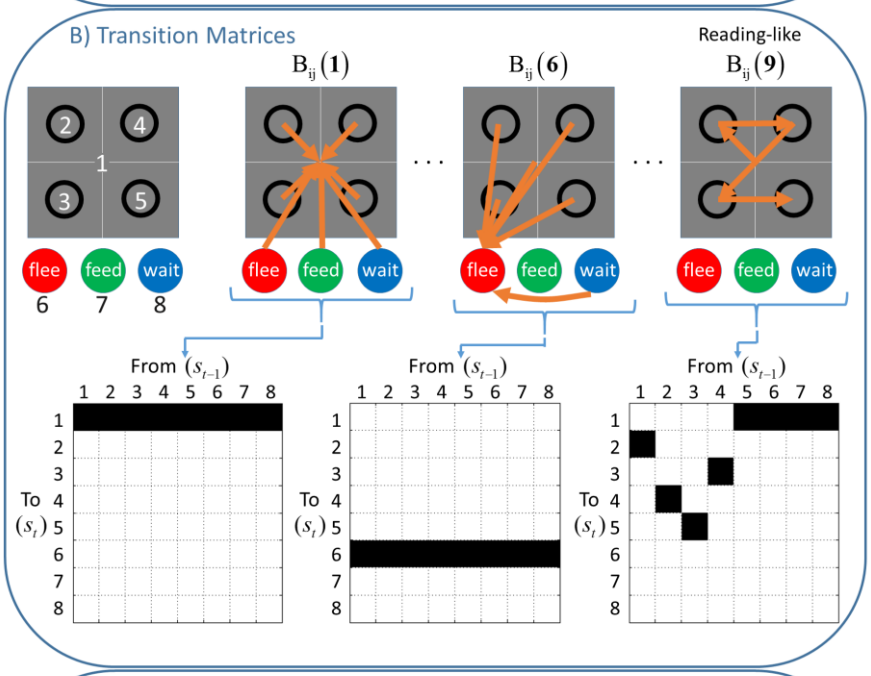
- First, I assessed the evidence that subjects engaged in epistemic searches – as described by minimising expected surprise (i.e. free energy), under ideal Bayesian assumptions. I therefore compared models of each subject's responses (during the last blocks of each session), under models that did – and did not – contain a salience or uncertainty-reducing term (i.e., epistemic or intrinsic value). Removing this epistemic value from expected free energy reduces it to an expected utility, scored in terms of prior preferences or cost (K. Friston et al., 2015). The evidence

for epistemic imperatives in visual searches was assessed for each subject individually in terms of the difference in log evidence for the two models – and then pooled (i.e. summed) to provide inference based on all the subjects' data.

- Second, I asked which parameter combinations account for the exploratory behaviour the best by evaluating the model evidence obtained under each model by using Bayesian model reduction and averaging, where each parameter combination corresponds to a model. This was assessed for each subject individually and then the model evidence under all models was pooled together over all subjects to produce an overall Bayes factor to find which model best accounts for subjects' behaviour overall.

- Finally, to characterise intersubject variability, I used canonical correlation analysis (CVA) to see whether there were significant relationships between the behaviour of our subjects and their prior beliefs, as estimated in terms of the parameters of the MDP model. This involved creating a matrix of independent or explanatory variables corresponding to the free parameters for each subject and trying to explain the corresponding dependent or response variables based upon subjects' performance. In this instance, I summarised behaviour in terms of their accumulated score over all trials and performance improvement from the first to the last block. These behavioural measures were supplemented with (partially redundant) performance measures; reflecting the percentage correct categorisations and the number of saccades emitted on average over trials. This analysis returned significant pairs of canonical vectors and variates describing how prior parameters or beliefs are manifest behaviourally. Note that these performance scores are distinct from the scan-path data used to estimate the prior beliefs of each subject.


A) Likelihood Matrices
Spatial transformations
Flee
Feed
Wait
$A^1 \in \mathbb{R}^{6\times(3\times8\times2\times2)}$
What
Where
$A^2 \in \mathbb{R}^{8\times(3\times8\times2\times2)}$
Sampling Locations
$A^1(k=2)=$
$=A^2(k=2)$


B) Transition Matrices
Reading-like
$B_{ij}(1)$
$B_{ij}(6)$
$B_{ij}(9)$
flee
feed
wait
From (s_{t-1})
To (s_t)

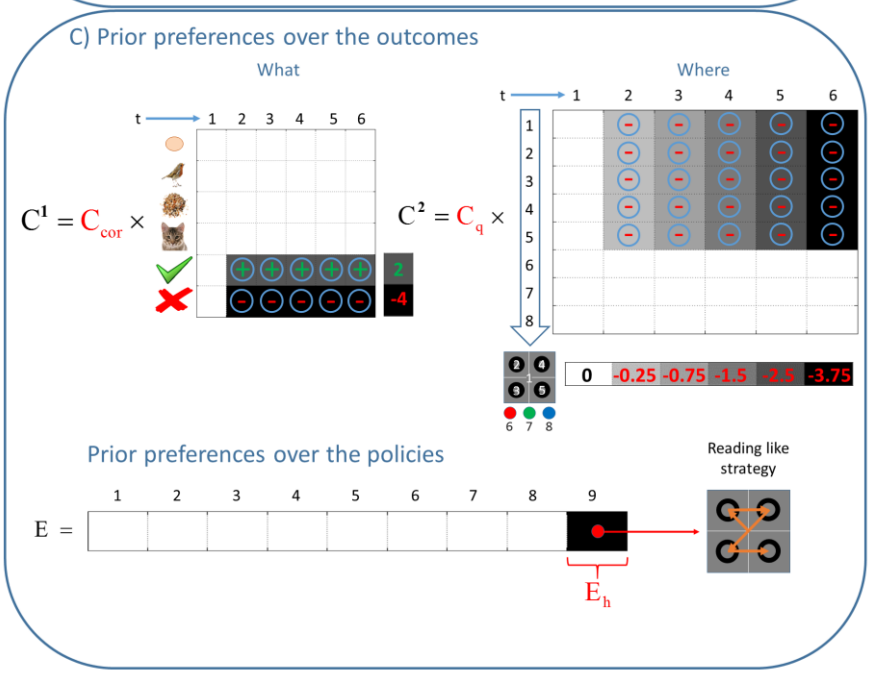

C) Prior preferences over the outcomes
What
Where
$C^1 = C_{cor} \times$
$C^2 = C_q \times$
t
0
-0.25
-0.75
-1.5
-2.5
-3.75
Prior preferences over the policies
Reading like strategy
E =
E_h

Figure 4. 1 ABCDE of generative model

A) This panel shows the likelihood matrices for the top left quadrant (location 2), under the vertical (but not horizontal) spatial transformation. The likelihood matrices encoded the probability of outcomes given (the four dimensions of) hidden states. **B)** This panel shows the action-dependent state transition matrices. In this scene construction task, only the sampling location (*where*) hidden state is action-dependent. The other transition matrices associated with the context and spatial transformations are identity matrices. The action-dependent transition matrix maps to the sampling locations associated with each action (for the first eight policies). However, this mapping changes in the case of a fixed-form policy, which corresponds to a repeating the ninth action in the MDP model. The mapping between sampling locations for the reading-like policy is shown on the rightmost side of the middle panel. **C)** This panel shows two priors. Firstly, the prior preferences over the *what* and *where* outcome modalities in the first six saccades are shown. Here, the columns of the matrices show the preferences (or utilities) over successive time steps; whereas the rows designate the outcomes (six cues under the *what* modality and eight locations under the *where* modality). The utilities in the first time step are zero (shown with the white colour), under both modalities; since the sampled location in the first time step is the central fixation. Different shades of grey indicate the absolute value (intensity) of the utilities, where the darker shades are associated with higher utilities. The prior preference matrix under the *what* modality equips the agent with beliefs that it expects to categorize correctly. The increasing utility over the columns in the prior preference for the *where* modality means that the tardy sampling (i.e., being undecided) becomes costly. Plus and minus signs indicate the valence of the utilities. The prior preferences over policies are shown on the right. One can define a propensity for a policy in the vector **E**, which encodes the prior preferences over the policies. An example **E** is shown at the rightmost side. The utility of the ninth policy (heuristic strategy, $\mathbf{E}_{\mathbf{heuristic}}$) is defined as $\log(2)$, relative to a value of zero for the remaining (eight) policies. The first eight policies correspond to the policies that take the agent to the locations associated with the central fixation, four quadrants and three choice locations. This renders the ninth policy ≈ 3 times more likely.

The number of significant canonical correlations defines the dimensionality of a phenotypic space in which different subjects lie. In other words, it provides a way of characterising the 'type' of each subject along different dimensions. For example, one type of subject may have very precise (hyperprior) beliefs about policy selection and therefore be relatively confident in how they prosecute the visual search. Furthermore, these subjects may adopt a fixed-form (heuristic) search strategy and consequently take a longer time to resolve uncertainty – but will, on average, be more accurate in their decisions. Another type of subject may be more epistemic in nature; reducing their uncertainty about the scene category more efficiently; thereby using shorter scan-path. By simulating responses for characteristic parameter values within the canonical

correlation space, one can illustrate the impact of different prior beliefs on behaviour and underlying confidence in decisions and uncertainty about the scene category. In this chapter, I focus on intersubject variability in a healthy human population. My hope is to show that there are systematic differences in prior beliefs and salience attribution (i.e., the ability to identify salient or epistemically valuable saccadic targets).

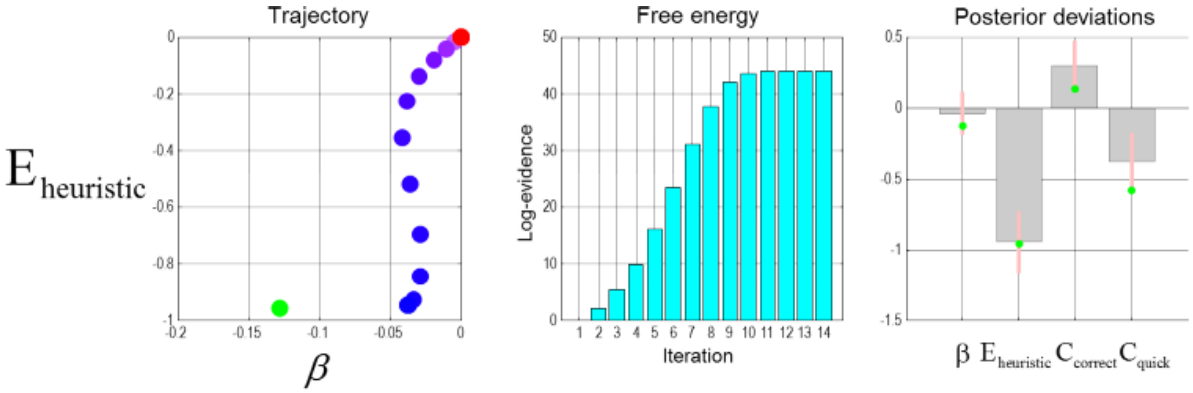


Figure 4. 2 Parameter estimation, simulation and re-estimation

This simulation shows that the estimates of the model parameters, given the saccadic scan-paths, can be recovered accurately. The estimated parameters are prior inverse precision, heuristic bias and prior preferences – **β**, $\mathbf{E_{heurisitc}}$, $\mathbf{C_{correct}}$ and $\mathbf{C_{quick}}$. Here, a sequence of saccadic choices was simulated using the estimated **β** and $\mathbf{E_{heuristic}}$ parameters from fitting a model to the saccadic choices of the first subject on the third testing block (shown with green dots). The left panel shows how the estimated **β** and $\mathbf{E_{heuristic}}$ parameters change with each iteration, when a model is fit to the simulated data (from red to blue). The middle panel shows how the free energy changes with each iteration during the model inversion. The right panel shows the means (grey bars) and the variances (pink bars) of the probability distributions over the estimated parameters.

4.2. Empirical Methods

The experimental design allowed participants to explore the scene by disclosing objects placed at each quadrant of the visual field using eye movements. Upon arriving at a decision, the participants reported their categorisation of the scene using a button box that was placed either to the right or to the left of the head-mount, depending on whether the individuals were right or left-handed.

Each subject underwent a pre-training phase comprising twenty trials; ensuring that they were accustomed with the experimental setup, head-mount, controller *etc*. They then performed five blocks of the task: two training blocks and then three test blocks.

Between each block the subjects rested for a few minutes. Each block consisted of a hundred trials. A fixation cross was displayed on the screen prior to the beginning of a trial. Looking at the fixation cross triggered the trial. After the beginning of the trial, the visual stimuli were displayed in a two-by-two grid; in which each square consisted of a grey dot within a black circle (see Fig 4.3). The visual display was gaze-contingent; in other words, the grey dots turned into objects (*null*, *bird*, *seed* and *cat*) when looked at. This allowed the subjects to accumulate evidence as they explored the scene within a given trial. At the beginning of a block, the subjects were given 100 points. Subjects were rewarded two points for making a correct categorization and penalized four points for an incorrect one. Both correct and incorrect categorizations were followed by auditory and visual feedback.

I incentivised the participants to sample locations that were more informative using a sampling cost. The penalty of attending to the *n-th* square was given by $-0.25 \times n$. The cost of exploration stacked cumulatively as the exploration proceeded; i.e., looking at two squares would cost $-0.25 + (-0.5) = -0.75$. These task instructions instantiate a particular task set or prior belief that was included in the model or prior preferences (i.e., the probability of not making a decision became less likely with the number of saccades). This introduces a distinction between $\mathbf{C}_{\mathbf{correct}}$ and $\mathbf{C}_{\mathbf{quick}}$ that encode preferences about being *right* or *wrong* and preferences about being undecided as time progresses. See Fig 4.1C.

There was a fixed time limit for each subject of between two and four seconds on each trial. Exceeding this time limit (without choosing) cost the subjects four points. Time limits for each subject were obtained using a staircase procedure during both the training and testing blocks. This staircase procedure was a function of the minimum number of saccades necessary for an efficient categorisation. For instance, if the first object was *bird*, then the most efficient way to explore the scene is to look at the square next to the bird. One only needs to make two saccades to categorize the scene in this case. The number of saccades was summed over 10 trials. Making 10% more saccades than was necessary increased the time limit by 200 ms: it decreased by 200 ms otherwise.

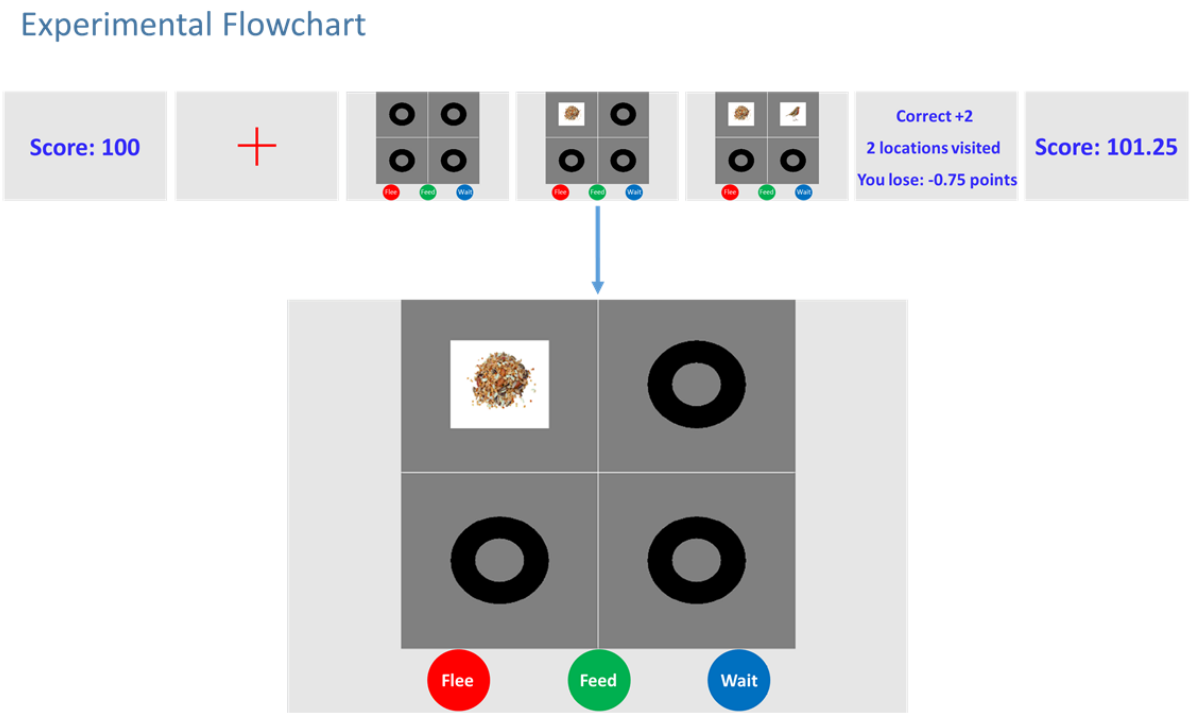


Figure 4. 3 Experiment flowchart

This flowchart shows how a single trial evolves. Firstly, the total score is displayed. Then a fixation cross appears in the centre of the screen. Upon looking at the fixation cross the trial begins. In this particular trial, the participant looks at the top left quadrant and observes a *seed*, and then looks at the top right quadrant and observes a *bird*. At this point it is obvious that the category of the scene is *Feed*. On the fifth action, the participant chooses the *Feed* category by pressing the green button on the controller. This is followed by an auditory feedback, associated with the correct decision. Consequently a feedback screen shows whether the chosen category was correct, the number of quadrants one looked at and the points lost by looking at those quadrants. Finally, the total score is displayed.

The training and testing blocks differed in two ways. In the training phase, the colours of the buttons on the controller were displayed as dots in the lower half of the screen (below the two-by-two grid scene) with the same colours (and in the same order) as the button press box, to ensure subjects learned to press the correct buttons as quickly as possible. In the testing phase, the coloured dots were removed from the screen and the grid scene was centred on the screen. The sequence of frames in Fig 4.3 shows the steps of this gaze-contingent paradigm. Stimuli were delivered using Cogent 2000 *(developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience)* and Psychtoolbox (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli, 1997).

Subjects: In total 22 subjects were recruited (9 males, 13 females) through the Institute of Cognitive Neuroscience subject database. All subjects gave informed written consent, and the study received ethical approval from the UCL ethics committee (4356/002). The majority of the subjects were students of University College London. The age of the participants ranged between 19 and 57, with mean 25.7 years and standard deviation 9.3 years.

Recording method: Subjects were seated 70 cm from the screen on which visual stimuli were displayed – and they rested their chins on a head-mount. Using the Eyelink 1000 eye-tracker their gaze coordinates were recorded as they performed the task. The grid scene was displayed on a $408mm \times 306mm$ screen with a resolution of 1600x1200. The angle of sight was $\approx 32.5°$ visual angles horizontally and $\approx 24.4°$ vertically. The angle of sight of the two-by-two grid scene was $\approx 20.3°$ during the training phase and $\approx 24.4°$ during the testing phase, both horizontally and vertically. The size of each object in each square was $\approx 5°$ and the centre of each object was $\approx 8.5°$ from the centre in terms of visual angles.

4.3. Results

4.3.1. Behavioural results

I first characterised performance across training and testing blocks in terms of their mean score per trial, percentage correct categorizations, mean time interval between saccades and mean number of saccades per trial. These performance measures were averaged over all subjects and are shown across the five blocks in Fig 4.4. The panels in this figure show that the score per trial increases over blocks and the percentage correct on the fifth block is greater compared to the first block. The middle panels show that both mean saccades per trial and mean time between saccades follow a decreasing trend across blocks. Separate two sample t-tests for these behavioural measures (between the fifth and the first blocks) show that the performance measures in the first and fifth blocks are significantly different. Finally the histogram in the lower panel shows the probability distribution over the saccades.

Given that there are four locations in the task and given the subjects do not revisit locations, there can be $4! = 24$ different ways of exploring the scene. Diverse patterns

of exploratory behaviour were observed. Some patterns can be described as heuristics, in that they were used repeatedly within subjects, independently of the context. Other subjects attended to different locations in a seemingly random fashion and some explored in a way to reduce uncertainty about the scene efficiently. Prominent among the heuristics were reading-like and clockwise strategies.

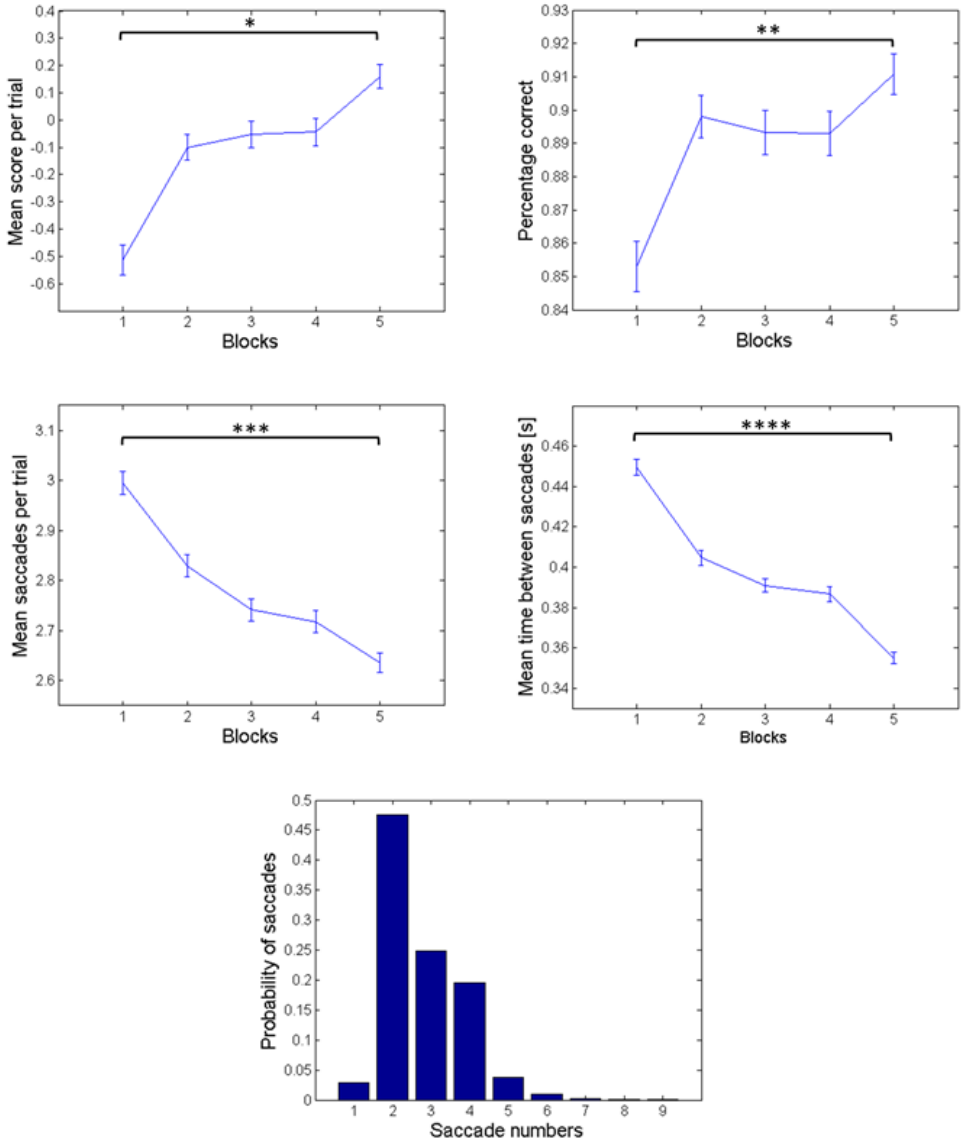


Figure 4. 4 Performance measures

The performance measures in the panels of this figure are averaged over all blocks and subjects. The means and standard error of the means (error bars) are plotted per measure. The top left and right panels show the mean score per trial and the percentage correct categorization over five blocks, respectively ($t(df) = 4388$, $^{*}p < 0.001$; $t(df) = 4393$, $^{**}p < 0.001$). The middle left panel shows

the mean saccades per trial (before categorizing the scene) and the middle right panel shows the mean time between saccades in seconds across five blocks ($t(df) = 4392$, $^{***}p < 0.001$; $t(df) = 7774$, $^{****}p < 0.001$). The histogram on the bottom panel reports the probability distribution over the number of saccades the participants made before arriving at a decision about the category of the scene.

There are commonalities in the heuristic strategies: e.g. the first two quadrants under the reading and clockwise policies (see the two rightmost scan-paths in Fig 4.5) are the same. Fig 4.5 shows the proportion of all trials in which the individual subject's most commonly used heuristic strategy was employed. There were six distinct heuristic strategies used by 22 subjects. By far the commonest policies are reading-like and clockwise strategies (used 47% and 42% of the time respectively), whereas the next most commonly used heuristic was at 12% (shorter scan-paths can be explained by multiple heuristic policies; hence these percentages do not add up to 100).

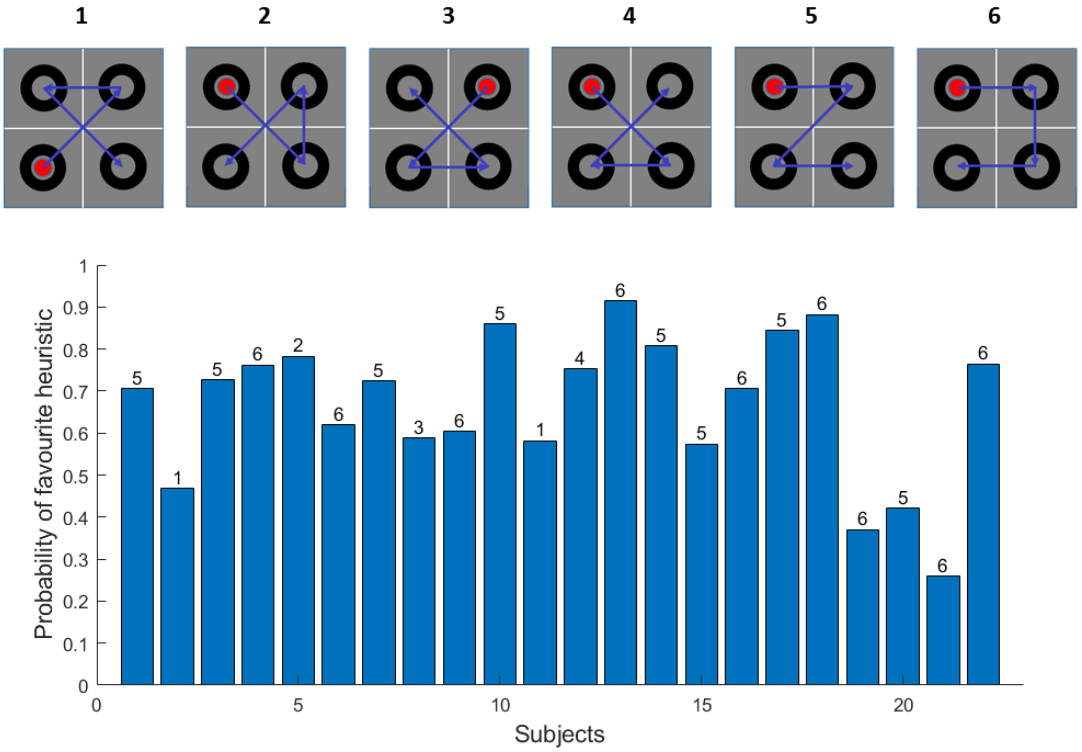


Figure 4. 5 Probability of subjects' favourite heuristics

The bar plot shows the probability of each participant's favourite heuristic in terms of the frequency with which the scan-path on a given trial accords with the scan-path of a heuristic strategy. This was repeated over all trials and blocks in each subject. The participants used six distinct heuristic strategies

and the scan-paths of these strategies are plotted in the upper panel and linked with subjects that used these strategies (see the number above each bar for each subject's favourite heuristic).

An epistemic policy can be defined as a policy (based on the current beliefs about the hidden states) that causes the greatest change in beliefs (aka information gain, Bayesian surprise or salience) about the hidden states. In contrast to the heuristic policies, the sequence of actions is not predetermined in the epistemic policies, and they can therefore accumulate information more efficiently, because they are belief based.

4.3.2. Scan-path results

I first tested whether subjects' scan paths evidenced the use of epistemic policies. The upper panel of Fig 4.6 shows the difference between the log evidences obtained with the models that did – and did not – contain epistemic value. The model that incorporates epistemic value had substantially more evidence for every subject. Pooling the log evidence over all subjects the epistemic model scored ≈ 888 more log evidence than the model that contained extrinsic value (i.e., prior preferences) only. This result suggests that the subjects indeed engaged in epistemic visual foraging – and that the epistemic affordance or salience of visual targets was necessary to explain their eye movements.

I then tested whether subjects showed evidence for changes in their prior beliefs from block to block. Fig 4.6 shows the results of Bayesian model comparison of these between-block or experience-dependent effects. This analysis used parametric empirical Bayes (Friston et al. 2007) to test for systematic (monoexponential) changes over blocks prior beliefs; namely, prior inverse precision, heuristic bias and prior preferences – $\boldsymbol{\beta}$, $\mathbf{E}_{\mathbf{heuristic}}$, $\mathbf{C}_{\mathbf{correct}}$ and $\mathbf{C}_{\mathbf{quick}}$. This model of exposure-dependent changes assumes that the greatest change in prior beliefs occurs at the start (in the first blocks) and then plateau in the last blocks. $\boldsymbol{\beta}$ (shown in red in Fig 3.1A) determines the confidence subjects place in their prior beliefs about policies. $\mathbf{C}_{\mathbf{correct}}$ and $\mathbf{C}_{\mathbf{quick}}$ (shown in red Fig 4.1C) are scaling coefficients on the log prior preferences about outcomes (in *what* and *where* modalities respectively) that tune the precision of preferences. The higher these parameters, more precise the preferences become. $\mathbf{E}_{\mathbf{heuristic}}$ (shown in red in Fig 4.1C) is the final element in the vector of prior preferences

over the policies. This encodes the prior propensity for a fixed-form policy (e.g. reading like strategy) specified relative to a value of zero for the remaining (eight) policies.

The changes in these prior beliefs were modelled by specifying a simple general linear model at the between block level that comprised a constant term and a monoexponential decay with a time constant of one block. The between-block parameters of this hierarchical model comprised a constant and decay parameter for each of the four (prior) parameters at the within-block level. Bayesian model reduction was then used to compare all combinations of the ensuing 4 x 2 = 8 between-block effects, for each subject. There are in total $2^8 = 256$ models.

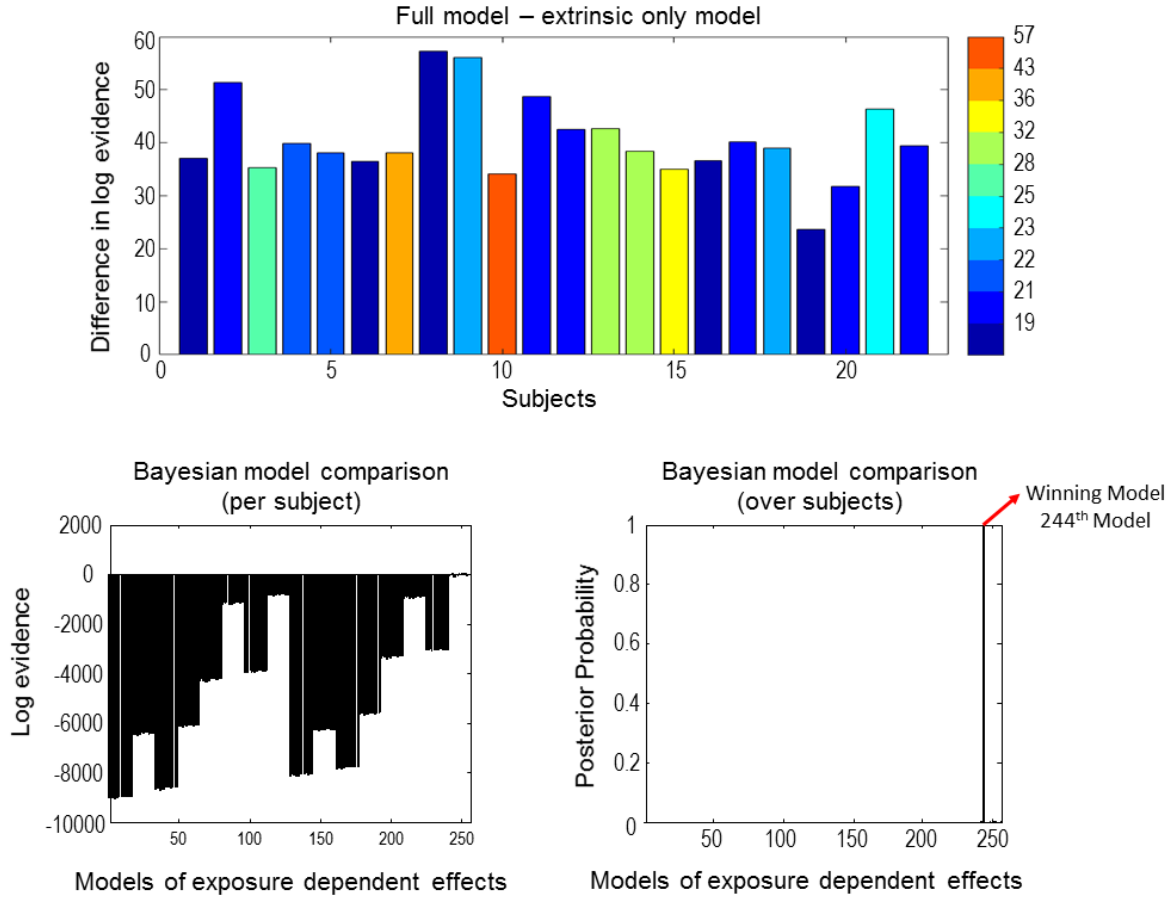


Figure 4. 6 Bayesian model comparison and reduction

The top panel shows the difference between free energies obtained with the full model (epistemic and extrinsic values) and the model with only extrinsic value. The model that contains the epistemic value (full model) has more evidence than the model that does not on an individual basis and over all subjects. The colours of the bars indicate the ages of the subjects. Monoexponential changes in the parameters over blocks where tested using parametric empirical Bayes, implemented with a general linear model that consists of a constant and a decay term. This means that there are $4 \times 2 = 8$ between-block effects.

The lower left panel shows the log evidences for all combinations of between-block effects ($2^8 = 256$). Here, the model that excludes the change in **β** and $\mathbf{E}_{\text{heuristic}}$ (244^{th} model) scores the greatest log evidence. The lower right panel shows the most likely model when a softmax function is applied to the log evidences on the top left panel.

The results of this analysis are shown in terms of log evidence (pooled or summed over subjects) over the (most likely) 256 models in the lower row, left panel of Fig 4.6. These results show that full models (to the right of the bar plot) have much greater evidence than reduced models, with fewer parameters. To assess the most likely model over subjects, I applied a softmax function to the pooled log evidence. The resulting marginal likelihood or model evidence over models and subjects is shown in the lower row, right panel of Fig 4.6. This model likelihood suggests that one can be almost certain a nearly complete model is necessary to account for the data. The winning model identified the *changes in* prior precision (***β***) and heuristic bias ($\mathbf{E}_{\text{heuristic}}$) over blocks as redundant. This was a little surprising because it suggests that systematic changes in subjects' preferences – with increasing experience of the paradigm – are expressed largely in terms of their prior preferences (**C**) for being correct or for responding quickly.

The upper panel of Fig 4.7 shows the Bayesian model averages of the (four) parameters for each subject. These parameter averages account for uncertainty about the model of between-block effects. The bar plot groups the Bayesian model averages for each parameter over subjects. The first four parameters correspond to the mean or constant effect, while the last four correspond to experience-dependent changes. One can see that there is a remarkably consistent profile of deviations from the prior mean over subjects (first four parameters). However, the experience-dependent effects are less consistent. As would be expected from the Bayesian model comparison, the changes in prior precision and heuristic bias are small; with the Bayesian model averages of heuristic bias ($\mathbf{E}_{\text{heuristic}}$) over subjects shrinking to almost zero. The interesting results here are the more consistent and negative parameters controlling the monoexponential decay of prior preferences (**C**). As the subjects become more familiar or experienced with the paradigm they increase the precision of their prior preferences; especially the prior belief that they will respond more quickly.

These effects are shown in terms of the expected changes in prior preferences over blocks based on subject specific estimates (dotted lines) and the group mean (solid red lines) for prior preferences about being correct (lower left panel) and being quick (lower right panel) in the lower row of Fig 4.7. These results suggest that as blocks progress, subjects increase their prior beliefs that they will avoid sampling further (unnecessary) information later in the trials, which can be seen by the increase in $\mathbf{C}_{\mathbf{quick}}$ over blocks. A key aspect of these subject specific effects is that there is a large intersubject variability that I characterised using canonical correlation analysis.

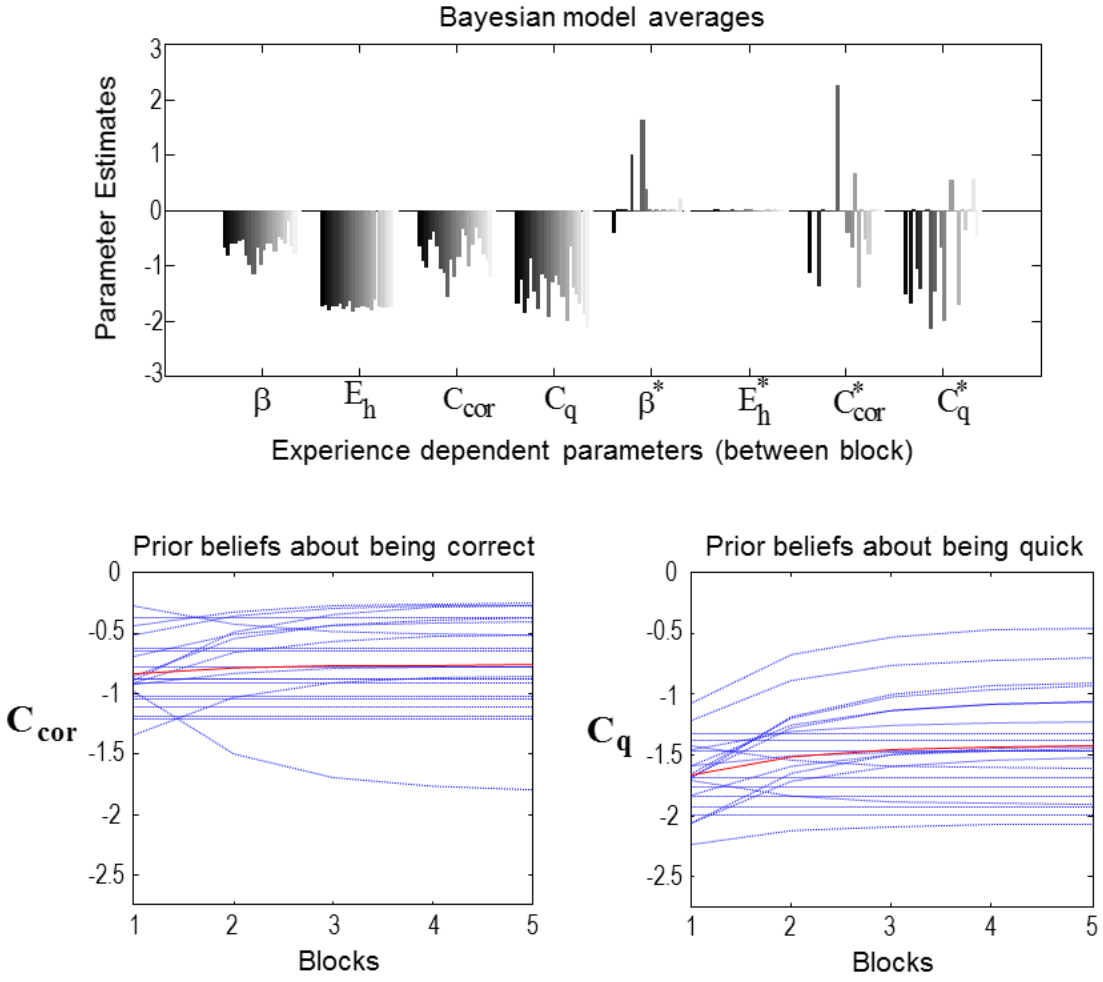


Figure 4. 7 Bayesian model averaging

The upper panel shows the Bayesian model averages for each parameter over subjects after Bayesian model reduction was applied to all 256 models (i.e., redundant parameters were eliminated). The left and right lower panels show the posterior estimates of the scaling coefficients that control the precision of the prior preference matrices about being correct (left) and quick (right).

4.3.3. Canonical correlation (variates) analysis of between subject effects

Fig 4.8 illustrates between-subject effects; specifically, the canonical correlations between mixtures of behavioural scores and mixtures of subject specific prior beliefs and experience-dependent changes in those beliefs. This analysis summarised behaviour using four behavioural scores for each subject (normalised to a mean of zero and a sum of squares of one). These scores were as follows:

- Trial performance (mean score per trial).
- Percentage correct (how accurate they were at categorising scenes)
- Mean saccades per trial (number of made saccades before categorizing a scene)
- Mean time between saccades (time period between sampling two consecutive locations in seconds)

These behavioural measures were correlated with the six (normalised) subject-specific Bayesian model averages of the prior beliefs (excluding the experience dependent changes in $\boldsymbol{\beta}$ and $\mathbf{E}_{\text{heuristic}}$) as explained in Fig 4.6 and Fig 4.7. These estimates correspond to a computational phenotype of each subject.

Canonical correlation analysis (equivalently, canonical variates analysis) established that there were three very significant canonical correlations. In other words, there were three pairs of orthogonal mixtures that could not be accounted for by chance. The significance of these canonical correlations is shown in the upper left panel of Fig 4.8, in terms of the log probabilities of the four canonical correlations. There are four, because this is the minimum dimensionality of the multivariate variables (i.e., the behavioural measures). It can be seen that the first three canonical correlations are extremely significant. This is reflected in the tight correlations between the predicted and observed behavioural factors (shown on the upper right). The amount of behavioural variance that could be accounted for – in terms of the computational modelling – was incredibly high: 96% for the first canonical pairs of vectors, 92% for the second and 70% for the third. (There is minimal contribution of the score per trial to any of the canonical correlations because this score is explained entirely by two other factors – being correct or not and the number of saccades used.)

The canonical vectors themselves are shown in the lower panels. These correspond to the weights of linear mixtures of the behavioural and computational scores that show the greatest correlation. I have limited these to the three significant correlations (black, grey and white bars). The canonical vectors for the behavioural scores define three behavioural phenotypes (noting that the signs of the canonical vectors are arbitrary):

- The first canonical correlation is driven largely by a correlation between the second behavioural score (percentage correct) and a prior bias towards the heuristic policy. If the signs of the first (black) canonical vectors are flipped, these results suggest that more *accurate* and *quick* subjects are those subjects who, computationally, have a lower prior bias towards heuristic use ($\mathbf{E}_{\mathbf{heuristic}}$). In other words, subjects who rely more on epistemic policies tend to categorise scenes more accurately and use fewer saccades.

- The second canonical correlation is dominated by the prior bias towards the heuristic policy that is expressed largely in the percentage correct and, to a certain extent, the mean saccades per trial and time between saccades. This behavioural phenotype is of a *careful* subject who is accurate and takes her time between eye movements. This sort of subject has a large negative heuristic bias ($\mathbf{E}_{\mathbf{heuristic}}$); in other words, a careful subject will not appeal to the heuristic search strategies and will prioritise being correct ($\mathbf{C}_{\mathbf{correct}}$) over using fewer saccades ($\mathbf{C}_{\mathbf{quick}}$).

- The third canonical correlation involves the most pronounced experience-dependent changes in beliefs during exposure to the paradigm. This appears to be expressed behaviourally in the percentage correct and time between saccades. This behavioural phenotype is dominated by a negative loading on mean time between saccades and can be regarded as a *hasty* subject, who trades-off between prior beliefs about being correct and being quick in the opposite direction to the careful subject. Crucially, these hasty subjects are the only sort of subjects that change their prior preferences from block to block.

In summary, there is clear evidence that both subjects' beliefs and their ability to change those prior beliefs with experience have predictive validity in relation to behavioural performance; enabling the prediction of most of the behavioural variance between subjects, given their computational phenotyping under the active inference (MDP) scheme.

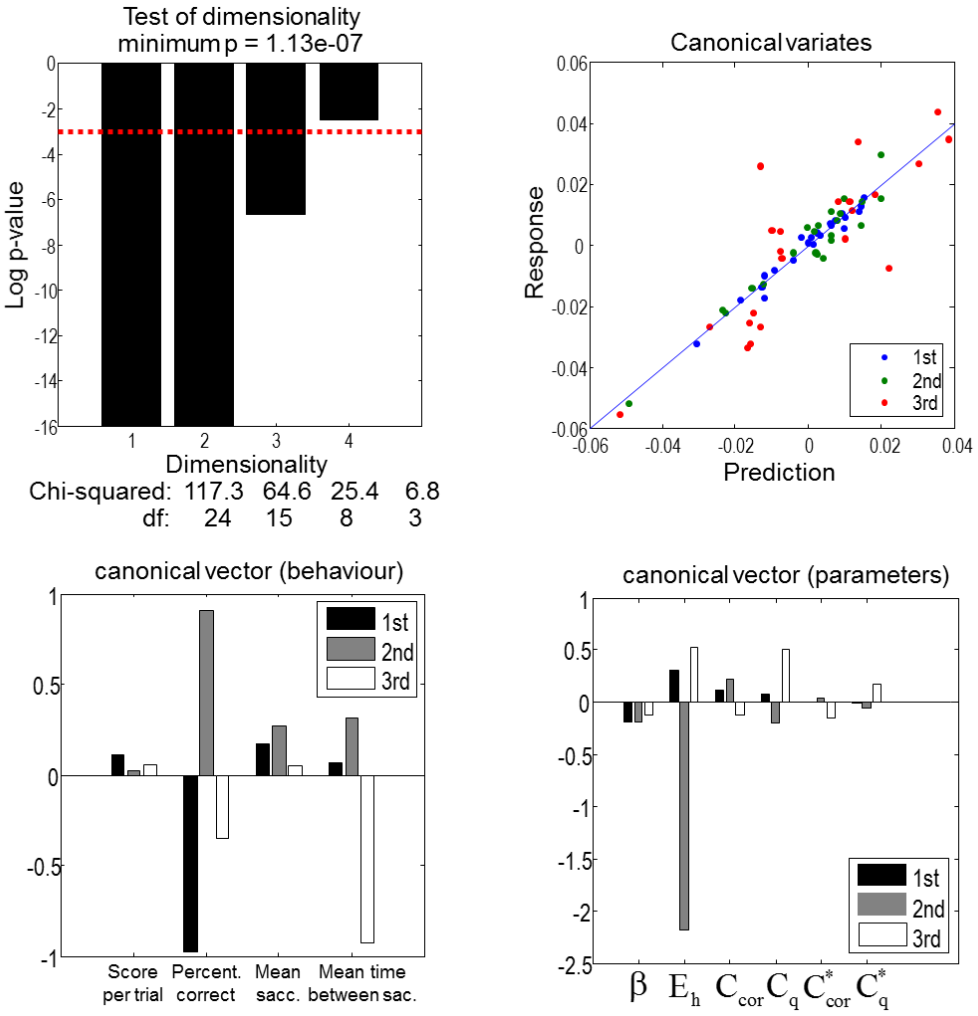


Figure 4. 8 Results from Canonical correlation analysis

The correlations between the Bayesian model average of parameters (among 256 models) and the behavioural measures were analysed using CVA. These parameters are means of the prior inverse precision **β**, heuristic bias $\mathbf{E_{heuristic}}$, scaling coefficients $\mathbf{C_{correct}}$ and $\mathbf{C_{quick}}$, and the decay terms $\mathbf{C^*_{correct}}$ and $\mathbf{C^*_{quick}}$. The behavioural measures are: mean score per trial, percentage correct, mean number of saccades per trial and the mean time interval between saccades. The top left panel shows the results of a chi-squared analysis of the canonical correlations. The first three of four canonical correlations are statistically significant; whereas the fourth canonical correlation is not. The top right panel shows the predicted and observed behavioural canonical variates. The bottom left and bottom right panels show

the corresponding canonical vectors. The three canonical variates are each composed of a pair of canonical vectors whose scores on behavioural measure and parameter are illustrated on the left and right bar charts respectively.

4.4. Discussion

In this chapter, I have shown that healthy subjects' visual exploration – of even simple scenes – provides substantial evidence for the use of epistemic affordance or salience in visual exploration (upper panel of Fig 4.6) by fitting models that did and did not contain epistemic, uncertainty-resolving imperatives for policy selection to saccadic behaviour.

Strikingly, a bias towards using heuristic policies to explore visual scenes was associated with lower accuracy (i.e. percentage of correctly categorised trials) in all three canonical variates relating model parameters to behaviour. Note that while one might expect heuristic policies would reduce the time between saccades (as in the 2nd and 3rd canonical variates) at the expense of increasing the mean number of saccades (as in the 1st canonical variate), there is no *a priori* reason accuracy should be affected by heuristic use. The association between diminished heuristic use and improved accuracy indicates that having a model of the task structure to direct one's behaviour not only permits epistemic foraging but also improves overall performance.

I have further shown that one can use canonical correlation analysis to quantify behavioural phenotypes and their underlying computational bases – in this case, the tendency to use efficient epistemic search, the tendency to be careful (using epistemic search but also extra saccades), and the tendency to be hasty but also refine one's behaviour over blocks.

Finally, I have demonstrated the use of parametric empirical Bayes to infer changes in parameters within subjects over the course of the experiment. Interestingly, subjects did not change their prior beliefs about the inverse precision parameter $\boldsymbol{\beta}$ and the heuristic bias $\mathbf{E}_{\mathbf{heuristic}}$ consistently. The change in the subjects' prior preferences about being 'correct' and 'quick' best accounts for performance improvements from the first to the last block (lower panel of Fig 4.7). In other words, a simple change in the way that people expected themselves to behave was sufficient to explain changes in behaviour. This does not mean to say that the subjects were more confident about

their policy selection; rather, they were more confident about the consequences or outcomes of their selected policy.

In recent decades several models have been introduced to explain what may drive visual attention. Some of these models map the features of objects (or image patches) such as colour, intensity, orientation (Itti and Koch, 2000; Parkhurst et al., 2002), motion (Rosenholtz, 1999), local contrast and two point intensity correlation (Parkhurts & Neibur, 2003) onto a saliency map. There are crucial differences between these formulations of salience and the one I used in this work: First, my model is not designed to process the visual features of the objects – it deterministically knows where it looks (*where*) and what it sees (*what*). Second, unlike early formulations, my model is endowed with a natural curiosity about the hidden causes of the world that drives its visual search and the prior preferences that encourage it to make accurate and timely choices. In short, salience in active inference is an attribute of a policy that has outcomes – not an attribute of stimuli. This is not to say that stimuli do not have an epistemic affordance but it is the sampling of that affordance that is underwritten by salience.

Another perspective on visual attention suggests that cognitive control processes drive visual search behaviour. Under this hypothesis the context in which the visual search tasks are performed drives exploratory behaviour (Chun and Jiang, 1998; Chun, 2000; Neider and Zelinsky, 2006; Yarbus, 1967). In my paradigm, however, the context is revealed as a *result* of gathering information; in other words, it has to be inferred. Thus a deterministic knowledge of potential contexts, but not the actual context, guides the agent's search behaviour.

It has also been shown that in a set of expected stimuli, the abrupt appearance and disappearance of an object (Brockmole & Henderson, 2005a, 2005b) or the presence of improbable stimuli given a context (Loftus & Mackworth, 1978) can drive visual attention. These results suggest that novelty or information in Shannon's terms attract visual attention. However, Itti & Baldi (2009) showed that areas of high Bayesian surprise (i.e. that cause greater shifts in beliefs) are more potent attractors of human visual attention; i.e., more salient than informative areas in Shannon's terms. One difference between the various approaches above is that in Itti and Baldi's work, Bayesian surprise is computed over low level visual features, rather than hidden states

of the world as in active inference. Nevertheless, the principle of Bayesian surprise underwriting visual salience is likely to hold throughout the cortical hierarchy.

The work in this chapter has some limitations. The small size of the grid scene in the simple visual task used in this work limits the potential benefit of using an epistemic strategy – as it is possible to explore all the quadrants in the time allotted. In a larger grid, the contribution of epistemic strategies to exploration may be even more pronounced. For the reasons of simplicity I did not use the explicit distractors (that are uninformative about the scene category) – I simply used a null or grey background. Under the aberrant salience hypothesis of schizophrenia (Kapur, 2003), one might predict that subjects with schizophrenia may sample stimuli of no epistemic value. A more complex visual task may incorporate sub-goals; i.e., utilities attached to objects and not only to *right* and *wrong* feedback. Such tasks may allow more thorough investigation of the exploration/exploitation trade-off. Finally, the visual search model that I used here does not explicitly model the processing of visual features of the objects. More ecological paradigms might also incorporate Bayesian surprise about lower level visual features; e.g., uncertainty in the identification of objects themselves.

In summary, chapter 3 has shown how (synthetic) subjects can evaluate (expected) Bayesian surprise – i.e. epistemic value – and use it to drive Bayes optimal search behaviour or 'epistemic foraging'. In this chapter, I have demonstrated that even in a very simple task, model comparison indicates strong evidence for epistemic foraging (alongside the use of fixed-form or heuristic policies) in healthy subjects. Furthermore, this epistemic foraging is associated not just with more efficient exploration but also with more accurate scene categorisation. In addition, I have shown how canonical correlation analysis can distinguish different behavioural phenotypes and their underlying computational parameters.

5. Contextual exploration and active inference

We are living in an age characterised by an overwhelming access to information. However our survival does not depend on seeking out any information but seeking out that which is relevant to our survival. The relevance of information depends on one's situation, or context. This means that the correct recognition of a context is essential to seek out the "correct" information.

Context can drive visual attention. In his classic study investigating exploratory eye movements, Yarbus asked his participants to look at the very same painting of a family while changing the participants' instructions (or the rule). These instructions were to either evaluate the material circumstances of the people in the picture or to guess their ages. Under the first instruction the participants paid more attention to the clothing of the people and the furniture whereas under the second they paid the most attention to people's faces. Yarbus concluded that what attracts human visual attention is information that is useful (Yarbus, 1967). This context-dependence in visual exploration is now a well-established phenomenon (Castelhano, Mack, & Henderson, 2009). The scene context can be a background that is consistent or inconsistent with a foreground object (Biederman, Mezzanotte, & Rabinowitz, 1982; Davenport & Potter, 2004; Neider & Zelinsky, 2006), or it can be defined in terms of the spatial layout of the objects (Chun, 2000; Chun & Jiang, 1998; Peterson & Kramer, 2001). Visual search performance has been shown to benefit from this contextual cueing as in each case, some parts of a scene become task relevant and contain more information.

This raises the question: What is information? Shannon proposed that an outcome contains more information if it is less predictable (Shannon, 1948). Itti and Baldi argue that regardless of how unexpected an outcome is, only the observations that causes a significant shift in prior belief distributions yield information *gain*. This notion, known as Bayesian surprise, (Itti & Baldi, 2009) conceptualises a unit of surprise – a "*wow*" – in terms of the KL divergence between the prior and posterior beliefs about the world. This forces us to think in terms of the *mutual* information between an observation, and the unobservable (hidden) states of the world that give rise to it. A new observation is presumed to be more surprising if the posterior distribution (about hidden states) is more dissimilar to the prior distribution. Observations that yield high amounts of

Bayesian surprise attract human visual attention, but also note that “The same data may carry different amounts of surprise for different observers, or even for the same observer taken at different times” (Itti & Baldi, 2009). However it has not been shown how Bayesian surprise can orient attention to different observations under different contexts. Here, I show, using active inference, how contextual exploration can occur – using Bayesian surprise – if beliefs about context influence beliefs about the mutual information between certain kinds of hidden state and sensory data.

In previous work, it has been suggested that perception corresponds to inference about hidden states and attention corresponds to optimisation of the precision of sensory inputs and their causes (Feldman & Friston, 2010; K. Friston, 2009). In this work I consider a generative model that can change the precision of different sensory signals depending on the states of the world. Active inference implies that we weight sensory inputs from different sensory channels in proportion to the precision, given our goals. In the context of rule based or contextual exploration this entails down-weighting the sensory precision of the stimuli (objects) and their causes irrelevant to the context. In this way, epistemic exploration guides the sensory organs to the stimuli that matter in a given context. My objective here is to introduce a computational model that can selectively attend to the task relevant stimuli and acquire useful information under a context.

This paper comprises five sections. In the first, I explain the computational mechanism that may underlie attention. In the second, I introduce a contextual exploration task called colour/shape task. In this task a scene can be categorised in different ways depending on what the rule is. The relevant information that can be acquired by exploring the scene depends on the rule. In the third section I describe the MDP model of this task. In the fourth section I explain how contextual exploration may arise by appealing to mechanisms of attention in an MDP model of active inference. In the final section I show the simulations of this task. Using the same principles described for this task, I also show that this computational mechanism may explain the scan-paths of the people who performed free scene exploration tasks under different instructions. I conclude with a discussion of how the model presented here relates to/diverges from the models of visual search in the literature and how this model could be used in computational psychiatry.

5.1. Attention

Attention has been suggested as the inference about the precision of the causes of sensory inputs (Feldman & Friston, 2010; K. Friston, 2009). In an MDP setup this precision corresponds to the precision of the mapping from hidden states to the observations. The probability of an observation given the hidden states is encoded by the likelihood matrix $A_{ij}^m = P\left(o_\tau^m = i \mid s_\tau = j\right)$, where the superscript m indicates the $m-th$ outcome modality. I now introduce a precision term to modulate the uncertainty of this mapping in the generative model (but not in the generative process).

$$A_{ij}^m = P\left(o_\tau^m = i \mid s_\tau = j, \zeta_j^m\right) = \frac{\left(\bar{A}_{ij}^m\right)^{\zeta_j^m}}{\sum_k \left(\bar{A}_{kj}^m\right)^{\zeta_j^m}} \tag{12}$$

The right-hand side of this equation can be referred to as a Gibb's distribution, where the denominator normalises each entry in the likelihood matrix to a range of $A_{ij} \in [0\ 1]$ and the sum over rows to $\sum_k A_{kj}^m = 1$. Here ζ_j^m is a term that modulates the precision of the likelihood matrix. This plays the role of an inverse temperature parameter. There exists such a precision term for each outcome modality $m \in M$ and each level of hidden states $j \in J$. When $\zeta_j^m \approx 0$ the mapping between the hidden states and observations becomes very ambiguous (low sensory precision), whereas when $\zeta_j^m \to \infty$ this mapping becomes very precise (high sensory precision). This precision term changes the state estimation equations introduced in Fig 2.4 in a way it is shown in Fig 5.1.

Perception (state estimation)

$$s^* = \arg\min_s F(\pi) = \sigma\left(\ln \mathbf{B}_{\tau-1}^\pi \mathbf{s}_{\tau-1}^\pi + \ln \mathbf{B}_\tau^\pi \cdot \mathbf{s}_{\tau+1}^\pi + [\tau \le t] \cdot \zeta \ln \mathbf{A} \cdot o_\tau\right)$$

$$\boldsymbol{\varepsilon}_\tau^\pi = \ln s^* - \ln \mathbf{s}_\tau^\pi$$

$$\mathbf{s}_\tau^\pi = \sigma\left(\mathbf{v}_\tau^\pi\right) : \dot{\mathbf{v}}_\tau^\pi = \boldsymbol{\varepsilon}_\tau^\pi$$

Figure 5. 1 Computational mechanism underlying attention

State prediction errors are used to infer the most likely hidden states of the world in the perception phase of the variational updates. The precision term ζ (shown in red) multiplies the logarithm of the likelihood matrix $\ln A$, in the first equation under perception which shows that when the sensory precision is very low $\zeta = 0$ the observation o_τ does not contribute to state prediction errors in the second equation and do not influence the inference stage of the variational updates.

In order to understand how this precision term works, consider the following example in which a distinction between a generative process and a generative model is made. Assume that you observe a series of coin flip outcomes, not seeing the coin that is being flipped. The process that generates the coin flip outcomes (generative process) is such that one of two coins is flipped at a time. These coins are unfair coins such that coin 1 is responsible for the heads outcomes and coin 2 is responsible for the tails outcomes (see the top panel of Fig 5.2). These two coins are the hidden states that generate the observations heads and tails. Now assume that you know the exact relation between the outcomes and the coins, i.e. the generative process and generative model are identical (see the bottom left panel of Fig 5.2). If you observe another coin flip outcome then you would know which coin is responsible for the observed outcome. In this case the observed outcomes can acquire information about the hidden states, e.g. if a heads is observed, the hidden state responsible for the outcome is coin 1. If the process that generates the coin flip outcomes uses the very same unfair coins but this time you believe that it is equally likely to observe heads and tails under both coins (i.e. the generative process and generative model are different), you can no longer tell which coin is responsible for the outcomes (see the bottom right panel of Fig 5.2). With the beliefs that there is an imprecise mapping between the hidden states and the outcomes the agent cannot resolve uncertainty about the hidden states of the world (i.e. inducing irreducible uncertainty). In the active inference formalism an agent would not pursue the policies that cannot acquire information about the hidden states because epistemic exploration seeks out the policies that can change the prior beliefs about the hidden states substantially.

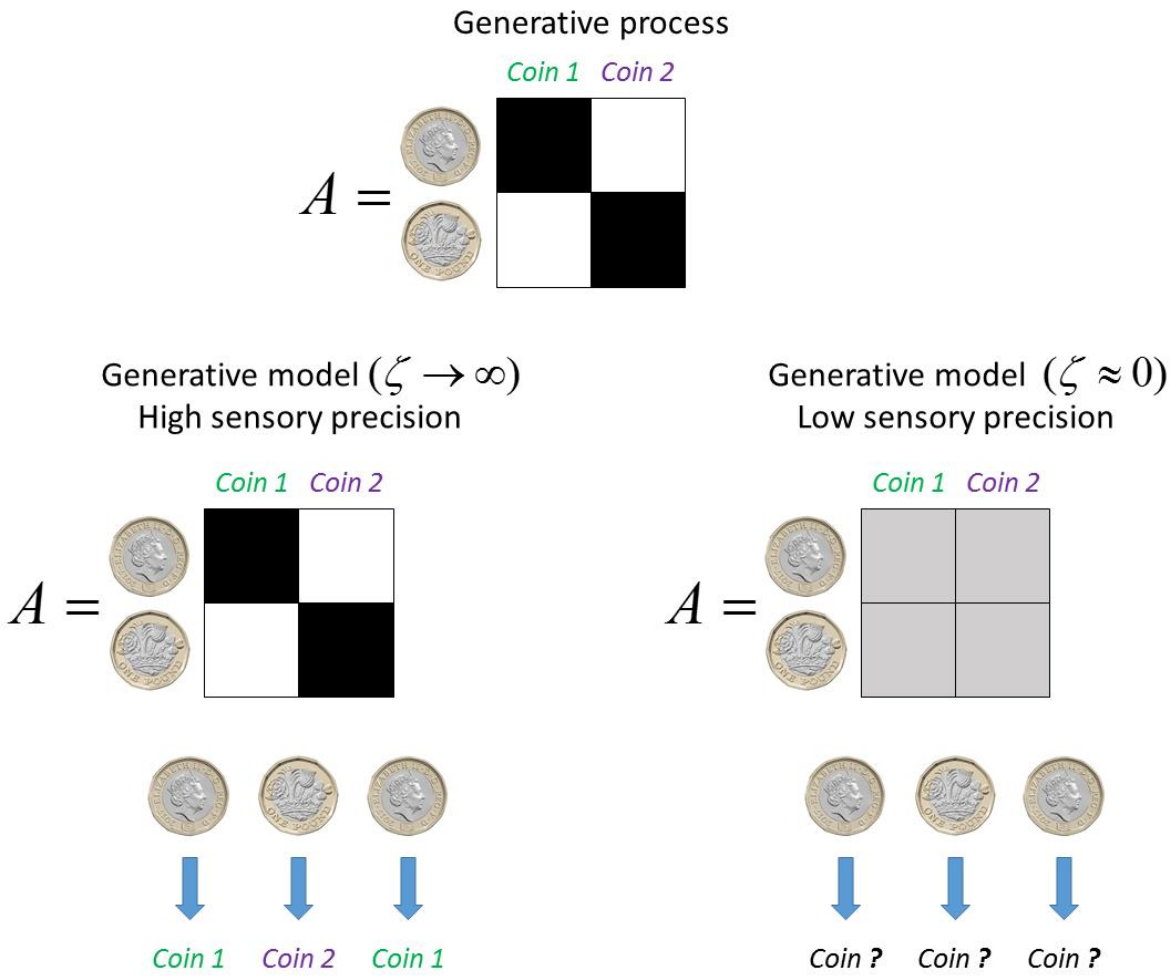


Figure 5. 2 Precise vs imprecise likelihood matrices

This figure shows how manipulating the precision of the likelihood matrix in the generative model changes the inference about the hidden states under the same generative process. The top panel specifies the likelihood matrix in the generative process for a coin flip, i.e. there are two different unfair coins where coin 1 is responsible for the heads outcomes and coin 2 is responsible for the tails outcomes. The lower left and right panels show what the beliefs in the generative model are like, namely very high $\zeta \to \infty$ and very low $\zeta \approx 0$, respectively. When the sensory precision is very high $\zeta \approx 0$, one would know coin 1 and coin 2 are responsible for heads and tails outcomes, respectively. However, when the sensory precision is very low $\zeta \approx 0$, one would no longer know which coin is responsible for the observed heads or tails, even when the generative process is exactly the same as in the top panel.

5.2. A contextual exploration task

In this section I introduce a contextual visual search task called the colour/shape task. The colour/shape task is performed on a two-by-two grid scene whose quadrants are masked in the beginning. Attending to a quadrant unmasks the object in that quadrant

(see Fig 5.3). In this task certain objects are associated with certain contexts. These contexts can be seen as rules that state what information should be sought out, very much like the instructions given in Yarbus' experiment (Yarbus, 1967). The goal in this task is to categorise the scene that is being explored. Crucially a scene can be categorised either in terms of its *colour* or its *shape*. The agent is informed about the categorisation rule before performing the task. The colour category is determined by mixtures of colours in top left and top right quadrants (the top left is always red, and the top right can be red, yellow or blue). The category of the scene is red, orange or purple if the object in the top right quadrant is red, yellow or blue, respectively given that the rule is *categorise colour*. The shape category of the scene is square, circle or triangle, if the object in the bottom left quadrant is square, circle or triangle, respectively given that the rule is *categorise shape* (see Fig 5.4). Beliefs about the category of the scene are reported by attending to one of the three choice locations at the bottom of the scene. These choice locations either correspond to *colour* or *shape* categories depending on what the *rule* is (see the rightmost panels in Fig 5.3). Upon making a categorisation, feedback of *right* or *wrong* is given.

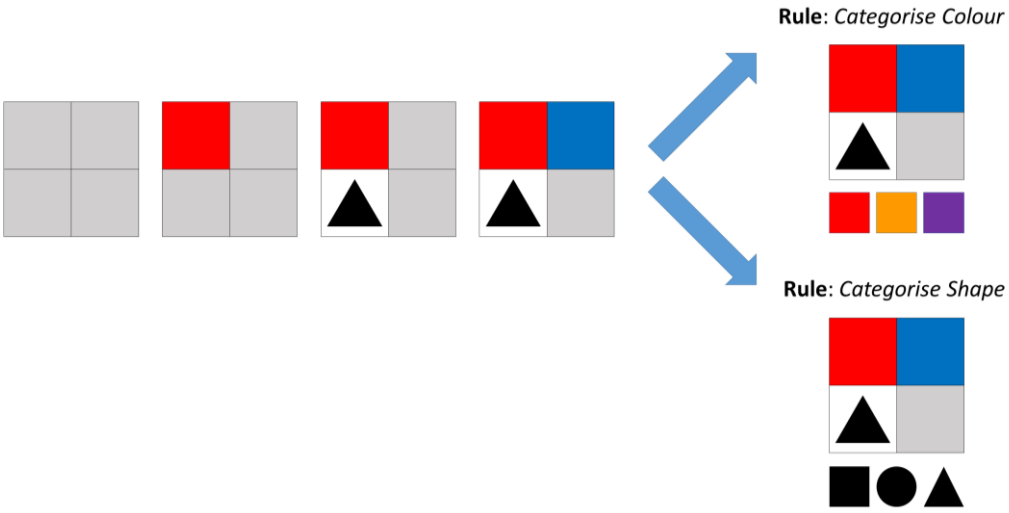


Figure 5. 3 Sequence of observations

The sequence of scenes shows an example scene exploration. In the beginning each quadrant is greyed out. Attending to each quadrant reveals the content of each quadrant. In this case the order of explored locations are top left, bottom left and top right quadrants.

Epistemic exploration of a scene would continue until all uncertainty about the hidden states of the world (here, colour and shape categories) is resolved. However, rule based exploration requires one to resolve uncertainty only about the relevant hidden states. There is no reason why one should resolve uncertainty about the *shape* category when the rule is *categorise colour*.

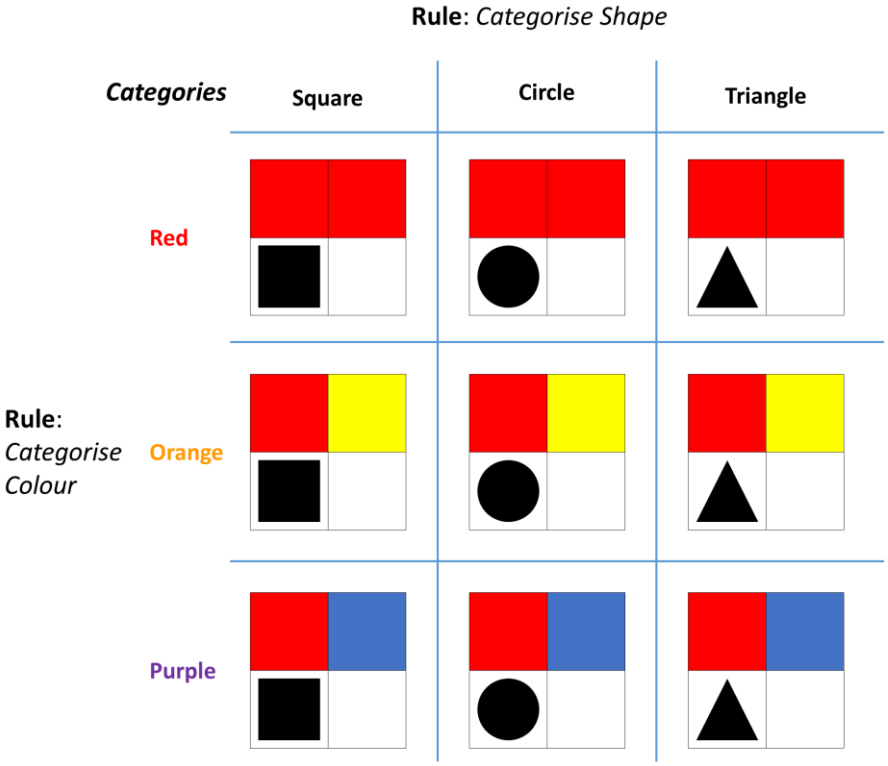


Figure 5. 4 Scene categories

A scene can be categorised in two different ways, either by its colour or its shape. The categories of the scenes are shown on the left and on top when the rule (context) is to categorise the scene in terms of its colour and shape, respectively.

In the next section I introduce the MDP model of the colour/shape task.

5.3. MDP model of the colour/shape task

This section describes the MDP model for the colour/shape task. In this MDP model I considered four sets of hidden states, namely *Rule*, *Where*, *Category: colours* and *Category: shapes*. The first set of hidden states *Rule* defines the context in which the scene will be categorised. A scene can be categorised in two ways, either in terms of its *colour* or its *shape,* depending upon the rule. The second set of hidden states

Where corresponds to the locations in the scene. There are eight locations in this task: central fixation (location 1), the four quadrants (locations 2-5) and three choice locations (locations 6-8) at the bottom. The choice locations are associated with the categories *red*, *orange* and *purple* when the rule is *categorise colour*, and *square*, *circle* and *triangle* when the rule is *categorise shape*. The third set of hidden states *Category: colours* controls what colours will appear on the top left and top right quadrants under the colour categories *red*, *orange* and *purple*, e.g. if the colour category is *purple* then the colours *red* and *blue* will be in these two locations. The fourth set of hidden states *Category: shapes* determines which shape will be in the bottom left quadrant under the shape categories *square*, *circle* and *triangle*.

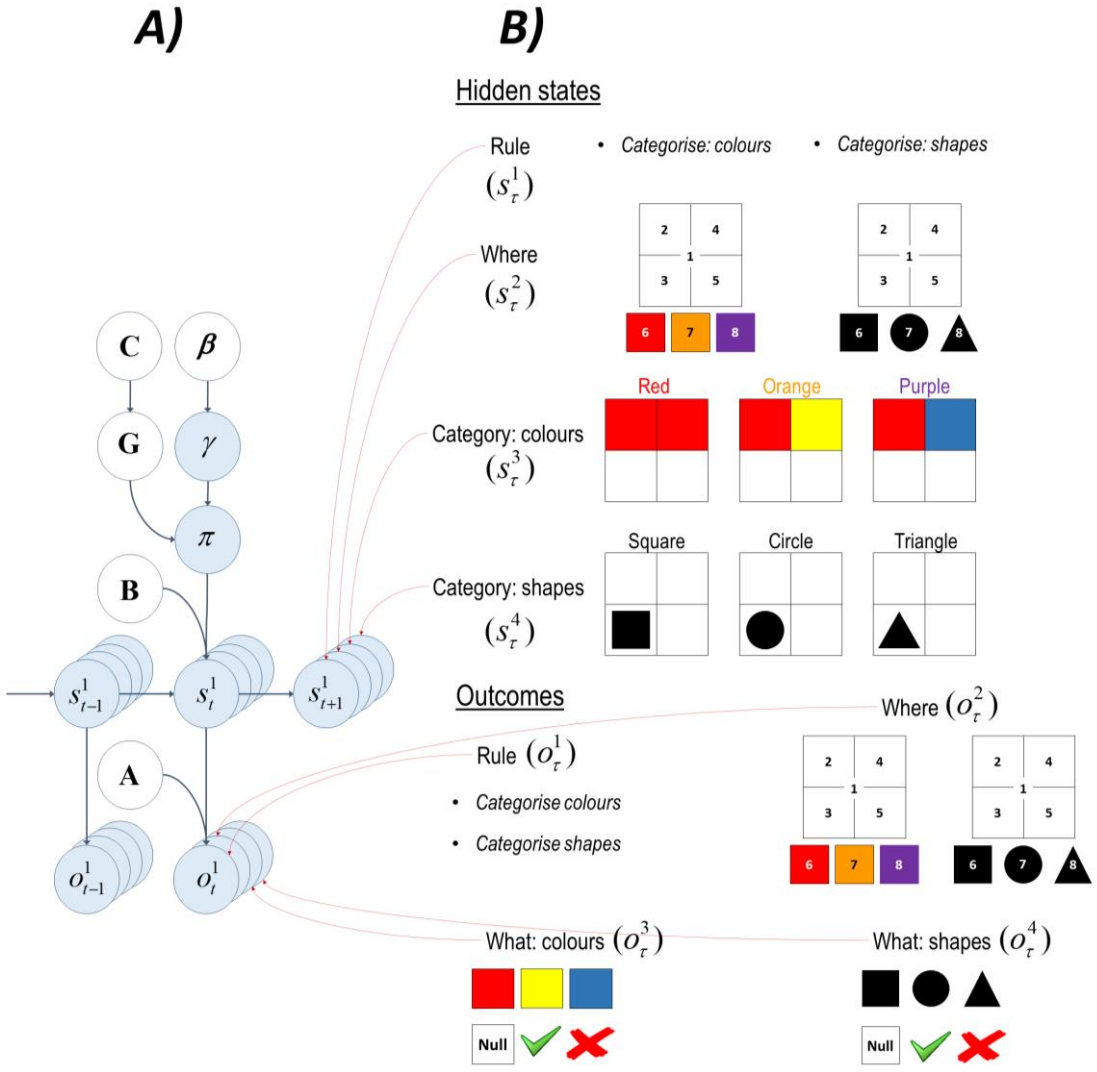


Figure 5. 5 Structure of the generative model

A) This panel shows the graphical representation of the MDP model and the conditional dependencies among the terms in the model. See Fig 3.1A for details. **B)** This panel shows the four sets of hidden

states and outcome modalities in the colour/shape task. There are four sets of hidden states, namely *Rule*, *Where*, *Category: colours* and *Category: shapes*. There are four outcome modalities, namely *Rule*, *Where*, *What: colours* and *What: shapes*.

I considered four outcome modalities, namely *Rule*, *Where*, *What: colours* and *What: shapes*. The first outcome modality *Rule* unambiguously cues the context, either categorise in terms of *colour* or *shape*. The second outcome modality *Where* signals the sampled location in the scene (one of eight locations). This can be thought of as a proprioceptive (or dorsal visual) signal. The third outcome modality *What: colours* signals which colour is observed in the sampled location. It can be *red*, *yellow*, *blue* or *null* (no colour). The fourth outcome modality *What: shapes* signals which shape is observed in the sampled location. It can be *square*, *circle*, *triangle* or *null* (no shape). Under both *What: colours* and *What: shapes* modalities there are two additional feedback outcomes, *right* and *wrong*. An agent can report its beliefs about the category of the scene by choosing one of the three choice locations associated with the categories under the rules *categorise colour* or *categorise shape* and obtain feedback about whether its choice was *right* or *wrong*. See Fig 5.5 for the hidden states and outcome modalities.

In this setup the *Rule* and *Where* hidden states directly map to the *Rule* and *Where* outcomes. The hidden states *Category: colours* and *Category: shapes* map onto *What: colours* and *What: shapes* objects as a function of *Where* and *Rule* hidden states, e.g. sampling location 8 when the rule is *categorise colours* would generate a *right* feedback if the scene category is *purple*. All the transition matrices are identity matrices except for the action dependent *where* transition matrix. The identity matrices indicate that the rule and the scene category do not change in the course of a trial. The action dependent *where* transition matrix specifies that the agent would attend to the location indicated by the action, e.g. if the sampled action is 4 then the agent would go to the top right location. In this setup, I defined prior preferences over *right* (utility or relative log probability of 2 nats) and *wrong* (utility of -4) outcomes under both *What: colours* and *What: shapes* modalities. With these utilities the agent avoids categorising the scene prematurely and categorises only once it has accumulated sufficient evidence. See Fig 5.6 and Fig 5.7 for the likelihood, prior preference and transition matrices.

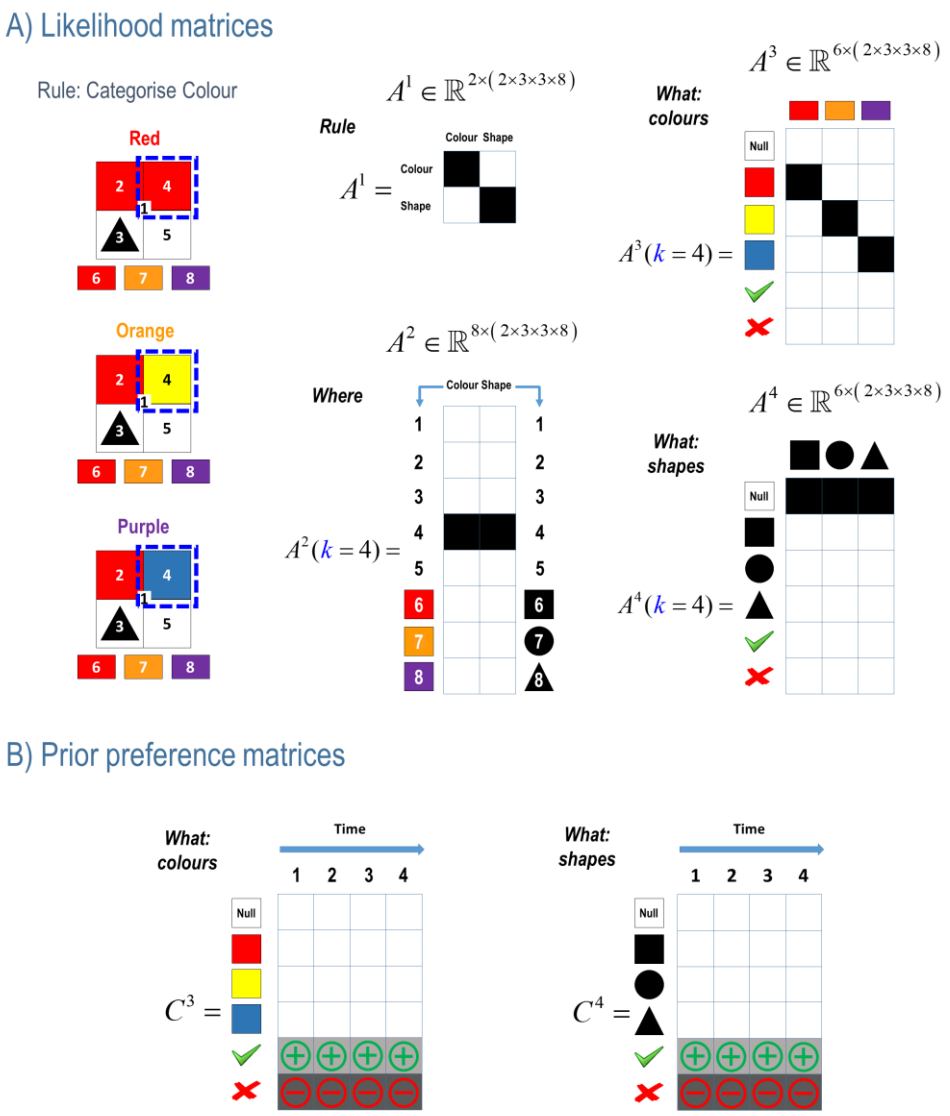


Figure 5. 6 Likelihood and prior preference matrices

This figure shows the likelihood and prior preference matrices used in the colour/shape task. **A)** The colour category of the scenes shown on the left is purely determined by the colour in location 4 (top right quadrant). The panels on the right show the likelihood ($\mathbf{A}$) matrices for location 4. The likelihood matrices encode the probability of outcomes (o_t) given the hidden states (s_t). The first likelihood matrix A^1 (*Rule*) signals what the rule is, either categorise *colour* or *shape*. The second likelihood matrix A^2 (*Where*) signals the sampled location on the scene, one of *eight* locations. The third likelihood matrix A^3 (*What: colours*) encode the probability of colours *red*, *yellow* and *blue* under different colour categories *red*, *orange* and *purple*. The final likelihood matrix A^4 (*What: shapes*) encode the probability of shapes *square*, *circle* and *triangle* under different shape categories *square*, *circle* and *triangle*. Because the colour and shape are mutually exclusive modalities, the probability of colour and shape objects are encoded by separate likelihood matrices A^3 and A^4. The likelihood matrix under the colour

modality for location 4 $A^3(k=4)$ shows that the colour category of a scene is purely determined by the colours in this location, however under the shape modality $A^4(k=4)$ the object in this location does not give any information about the shape category, i.e. *null*. **B)** The prior preference matrices are shown in this panel. The prior preference matrices encode how much one outcome is preferred relative to another outcomes as a function of time. The only preferences are defined over the columns of C^3 (*What: colours*) and C^4 (*What: shapes*). Under both C^3 and C^4 the utility of making a right categorisation and wrong categorisation is $+2$ and -4, respectively. With these utilities the agent expects to categorise a scene correctly while avoiding to make an incorrect categorisation.

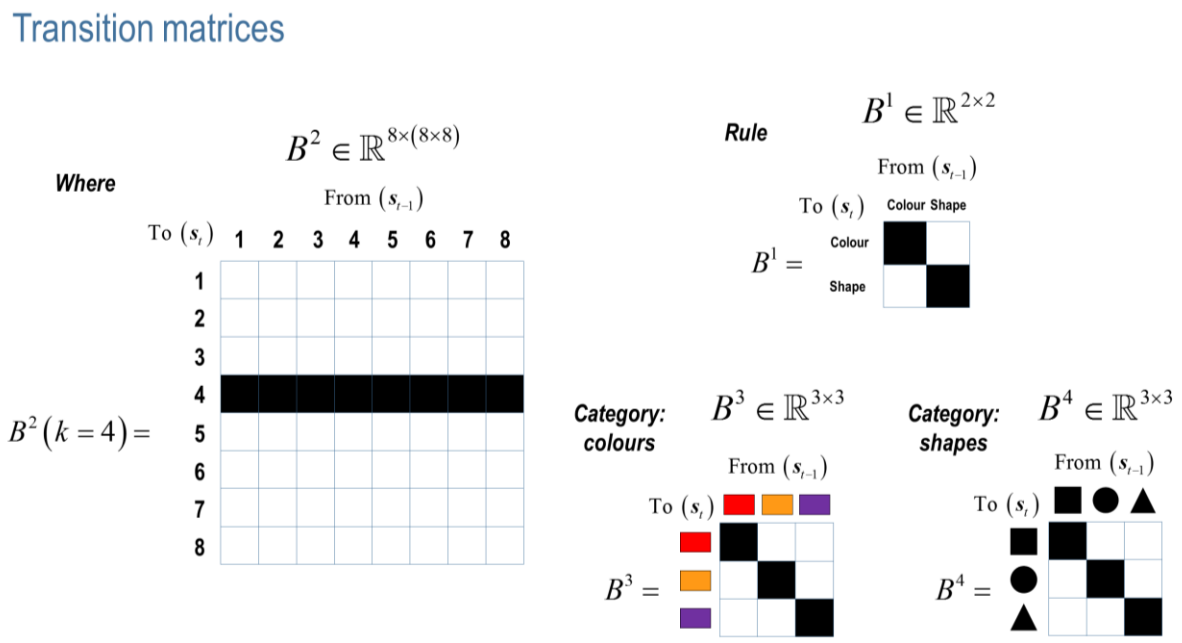


Figure 5. 7 Transition matrices

This panel shows the transition matrices. All the transition matrices are identity matrices except for the action dependent transition matrix B^2, which encodes the most likely location to be sampled as a function of action, e.g. $B^2(k=4)$ shows that under action **4**, the top right quadrant is the most likely location to be sampled in the next time step. The identity transition matrices B^1 (*Rule*), B^3 (*Category: colours*) and B^4 (*Category: shapes*) express that the *rule* and the *colour* and *shape* objects in the scene do not change in the course of a trial.

Epistemic exploration seeks out the information that can be acquired about an environment. However more often than not, the information out there is not useful to the task at hand. In the next section I show that attentional mechanisms need to be in

play for contextual exploration to occur and how information that is task relevant can be acquired.

5.4. Contextual epistemic exploration

In the model described above, the uncertainty that can be resolved through exploration is about the scene category in terms of its *colour* and *shape*. Epistemic exploration favours saccades to the locations that offer information about *colour* and *shape* categories of the scene regardless of what the *Rule* is. Rule based (contextual) exploration requires an agent's attention to be directed such that only relevant information under a context matters. The most salient actions are then those that yield observations (in this case *colour* and *shape* modalities) that are generated by hidden states (objects under *colour* and *shape* categories) with a high fidelity (precision). Here, I show that beliefs about the uncertainty in the mapping from the hidden states of the world s_t to sensory observations o_t can modulate the salience associated with saccades to each location (T. Parr & Friston, 2017a).

In the colour/shape task, the precision of the sensory signals is modulated as a function of the *Rule* hidden state dimension in the generative model. This works such that when the *Rule* hidden state is *categorise: colour* the sensory precision of the *shape* objects is set very low while the sensory precision of the *colour* objects is set very high, and vice versa for hidden state *categorise: shape*. This can be formally expressed with the equations below.

$$A^m_{nijkl} = P\left(o^m = n \mid s^1 = i, s^2 = j, s^3 = k, s^4 = l\right) \tag{13}$$

$$P\left(o^m = n \mid s^1 = i, s^2 = j, s^3 = k, s^4 = l\right) = \sigma\left(\zeta^m_i \ln \bar{A}^m_{nijkl}\right) \tag{14}$$

$$\zeta = \boldsymbol{m} \begin{array}{c} \\ \textit{Rule} \\ \textit{Where} \\ \textit{What}:\textit{Colour} \\ \textit{What}:\textit{Shape} \end{array} \begin{array}{c} \boldsymbol{i} \\ \begin{array}{cc} \textit{Categorise} & \textit{Categorise} \\ \textit{colours} & \textit{shapes} \end{array} \\ \begin{bmatrix} \infty & \infty \\ \infty & \infty \\ \infty & z \\ z & \infty \end{bmatrix} \end{array} \tag{15}$$

Equation 13 expresses the likelihood of the outcome $o^m = n$ in the *generative process* given the hidden states $s^1 = i$, $s^2 = j$, $s^3 = k$ and $s^4 = l$ with $m \in M$, $n \in N$, $i \in I$, $j \in J$, $k \in K$, $l \in L$ where M is the number of different outcome modalities (*Rule*, *Where*, *What: colours* and *What: shapes*), N is the number of outcomes in an outcome modality (e.g. under the *What: colours* modality red, yellow and blue colours) and, $I = \{ categorise\ colour, categorise\ shapes \}$, $J = \{ location\ 1, \ldots, location\ 8 \}$, $K = \{ red, orange, purple \}$ and $L = \{ square, circle, triangle \}$ are the number of states under different hidden state dimensions (e.g. under the first hidden state dimension s^1, *Rule* states *categorise: colours* and *categorise: shapes*).

Equation 14 expresses the likelihood of the same outcome in Equation 13 but this time for the *generative model*. This likelihood mapping is subject to the precision terms ζ. The precision term ζ_i^m is applied to the logarithm of the likelihood matrix for the *m*-th outcome modality A^m given the *i*-th level of the first hidden state $s^1 = i$. Finally a softmax function is applied to the resulting term to normalise the columns of the likelihood matrix to the range of probabilities.

The matrix in Equation 15 is a precision matrix which shows the values of the precision terms ζ_i^m for different outcome modalities m and different levels of the first hidden state dimension $s^1 = i$ which specifies what the rule is (*categorise: colour* or *categorise: shape*). This matrix shows that this precision term ζ_i^m is infinitely big under the modalities *Rule* and *Where* for $\forall i$, which means that there is a deterministic mapping to the outcomes under these two modalities. The crucial manipulation is implemented

under *What: colours* and *What: shapes* modalities. When the first hidden state dimension s^1 is on the $i = categorise : colours$ level (see the first column of the precision matrix) there is a very precise mapping to the colour objects ($\zeta_i^m = \infty$) under the *What: colours* modality. With the beliefs of this precise mapping agent thinks that it can resolve uncertainty about the *Category: colours* hidden state (see the left panel of Fig 5.8). The precision of the mapping from $i = categorise : colours$ to shape objects under *What: shapes* modality is expressed as a function of z. When $z = 0$ the mapping to the shape objects become very imprecise which makes the agent believe it cannot resolve uncertainty about *Category: shapes* hidden state. Therefore the agent's attention would be focused only on task relevant objects, namely colour objects (see the right panel of Fig 5.8). When $z \to \infty$ the mapping to the shape objects are very precise which means that the agent's attention would be divided between two different outcome modalities, namely *What: colours* and *What: shapes*, that could resolve uncertainty about two different hidden states, namely *Category: colours* (task relevant) and *Category: shapes* (task irrelevant) categories (see the middle panel of Fig 5.8). A similar formulation is expressed in the second column of the precision matrix when the rule is *categorise: shapes*.

In the next section I show how changing the precision terms introduced in this section can give rise to contextual exploration with simulated responses.

5.5. Simulations of the colour/shape task

5.5.1. Colour/shape task

The information gained from observing a stimulus depends critically on the precision ζ of the likelihood mapping between that stimulus and an uncertain hidden state, e.g. the degree to which seeing 'blue' means the scene must be a 'purple' category. By reducing the ζ of the task irrelevant likelihood, an agent can reduce the expected information gained from observing task irrelevant objects, and thus 'attend away' from them.

Let us assume that the rule is to categorise the scene in terms of its colour. Under this rule, colour objects are task relevant and shape objects are task irrelevant. Thus, when

performing the contextual scene categorisation task, an agent would only attend to the colour objects if ζ^{colour} is maximised (Fig 5.8, left panel) and ζ^{shape} is minimised, i.e. $\zeta^{shape} = 0$ (Fig 5.8, right panel). If ζ^{shape} is maximised (i.e. $\zeta^{shape} \to \infty$, see Fig 5.8, middle panel), the agent becomes more likely to attend to the task irrelevant objects.

Note that the agent's beliefs about the likelihood mapping and the mapping in the real world may not be the same: i.e. the generative model (internal beliefs) and generative process (real-world) may be different. In the right panel of Fig 5.8 the agent believes that there is an imprecise mapping between *shape categories* and objects (generative model) but this mapping is very precise in the process that generates outcomes (generative process). In fact the middle and right panels of Fig 5.8 show the cases when the generative model and generative process are identical and different, respectively. These panels show how varying ζ^{shape} changes the mapping between the task irrelevant *shape* category and objects when the rule is *categorise colour*. The mapping between *colour* category and objects are changed in the very same way using ζ^{colour} when the rule is *categorise shape*.

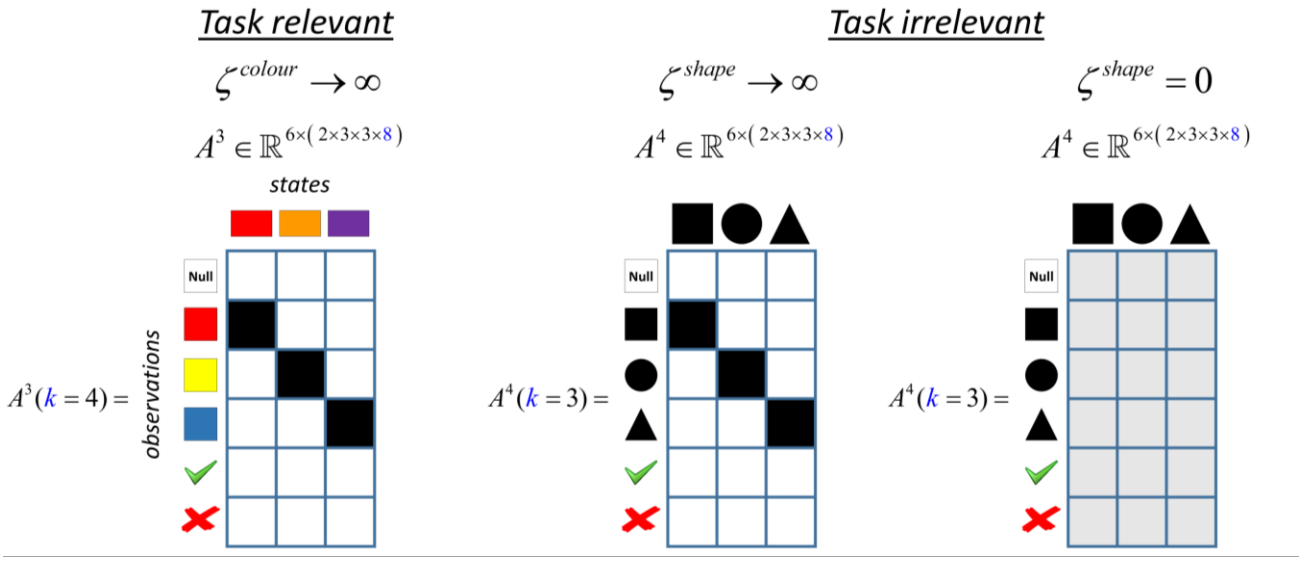


Figure 5. 8 Likelihood matrices under different precisions

In this figure the rule is to categorise a scene in terms of its *colour*. The objects that resolve uncertainty about the *colour* category are task relevant, and the objects that resolve uncertainty about the *shape*

category are task irrelevant. The likelihood matrix on the left shows the mapping between the task relevant objects and categories (in this case *colour* objects and categories) under high precision $\zeta^{colour} \to \infty$. The likelihood matrices in the middle and right panels show how the mapping between the task irrelevant objects and the categories (in this case *shape* objects and categories) change under two levels of precision ζ^{shape}. Under a high precision $\zeta^{shape} \to \infty$ this mapping is very precise, however under a low precision $\zeta^{shape} = 0$ it becomes very ambiguous. When this mapping is very imprecise the agent no longer acts to resolve uncertainty about the task irrelevant category. The task relevant and irrelevant likelihood matrices are illustrated for locations 3 and 4, because these are the only locations that hold the objects that can resolve uncertainty about the *shape* and *colour* categories, respectively. The precision term ζ^{colour} is used to change the mapping between *colour* objects and categories in the same way when the rule is *categorise shape*.

The upper and lower left panels of Fig 5.9 show the quadrants that the agent attended to in the course of a trial under high $\zeta^{shape} \to \infty$ and low $\zeta^{shape} = 0$ levels of task irrelevant precision, while keeping the task relevant precision high $\zeta^{colour} = 0$ for both trials. The heat maps in the right panels show how likely the agent is to attend to a particular location in the scene, expressed in terms of a softmax function of expected free energy under eight policies (i.e. visiting one of the eight locations in the scene). On the trials shown in Fig 5.9, the rule is to categorise the scene in terms of its *colour*. When the agent believes that it can acquire information about the task irrelevant *shape* category (i.e. $\zeta^{shape} \to \infty$; upper panel), it finds that it is equally likely to attend to the top right (colour) and bottom left (shape) quadrants at $t = 1$, even though the only object that can resolve uncertainty about the *colour* category is in the top right quadrant. The agent chooses between the two randomly, in this case the bottom left (shape) quadrant, and only then attends to the top right (colour) quadrant at $t = 2$, successfully categorising the scene as *purple* at $t = 3$. Conversely, when the agent does not believe that it can resolve uncertainty about the task irrelevant *shape* category $\zeta^{shape} = 0$ (lower panel), it ignores the bottom left (shape) quadrant and categorises the scene as *purple* one timestep earlier.

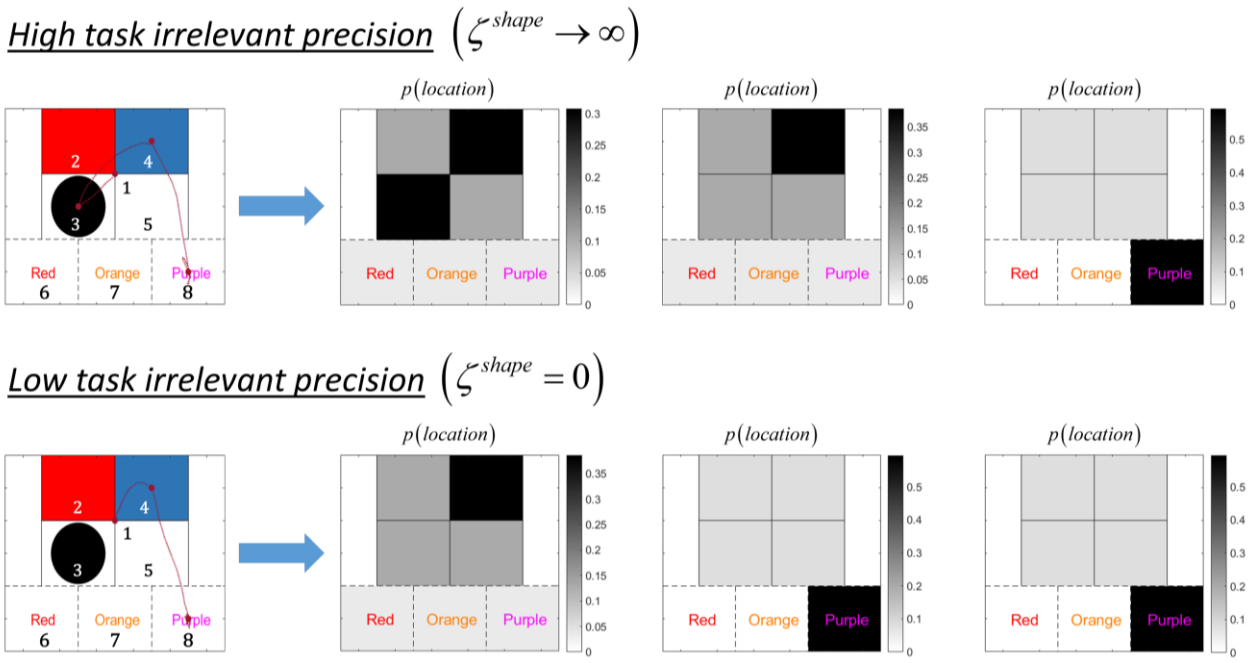


Figure 5. 9 Exploratory behaviour under different precisions

The upper and lower left panels show how the exploratory behaviour changes under two different levels of task irrelevant precisions $\zeta^{shape} \to \infty$ and $\zeta^{shape} = 0$, while keeping taks relevant precision high $\zeta^{colour} \to \infty$. The subsequent panels show how likely an agent is to sample a location in the course of a trial, expressed in terms of prior probabilities for each policy. At the beginning of each trial, the agent fixates at the centre of the screen (location 1). Under high task irrelevant $\zeta^{shape} \to \infty$ and task relevant $\zeta^{colour} \to \infty$ precisions the agent finds that it is equally likely to attend to the task relevant *colour* objects in location 4 as the task irrelevant *shape* objects in location 3, in the beginning. The agent first attends to location 3 where it finds a *circle* and then it attends to location 4 where it finds the colour *blue*. Subsequently the scene is categorised as *purple*. Under a low task irrelevant precision $\zeta^{shape} = 0$ and high task relevant precision $\zeta^{colour} \to \infty$ the agent finds that the only location that matters is location 3 which holds the task relevant *colour* objects. In the next time step it attends to location 3 and finds the colour *blue* and subsequently categorises the scene as *purple*.

5.5.2. Yarbus' free exploration task

The same principles that are applied in the task above can be applied to Yarbus' free exploration task. Just like in the colour/shape task, the scene can be explored in two different ways depending on what the instruction is. When the instruction is *estimate the family's material circumstances*, the only objects that matter are the furniture and

the people's clothing, whereas when the instruction is *give the ages of the people*, the faces of the people hold the most information.

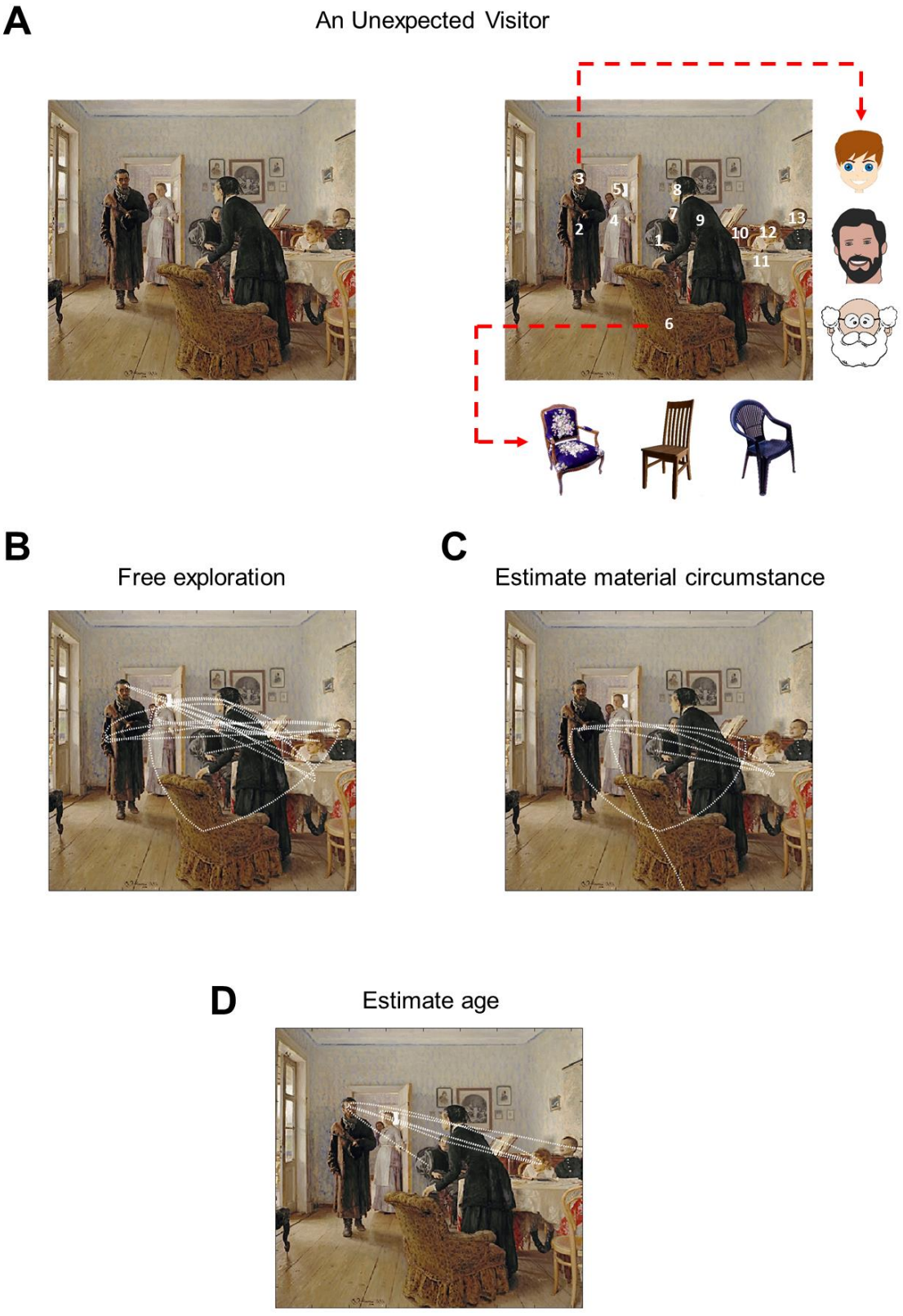


Figure 5. 10 Yarbus' free exploration task and simulations

A) The painting *An Unexpected Visitor* by Ilya Repin is shown on the left. The panel on the right highlights 13 locations that can give information about either *material circumstances* of the family or *the ages* of the people in the painting. The first location is the centre of the scene. Furniture and people's clothing appear on locations 2, 4, 6, 9, 10 and 11. People's faces appear in locations 3, 5, 7, 8, 12 and 13. For illustrative purposes the location 3 has been chosen to show that one can see *young*, *middle aged* or *old* faces in locations a face can appear. Location 6 has been chosen to show that one can see an *antique*, a *modest* or a *common chair* in locations a piece of furniture (or man's/woman's clothing) can appear. **B)** This panel shows the simulated exploratory behaviour of a context naïve agent (unaware of the instructions and thus exploring the scene freely). The agent starts exploring from the centre of the scene (location 1). **C)** This panel shows how the agent explores the painting under the instructions *estimate material circumstances* of the family. **D)** This panel shows the responses of the agent when exploring the painting under the instructions *give the ages of the people*. Please compare the simulated scanpaths in panels B, C and D with the scanpaths of the subjects from Yarbus' work (see the first three scanpaths in Fig 109 from Yarbus (1967)). The painting *Unexpected Visitors* (or *They did not expect him*) by *Ilya Repin* has been downloaded from https://en.wikipedia.org/wiki/File:Ilya_Repin_Unexpected_visitors.jpg (Repin, 1884-1888). Wikipedia contributors. (2019, January 21). Ilya Repin. In *Wikipedia, The Free Encyclopedia*. Retrieved 14:37, February 15, 2019, from https://en.wikipedia.org/w/index.php?title=Ilya_Repin&oldid=879516824. The simulated scanpaths are superimposed on this painting to produce the responses seen in the panels B, C and D.

The subjects in Yarbus' study knew where to expect certain objects in the painting because they were instructed to explore the scene freely before exploring the same scene under different instructions, and secondly one expects the objects in the world to appear in certain locations (Biederman, 2017) e.g. furniture and faces tend to appear at different heights, and different positions relative to other objects. These locations are highlighted with numbers between 1 and 13, where location 1 corresponds to the centre of the scene (see the right panel of Fig 5.10A). The furniture and clothes appear in locations 2, 4, 6, 9, 10, 11 whereas the faces appear in locations 3, 5, 7, 8, 12, 13.

I considered that each of these locations could hold a number of objects: i.e. faces of different ages at the higher locations (e.g. location 3, Fig 5.10A), and a variety of furniture or clothes at the lower locations (e.g. location 6, Fig 5.10A). The presence of an *antique*, a *modest* or a *common chair* at location 6 cues the *material circumstances* of the family: *wealthy*, *middle class* or *poor*. The presence of different faces cues the

average age of the people in the picture: *young*, *middle aged* or *old*. In this setting, a very low precision $\zeta = 0$ would induce imprecise likelihood matrices for the task irrelevant objects and categories, e.g. when the instruction is to estimate *material circumstances* of the family, the likelihood matrix for the *ages* becomes very imprecise.

Fig 5.10 shows the simulated saccadic patterns generated using the same principle described for the colour/shape task. In Yarbus' work the participants were first asked to explore the painting freely, then to estimate the material circumstances of the people, or their ages (see the first three scanpaths in Fig 109 from Yarbus (1967)). Like Yarbus' participants, the agent attends to all the faces and most of the furniture (and clothing) in the scene during free exploration (Fig 5.10B), i.e. when the agent is unaware of the instructions, but when the instructions are *estimate the material circumstances of the family* or *give the ages of the people* the agent selectively attends to the furniture and clothing (Fig 5.10C) or faces (Fig 5.10D) respectively.

Model behind the simulations of Yarbus' task

The hidden state and outcome spaces used to generate the simulated saccadic patterns in Yarbus' experiment (see Fig 5.10) is fairly similar to the model described above. The hidden state space consists of five dimensions, namely *Instruction*, *Where*, *Category: wealth*, *Category: age, and Scene type*. *Instruction* is either *estimate material circumstances* or *give the ages of the people*. *Where* encodes one of thirteen locations in the scene. *Category: wealth* encodes the material circumstances, which could be *wealthy*, *middle class* and *poor*. *Category: age* encodes the average age of the people in the scene and these are *young*, *middle aged* or *old*. *Scene type* consists of a number of different scenes that map onto the same states in *Category: wealth* and *Category: age* state dimensions. Essentially *young*, *middle aged* and *old* faces can appear under different age categories. A scene whose category is *young* contains predominantly *young* faces. Each scene under the category *young* can contain other type of faces, e.g. *middle aged* and *old*. *Scene type* encodes the number and locations of these faces. *Scene type* encourages exploration of the scene. There are four outcome modalities, namely *Instruction*, *Where*, *What: object* and *What: face*. *Instruction* and *Where* states directly map onto *Instruction* and *Where* outcomes. *What: object* outcome contains *antique*, *modest* and *common* chairs, whereas *What:*

age outcome contains *young*, *middle* aged and *old* faces. See Fig 5.11 for the generative model used to simulate the scan-paths in Fig 5.10.

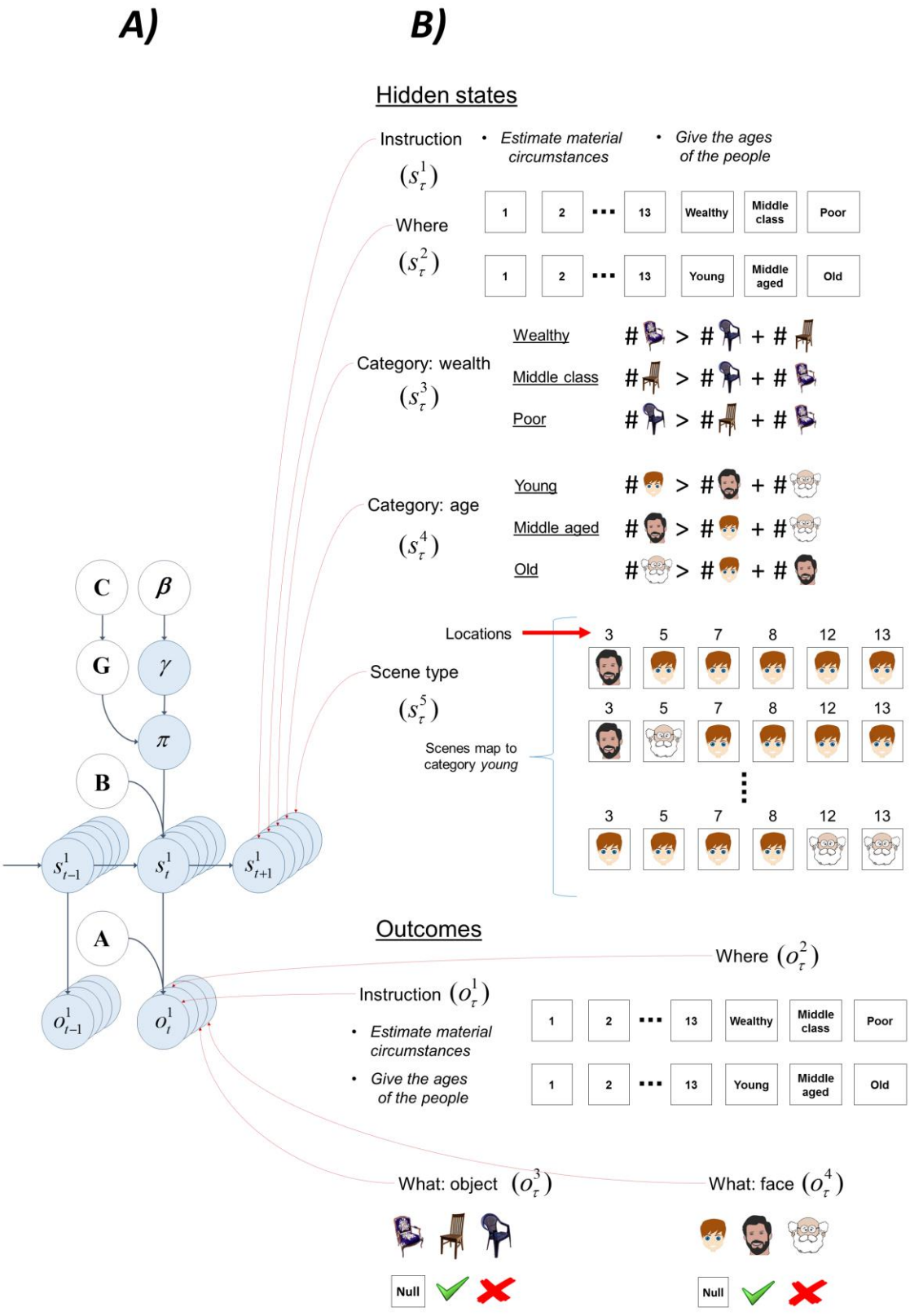


Figure 5. 11 Structure of the generative model – Yarbus' task

A) This panel shows the graphical representation of the MDP model and the conditional dependencies among the terms in the model. See Fig 3.1A for details. **B)** This panel shows the five sets of hidden states and four outcome modalities in our MDP version of Yarbus' task. There are five sets of hidden states, namely *Instruction*, *Where*, *Category: wealth*, *Category: age* and *Scene type*. There are four outcome modalities, namely *Instruction*, *Where*, *What: object* and *What: face*. See the main text for details.

5.6. Discussion

In this chapter I showed that a computational agent can selectively attend to the information that is useful under a context by inferring the appropriate attentional targets. Computationally, this corresponds to modulating the precision of the mapping (encoded by the likelihood matrix) between task irrelevant sensory inputs (stimuli that aren't informative in a certain context) and their hidden causes. When the precision of the task irrelevant likelihood is very low, an agent only attends to the task relevant stimuli. This model reproduces the saccadic patterns in empirical studies of context-dependent human exploratory behaviour (Yarbus, 1967).

The exploratory behaviour of the agent described in this chapter is driven by epistemic value (Mirza et al., 2016) a.k.a. Bayesian surprise. Bayesian surprise attracts human attention (Itti & Baldi, 2009): in other words, a stimulus attracts attention if it changes an observer's beliefs significantly. Clearly, this depends upon what beliefs an observer currently holds. I demonstrated the capacity for an agent to revaluate beliefs about context, given a cue, such that the same stimulus can carry different levels of surprise in different contexts.

Most computational models of visual search are bottom-up models of visual attention that do not take into account the contextual information inherent in visual scenes. These models usually create a 'saliency map' based on the features of the objects in the scene. These features include orientation, intensity, colour information (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002), luminance (Achanta, Hemami, Estrada, & Susstrunk, 2009), contrast (Ma & Zhang, 2003) and motion (Rosenholtz, 1999). Typically, the locations in these saliency maps are attended in order of decreasing salience – often requiring an inhibition-of-return rule to be explicitly introduced. Although these models provide relatively good

predictions of where visual attention will be deployed in pop-out visual search tasks, they do not incorporate contextual information. There is no reason why a bottom-up visual search model would find the faces of people more salient when an instruction such as ‘give the ages of the people in the scene’ is given. Only models with a top-down element have the potential to make use of such instructions.

There are a number of visual attention models that can incorporate top-down knowledge during visual search. Top-down instructions in these models are usually given in the form of prior knowledge about the features of an object of interest. While some top-down models evaluate the similarity (or dissimilarity) of the features of the object of interest with the features in the scene that is being explored (Torralba, Oliva, Castelhano, & Henderson, 2006), there are other models that either modulate or select feature outputs such that the features of the object of interest during visual search become more salient (Navalpakkam & Itti, 2006; Wolfe, 1994). A noteworthy model defines image categories as different visual patterns and approaches the scene categorisation problem by maximising the mutual information between scene categories and pixel values at possible fixation locations (Yang, Lengyel, & Wolpert, 2016). A similar approach maximises the pointwise mutual information between a target object and visual features (Zhang, Tong, Marks, Shan, & Cottrell, 2008). There are also other top-down models that either use iconic scene representations to predict the location that holds the object of interest (Rao, Zelinsky, Hayhoe, & Ballard, 2002) or models that equate salience to discrimination and consider the features that best distinguish the object of interest from the other objects as salient (Gao & Vasconcelos, 2005).

Similar to some of the models above, the model I introduced in this work is equipped with an information acquiring component, namely epistemic value. Epistemic value resolves uncertainty about the hidden states of the world and it is defined as the mutual information between the hidden states and the observations. This model diverges from the above models in a number of ways. In this model contextual exploration arises due to entertaining imprecise beliefs about context irrelevant objects, which precludes information gain about context irrelevant hidden states (scene categories). This is the first model that can explore scenes under a context using this mechanism to my knowledge. Furthermore, this model can successfully report its beliefs about the scene

category by exploiting extrinsic value (expected utility). I emphasise the top-down inferential processes that use relatively abstract semantic representations. This contrasts with the lower level representations used in other models to describe features of visual scenes. While I have not addressed the contributions of early visual pathways here, I could interpret the sensory outcomes as alternative hypotheses about the continuous variables describing simple visual features of the objects (K. J. Friston, Parr, & de Vries, 2017). Furthermore, my approach makes use of an explicit generative model that depends upon prior beliefs. This is important, as a number of clinical conditions have been associated with abnormalities in prior beliefs, and this paradigm might afford an opportunity to investigate these changes quantitatively.

People with a diagnosis of autism spectrum disorder (ASD) are known to explore visual scenes (especially faces) differently than neurotypicals. In free visual search tasks that contain pictures of faces, people with ASD attend less to the core features of faces (e.g., eye, nose and mouth) and more to the non-feature regions of faces (Pelphrey et al., 2002) and they are slower at discriminating faces in face discrimination tasks (Behrmann et al., 2006). In contrast, people with ASD have been shown to be superior to controls on visual search tasks that involve visual illusions (Happé, 1996) and faster on tasks that involve spotting a target object that shares certain features with the distractors (Plaisted, O'Riordan, & Baron-Cohen, 1998). The exploratory behaviour of people with ASD may be due to one or more perturbations in this model: altered model structure (not knowing the mapping from e.g. gaze to mental states), reduced recognition of context (e.g. not realising that a given situation warrants information gathering about mental states) where context can be defined as the global configuration of features and objects, or a difficulty in down-modulating the precision of task irrelevant object mappings. Paradoxically, the latter would imply a more accurate generative model – consistent with superior (pop-out) visual search performance in autism (O'riordan, Plaisted, Driver, & Baron-Cohen, 2001).

Autism is not the only condition that has been associated with abnormal precision weighting. The aberrant salience hypothesis of schizophrenia proposes that altered attribution of salience to sensory stimuli may underwrite perceptual and attentional changes in psychosis (Kapur, 2003). Aberrant attribution of salience may be exacerbated by deficits in context processing (J. D. Cohen, Barch, Carter, & Servan-

Schreiber, 1999), but there is thought to be much more impairment in the control of attention (i.e. feature selection) in schizophrenia than in the subsequent inference using those features (Luck & Gold, 2008). Indeed, whilst subjects with schizophrenia may be unimpaired or even show enhanced performance in simple attentional cueing tasks (Spencer et al., 2011), in more complex tasks, such as viewing natural images, they consistently fixate less on informative areas (Beedie et al., 2011). The model I have presented here illustrates a potential computational mechanism behind the intimate connection between attentional control and the efficient sampling of information.

6. Patch leaving paradigm and active inference

Our everyday lives present us with different paths that lead to different outcomes. When choosing among alternative courses of action, we take into account the overall reward we are likely to get, if we were to follow a certain path – and the time it would take to obtain the reward. While some of us care more about long term goals, others have a tendency to act for immediate gratification; even when the latter is less beneficial in the long run (Logue, 1995; Strotz, 1955). This sort of behaviour can be characterized as impulsive. More precisely, impulsive behaviour can be operationally defined as seeking proximal rewards over distal rewards. A common theme in many impulsivity scales (S. B. Eysenck & Eysenck, 1978; Patton, Stanford, & Barratt, 1995; Whiteside & Lynam, 2001) is a failure to plan ahead. In this chapter I show that at least three different factors can lead to impulsive behaviour. To show this formally, I use a Markov decision process formulation of active inference in a patch leaving paradigm.

In the patch-leaving paradigm (Charnov, 1976; Gibb, 1958; MacArthur & Pianka, 1966) the problem is deciding when to leave an environment with exhaustible resources. In my version of this task, there are several patches with unique reward-probability decay rates. Although a general notion, I can make it more intuitive with an example. A patch can be thought of as a bag of chocolates and stones, where chocolate is a rewarding and stone is a non-rewarding outcome. One can successively draw single items from the bag. Crucially, there is a hole at the bottom of the bag and the chocolates are falling from the bag faster than the stones. This means that the probability of drawing a chocolate decreases with time. At each time point one is presented with the choices *stay* and *leave*. Choosing to *stay* entails drawing a chocolate from the same bag that one has been foraging in. Choosing to *leave* entails moving onto a new bag that might have more chocolates. However, *leaving* has a cost – and the cost (i.e. switching penalty) is to forfeit attempts at drawing a chocolate for the time taken to find the new bag. The new bag can be a new kind of bag or the same kind of bag as the previous. Crucially, the holes at the bottom of each kind of bag have different sizes. This means that the chocolates are dropping from each kind of bag with a different rate. This task requires one to decide when to leave a patch to maximize reward. In this task, I equate staying longer in a patch (compared to a simulated reference subject) with more impulsive behaviour. Intuitively, a greater emphasis on proximal outcomes means a

greater reluctance to accept the switching penalty compared to accepting a small probability of immediate reward.

In the next section, I describe the Markov decision process used to define the patch leaving paradigm. Through simulation I illustrate the different deficits that can lead to impulsive behaviour. This illustration entails manipulating how deeply a synthetic subject looks into the future (expressed in terms of her *policy depth*), its capacity to maintain and process sequential information (expressed in terms of the *precision* of beliefs about transitions) and how much immediate rewards and penalties are discounted compared to distant ones (expressed in terms of a *discount slope* of preferences over time). These manipulations will be unpacked in subsequent sections and their effects on the simulated responses will be compared to a MDP model that serves as a point of reference (a *canonical* model).

This chapter comprises three sections. The first section describes a parameter of the MDP model, namely policy depth. I will explain how this parameter affects variational message passing. I will also explain the simulated electrophysiological responses as a consequence of this message passing scheme in greater detail. The second section describes a Markov Decision Process formulation of active inference for the patch leaving task. The third section illustrates how manipulating the three components of the MDP model (see above paragraph) can produce impulsive behaviours. These manipulations will underline the prior beliefs that can lead to impulsive behaviours. I present the associated (simulated) electrophysiological responses and how these responses change with the above manipulations. I conclude with a brief discussion.

6.1. Variational message passing and policy depth

The belief update equations introduced in Fig 2.4 shows that a policy is more likely if it minimises the expected free energy $\mathbf{G}$ (see policy evaluation in Fig 2.4). Computing the variational free energy, under competing policies, requires an agent to have expectations about the past and future states of the world. Optimising these (posterior) expectations entails minimising the variational free energy under a policy, given the current observations. These posterior expectations are then projected into the future to obtain the expected states (and outcomes). How far into the future the posterior expectations are projected depends on the *policy depth*.

The *policy depth* (shown with the subscript PD in $s^{\pi}_{\tau+PD}$ in the lower half of Fig 6.1) determines how many epochs beliefs about hidden states are projected into the future. An important feature of this scheme is that a synthetic subject holds beliefs about 'epochs' in both the past and the future. This means that there are two sorts of times. The first is the actual time that progresses as the subject samples new observations. The second (epoch) time is referenced to the onset of a trial and can be in the past or future, depending on the actual time. Posterior expectations about the hidden states of the world can change as the actual time progresses and are projected to both future and past epochs. In this (variational message passing) scheme, it is assumed that beliefs at the current epoch are projected: i) back in time to all epochs from the current epoch to the initial epoch, and ii) forward in time (to form future beliefs) to a number of epochs corresponding to the *policy depth*.

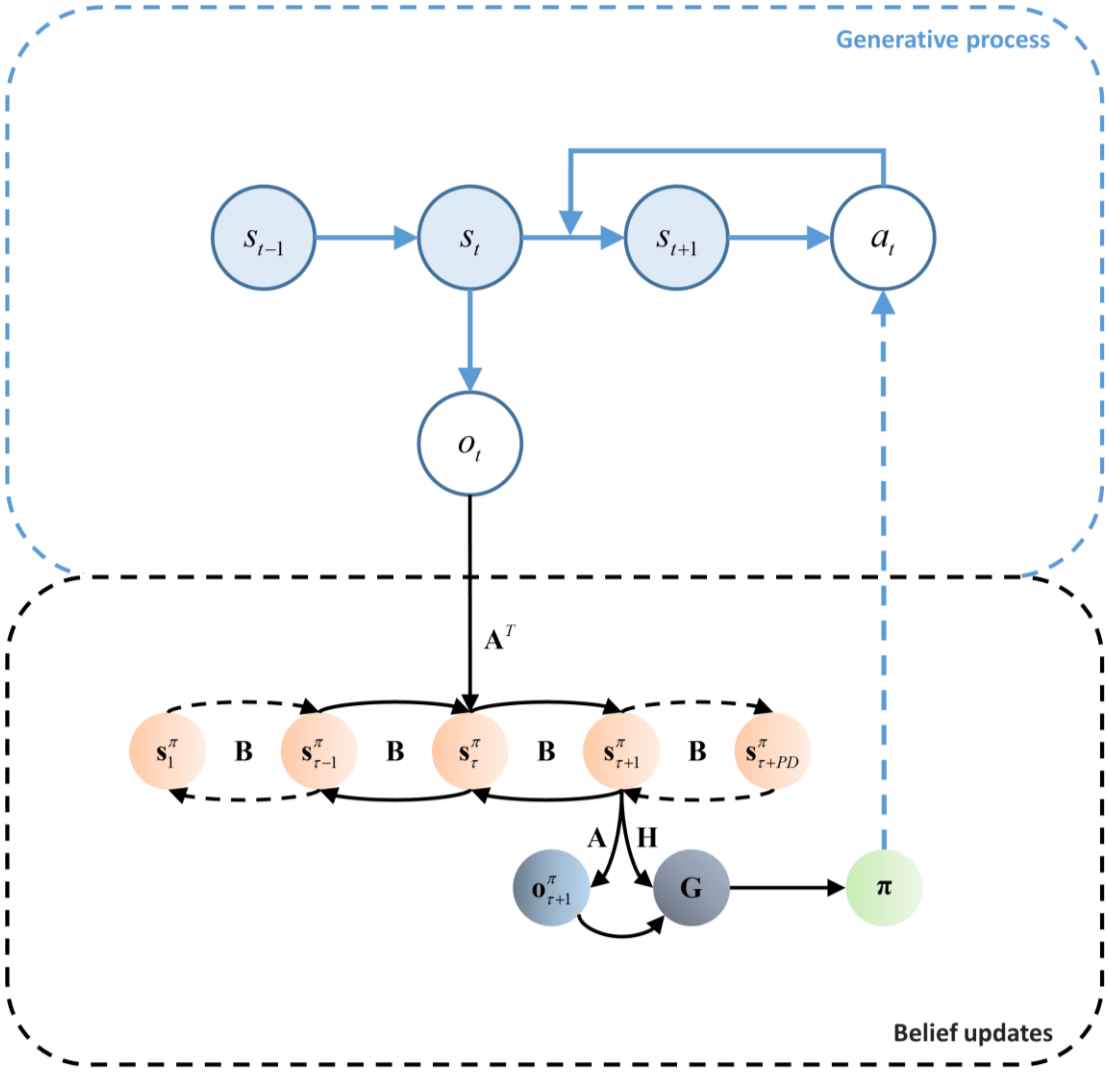


Figure 6. 1 Generative process and belief updates

The upper half of this panel shows the generative process. This process specifies that the hidden state of the world in the current epoch (s_t) depends on the hidden state in the previous epoch (s_{t-1}) and the action (a_t). The hidden state in the current epoch then produces a new observation (o_t). The lower half of this panel shows the Bayesian belief updates (variational message passing). The new observations are used to infer the most likely causes ($\boldsymbol{s}_\tau$) of the observations. The beliefs about the hidden states ($\boldsymbol{s}_\tau$) are then projected backwards ($\boldsymbol{s}_{\tau-1},\ldots,\boldsymbol{s}_1$) and forwards ($\boldsymbol{s}_{\tau+1},\ldots,\boldsymbol{s}_{\tau+PD}$) in time. Here, PD is a variable that specifies how far into the future these beliefs should be projected. This term will be used later in the simulations. The expected hidden states in the future ($\boldsymbol{s}_{\tau+1},\ldots,\boldsymbol{s}_{\tau+PD}$) are used to specify expected observations in the future ($\boldsymbol{o}_{\tau+1},\ldots,\boldsymbol{o}_{\tau+PD}$). Only $\boldsymbol{s}_{\tau+1}$ and $\boldsymbol{o}_{\tau+1}$ are shown for simplicity. Then these expectations are used along with the entropy of the likelihood matrix ($\boldsymbol{H}$) to compute the (path integral of) expected free energy ($\boldsymbol{G}$) under all policies. A softmax function of expected free energies under all policies provides the posterior distribution over policies. Finally, an action is sampled from the posterior distribution over the policies. The conditional dependencies in the generative process are shown with blue arrows, whereas the message passing – implementing belief updates – is shown with black arrows.

The ensuing belief updates are used to mimic electrophysiological responses obtained in empirical studies. I have previously used a similar approach to simulate electrophysiological responses during a scene construction task (see Fig 3.5). I will now explain these responses greater in detail. Fig 6.2 shows how beliefs about hidden states change at different epochs as new observations are made, and how these beliefs are passed to other epochs. The actual time that progresses as new observations are made is shown on the x-axis. After each observation, expectations about the hidden states are optimised. In this case, there are four hidden states. Each set of four units on the y-axis corresponds to expectations about these hidden states on different epochs (e.g. 1st, 5th, 9th and 13th rows show the expectations about the first hidden state in epochs one to four). Expectations about hidden states in each epoch are updated as new observations are made. In Fig 6.2, the current time is shown on the diagonal (with red squares), whereas the past and future epochs are shown above and below the diagonal, respectively. In this example the *policy depth* is one, which means that expectations about hidden states at the current time are projected one epoch into the future (i.e. there is only one epoch represented below the diagonal in

each column). This shows that beliefs about hidden states reach one epoch into the future.

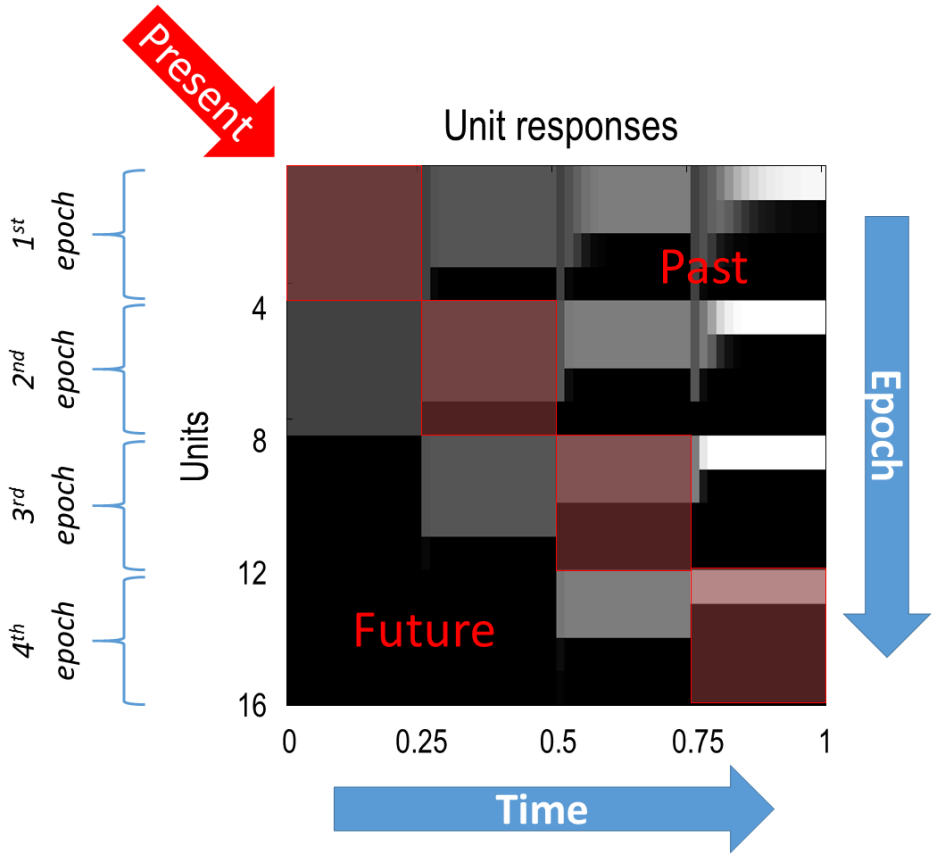


Figure 6. 2 Variational message passing

This figure shows how the expectations about hidden states are optimised at the current time and projected to (past and future) epochs. The actual time – that progresses as new observations are made – is shown on the x-axis. Epochs occupy a fixed time frame of reference and are shown along the y-axis. In this example, there are four hidden states that repeat over epochs on the y-axis. This figure shows that expectations about hidden states at the present time (shown on the diagonal in blue squares) are projected backwards to the past (above diagonal) and forwards into the future (below the diagonal) epochs.

The panels in Fig 6.3 shows the how far the beliefs are projected into the future when different *policy depths* are used. From left to right the *policy depths* are two, four and six. One can see that the number of epochs current beliefs are projected to, is two, four and six from left to right, respectively. Later I will show how the *policy depth* changes the simulated electrophysiological responses mentioned above – and can have a substantial effect on policy evaluation and subsequent choice behaviour.

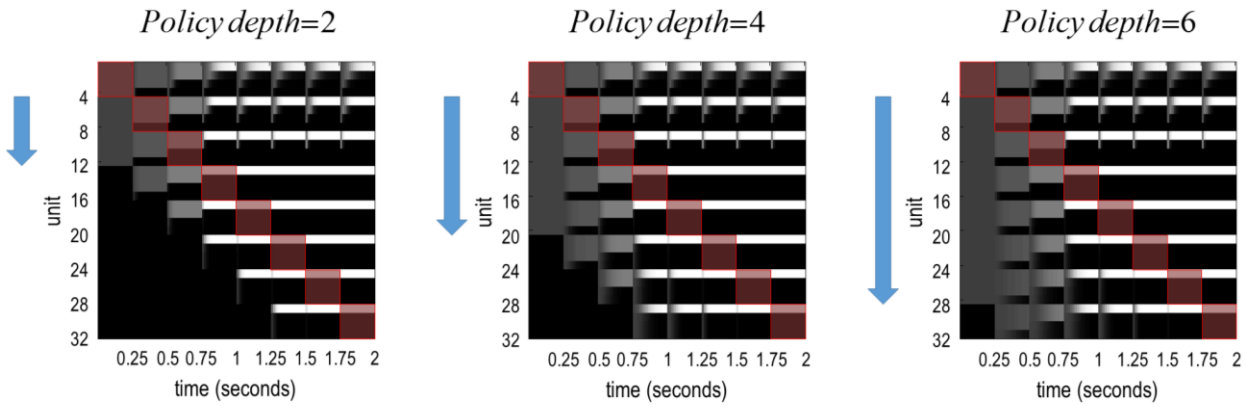


Figure 6. 3 Varying the depth of policy

These panels show that using different *policy depths* project expectations about hidden states to *n* number of epochs in the future, where *n* is chosen as 2, 4 and 6 from left to right, respectively.

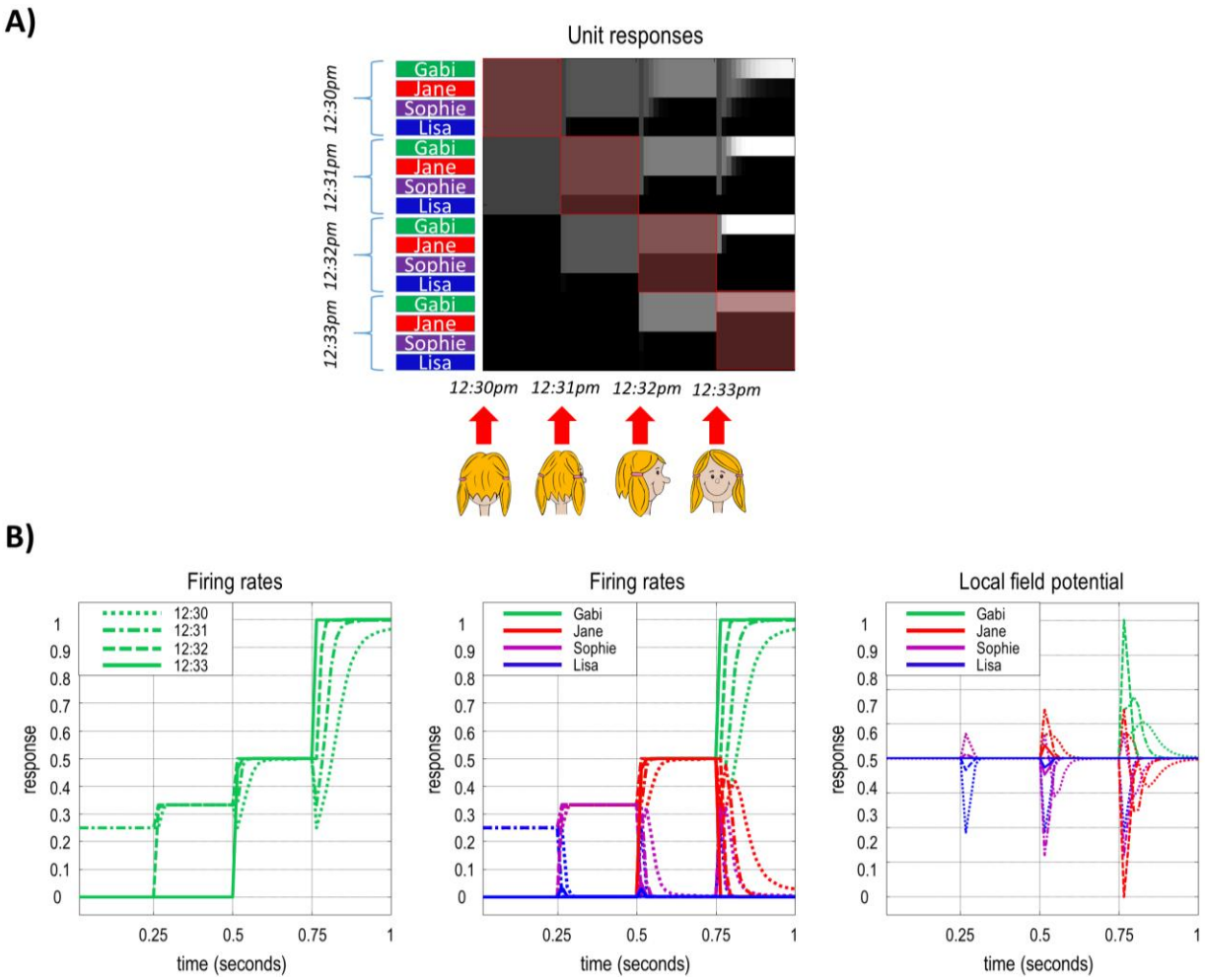


Figure 6. 4 Simulated electrophysiological responses

A) This panel shows the variational message passing in the context of identifying someone by accumulating evidence in a sequential manner across different epochs. Here, one sees someone that resembles one of four people at 12:30pm. These four identities are Gabi, Jane, Sophie and Lisa. Over time the identity is disclosed as one gets a better view of the person. Finally at 12:33pm the person that was seen is identified as Gabi. In this example the *policy depth* is one. This means that expectations about hidden states are projected one epoch into the future. **B)** The left panel shows the expectations of hidden state that encodes the identity of Gabi over different epochs, using curves rather than using a raster plot (as shown in the panels above). These epochs correspond to 12:30, 12:31, 12:32 and 12:33pm. The middle panel shows this for all possible identities. Each colour in the legend corresponds to the identity of each person in this panel. The right panel shows the local field potentials, defined in terms of rate of change of expectations about hidden states; i.e., the gradient of each curve in the middle panel.

To gain further intuition about this way of how sequences of states and actions might be modelled, consider the example in Fig 6.4A. Assume that you are walking behind someone that you think you recognise. At 12:30pm you can only see this person from behind – and she resembles one of four people you know; e.g., Gabi, Jane, Sophie, Lisa. These identities are the four hidden states in this case. At 12:31pm you get closer and now you are sure that she is not Lisa. At 12:32 you catch up and see her from the side. Now you are convinced this person is not Sophie either. At 12:33pm you finally see the person's face and you recognise her as Gabi. This resolves all uncertainty over the identity of the person. The belief that the person you see at 12:33pm is projected backwards in time to 12:30pm – this can be seen clearly in the final column. Intuitively, at 12:33 you know that the person you saw at 12:30pm was Gabi.

The left panel of Fig 6.4B shows the same expectations about the hidden states that encode Gabi's identity as in Fig 6.4A over all epochs (see the 1st, 5th, 9th and 13th rows in Fig 6.4A). The figure in the middle panel of Fig 6.4B shows the same as the left panel but for each identity. The right panel of Fig 6.4B shows the simulated local field potentials in terms of the rate of change in the expectations about the hidden states (shown in the middle panel of Fig 6.4B). See section 3.3.1 for a brief discussion on what type of cells might be involved in generating these responses.

6.2. MDP model of the patch leaving paradigm

This section describes a Markov decision process model of active inference for the patch leaving paradigm. The model is used to simulate behavioural responses (i.e. choosing to *stay* or *leave*) when the reward probability in a patch declines exponentially as one stays in a patch. In this paradigm, there are several patches with their own unique reward probability decay rates. Choosing to *leave* a patch warrants one epoch to be spent in a reward-free state (i.e., a switch state). In the next epoch, one enters a patch randomly and all reward probabilities reset to their initial values. This means one needs to consider how many epochs to spend in a patch before leaving to realise prior preferences; i.e., being rewarded as much as possible.

In this MDP (see Fig 6.5), I considered two dimensions of hidden states, namely *where* and *when*. The first hidden dimension, *where*, corresponds to the *patch identity*. There are four hidden states under this dimension, namely *patch 1*, *patch 2*, *patch 3* and a *switch* state. Under the action *stay,* the *where* state does not change unless it is in the *switch* state. Under the action *leave,* the *where* state changes to the *switch* state, except for the *switch* state itself. Under both *stay* and *leave* the *switch* state transitions to one of the first three patches with equal probabilities. The second hidden state dimension, *when*, keeps track of the number of time-steps since a *switch* state. The *time since a switch* state is represented by t_s. This state t_s increases by one up to a maximum of four. The hidden state associated with the *fourth epoch since a switch state* $t_s = 4$ is an absorbing state and does not change over subsequent epochs. The reward probability in a given patch declines with t_s and does not change after $t_s = 4$, even if one chooses to *stay* after the fourth epoch, i.e. reward probability under a patch is the same for $t_s > 4$ as $t_s = 4$. Choosing to *leave* at any point in time resets t_s to one; i.e., $t_s = 1$.

There are two outcome modalities. The first modality signals the *feedback* (*reward* or *no reward*). The probability of reward declines exponentially under all patches as t_s increases (up to a maximum of four). There are three different patches with unique rates of decline in reward probability. The rate at which the reward probability declines under the first patch $\exp((1-t_s)/16)$ is slower than the second $\exp((1-t_s)/8)$ and the

third $\exp((1-t_s)/4)$ patches, where $t_s \in \{1,2,3,4\}$ respectively. The reward probabilities under different patches are shown on the left panel in Fig 6.6A. The second outcome modality, *where*, signals the patch identity. Notice that the patch identity (*where*) appears both as an outcome and as a hidden state. This is because *where* (patch identity) as an outcome is used to inform the agent about the *where* hidden state.

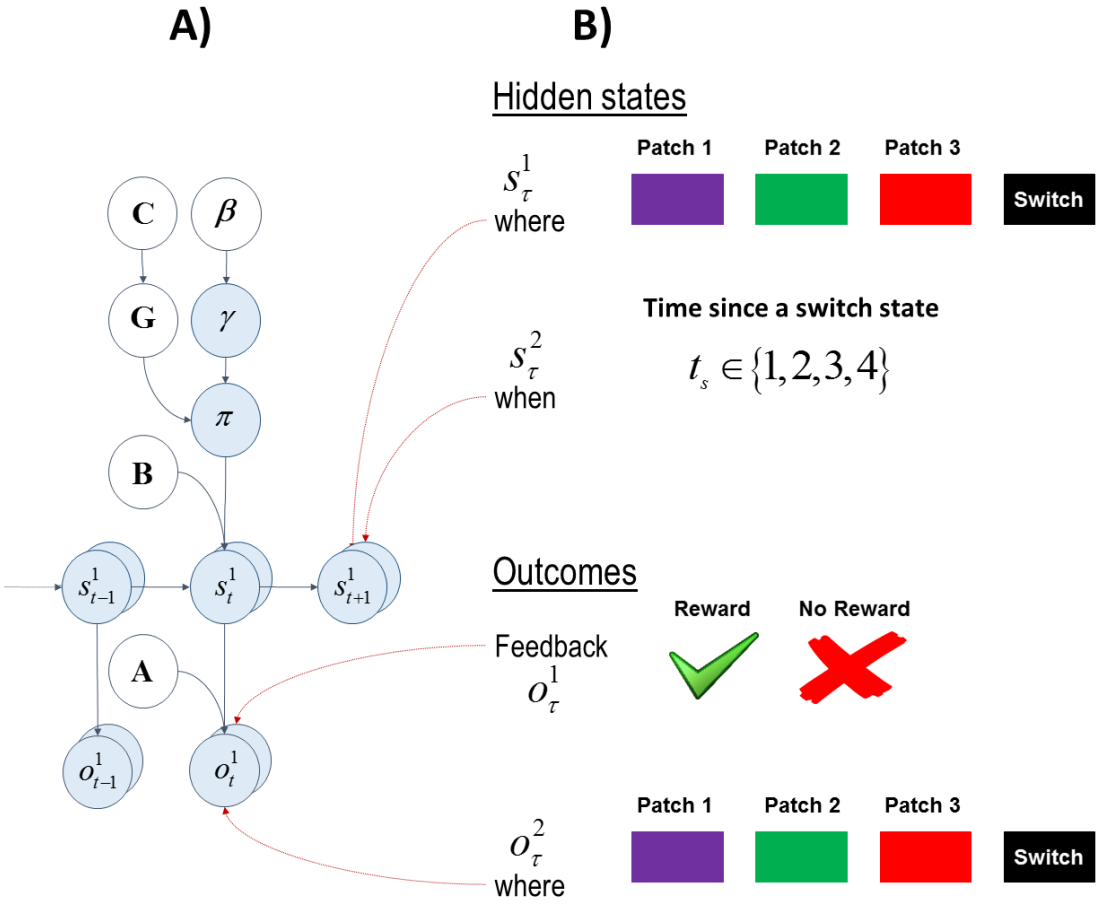


Figure 6. 5 Graphical representation of the generative model

A) This panel shows the graphical representation of the MDP model and the conditional dependencies among the terms in the model. See Fig 3.1A for details. **B)** The right panels show different sets of hidden states and outcome modalities in the patch leaving task. There are two sets of hidden states, namely the *patch identity* and the *time since a switch state* t_s (*where* and *when* respectively). There are two outcome modalities, namely the *feedback* and *where*. The *feedback* modality signals whether an agent receives a reward or not, whereas the *where* modality signals on which patch an agent is in.

In this MDP scheme, I consider prior preferences over only the *feedback* modality such that the agent expects *reward* (utility or relative log probability of 2 nats) more than *no reward* (utility of -2 nats). I defined no prior preferences over the *where* modality, which

means that there were no preferences over patch identity. See Fig 6.6 for the likelihood, transition and prior preference matrices provide a complete specification of this patch leaving paradigm.

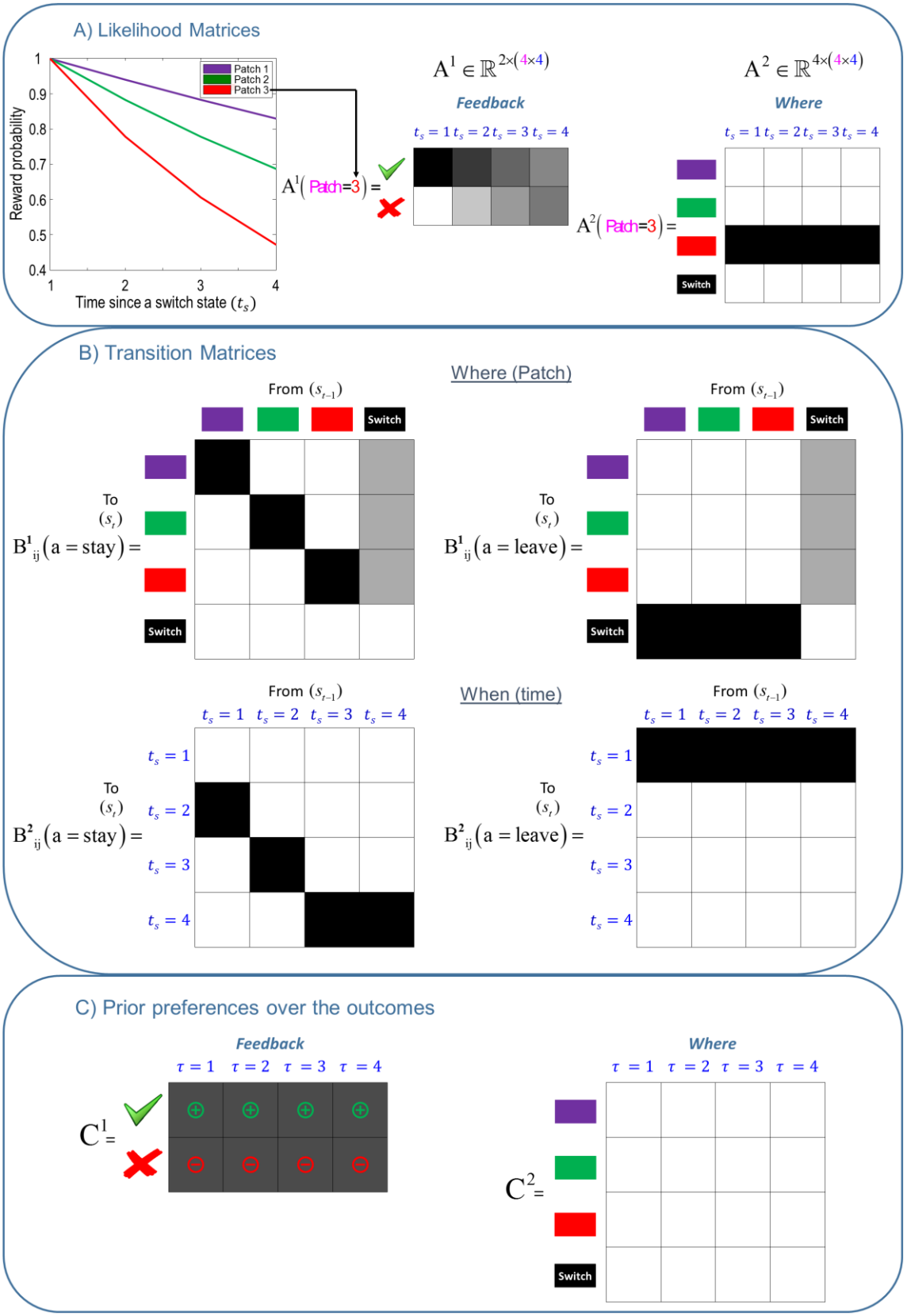


Figure 6. 6 ABC of generative model

A) The panel on the left shows how the reward probability decreases in different patches as a function of *time since a switch state* t_s. The subsequent two panels show the likelihood (**A**) matrices. The likelihood matrices specify the probability of outcomes given two sets of hidden states, namely *where* (the patch the agent is in, shown with magenta colour) and *when* (t_s is shown with blue colour). Here the likelihood matrices are shown for patch 3 (shown with red colour) as a function of *when* hidden state t_s. The first likelihood matrix A^1 shows that the probability of reward (shown with green tick) decreases as $t_s$increases. The second likelihood matrix A^2 signals the patch the agent is in (in this case patch 3) with respect to the *when* hidden state. **B)** This panel shows the transition matrices for *where* and *when* (t_s) dimensions of hidden states. The state transitions depend on the actions. The first (*where*) transition matrix, shows that under the action *stay* $B^1(a = \text{stay})$, the agent stays in the same patch it is in currently; except when the agent is in the *switch state*. Under the action *leave* $B^1(a = \text{leave})$, the agent enters the *switch* state, given that the agent is not in the *switch* state. The probability of entering one of the three patches is equally likely when the agent takes the actions *stay* or *leave* given it is in the *switch* state. The second (*when*) transition matrix under a *stay* $B^2(a = \text{stay})$ increases by one – up to a maximum of four. The fourth epoch is an absorbing state – and an agent would have to take the action *leave* to leave this state. Under a *leave* $B^2(a = \text{leave})$, t_s is reset to one, i.e. $t_s = 1$. **C)** This panel shows the prior preferences over outcomes as a function of time (relative to the current time). I only define a prior preference over *reward* and *no reward* outcomes, under the feedback modality and do not define any preference over the patches (*where* modality). Plus and minus signs show the valence of the utilities; whereas different shades of grey indicate their magnitude. The model described in this figure is the *canonical* model. The *policy depth* in this model is chosen as 4.

6.3. Simulated responses

6.3.1. Simulating impulsivity

Impulsivity can be characterised as a tendency to act to require immediate rewards, rather than planning to secure rewards in the long run. In the patch leaving paradigm, one is always presented with the choices *stay* and *leave*. The experimental design for this paradigm is such that it requires one to spend one epoch in a reward-free *switch* state upon leaving a patch (i.e., switching penalty). However, staying in a patch always has the prospect of reward. Acting on the proximal reward requires one to choose

stay, whereas acting on the distal reward requires one to choose *leave* at some point. Here, I operationally define *impulsivity* as staying longer in a patch because only *stay* has the prospect of an immediate – if less likely – reward. This raises the question 'longer than what?' To address this, I introduce an agent who serves as a reference or *canonical* model.

In this section, I show how impulsive behaviour can be underwritten by changes in prior beliefs about the different aspects of the MDP model. For this purpose, I use the MDP described in Fig 6.6 to as a *canonical* model. The simulated responses obtained under the *canonical* model will be compared with the models that deviate from this reference, in terms of the *policy depth*, *precision* of the transition matrices and the *discount slope* of the prior preferences over time (i.e., time discounted reward sensitivity). These models will be compared with the *canonical* model in terms of *dwell times*. *Dwell time* is the average time spent in a patch upon entering it. The models that induce an agent to stay longer than the *canonical* model are considered to exhibit impulsive behaviour. These models are as follows:

- **Varying the *policy depth*.** The *policy depth* of the canonical model is four. This model is compared with the models where the *policy depth* is varied over three levels; namely, $PD = 3$ (deep policy), $PD = 2$ (intermediate policy) and $PD = 1$ (shallow policy) models. See Fig 6.7 for a comparison between the canonical model and the models above. The policy depth for all remaining models was $PD = 4$.

- **Varying the *precision* of the transition matrices.** Here, the *precisions* of state transitions were rendered less precise. In other words, I modelled a loss of confidence in beliefs about the future. Operationally, this is implemented by multiplying the columns of the (log) transition matrices (shown on Fig 6.6B) with a constant, $b_{ij} = \omega \ln \mathbf{B}_{ij}$ and then applying a softmax function. This ensures each column corresponds to a probability distribution, $\mathbf{B}_{ij} = e^{b_{ij}} \Big/ \sum_k e^{b_{kj}}$. The precision, also known as an inverse temperature, was varied over three levels: $\omega = 16$ (*high precision*), $\omega = 8$ (*medium precision*) and $\omega = 0$ (*low precision*). The lower the precision, the more uniform the distributions over state transitions

become from any given state. This manipulation is only applied to the transition matrices in the generative model (i.e. the subject's beliefs about transitions) and not to the generative process (that actually generates the data presented to the subject). See Fig 6.8 for the difference between an example transition matrix with a low precision. In this figure, although only one transition matrix is shown (transition matrix for *where* under the action *stay*), the precision of all transition matrices under all actions are subject to the same manipulation. The *precision* was $\omega \gg 16$ in all other models.

- **Varying the *discount slope*.** In this model, the prior preferences over outcomes are equal to the prior preferences in the canonical model on average. In the canonical model, the utilities for *reward* and *no reward* are fixed at 2 and -2, respectively. These utilities are not discounted as the agent plans into the future. However, in models where I manipulate the slope of prior preferences, they change in the following way:

  $C_{reward}(\tau) = 2 + slope \times x(\tau)$ and $C_{No\ reward}(\tau) = -2 - slope \times x(\tau)$,

  where $x = [2.25, 0.75, -0.75, -2.25]$ and $\tau \in \{1, 2, 3, 4\}$. Here τ represents the future epochs; e.g. $\tau = 1$ means 1 epoch in the future. These equations show that the agent discounts utilities as it plans into the future. The term *slope* took the following values: 0.75 (*high slope*), 0.5, (*medium slope*), or 0.25 (*low slope*). Manipulating the slope makes the utility of *reward* in the near future appear larger (and *no reward* smaller), and the opposite effect for the distant future. This means that proximal rewards will always be regarded as more valuable and distal rewards as less valuable, compared to the canonical model (this comparison is illustrated in Fig 6.9). The *slope* term was $slope = 0$ in all other models.

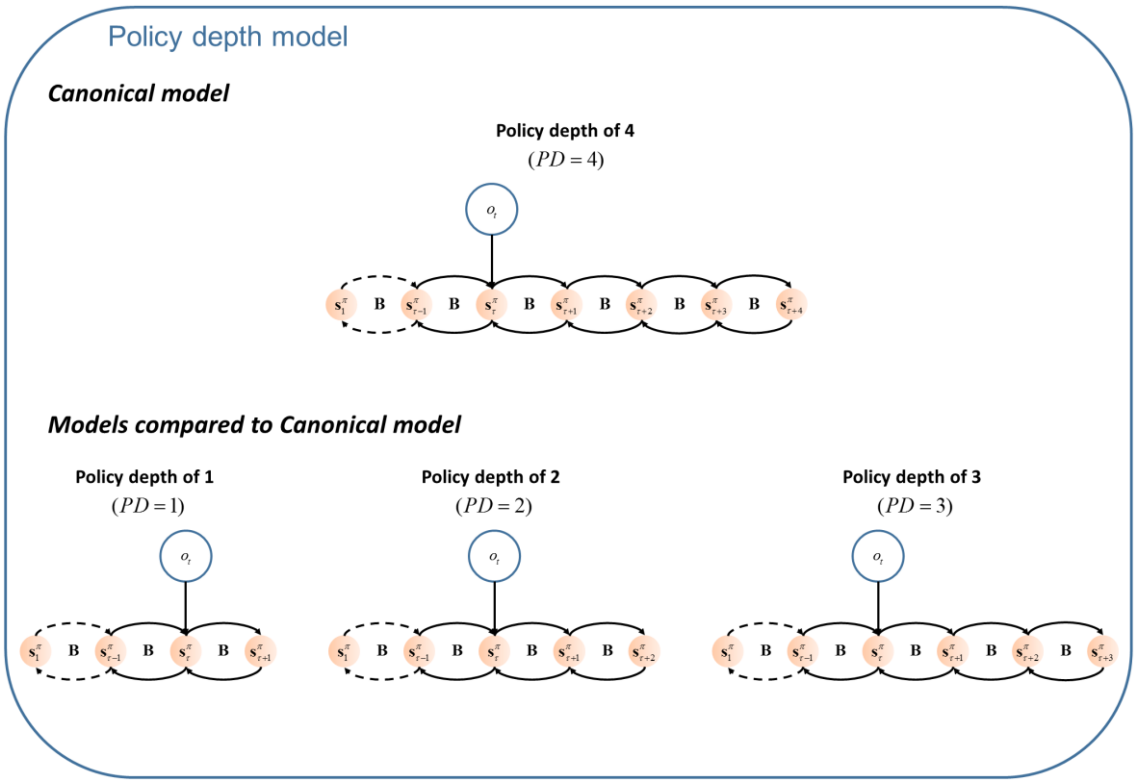


Figure 6. 7 Varying policy depth

This figure shows the difference between the canonical model and the model in which the *policy depth* is changed. The *policy depth* in the canonical model is four. The *policy depths* in the models that are compared with the canonical are one, two and three.

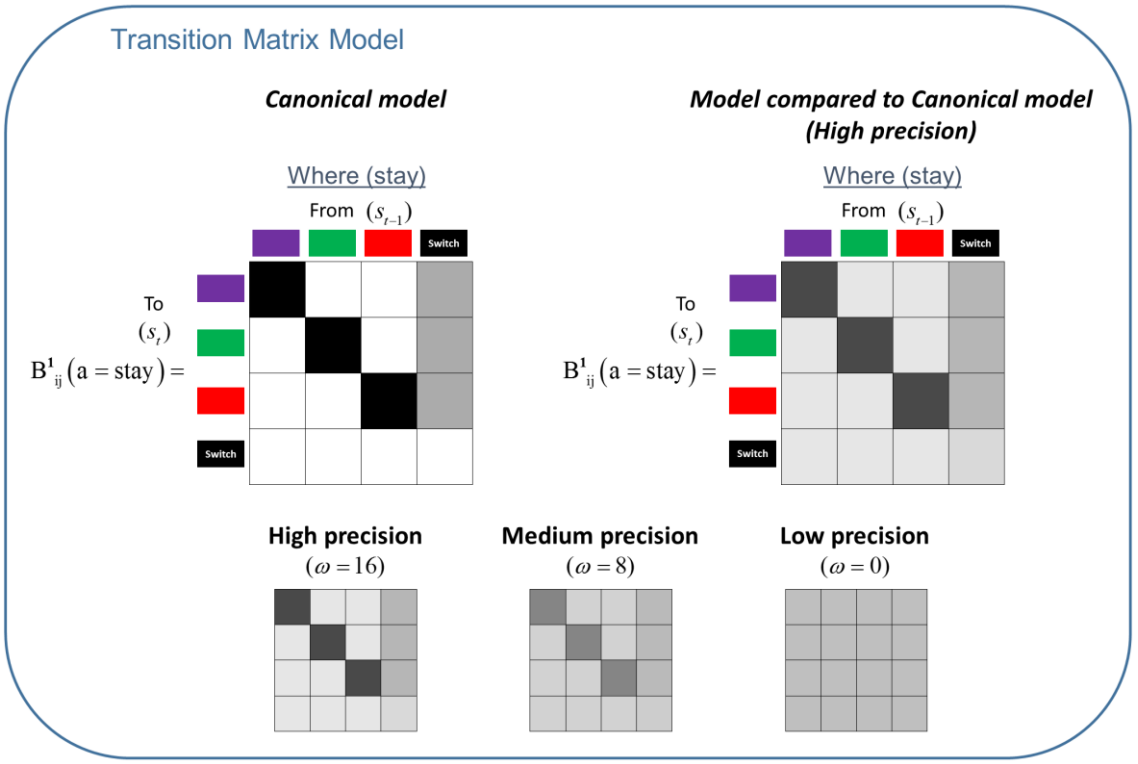


Figure 6. 8 Varying the precision of the transition matrices

This figure shows the difference between the canonical model and the model in which the *precision* of transition matrices is changed. For illustrative purposes only the transition matrix for *where* under the action *stay* is used in this panel; however, the changes are applied to all transition matrices under all actions. The precision of the transition matrices are changed over three levels. These are high, medium and low levels of precisions. The higher the precision the more similar the transition matrices approach those of the canonical model. With lower precisions the uncertainty in the probability distributions over the columns of the transition matrices increases.

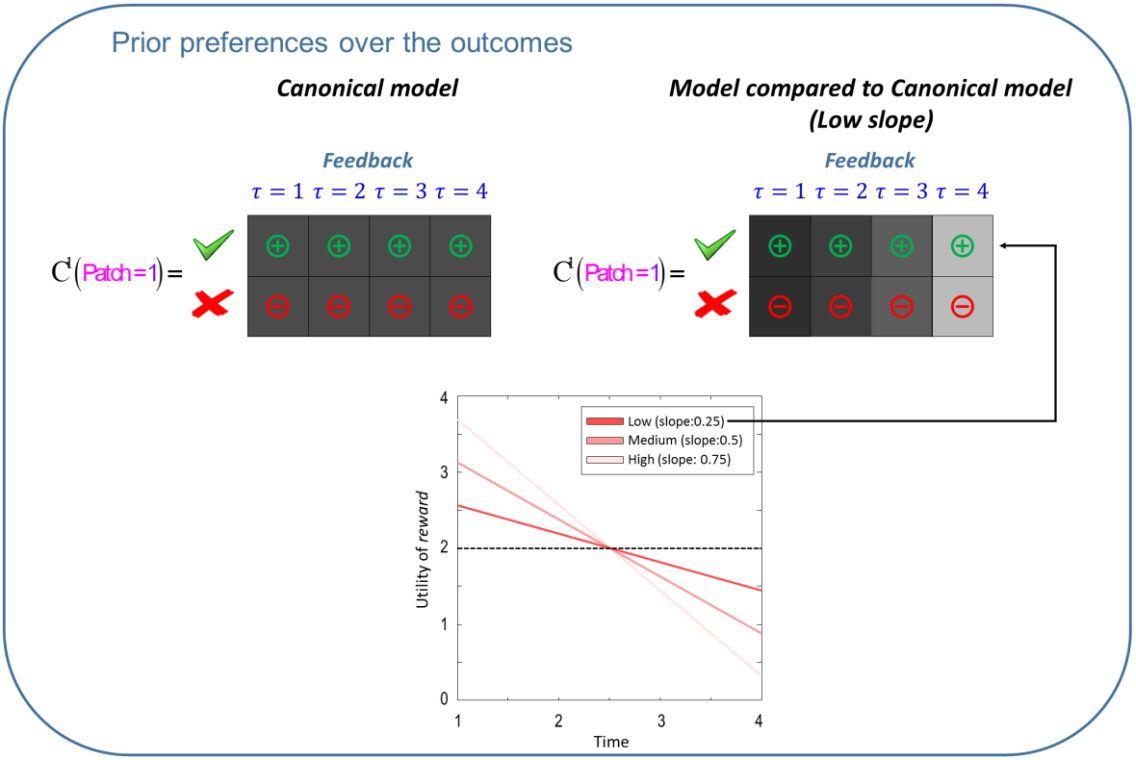


Figure 6. 9 Varying the discount slope

This figure shows the difference between the canonical model and the model in which the *discount slope* is changed. In the canonical model, the prior preferences over a *reward* and *no reward* are fixed at 2 and -2 (i.e., they are not time sensitive). However, the model in which the *discount slope* is changed is subject to the following equation $C_{reward}(\tau) = 2 + slope \times x(\tau)$ and $C_{No\ reward}(\tau) = -2 - slope \times x(\tau)$, where $x = [2.25, 0.75, -0.75, -2.25]$ and $\tau \in \{1, 2, 3, 4\}$. Here τ represents the future the box, e.g. $\tau = 1$ means 1 epoch into the future. The intercepts of these equations are set to the prior preferences over *reward* (and *no reward*) in the canonical model, which is 2 (and -2). The *slope* term endows prior preferences with time sensitivity, when planning future actions. The *slope* is changed over three levels; namely, high slope (0.75), medium slope (0.5) and low slope (0.25). The lower panel shows how the utility of *reward* changes over future epochs with

different slopes. The utility of *no reward* (under different slopes) is just a mirrored version of this figure (since the utility of *no reward* is negative). With these equations the agent discounts the utility of *reward* and *no reward* outcomes as it plans further into the future.

Comparing the simulated behaviour of the canonical model and the above models shows that all manipulations resulted in longer *dwell times*. In other words, all of the above manipulations induced more impulsive, short-term, behaviour; in which synthetic subjects found it difficult to forego the opportunity for an immediate reward – and overcome the switching cost of moving to a new patch. The bar plots in Fig 6.10 show the increase in *dwell times* under the three models (over three different levels of each model) compared to the canonical model. The average increase in *dwell times* over all patches is shown on the top panel of Fig 6.10. The three panels below show the same results for each patch separately.

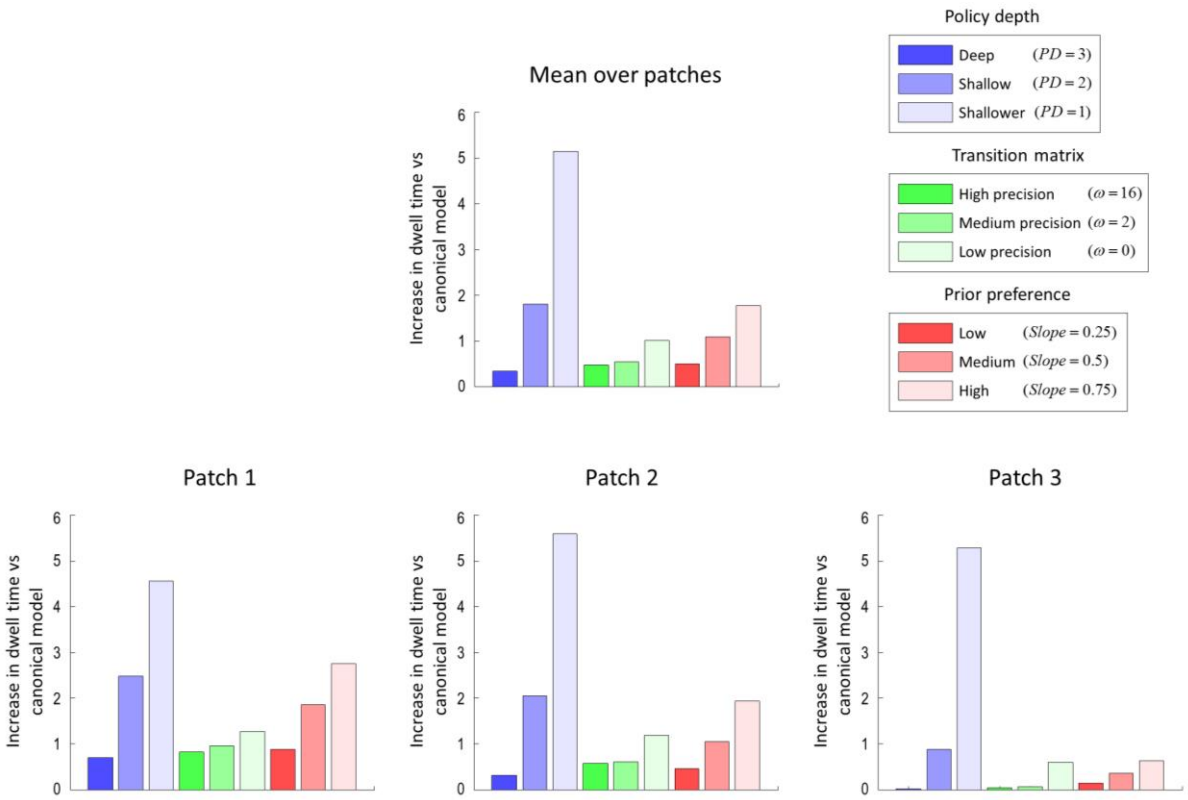


Figure 6. 10 Average time spent in patches under different models

This figure show the increase in *dwell time* under the alternative models, compared with the canonical model. In the alternative models, the *policy depth*, *precision* of transition matrices and the *slope* of prior preference matrices are changed (over three levels) with respect to the canonical model. The *policy depth* in the canonical model is chosen as four. In the models compared to the canonical model the *policy depth* is varied over three levels, namely deep ($PD = 3$), intermediate ($PD = 2$) and shallow (

$PD = 1$), respectively. The *precision* of the transition matrices is varied over three levels, namely high ($\omega = 16$), medium ($\omega = 8$) and low ($\omega = 0$). The *discount slope* are changed over three levels, namely high ($slope = 0.75$), medium ($slope = 0.5$), low ($slope = 0.25$).The top panel shows the increase in *dwell time*, averaged over patches; whereas the below three panels show the increase in *dwell times* in each patch separately. The panels of this figure show that manipulating the *policy depth*, *precision* of the transition matrices and the *discount slope* all cause the *dwell time* to increase.

The *policy depth*, *precision* of the transition matrices and the *slope* of the prior preference matrix have similar kinds of effects on *dwell times*. With deeper policies, the agent leaves the patches earlier in order to exploit the distal rewards. With shallow policies the agent stays longer in the patches and exploits proximal rewards (see blue bars in Fig 6.10). With less precise transition matrices the agent remains longer in any patch. This is because imprecise transition matrices mean that the further one looks ahead, the less precise one's beliefs become and the future becomes uncertain. These beliefs are about both *where* (which patch) and *when* t_s the agent is. With uncertainty over *where* and *when*, the agent prefers proximal rewards, rather than risking leaving a patch for an uncertain outcome (see green bars in Fig 6.10). With more time sensitive prior preferences, the agent discounts the utility of *reward* more steeply over time. This means that the agent prefers proximal *rewards*, however unlikely they may be, over distal *rewards*: hence, the agent stays longer in each patch to exploit rewards in the near future (see the red bars in Fig 6.10).

In the following, I ask whether the different models examined above can be distinguished by observing their choice behaviour. This entails fitting models to the simulated choice behaviour and using the resulting Bayesian model evidence to perform Bayesian model selection (assuming uniform priors over models). (K. Friston, Mattout, Trujillo-Barreto, Ashburner, & Penny, 2007; Mirza et al., 2018; Schwartenbeck & Friston, 2016). The models that were used to generate (synthetic) behavioural data were the above models, in which the *policy depth*, *precision* of the transition matrices and the *discount slope* varied over 3 levels (see Fig 6.7, Fig 6.8 and Fig 6.9) and the *canonical* model (10 models in total). These 10 models were then fit to the data generated with each model to create a confusion matrix of model evidences (i.e., the probability that any one model was evidenced by the data from itself or another). The posterior distributions over the models suggest that these models can indeed be disambiguated in terms of their Bayesian model evidence (see

Fig 6.11). This shows that although the resulting behaviour under these models looks similar – namely, staying longer in patches (greater *dwell times*) – subtle differences in choice behaviour can still inform model comparison.

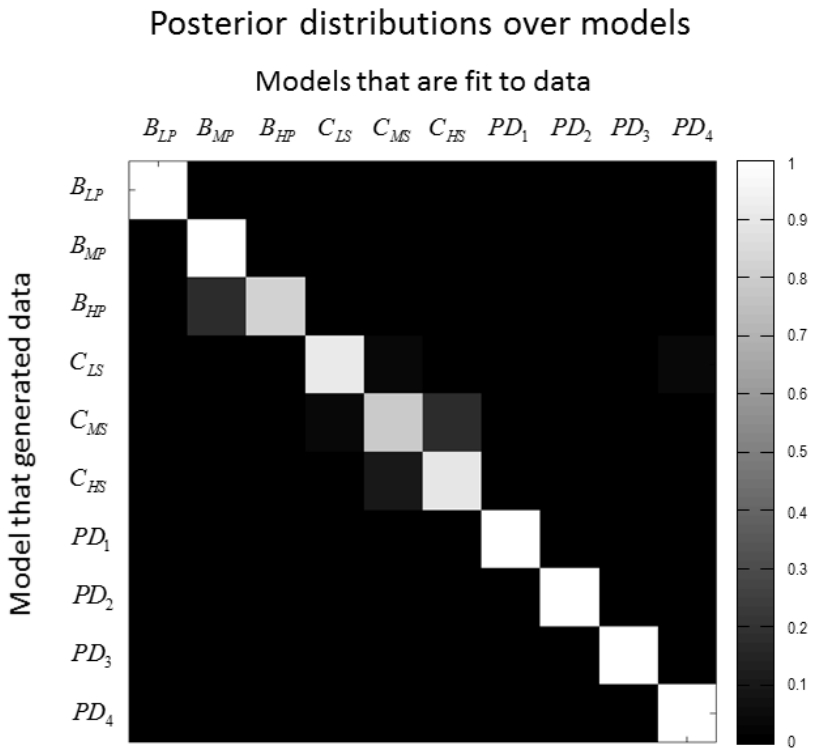


Figure 6. 11 Bayesian model comparison

This figure shows the posterior distribution over models, when these models are fit to data generated by the same models. The simulated data are generated with the models on the y-axis. The models shown on the top are fit to the data to estimate the log-evidence for each model. These simulations show that these models considered (see previous figures) can be distinguished in terms of their model evidence. In this figure B_{LP} , B_{MP} and B_{HP} correspond to low, medium and high precision transition matrices, respectively. C_{LS} , C_{MS} and C_{HS} correspond to low, medium and high slopes over the prior preferences, respectively. PD_1 , PD_2 and PD_3 correspond to *policy depths* 1, 2 and 3, respectively. The *canonical* model PD_4 is included in these simulations.

In summary, I have shown distinct differences in the form and nature of prior beliefs that underlie generative models of active inference can all lead to impulsive behaviour. In the next section I will simulate and characterise the electrophysiological responses I would expect to observe under these distinct causes of impulsivity.

6.3.2. Simulated electrophysiological responses

In this section, I show how simulated electrophysiological responses vary with the *policy depth*, *precision* of the transition matrix and the *slope* of prior preferences. The simulated responses under question are local field potentials (LFPs). As new observations are made, evidence for the competing hypotheses (hidden states) is acquired. Variational message passing that mediates belief updates over these hypotheses, where I assume that activity in different neural populations reflects belief updating over different hypotheses. The simulated depolarisation of these 'neural populations' are combined to simulate LFPs. In this setting, a simulated LFP is defined as the rate of change in expectation about a hidden states over an epoch; i.e. the rate of change in v_τ^π (see Fig 2.4). There are 16 epochs in each trial and on each epoch the expectations are updated with 16 variational iterations of the above gradient descent.

The local field potentials can be characterised by their *amplitude* and *convergence time*. Higher amplitudes are associated with greater belief updates that can be thought of in terms of larger state prediction errors. Convergence time can be defined as the time it takes before the local field potentials returned to zero, as belief updating converges on a new posterior belief. These two characterisations speak to the confidence in beliefs about hidden states and how quickly that confidence is manifest.

I characterised the responses of units encoding the hidden state dimension *where* (patch identity). First, I examined belief updates when the agent stays in the three patches for four consecutive epochs. The corresponding LFPs are shown in Fig 6.12. Smaller LFPs are generated when the reward probability decreases at a greater rate with t_s (compare patches 1 to 3 from left to right in Fig 6.12). This follows because the subject's belief about staying in a patch reaches a higher level of confidence when the reward probability declines at a slower rate (e.g., patch 1). This results in larger LFPs being generated under that patch. A second observation here is that the LFPs at the first epoch are greater than the LFPs in the subsequent epochs under all patches. This is because before entering a patch, the agent has uniform beliefs about what patch it will end up in. This means that once a patch is entered, there will be more belief updates initially; whereas later epochs just modify those beliefs already held.

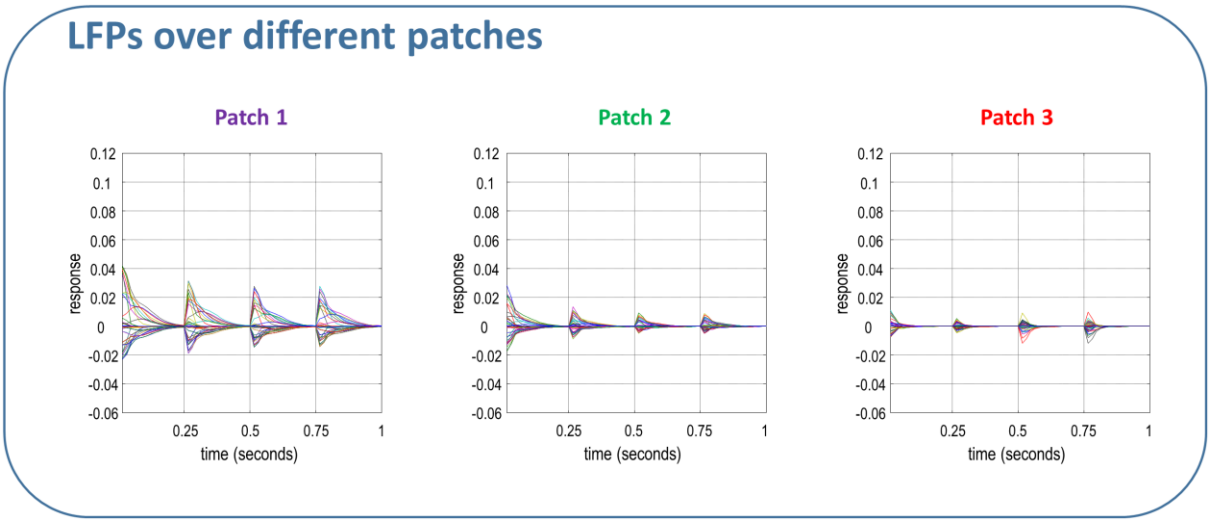


Figure 6. 12 Simulated LFPs over different patches

This panel shows the updates over expectations about (*where*) hidden states when the agent stays in different patches for four consecutive epochs. As the reward probability decreases faster with t_s the LFP peaks are attenuated and it takes longer for them to converge. The inconsistency in the degree of belief updating in later epoch – in patch 3 compared with the other patches – is because the agent expects to leave this patch; however, it ends up staying in it due to an unlucky sampling of the action *stay* (sampling low probability *stay* rather than high probability *leave*), which induces more belief updating in later epochs.

Secondly, I examined how the LFPs change with different *policy depths*. The LFPs have higher peaks, when the agent entertains a shallow representation of the future ($PD = 2$) and peak less when it looks deeper into the future ($PD = 4$): see Fig 6.13. With deeper policies the beliefs (expectations) about the hidden states are projected further into the future, causing future epochs to be informed by the expectations over the hidden states at the present time. This causes the beliefs about being in a certain patch during an epoch to change less over time. A second observation here is that the expectations converge faster under shallow policies. Before these expectations are projected to any future epochs, the agent maintains uniform distributions over the hidden states. The further the expectations about the hidden states in the current epoch are projected to future, the more imprecise these expectations become, taking longer to converge; especially in the epochs in the distant future. This is why deeper policies require longer for expectations to converge.

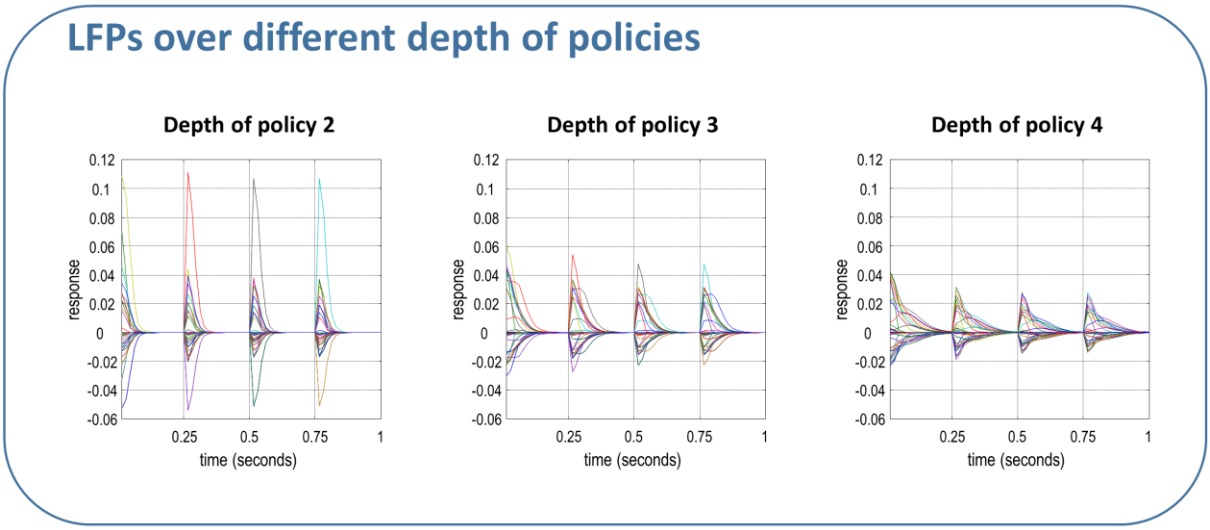


Figure 6. 13 Simulated LFPs over different depth of policies

This figure shows the effect of the *policy depth* on LFPs: with deeper policies the LFPs peak less and it takes longer for them to converge.

Thirdly, the effect of the *precision* of the transition matrices on the LFPs is characterised. With precise transition matrices, the LFPs have greater amplitude – and it takes less time for these expectations to converge (see Fig 6.14). This follows because – with precise transition matrices – the expectations about the hidden states in the current epoch are projected forwards with greater fidelity than with less precise transition matrices. This induces large updates over expectations and more rapid convergence.

Finally, the effect of the *discount slope* on the LFPs is shown on Fig 6.15. As shown in Fig 6.9, the utility over *reward* declines at different rates under different *slopes*, while the average over future times is conserved. When the discount slope is high, the agent values rewards in the immediate future more than the distant future. With a high slope over the prior preferences, the agent believes that it will stay in the same patch with a greater degree of confidence than with lower slopes. This causes the LFPs to peak higher. However, this does not affect the convergence time.

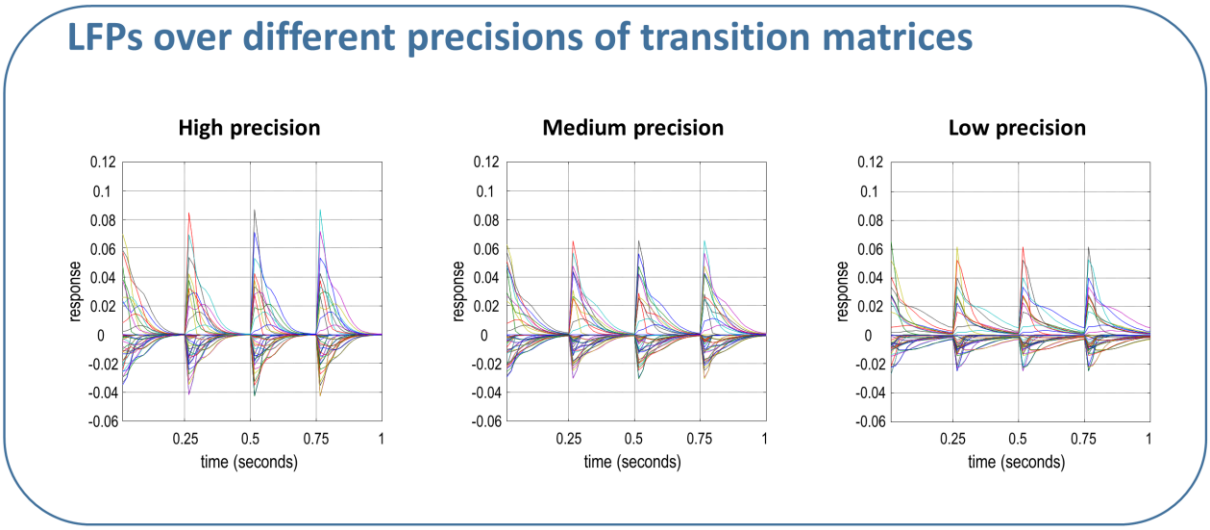


Figure 6. 14 Simulated LFPs over different precisions of transition matrices

In this figure, the LFPs obtained with different *precisions* of transition matrices are shown: with more precise transition matrices, the LFPs peak higher and converge more quickly

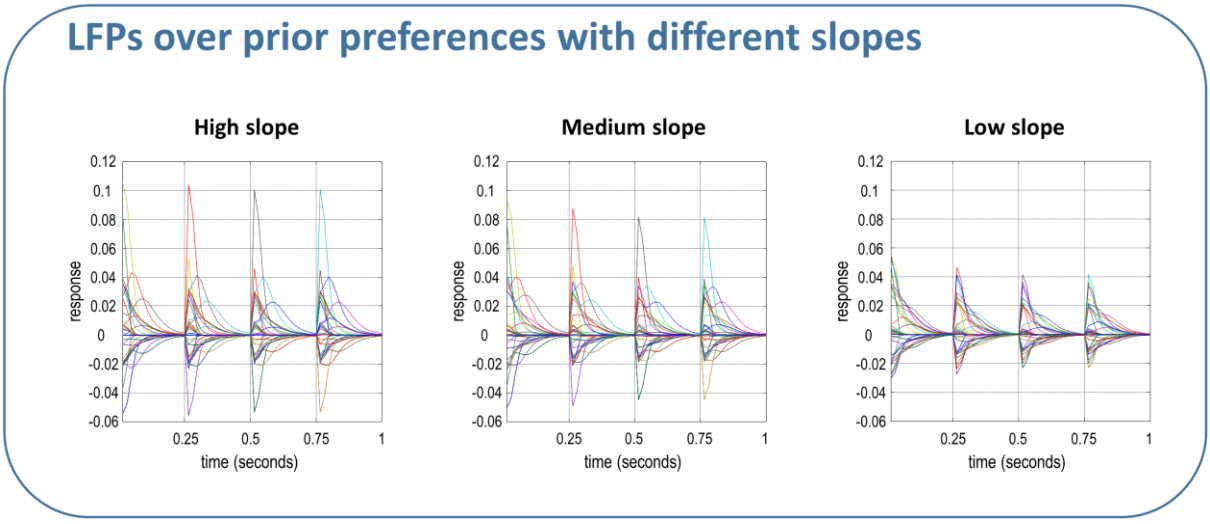


Figure 6. 15 Simulated LFPs over different prior preferences with different slopes

This figure shows how the LFPs change when the *discount slope* varies over three levels; while keeping the average utilities over time fixed. With higher slopes the LFPs peak at higher levels, while the convergence does not appear to be sensitive to the different slopes.

6.4. Discussion

The work in this chapter diverged from the work in the previous chapters in a way that the task that was used, namely the patch leaving task, was not tailored for a visual

search paradigm. However this paradigm is still relevant to computational modelling of information gathering in the context of exploration and exploitation dilemma, as it balances exploitation of the current patch against exploration of other patches.

In this chapter, my objective was to show that there are different computational mechanisms that can lead to impulsive behaviour: lower depth of planning, poor maintenance of information, and preference for immediate rewards. For this purpose, I introduced a MDP formulation of active inference for the patch-leaving paradigm. I defined impulsivity as acting to gain temporally proximal rewards at the expense of more distal rewards. The patch-leaving task allows impulsivity to be addressed, as it places proximal and distal rewards in conflict. Although the reward probability declines as one stays in the same patch, only choosing to stay can deliver an immediate reward, however unlikely it may be. This means that acting to secure proximal rewards requires one to stay in a patch for longer.

I introduced a *canonical* model that serves as a point of reference for the *dwell time* in various patches. This model was compared with deviant models in which the *policy depth*, *precision* of the transition matrix and the *discount slope* were manipulated. With shallow policies, the agent stays longer in each patch (see the light blue bars in Fig 6.10). An agent that uses deep policies realizes how quickly (or slowly) the reward probabilities decline (see dark blue bars in Fig 6.10). This realization causes the agent to leave before the reward probability declines a great deal under the prospective belief it will secure rewards elsewhere.

With imprecise beliefs about probability transitions, the agent places less confidence in its beliefs about future hidden states and outcomes. This means that it is difficult to infer what might happen after leaving a patch; since this requires the subject to look at least two epochs into the future to see if reward can be obtained.. In comparison, the expected outcome of staying in the same patch requires the agent to consider only one epoch into the future (anticipating the reward probability in the very next outcome). Since the agent is relatively more confident about the outcome of staying in a patch (and thus more certain about getting a reward upon staying in a patch) it chooses to stay for longer under less precise transition matrices than more precise transition matrices (see light and dark green bars in Fig 6.10). This result suggests that impulsivity can result from not being able to anticipate the future confidently.

Finally, manipulating the *discount slope* over time proves to have a profound effect on *dwell times* as well. When the time sensitivity of preferences is high, the agent values the immediate future much more – and hence dwells longer – than when the slope is low (see the light and dark red bars in Fig 6.10). This causes the agent to value proximal rewards more, even when they are less likely.

The underlying causes of impulsivity under the three models mentioned above speak to different personality traits. The explanation for impulsivity under shallow policies is due to steep discounting of the future (Alessi & Petry, 2003), which may be due to a lack of future planning (Patton et al., 1995). Imprecise beliefs about environmental transitions impair an agent's ability to maintain and process information when planning its future actions (T. Parr & Friston, 2017b). The kind of response obtained here is similar to acting impulsively due to high working memory load (Hinson, Jameson, & Whitney, 2003). The high temporal sensitivity of prior preferences causes the agent to act impulsively; despite an ability to plan deep into the future. This is because it prefers immediate rewards more than rewards in the distant future. These prior preferences can lead to risk-taking behaviour (Leigh, 1999; Lejuez et al., 2002) or 'venturesomeness' (B.G. Eysenck, Pearson, Easting, & Allsopp, 1985; S. Eysenck, 1993).

I have also shown how the belief updates relate to (simulated) LFPs under these different models. Comparing the LFPs obtained with the canonical model on the first and subsequent epochs, I showed that the LFPs peak less as time progresses (see Fig 6.12). Comparing different patches, the LFPs peak less as the reward probability declines faster in a patch (compare patches 1 to 3 in Fig 6.12). This suggests that the amplitude of the LFPs correlate positively with the reward probability. Comparing different *policy depths*, the LFPs peak higher with shallow policies (compare $PD = 2$ with $PD = 4$ in Fig 6.13). The LFPs peak higher with more precise transition matrices than less precise transition matrices (compare high to low precision in Fig 6.14). Finally, with high slopes over the prior preferences, the LFPs peak higher (compare high to low slope in Fig 6.15). The findings in the event related potential (ERP) literature show that the different components of ERPs can indeed be manipulated by reward probability (M. X. Cohen, Elger, & Ranganath, 2007; Eppinger, Kray, Mock, & Mecklinger, 2008; Walsh & Anderson, 2012) and reward magnitude (Bellebaum,

Polezzi, & Daum, 2010; Goldstein et al., 2006; Meadows, Gable, Lohse, & Miller, 2016). Using the simulated LFPs I have shown that similar reward probability and magnitude effects are an emergent property of belief updating and neuronal (variational) message passing in synthetic brains.

These simulated electrophysiological responses show that although the observed behaviours under different models (i.e. staying longer in a patch) are similar, different LFPs are generated. Comparing a shallow policy model (see the left panel of Fig 6.13) with the model in which the slope of the preferences is high (see the left panel of Fig 6.15), the amplitude of the LFPs look similar; however, the LFPs in the model with shallow policies converge sooner. Comparing the model with low precision transition matrices (see the right panel of Fig 6.14) with the above two models, the LFPs neither peak as high nor do they converge as quickly.

An influential model (Gläscher, Daw, Dayan, & O'Doherty, 2010) assesses the degree to which subjects are 'model-based' (i.e. learn the transition matrix and then use it to plan) versus 'model-free' (i.e. just repeating previously rewarded actions). It has been shown that various disorders of compulsivity (e.g. OCD, binge eating, drug addiction) are less 'model-based' in this task (Voon et al., 2015), as are high impulsivity subjects (Deserno et al., 2015), and that compulsivity in a large population sample also relates to this task measure (Gillan, Kosinski, Whelan, Phelps, & Daw, 2016). However, this model does not explain why the subjects are less model-based. My formulation suggests that one possibility for this is a less precise transition matrix and another is lower *policy depth*.

This work has some limitations. The *policy depth*, the *precision* of the transition matrices and the *discount slope* cannot be manipulated experimentally in a straightforward way using the patch leaving paradigm described in this work. This means model selection given empirical choice behaviour can only be validated in relation to independent variables; e.g., correlations between working memory measures and transition matrix precision. Furthermore, I have only looked at model features that explain impulsivity relating to depth of planning, working memory and value discounting: I have not considered other causes, e.g. motor disinhibition or effort cost (Klein-Flugge, Kennerley, Saraiva, Penny, & Bestmann, 2015).

7. Discussion

In summary, I have used active inference to study Bayes optimal behaviour in different task setups, with a special emphasis on information gathering behaviour. Active inference provides computational accounts of perception, attention and action. All three of these mechanisms are essential for the adaptive exchange of biological systems (or agents) with their environments in the face of volatility and abnormalities in these mechanisms might have severe consequences on a biological agent's existence. Perception allows the changes in the environment to be recognised, whereas attention allows an agent to selectively attend to different sensory stimuli under different contexts. Perception of changing contexts is essential for attending to the information that is relevant under a context. Finally, action is necessary to keep a biological agent in the environmental states that it can exist. In active inference these three mechanisms minimise variational free energy, which is an alternative to (negative) Bayesian model evidence, thus maximising an agent's evidence for its existence.

In active inference perception corresponds to inference about the hidden states of the world, whereas attention is inference about the uncertainty of sensory signals and their causes. Action arises as a result of minimising the expected free energy in the future. In this framework an agent *a priori* believes that it will minimise the expected free energy in the future, thus maximising the evidence for its existence in the future. Expected free energy is comprised of two main components, namely *extrinsic value* and *epistemic value*. Extrinsic value is the expected utility in the future, whereas epistemic value is the expected information gain about the hidden states of the world. It is important to express that both extrinsic and epistemic value are computed based on an agent's beliefs. These beliefs include agent's preferences for different sensory signals and its beliefs about the hidden states of the world. These hidden states can be seen as the hidden aspects that define an environment. Biological agents do not possess complete knowledge about the real world process that generates the outcomes that can be observed. Agents require a *generative model* to infer the most likely states of the world that the outcomes originated from and uses this model to form beliefs about the policies. Policies are sequences of actions that the agent can take in the future and each policy defines a different trajectory into the future. An agent is

more likely to pursue a policy if it believes that policy provides more evidence for its future existence (see chapter 2).

I have used active inference to describe information gathering behaviour. For this purpose I created a number of tasks that require information gathering and evidence accumulation. Firstly, I introduced a new task called scene construction task. This task was performed on a two-by-two grid scene. The relative locations of the objects in this grid scene defined the category of the scene and the goal was to categorise the scene correctly attending to minimum number locations in the scene. The quadrants on the grid scene were masked and attending to them would disclose their contents. Essentially, seeing a particular object would make a certain location more salient than the others. I placed special emphasis on epistemic value on this task because successful categorisation of the task depends on the resolution of uncertainty about the hidden states of the scene. Once all the uncertainty about the hidden states is resolved, the scene category becomes apparent and the scene can be categorised. I have shown that an active inference agent attends to the locations that are epistemically valuable before categorising the scene (see chapter 3).

In the follow up study the scene construction task was performed by healthy humans. Using an eye-tracker the saccadic choices of the people have been collected as they performed the task. Analysing the scanpaths of the participants revealed that they may have used heuristic strategies to explore scenes. The difference between heuristic and epistemic policies is that the epistemic policies take the agent's beliefs into account and resolves uncertainty about the hidden states whereas the heuristic strategies are fixed in the sense that the order of locations that are visited do not change from one trial to the next. In a sense heuristic policies are non-epistemic policies that do not resolve uncertainty about the scene efficiently. Each subject's favourite heuristic policy was included in the form of a prior preference in their repertoire of policies. In the next, I asked whether healthy people's saccadic exploration while performing the task afforded evidence for epistemic exploration. I fitted two separate models to the saccadic choices of the subjects. Both of these models included the subject's preferences (including a heuristic policy). The first model also included the uncertainty resolving component about the scene, namely epistemic value. Comparing the evidence under these models revealed that the model with

epistemic value yielded more evidence. This result suggests that the healthy people do resolve uncertainty (acquire information) about the hidden states of the visual scene. I then moved onto characterising different exploratory behaviours by using canonical correlation analysis. This entailed maximising the correlation between the linear mixtures of performance measures of subjects and the estimated parameters of the model by fitting models to the saccadic choices of the participants. I found that there are three distinct behavioural phenotypes (see chapter 4). In each of these behavioural phenotypes a preference for the use of heuristic policies was linked with a tendency to categorise the scene incorrectly. There is no reason for categorising the scene incorrectly under heuristic and epistemic policies, however in order to exploit epistemic policies one needs to have a model of the task. This result suggests that having a model of the task pays off.

I then showed how contextual exploration of a visual scene can occur by appealing to a computational mechanism that may correspond to attention. To show this I used a visual search task that resembled the scene construction task. In this task a scene could be categorised in two different ways, depending on the context. This meant that a piece of information that is salient under one context could be completely irrelevant under another. This was accomplished by suppressing expected information gain (epistemic value) about the hidden states that are irrelevant to the current context while allowing information gain about the hidden states that are relevant to context. Crucially, epistemic value can acquire information only when there is a precise mapping between the sensory signals and their causes. Making this mapping very precise for the task relevant sensory signals and very imprecise for the task irrelevant sensory signals can explain exploratory behaviours under different contexts. By applying this principle on Yarbus' task I showed that this model can generate saccadic patterns that resemble the ones seen in Yarbus' empirical visual exploration paradigm (see chapter 5). In summary I showed how a global context can drive visual attention and relevance of information. This is important as a number of clinical conditions such as autism, schizophrenia and anxiety have been associated with an imbalance of precisions between top down and bottom up processes.

In the final study, I showed that there could be distinct causes of impulsive behaviour by using an active inference model of a patch leaving paradigm. This paradigm

diverged from the above paradigms as it was not a visual search paradigm. This is still relevant to computational modelling of information gathering as this paradigm is concerned with the exploration-exploitation dilemma. The question in this paradigm is to decide when to leave an environment with limited resources for a new one to maximise reward in the long run. I showed that there are at least three distinct causes of impulsivity by manipulating the parameters of the MDP model, namely the depth of planning, the capacity to maintain and process information, and the perceived value of immediate rewards (see chapter 6). In addition to the simulated behavioural responses, I have also shown how the belief updates introduced in Fig 2.4 may relate to simulated local field potentials (see chapter 3 and 6). These simulated electrophysiological responses are obtained based on the assumption that the activity in different neural populations reflects belief updating over different hypotheses. Simulated LFPs were defined as the rate of change in expectations about the hidden states. Under different prior beliefs these simulated LFPs vary drastically (see Fig 6.12 - 6.15). In theory one can fit a model to the electrophysiological responses in imaging studies and estimate the prior beliefs of individuals. This could be an effective method for characterising behaviour of the individuals (or groups) in terms of electrophysiological responses, and vice versa.

Although the tasks described in this thesis are tailored for humans, they could be changed in various ways and used in rodent and non-human primate studies. Information gathering tasks have been extensively used in monkey and rodent studies (Deacon & Rawlins, 2006; Foley, Kelly, Mhatre, Lopes, & Gottlieb, 2017). The scene construction task can be translated in a way that it could be used in monkey studies by exposing monkeys to different scene categories and spatial transformations (the hidden state space). In this setup monkeys would use their eyes to report their saccadic choices and might learn the hidden state space by associating correct categorizations with a rewarding outcome such as food (or fluid). The choice about the scene category can be reported using saccades to separate choice locations (similar to choice locations shown in Fig 4.3). Adapting this task for rodents to perform might be more challenging. The scene construction task might be translated to a two-room by two-room maze that the rodents can enter and explore. Each room in the maze would be accessible through the other three rooms. Each room would contain a cue about the scene category. Including a decision room that contains three levers

(associated with different scene categories) that the rodents can use to report their beliefs about the scene category and get feedback (right or wrong categorization) might enable them to learn scene categories (and the hidden state space). Successful categorizations would yield food or fluid reward.

Patch leaving paradigm has been employed in animal studies in various forms. Animals act in their ecological niche in a way to ensure they acquire enough food to survive and the exploration-exploitation dilemma, which lies at the heart of the patch-leaving paradigm, is a part of their everyday life. Adaptations of the patch leaving paradigm usually involve a training period in which the primates are trained to report their decisions with their eyes by choosing one of two different sized objects (e.g. vertical bars) that shrink upon looking at them. In these versions of the patch leaving paradigm the delay in choosing different options is proportional to the length of the vertical bars. See (Blanchard & Hayden, 2015) for an example of this. In these versions of the patch leaving paradigm the rewarding outcomes are delivered in the form of fluid after the vertical bar associated with *stay* disappears. Once the chosen bar disappears, the same bars are displayed on the screen again but this time choosing the stay option delivers a smaller amount of reward. Leaving a patch requires choosing the other bar on the screen and upon doing so sets the reward amount to its maximum. Crucially the vertical bar associated with the leave option is longer than the vertical bar associated with stay option and looking at these bars shrinks them at the same rate. This means that choosing to leave has a cost (similar to a travel cost) which is to wait without having any reward until the bar associated with leave option disappears. Patch leaving paradigm can be adapted in a way that could be performed by rodents as well. One can have multiple T-maze like structures that has two alternative choices. One arm would be associated with the stay option, and the other would be associated with the leave option. With repeated stays the reward amount/probability would decrease. The arm associated with leave would connect the current T-maze to another T-maze with stay and leave arms. This way one can introduce travelling costs in a more ecological way.

As opposed to the other models, in the MDP formulation of active inference there is a distinction between the beliefs about the real world dynamics that generates outcomes (generative process) and the beliefs about these real world dynamics (generative

model). Crucially the generative model depends upon prior beliefs and the different combinations of prior beliefs would allow for different optimal policies. This makes the active inference framework more suitable to studying differences between individuals in terms of their prior beliefs (see chapter 4 for details). Moreover, a number of clinical conditions have been hypothesised to be associated with abnormal prior beliefs. Different combinations of prior beliefs may lead to the same abnormalities in the observed behaviour. The model inversion methods used in chapter 4 may enable us to distinguish between these different prior beliefs. This is important as estimating these prior beliefs for different patient populations may give an opportunity to study psychopathologies. Moreover, estimating the subject specific prior beliefs may allow for more customized therapies to be administered to patient populations. Clinical trials that target specific populations can benefit from estimating the prior beliefs to validate the population that is intended to be studied.

The hidden state spaces in the models used in this work are set in a way that they are not dissimilar to the functionally segregated brain regions. For example the visual search paradigms introduced in this thesis make a distinction between ‘what’ and ‘where’ attributes of a visual scene, which is similar to the functionally segregated ventral (what) and dorsal (where) streams in the brain. These models set an example to show how neurobiologically plausible models can be created. Using these models one can simulate both behavioural and electrophysiological responses. Furthermore one can fit models to both behavioural and electrophysiological responses obtained in empirical studies to understand the computational goal of the brain and the neural mechanisms that underlie those computations.

Table 1 Glossary of expressions

Expression	Description
$o_\tau = (o_\tau^1, \ldots, o_\tau^M)$	Outcomes in *M* modalities
$\tilde{o} = (o_1, \ldots, o_t)$	Sequences of outcomes up until the current time
$s_\tau = (s_\tau^1, \ldots, s_\tau^N)$	Hidden states in *N* factors (dimensions)
$\tilde{s} = (s_1, \ldots, s_T)$	Sequences of hidden states up until the end of the current trial
$\pi = (\pi_1, \ldots, \pi_K),\ \boldsymbol{\pi} \in [0,1]$	Policies specifying action sequences and their posterior expectations
$a = \pi(t) = (a_1, \ldots, a_N)$	Action or control variables for each factor (dimension) of hidden states
$\gamma, \boldsymbol{\gamma} = \frac{1}{\boldsymbol{\beta}}$	The precision (inverse temperature) of beliefs about policies and its posterior expectation
β	Prior expectation of temperature (inverse precision) of beliefs about policies
$\mathbf{A}^m$	Likelihood array mapping from hidden states to the *m*-th modality
$\mathbf{B}^n(a),\ \mathbf{B}_\tau^{n,\pi} = \mathbf{B}^n(a = \pi(\tau))$	Transition probability for the *n*-th hidden state under each action
$\mathbf{C}_\tau^m$	Logarithm of the prior probability of the *m*-th outcome; i.e. preferences or utility
$\mathbf{D}^n$	Prior expectation of the *n*-th hidden state at the beginning of each trial
$\mathbf{F} : \mathbf{F}_\pi = F(\pi) = \sum_\tau F(\pi, \tau) = \sum_{n,\tau} F(\pi, \tau, n)$	Variational free energy for each policy
$\mathbf{G} : \mathbf{G}_\pi = G(\pi) = \sum_\tau G(\pi, \tau) = \sum_{m,\tau} G(\pi, \tau, m)$	Expected free energy for each policy
$\mathbf{H}^m$	Entropy of the *m*-th outcome

References

Achanta, R., Hemami, S., Estrada, F., & Susstrunk, S. (2009). *Frequency-tuned salient region detection.* Paper presented at the Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on.

Alessi, S. M., & Petry, N. M. (2003). Pathological gambling severity is associated with impulsivity in a delay discounting procedure. *Behav Processes, 64*(3), 345-354.

Andreopoulos, A., & Tsotsos, J. (2013). A computational learning theory of active object recognition under uncertainty. *International journal of computer vision, 101*(1), 95-142.

B.G. Eysenck, S., Pearson, P. R., Easting, G., & Allsopp, J. F. (1985). *Age norms for Impulsiveness, Venturesomeness and Empathy in adults* (Vol. 6).

Beedie, S. A., Clair, D. M. S., & Benson, P. J. (2011). Atypical scanpaths in schizophrenia: evidence of a trait-or state-dependent phenomenon? *Journal of psychiatry & neuroscience: JPN, 36*(3), 150.

Behrmann, M., Avidan, G., Leonard, G. L., Kimchi, R., Luna, B., Humphreys, K., & Minshew, N. (2006). Configural processing in autism and its relationship to face processing. *Neuropsychologia, 44*(1), 110-129. doi:https://doi.org/10.1016/j.neuropsychologia.2005.04.002

Bellebaum, C., Polezzi, D., & Daum, I. (2010). It is less than you expected: The feedback-related negativity reflects violations of reward magnitude expectations. *Neuropsychologia, 48*(11), 3343-3350. doi:https://doi.org/10.1016/j.neuropsychologia.2010.07.023

Bellman, R. (1952). On the Theory of Dynamic Programming. *Proc Natl Acad Sci USA, 38*, 716-719.

Biederman, I. (2017). On the semantics of a glance at a scene. In *Perceptual organization* (pp. 213-253): Routledge.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cogn Psychol, 14*(2), 143-177.

Blanchard, T. C., & Hayden, B. Y. (2015). Monkeys are more patient in a foraging task than in a standard intertemporal choice task. *PLoS One, 10*(2), e0117057. doi:10.1371/journal.pone.0117057

Bonet, B., & Geffner, H. (2014). Belief Tracking for Planning with Sensing: Width, Complexity and Approximations. *Journal of Artificial Intelligence Research, 50*, 923-970.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436.

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., . . . Knight, R., T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science, 313*(5793), 1626-1628.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of vision, 9*(3), 6-6.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theor Popul Biol, 9*(2), 129-136.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends Cogn Sci, 4*(5), 170-178.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn Psychol, 36*(1), 28-71. doi:10.1006/cogp.1998.0681

Cohen, J. D., Barch, D. M., Carter, C., & Servan-Schreiber, D. (1999). Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. *Journal of abnormal psychology, 108*(1), 120.

Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage, 35*(2), 968-978. doi:10.1016/j.neuroimage.2006.11.056

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychol Sci, 15*(8), 559-564. doi:10.1111/j.0956-7976.2004.00719.x

de Lafuente, V., Jazayeri, M., & Shadlen, M. N. (2015). Representation of accumulating evidence for a decision in two parietal areas. *J Neurosci, 35*(10), 4306-4318. doi:10.1523/jneurosci.2451-14.2015

Deacon, R. M., & Rawlins, J. N. (2006). T-maze alternation in the rodent. *Nat Protoc, 1*(1), 7-12. doi:10.1038/nprot.2006.2

Deserno, L., Wilbertz, T., Reiter, A., Horstmann, A., Neumann, J., Villringer, A., . . . Schlagenhauf, F. (2015). Lateral prefrontal model-based signatures are reduced in healthy individuals with high trait impulsivity. *Transl Psychiatry, 5*, e659. doi:10.1038/tp.2015.139

Donaldson, I. M. (2000). The functions of the proprioceptors of the eye muscles. *Philosophical Transactions of the Royal Society B: Biological Sciences, 355*(1404), 1685-1754.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*(5040), 90-92.

Eppinger, B., Kray, J., Mock, B., & Mecklinger, A. (2008). Better or worse than expected? Aging, learning, and the ERN. *Neuropsychologia, 46*(2), 521-539. doi:https://doi.org/10.1016/j.neuropsychologia.2007.09.001

Eysenck, S. (1993). The I7: Development of a measure of impulsivity and its relationship to the superfactors of personality. *The impulsive client: theory, research and treatment*.

Eysenck, S. B., & Eysenck, H. J. (1978). Impulsiveness and venturesomeness: their position in a dimensional system of personality description. *Psychol Rep, 43*(3 Pt 2), 1247-1255. doi:10.2466/pr0.1978.43.3f.1247

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

Fernandes, B. S., Williams, L. M., Steiner, J., Leboyer, M., Carvalho, A. F., & Berk, M. (2017). The new field of 'precision psychiatry'. *BMC Med, 15*(1), 80. doi:10.1186/s12916-017-0849-x

FitzGerald, T. H., Schwartenbeck, P., Moutoussis, M., Dolan, R. J., & Friston, K. (2015). Active inference, evidence accumulation, and the urn task. *Neural Comput, 27*(2), 306-328. doi:10.1162/NECO_a_00699

Foley, N. C., Kelly, S. P., Mhatre, H., Lopes, M., & Gottlieb, J. (2017). Parietal neurons encode expected gains in instrumental information. *Proc Natl Acad Sci U S A, 114*(16), E3315-E3323. doi:10.1073/pnas.1613844114

Frank, M. J. (2011). Computational models of motivated action selection in corticostriatal circuits. *Current opinion in neurobiology, 21*(3), 381-386.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn Sci, 13*(7), 293-301. doi:10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci, 11*(2), 127-138. doi:10.1038/nrn2787

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Front Psychol., 3*, 151.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Comput, 29*(1), 1-49. doi:10.1162/NECO_a_00912

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *J Physiol Paris, 100*(1-3), 70-87. doi:10.1016/j.jphysparis.2006.10.001

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage, 34*(1), 220-234. doi:10.1016/j.neuroimage.2006.08.035

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cogn Neurosci*, 1-28. doi:10.1080/17588928.2015.1020053

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making. *Philos Trans R Soc Lond B Biol Sci, 369*(1655). doi:10.1098/rstb.2013.0481

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Raymond J. Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Front Hum Neurosci., 7*, 598.

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: belief propagation and active inference. *Network Neuroscience, 1*(4), 381-414.

Frith, C. (2003). What do imaging studies tell us about the neural basis of autism? *Novartis Found Symp, 251*, 149-166; discussion 166-176, 281-197.

Gao, D., & Vasconcelos, N. (2005). *Discriminant saliency for visual recognition from cluttered scenes.* Paper presented at the Advances in neural information processing systems.

Gibb, J. A. (1958). Predation by Tits and Squirrels on the Eucosmid Ernarmonia conicolana (Heyl.). *Journal of Animal Ecology, 27*(2), 375-396. doi:10.2307/2245

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *Elife, 5*.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*(4), 585-595.

Goldstein, R. Z., Cottone, L. A., Jia, Z., Maloney, T., Volkow, N. D., & Squires, N. K. (2006). The effect of graded monetary reward on cognitive event-related potentials and behavior in young healthy adults. *Int J Psychophysiol, 62*(2), 272-279. doi:10.1016/j.ijpsycho.2006.05.006

Grossberg, S., Roberts, K., Aguilar, M., & Bullock, D. (1997). A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. *Journal of Neuroscience, 17*(24), 9706-9725.

Grossberg, S., Roberts, K., Aguilar, M., & Bullock, D. (1997). A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. *J Neurosci., 17*(24), 9706-9725.

Happé, F. G. (1996). Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. A research note. *Journal of Child Psychology and Psychiatry, 37*(7), 873-877.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends Cogn Sci, 11*(7), 299-306. doi:10.1016/j.tics.2007.05.001

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience, 14*(11), 6336-6353.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2003). Impulsive decision making and working memory. *J Exp Psychol Learn Mem Cogn, 29*(2), 298-306.

Howard, R. (1966). Information Value Theory. *IEEE Transactions on Systems, Science and Cybernetics, SSC-2*(1), 22-26.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295-1306. doi:https://doi.org/10.1016/j.visres.2008.09.007

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10), 1489-1506. doi:https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence, 20*(11), 1254-1259.

Jaakkola, T., & Jordan, M. (1998). Improving the Mean Field Approximation Via the Use of Mixture Distributions. In M. Jordan (Ed.), *Learning in Graphical Models* (Vol. 89, pp. 163-173): Springer Netherlands.

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry, 160*(1), 13-23. doi:10.1176/appi.ajp.160.1.13

Kira, S., Yang, T., & Shadlen, M. N. (2015). A neural implementation of Wald's sequential probability ratio test. *Neuron, 85*(4), 861-873. doi:10.1016/j.neuron.2015.01.007

Klein-Flugge, M. C., Kennerley, S. W., Saraiva, A. C., Penny, W. D., & Bestmann, S. (2015). Behavioral modeling of human choices reveals dissociable effects of physical effort and temporal delay on reward devaluation. *PLoS Comput Biol, 11*(3), e1004116. doi:10.1371/journal.pcbi.1004116

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol, 4*(4), 219-227.

Kojima, T., Matsushima, E., Ando, K., Ando, H., Sakurada, M., Ohta, K., . . . Shimazono, Y. (1992). Exploratory eye movements and neuropsychological tests in schizophrenic patients. *Schizophr Bull, 18*(1), 85-94.

Latimer, K. W., Yates, J. L., Meister, M. L., Huk, A. C., & Pillow, J. W. (2015). NEURONAL MODELING. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science, 349*(6244), 184-187. doi:10.1126/science.aaa4056

Leigh, B. C. (1999). Peril, chance, adventure: concepts of risk, alcohol use and risky behavior in young adults. *Addiction, 94*(3), 371-383.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., . . . Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *J Exp Psychol Appl, 8*(2), 75-84.

Lisman, J., & Redish, A. D. (2009). Prediction, sequences and the hippocampus. *Philos Trans R Soc Lond B Biol Sci, 364*(1521), 1193-1201. doi:10.1098/rstb.2008.0316

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron, 77*(6), 1002-1016. doi:10.1016/j.neuron.2013.03.007

Logue, A. W. (1995). *Self-control: Waiting until tomorrow for what you want today*. Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Luck, S. J., & Gold, J. M. (2008). The construct of attention in schizophrenia. *Biological psychiatry, 64*(1), 34-39.

Ma, Y.-F., & Zhang, H.-J. (2003). *Contrast-based image attention analysis by using fuzzy growing.* Paper presented at the Proceedings of the eleventh ACM international conference on Multimedia.

MacArthur, R. H., & Pianka, E. R. (1966). On Optimal Use of a Patchy Environment. *The American Naturalist, 100*(916), 603-609.

Meadows, C. C., Gable, P. A., Lohse, K. R., & Miller, M. W. (2016). The effects of reward magnitude on reward processing: An averaged and single trial event-related potential study. *Biol Psychol, 118*, 154-160. doi:10.1016/j.biopsycho.2016.06.002

Mirza, M. B., Adams, R. A., Mathys, C., & Friston, K. J. (2018). Human visual exploration reduces uncertainty about the sensed world. *PLoS One, 13*(1), e0190429. doi:10.1371/journal.pone.0190429

Mirza, M. B., Adams, R. A., Mathys, C. D., & Friston, K. J. (2016). Scene Construction, Visual Foraging, and Active Inference. *Front Comput Neurosci, 10*, 56. doi:10.3389/fncom.2016.00056

Mirza, M. B., Adams, R. A., Parr, T., & Friston, K. (2019). Impulsivity and Active Inference. *J Cogn Neurosci, 31*(2), 202-220. doi:10.1162/jocn_a_01352

Moutoussis, M., Trujillo-Barreto, N. J., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). A formal model of interpersonal inference. *Front Hum Neurosci, 8*, 160. doi:10.3389/fnhum.2014.00160

Navalpakkam, V., & Itti, L. (2006). *An integrated model of top-down and bottom-up attention for optimizing detection speed.* Paper presented at the Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614-621. doi:https://doi.org/10.1016/j.visres.2005.08.025

O'riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 719.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107-123. doi:https://doi.org/10.1016/S0042-6989(01)00250-4

Parr, T., & Friston, K. J. (2017a). Uncertainty, epistemics and active inference. *J R Soc Interface, 14*(136). doi:10.1098/rsif.2017.0376

Parr, T., & Friston, K. J. (2017b). Working memory, attention, and salience in active inference. *Sci Rep, 7*(1), 14678. doi:10.1038/s41598-017-15249-0

Parr, T., & Friston, K. J. (2018). The Computational Anatomy of Visual Neglect. *Cereb Cortex, 28*(2), 777-790. doi:10.1093/cercor/bhx316

Parr, T., Rees, G., & Friston, K. J. (2018). Computational Neuropsychology and Bayesian Inference. *Frontiers in Human Neuroscience, 12*(61). doi:10.3389/fnhum.2018.00061

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *J Clin Psychol, 51*(6), 768-774.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*(4), 437-442.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual Scanning of Faces in Autism. *Journal of Autism and Developmental Disorders, 32*(4), 249-261. doi:10.1023/a:1016374617369

Peterson, M. S., & Kramer, A. F. (2001). Attentional guidance of the eyes by contextual information and abrupt onsets. *Perception & Psychophysics, 63*(7), 1239-1249. doi:10.3758/bf03194537

Pezzulo, G., Rigoli, F., & Chersi, F. (2013). The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Front Psychol, 4*, 92. doi:10.3389/fpsyg.2013.00092

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced visual search for a conjunctive target in autism: A research note. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(5), 777-783.

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42*(11), 1447-1463.

Repin, I. (1884-1888). File:Ilya Repin Unexpected visitors.jpg. Retrieved from https://en.wikipedia.org/wiki/File:Ilya_Repin_Unexpected_visitors.jpg

Robinson, O. J., Charney, D. R., Overstreet, C., Vytal, K., & Grillon, C. (2012). The adaptive threat bias in anxiety: amygdala-dorsomedial prefrontal cortex coupling and aversive amplification. *Neuroimage, 60*(1), 523-529. doi:10.1016/j.neuroimage.2011.11.096

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research, 39*(19), 3157-3163.

Rudy, J. W. (2009). Context representations, context functions, and the parahippocampal–hippocampal system. *Learning & memory, 16*(10), 573-585.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., & Friston, K. (2014). The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes. *Cereb Cortex*. doi:10.1093/cercor/bhu159

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., & Friston, K. (2014). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cerebral cortex, 25*(10), 3434-3445.

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., & Friston, K. (2015). Optimal inference with suboptimal models: addiction and active Bayesian inference. *Med Hypotheses, 84*(2), 109-117. doi:10.1016/j.mehy.2014.12.007

Schwartenbeck, P., & Friston, K. (2016). Computational Phenotyping in Psychiatry: A Worked Example. *eNeuro, 3*(4). doi:10.1523/ENEURO.0049-16.2016

Seeck, M., Schomer, D., Mainwaring, N., Ives, J., Dubuisson, D., Blume, H., . . . Mesulam, M. M. (1995). Selectively distributed processing of visual object recognition in the temporal and frontal lobes of the human brain. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 37*(4), 538-545.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal, 27*(3), 379-423. doi:doi:10.1002/j.1538-7305.1948.tb01338.x

Shen, K., Valero, J., Day, G. S., & Paré, M. (2011). Investigating the role of the superior colliculus in active vision with the visual search paradigm. *Eur J Neurosci., 33*(11), 2003-2016.

Shirvalkar, P. R., Rapp, P. R., & Shapiro, M. L. (2010). Bidirectional changes to hippocampal theta-gamma comodulation predict memory for recent spatial episodes. *Proc Natl Acad Sci U S A, 107*(15), 7054-7059. doi:10.1073/pnas.0911184107

Spencer, K. M., Nestor, P. G., Valdman, O., Niznikiewicz, M. A., Shenton, M. E., & McCarley, R. W. (2011). Enhanced Facilitation of Spatial Attention in Schizophrenia. *Neuropsychology, 25*(1), 76-85. doi:10.1037/a0020779

Srihasam, K., Bullock, D., & Grossberg, S. (2009). Target selection by the frontal cortex during coordinated saccadic and smooth pursuit eye movements. *Journal of Cognitive Neuroscience, 21*(8), 1611-1627.

Srihasam, K., Bullock, D., & Grossberg, S. (2009). Target selection by the frontal cortex during coordinated saccadic and smooth pursuit eye movements. *J Cogn Neurosci., 21*(8), 1611-1627.

Stephan, K. E., & Mathys, C. (2014). Computational approaches to psychiatry. *Curr Opin Neurobiol, 25*, 85-92. doi:10.1016/j.conb.2013.12.007

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies, 23*(3), 165-180. doi:10.2307/2295722

Tanji, J., & Hoshi, E. (2001). Behavioral planning in the prefrontal cortex. *Current opinion in neurobiology, 11*(2), 164-170.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review, 113*(4), 766.

Tsunoda, M., Kurachi, M., Yuasa, S., Kadono, Y., Matsui, M., & Shimizu, A. (1992). Scanning eye movements in schizophrenic patients. Relationship to clinical symptoms and regional cerebral blood flow using 123I-IMP SPECT. *Schizophr Res, 7*(2), 159-168.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. .Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., . . . Sahakian, B. J. (2015). Disorders of compulsivity: a common bias towards learning habits. *Molecular psychiatry, 20*(3), 345.

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neurosci Biobehav Rev, 36*(8), 1870-1884. doi:10.1016/j.neubiorev.2012.05.008

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669-689. doi:10.1016/S0191-8869(00)00064-7

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review, 1*(2), 202-238.

Wurtz, R. H., McAlonan, K., Cavanaugh, J., & Berman, R. A. (2011). Thalamic pathways for active vision. *Trends Cogn Sci., 5*(4), 177-184.

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

Yarbus, A. (1967). Eye movements and vision. 1967. *New York*.

Zeidman, P., Lutti, A., & Maguire, E. A. (2015). Investigating the functions of subregions within anterior hippocampus. *Cortex, 73*, 240-256. doi:10.1016/j.cortex.2015.09.002

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature, 335*, 311-317.

Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). SUN: A Bayesian framework for saliency using natural statistics. *Journal of vision, 8*(7), 32-32.