

Published in final edited form as:

*Neural Comput.* 2015 February ; 27(2): 306–328. doi:10.1162/NECO\_a\_00699.

## Active inference, evidence accumulation and the urn task

Thomas HB FitzGerald<sup>1</sup>, Philipp Schwartenbeck<sup>1,2,3</sup>, Michael Moutoussis<sup>1</sup>, Raymond J Dolan<sup>1</sup>, and Karl Friston<sup>1</sup>

<sup>1</sup>The Wellcome Trust Centre for Neuroimaging, UCL, 12 Queen Square, London, WC1N 3BG, UK

<sup>2</sup>Centre for Neurocognitive Research and Department of Psychology, University of Salzburg, Salzburg, Austria

<sup>3</sup>Neuroscience Institute and Centre for Neurocognitive Research, Christian-Doppler-Klinik, Paracelsus Medical University Salzburg, Salzburg, Austria

### Abstract

Deciding how much evidence to accumulate before making a decision is a problem we and other animals often face, but one which is not completely understood. This issue is particularly important because a tendency to sample less information (often known as reflection impulsivity) is a feature in several psychopathologies, such as psychosis. A formal understanding information sampling may therefore clarify the computational anatomy of psychopathology. In this theoretical paper, we consider evidence accumulation in terms of active (Bayesian) inference using a generic model of Markov decision processes. Here, agents are equipped with beliefs about their own behaviour – in this case, that they will make informed decisions. Normative decision-making is then modelled using variational Bayes to minimise surprise about choice outcomes. Under this scheme, different facets of belief updating map naturally onto the functional anatomy of the brain (at least at a heuristic level). Of particular interest is the key role played by the expected precision of beliefs about control, which we have previously suggested may be encoded by dopaminergic neurons in the midbrain. We show that manipulating expected precision strongly affects how much information an agent characteristically samples, and thus provides a possible link between impulsivity and dopaminergic dysfunction. Our study therefore represents a step towards understanding evidence accumulation in terms of neurobiologically plausible Bayesian inference, and may cast light on why this process is disordered in psychopathology.

### Keywords

active inference; evidence accumulation; Bayesian; bounded rationality; free energy; inference; utility theory; dopamine; schizophrenia

---

**Correspondence:** Thomas FitzGerald, The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, 12 Queen Square, London, UK WC1N 3BG, thomas.fitzgerald@ucl.ac.uk., Telephone: +44 20 344 84366.

**Conflict of Interest:** The authors declare no competing financial interests.

## Introduction

This paper considers the nature of evidence accumulation in terms of active (Bayesian) inference using the urn or beads task (Huq et al., 1988; Moritz and Woodward, 2005). In particular, we cast decision-making in terms of prior beliefs an agent entertains about its own behaviour. In brief, we equip agents with the prior belief that they will make informed decisions and then sample their choices from posterior beliefs that accumulate evidence from successive observations. In this paper, we focus on the formal aspects of this (Bayes) optimal evidence accumulation – using a generic scheme that has been applied previously to limited offer and trust games (Friston et al., 2013; Moutoussis et al., 2014; Schwartenbeck et al., 2014). This scheme is based upon approximate Bayesian inference using variational Bayes. Importantly, the neurobiological correlates of belief updating implicit in this model can be linked to various aspects of neuroanatomy and, possibly, dopaminergic discharges. In what follows, we introduce the particular generative model that underlies evidence accumulation in the urn task and show how the ensuing performance depends upon key model parameters, like decision criteria and precision of – or confidence in – beliefs about choices.

Evidence accumulation, in its various forms, has been the subject of much interest in cognitive neuroscience (Gold and Shadlen, 2007; Yang and Shadlen, 2007; Moutoussis et al., 2011; Hunt et al., 2012). Incorporating it within the broader framework of Bayes-optimal inference is of intrinsic interest particularly with respect to pathological conditions characterized by increased ‘reflection impulsivity’ (a decreased sampling of evidence prior to making a decision; (Kagan, 1966)). This has been much researched in schizophrenia, where abnormal behaviour often includes increased reflection impulsivity (Huq et al., 1988; Garety et al., 1991; Fear and Healy, 1997; Moritz and Woodward, 2005; Clark et al., 2006; Averbeck et al., 2011; Moutoussis et al., 2011). It is, however also important in other conditions such drug abuse, Parkinson’s disease and frontal lobe lesions (Djamshidian et al., 2012; Lunt et al., 2012; Averbeck et al., 2013). By varying the parameters of the generative model entertained by the agent, our approach allows us to consider how inter-individual differences can arise in the context of approximately optimal inference – and to generate hypotheses relating behavioural abnormalities to underlying neuropathology (it also allows us to recover the underlying parameters governing observed behaviour). In particular, as disorders such as schizophrenia, Parkinson’s disease and drug abuse are characterised by aberrant dopaminergic function (Dagher and Robbins, 2009; Howes and Kapur, 2009), we focus on the effects of varying expected precision (which we assume to be encoded by dopamine) on reflection impulsivity.

We have chosen to focus on the urn task as a simple paradigm that emphasises evidence accumulation and has been used extensively to study impulsivity – especially in schizophrenia research. In the urn task, subjects are told that the urn contains beads or balls that are largely of one colour. The subject is allowed to make successive draws to decide which colour predominates. In the *draws to decision* version of the task, the subject has to answer as soon as they are certain. In the *probability estimates* version, the subject can continue to draw and change their answer. Interestingly, delusional patients ‘jump to conclusions’ in the first version, while they are more willing to revise their decision in light

of contradictory evidence in the second (Garety and Freeman, 1999). Jumping to conclusions may reflect a greater influence of sensory evidence, relative to the confidence in beliefs about the (hidden) state of the urn (Speechley et al., 2010; Moutoussis et al., 2011). It is this sensitivity to confidence or precision during evidence accumulation that we wanted to understand more formally.

## Methods

### An active inference model of evidence accumulation

We previously described a variational Bayesian scheme for updating posterior beliefs based upon a generic model of (finite horizon) Markovian decision processes (Friston et al., 2013). The inversion of this model – and the biological plausibility of the belief updates and message passing required for model inversion – has previously been discussed in terms of planning and perception (Friston et al., 2013; Schwartenbeck et al., 2014). A key aspect of this scheme is that choices are sampled from posterior beliefs about control states, which (by definition) possess a confidence or precision (inverse variability). Crucially, this means that both the beliefs and their precision have to be optimised to minimise variational free energy. In previous work, we showed how this optimisation can proceed using standard (approximate Bayesian) inference schemes and how these Bayesian update schemes might be realised in the brain. As we will see below, that the expected precision shares many features of dopaminergic firing and, heuristically, can be understood as the confidence that desired outcomes will follow from the current state.

In active inference, desired outcomes or goals are simply states that, *a priori*, an agent believes they will attain. This means that a particular policy is more likely (and therefore valuable) if it minimises the difference between expected and desired outcomes. We have previously discussed this difference or divergence in terms of novelty (entropy) and expected utility – where utility is defined as the logarithm of prior beliefs about final outcomes that an agent expects (Friston et al., 2013).

To understand how this formulation of choice behaviour pertains to evidence accumulation, all we need to do is to specify the generative model for an appropriate (e.g., urn) task. Once specified, one can then use standard (variational Bayes) inversion to simulate observed behaviour and examine the dependence of that behaviour on the key parameters of the generative model (the Matlab routines required to do this are available as academic freeware from <http://www.fil.ion.ucl.ac.uk/spm>). Furthermore, we can use the model to predict observed choice behaviour and therefore estimate model parameters to phenotype any given subject in computational terms (see appendix).

**The generative model**—The general form of the model is provided in Figure 1. This figure illustrates the dependencies among hidden states  $\{s_0, \dots, s_t\} = \tilde{s}$  and control states  $\{u_t, \dots, u_T\} = \tilde{u}$  that the agent uses to predict observations or outcomes  $\{o_0, \dots, o_t\} = \tilde{o}$ . The equations on the left specify the model in terms of the joint probability over observations, the states causing these observations and  $\gamma$ , the precision of prior beliefs about these states. The form of the equations rests upon Markovian or independence assumptions

about controlled state transitions and can be expressed in terms of the likelihood (top panel), empirical priors (middle panels) and full priors (bottom panel).

The meaning of these variables is summarised on the right in terms of a directed acyclic graphical model. This illustrates the dependencies among successive hidden states and how they depend upon past and future control states. Note that future control states (alternative policies) depend upon the current state. This is because (the agent believes that) policies depend upon the relative entropy or divergence between distributions over the final state that are, and are not, conditioned on the current state (between empirical priors and full priors). The resulting choices depend upon the precision of beliefs about control states, which, in turn depend upon the parameters of the model. Finally, observed outcomes depend on, and only on, the hidden states at any given time. Given this generative model, one can simulate inferences about observed outcomes – and consequent choices – using standard (variational Bayes) schemes for approximate Bayesian inference.

Figure 2 illustrates the cognitive and functional anatomy suggested by the variational Bayesian inversion of the generative model above. Here, we have associated the Bayesian updates of expected hidden states with perception, of expected control states with action selection and, finally, expected precision with evaluation. The forms of these updates show that sufficient statistics from each set of variables are passed among each other until convergence to an internally consistent (approximately Bayes optimal) solution. In terms of neuronal implementation, this might be likened to the exchange of neuronal signals via extrinsic connections among functionally specialised brain systems. In this (purely iconic) schematic, we have associated perception (inference about the current state of the world) with the prefrontal cortex, while assigning action selection to the basal ganglia. Crucially, precision has been associated with activity in dopaminergic projections from ventral tegmental area and substantia nigra that, necessarily, project to both cortical (perceptual) and subcortical (action selection) systems. See (Friston et al., 2013) for further details.

**The ABC of evidence accumulation**—Figures 1 and 2 show that the generative model (and associated belief updating) is specified by three sorts of matrices. The first **A** constitutes the parameters of the likelihood model and encodes the probability of an outcome, given a hidden state. The second, **B**( $u_t$ ) specifies the probabilistic transitions among hidden states that depend upon the current control state. Finally, the vector **C** encodes the prior probability that a final outcome will be observed – where the logarithm of this probability corresponds to the utility of each final (hidden) state. The **Q** matrix is a function of these parameters (Friston et al., 2013) and represents the value of a particular policy from the current state – as measured by the relative entropy or divergence between the distribution over final states under that policy and the desired distribution encoded by the utility. This leads to the question of what these three sets of parameters look like for the urn task.

The first thing that needs to be resolved is the nature of hidden states. In this paper, we will assume hidden states have four attributes or dimensions that contain all the necessary information to define the utility of final states. We will assume that all attributes of hidden states are available for observation apart from the (hidden) state of the urn, which can

contain a preponderance of red or green balls. These attributes include the total number of draws, the total number of green balls, the decision (undecided, red or green) and the true state of the urn (a red urn or a green urn). Notice that the size of this state space is much smaller than the number of possible outcomes; in the sense that it grows quadratically with the number of draws (as opposed to exponentially). Equipped with this state space, it is now relatively simple to specify the **(A,B,C)** of the model:

Figure 3 illustrates these contingencies. **(A) Upper panel:** this shows the likelihood mapping from hidden states to observations. This mapping simply says that for every observed state there are two hidden states – one for the red urn and one for the green urn. **(B) Middle panel:** this shows for every control state there is a probability transition matrix. Here, there are three control states: wait, choose red or choose green. If the agent waits, the number of drawn balls increases by one and the number of green balls can either stay the same or increase. The probability of these alternatives depends upon whether the hidden state is a red or a green urn. If the agent decides, then the hidden state jumps from an undecided to a red or green decision – having seen  $k$  green balls out of  $n$  balls. **(C) Lower panel:** prior beliefs about policies or sequences of control states are based upon goals encoded by their utility.

Here, utility reflects an informed decision – defined as a log odds ratio of greater than  $\tau = 3$ , when comparing the probability of observed outcomes under the appropriate hidden state, relative to the alternative. Simple binomial theory shows that this ratio is just  $\pm(2k-n)\rho$ , where  $\rho = \ln(p) - \ln(1-p)$  reflects the imbalance in the proportion  $p$  of red or green balls in the urn (in this paper  $p = 0.85$ ). Note that when  $2k = n$  or  $\rho = 0$ , the log odds ratio is zero. In other words, when half the drawn balls are green, or the urn contains an equal number of red and green balls, the log odds ratio is always zero. Here,  $\kappa = 4$  controls the sensitivity to a significant odds ratio which is taken to mean a log odds ratio of three or more, corresponding to an odds ratio of 20 to 1 – roughly analogous to a p-value of 0.05. Having specified the generative model, with standard gamma priors on precision of  $P(\gamma|m) = \Gamma(\alpha = 4, \beta = 1)$ , we can now use the variational Bayesian updates in Figure 2 to simulate responses to any sequence of draws.

Finally, sampling costs can be incorporated into prior beliefs by subtracting a penalty  $n \cdot w$  from the utility of each state in the **C** matrix, where  $w$  is the fixed cost of drawing a further ball. We ignore sampling cost ( $w = 0$ ) unless otherwise stated.

## Results

### A simulated game

Figure 4 shows the results of simulated eight draw game: the upper right panel shows the inferred hidden states as a function of trial or draws. In this example, all the draws returned green balls apart from the fourth ball, which was red. The image bar on the right of the graph indicates those final states that have a high utility or prior probability. Here, the agent was prevented from deciding until the last draw, when it chooses green. We delayed the decision to reveal the posterior beliefs about policies shown on the upper right: the solid lines correspond to the posterior expectation of waiting (blue) or deciding (red or green respectively). The dotted lines show the equivalent beliefs at earlier draws. We see here that

beliefs about decisions fluctuate as evidence is accumulated with an increasing propensity to choose green as more green balls are drawn. Notice that an initial increase in the posterior probability of choosing green decreases after a red ball is drawn on the fourth draw; however, it recovers quickly with subsequent green draws. These probabilities are based upon the posterior expectations about policies that are shown in the middle two panels. The allowable policies are shown on the left and, in this example, considered decisions at all eight draws. The corresponding posterior expectations are shown in the middle right panel in image format. It can be seen that there is a progressive increase in the expected probability of the eighth policy – that corresponds to a green choice on the penultimate draw.

The lower panels show the expected precision over trials. The lower left panel shows the precision as a function of variational updates (eight per trial), while simulated dopamine responses are shown on the lower right. These are the updates after deconvolution with an exponentially decaying kernel of eight updates. The interesting thing about these simulated responses is that there is a positive dopamine response to the occurrence of each subsequent green ball – with the exception of the fourth (red) ball that induces a transient decrease in expected precision. Intuitively, these changes in expected precision reflect the confidence in making an informed decision, which increases with confirmatory and decreases with contradictory evidence. This is reminiscent of dopaminergic responses to the omission of expected rewards, if we regard a green ball as the expected outcome under posterior beliefs about the world. Indeed, the most pronounced (phasic) increase in precision occurs after the eighth draw – after which the agent declares its decision.

**Cognitive phenotypes**—In the above simulations, we assumed particular values for the priors on precision, the threshold  $\tau = 3$  defining an outcome as informative and other model parameters. Clearly, there are a whole range of Bayes optimal behaviours (where optimality is defined relative to the agent's generative model), depending upon the values of these parameters (Mathys et al., 2011). The premise behind this sort of modelling is that one can characterise any given subject in terms of parameters – that generally encode their prior beliefs. In this section, we look at some characteristic behavioural expressions of changes in key model parameters.

Figure 5 illustrates the effects of increasing threshold criterion for decision – top row:  $\tau = (0, \dots, 8)$  and prior precision – middle row:  $\alpha = (2, \dots, 8)$  on the propensity to make an early decision. The upper panels show the probability of deciding as a function of the number of draws (latency) when increasing the log odds ratio criterion that is regarded as having a high utility. The left panel shows the effects in terms of the probability of deciding given no previous decision, the middle panel shows the corresponding distribution over latencies to decide, and the right panel shows the cumulative probability of making a decision. Increasing the threshold or evidence required for a confident decision in general delays the decision. Interestingly however, early on in the game, agents with very high thresholds are more likely to decide than agents with intermediate thresholds. This follows from a low expected precision, because (they believe) it is unlikely that they can make a confident decision. It is relatively unlikely, for example, that all eight balls will be one colour – and the agent knows that at the beginning of the game.

This highlights the nature of precision and its intimate relationship with utility. Put simply, precision reports the confidence in optimistic beliefs about outcomes and choice. Any evidence that precludes a high utility outcome renders those optimistic beliefs untenable and therefore their precision falls. Heuristically – in terms of minimising surprise or free energy – this is like accepting a goal cannot be achieved and telling yourself you did not really know if you could have achieved it anyway.

The middle panels show the equivalent effects upon beliefs when increasing the (scale parameter of) prior precision. As confidence about policies increase, there is again a tendency to wait for more evidence. This reflects the fact that the relative utility of making an informed and correct decision becomes greater, relative to making an incorrect decision. Notice that even with very low precisions, it is always better to wait for as many draws as possible – to ensure the greatest number of possible outcomes. This reflects the entropy or novelty part of the relative entropy or divergence that makes policies valuable (Schwartenbeck et al., 2013). The effect of varying prior precision on impulsivity is particularly interesting, given the proposal that precision is encoded by dopaminergic activity in the midbrain and the fact that dopaminergic abnormalities are implicated in a number of psychopathologies (Everitt and Robbins, 2005; Howes and Kapur, 2009; Steeves et al., 2010; Adams et al., 2013).

The costs of accumulating extra information are an additional factor that agents must consider when modelling their environment. These can be implicit, for example the extra effort or attentional resources required to perform the task for longer, or explicit, as in formulations of the urn task – in which subjects are penalised for each extra bead they choose to draw. From the perspective of active inference, this penalty is simply a prior belief that the agent will decide sooner rather than later. Figure 5 (bottom row) shows the effects of increasing costs on postponing decisions. Costs were varied from  $w = (0, \dots, 1/8)$ . As expected, when the cost of accumulating more information increases, the agent tends to choose earlier.

## Discussion

We have considered evidence accumulation from the perspective of active inference using a recently proposed framework for inference within the context of Markov decision processes (Friston et al., 2013). Here, agents expect to make informed decisions, and sample information in order to fulfil this belief. We show that by simply specifying three matrices (the ABC of evidence accumulation) it is possible to simulate plausible behaviour in the widely used urn task. Our model also provides testable predictions about the relationship between the outcomes of specific trials and expected precision, which we hypothesise is encoded by activity in the dopaminergic midbrain. In addition, we briefly show how varying the parameters of an agent's generative model leads to different (but, from the agent's perspective) equally optimal patterns of behaviour. This framework allows us to consider how changes in model parameters might induce reflection impulsivity, which characterises a number of psychopathologies such as schizophrenia (Huq et al., 1988; Garety et al., 1991; Fear and Healy, 1997; Moritz and Woodward, 2005; Clark et al., 2006; Corcoran et al., 2008; Moutoussis et al., 2011).

The distinction between optimal inference conditioned on an agent's model of the environmental contingencies and that conditioned on the 'true' contingencies (as known to the experimenter) is important to stress. This distinction allows normal and pathological variations in behaviour to result from inferential processes, which are themselves equally close to optimal. This reconciles a Bayesian view of cognition with the fact that human behaviour is, in many situations, manifestly non-optimal relative to the true nature of the environment. Prior work has considered how variations in observed behaviour can result from optimal inference, when the parameters governing estimates of environmental volatility (Mathys et al., 2011; Vossel et al., 2013) and sensory precision (Adams et al., 2012; Brown et al., 2013) are altered. The simulations presented here thus join an increasingly large literature seeking to explain inter-individual differences in terms of optimal inference.

Previous modelling work has considered behaviour on the urn task in terms of Bayesian inference (Averbeck et al., 2011; Moutoussis et al., 2011), or presented approaches to evidence accumulation during perceptual decision-making that bear similarities to that taken here. Huang and Rao formulate evidence accumulation as a Partially Observable Markov Decision Process, solved using dynamical programming (in this case value iteration) (Huang and Rao, 2013). Drugowitsch et al. take a similar approach, to produce a hybrid model combining dynamic programming with drift diffusion models to capture collapsing bounds (Drugowitsch et al., 2012). Deneve also considers evidence accumulation in the light of optimal inference, focussing particularly upon the necessity to infer the reliability of sensory information (a consideration that is less pressing in the urn task that we consider here, but crucial in perceptual decision-making paradigms such as the random-dot motion task) (Deneve, 2012). Our approach differs from – and extends – these models in a number of crucial respects. By considering behaviour in terms of active inference, we were able to specify planning and behaviour as purely inferential processes, obviating the need to include a value function over and above the agent's beliefs about the world. This effaces the distinction between reward and non-reward based decision-making, as exemplified by the time course of expected precision, which shows reward prediction error-like dynamics even in absence of reward. Our model also goes beyond previous work in that we consider how the agent actually performs inference, namely through (Variational Bayesian) message passing. Variational inference is both efficient and neurobiologically plausible, and also motivates natural, if simplified, hypotheses about how the brain regions encodes different quantities in the model, an essential step in relating impulsivity to underlying neuropathology.

Variational Bayes mandates a partition between different classes of model parameters (the mean field approximation), each of which are updated only according to the expectations of the parameters in the other subsets (Attias, 2000; Beal, 2003). This motivates and explains both functional segregation (Zeki et al., 1991) (since subsets are conditionally independent of one another) and functional integration (Friston, 2002) (since updating requires message passing between subsets) (Friston, 2010). The particular structure of our model suggests a natural (though necessarily caricatured) assignment of observed, hidden and control states to sensory, prefrontal and striatal regions respectively (Figure 2). It also requires that expected precision over control states be encoded by a system that projects widely and reciprocally to



– and has a modulatory (multiplicative) effect upon – regions supporting perception and action selection (Figure 2) (Friston et al., 2014). This profile fits easily with characteristics of the dopaminergic system, which heavily innervates the striatum and prefrontal cortex (Seamans, 2007), and exerts a modulatory influence on postsynaptic potentials (Smiley et al., 1994; Krimer et al., 1997; Yao et al., 2008; Friston et al., 2012). Despite its advantages, variational inference is only an approximation to Bayes-optimal inference, and under certain conditions may diverge significantly from true optimality. This has the potential to furnish interesting predictions about behaviour, something that could usefully be explored in future work.

At the same time we note that the fundamentally interactive nature of message-passing between functional modules in our model may relate to the multiple, specific brain modules known to be affected by disorders resulting in reflection impulsivity. In psychosis research, for example, it has long been hypothesized that dopaminergic abnormalities affect, and are exacerbated by, impaired prefrontal cortical function (Laruelle, 2008). Indeed, while gross measures of reflection impulsivity such as draws-to-decision are normalized by antidopaminergic therapy (Menon et al., 2008), the finer statistical structure of decision making in the Urns task is not (Moutoussis et al., 2011). This is likely to reflect the fact that remission from psychosis in people with schizophrenia compensates function by increasing some parameters, such as the decision threshold parameter  $\kappa$ , and possibly the gross statistics of precision ( $\gamma$ ), but it may not normalize the fine structure of the cortical-subcortical interactions that would normally compute  $\gamma$ . In support of this, recent imaging and lesion studies of the Urns task (Furl and Averbeck, 2011; Lunt et al., 2012) highlighted the importance of a lateral-prefrontal – parietal network in preventing hasty decisions, even in the absence of psychosis. This was contrasted with activation of a network involving the ventral striatum and a variety of cortical areas associated with the actual decision to stop sampling further information. Along similar lines, relatively subtle hasty decision making in the Urns task in Parkinson's disease has been attributed to ventral striatal hyperdopaminergia while more severe deficits associated with impulsive-compulsive syndromes in parkinsonism have been attributed to cortico (-striatal) pathology (Djamshidian et al., 2012).

The hypothesis that dopamine encodes expected precision is also supported by the characteristic time course of expected precision during evidence accumulation, which shows a pattern similar to that observed in dopaminergic neurons in response to omitted and expected reward (Schultz et al., 1997; Cohen et al., 2012). Consistent with this, it has been shown that activity in the ventral tegmental area increases during performance of the urn task – relative to a control condition – and that this activity peaks at the time of decision (resembling the peak phasic activity just prior to choice in our model (Figure 4)) (Esslinger et al., 2013). We have also recently shown, in the context of a limited offer game, that midbrain activity strongly reflects trial-by-trial fluctuations in expected precision (Schwartenbeck et al., 2014). A large body of work seeks to explain (phasic) dopamine function in terms of reinforcement learning, and reward prediction errors in particular (Schultz et al., 1997; Frank et al., 2004; O'Doherty et al., 2004; Waltz et al., 2007) though see (Shiner et al., 2012). Given that the urn task involves inference (on hidden states) rather than learning (about model parameters), a putative role of dopamine in learning is unlikely

to be of direct relevance here. Moreover, although expected precision will undoubtedly influence learning as well as inference, the model considered here does not treat learning, and it is thus difficult to draw direct comparisons with existing literature on dopamine and reward learning. This is an issue we intend to explore in future, by incorporating parameter learning within our framework.

A link between dopamine and precision is of particular interest here given that both dopaminergic abnormalities and increased reflection impulsivity are known to be features of schizophrenia (Moritz and Woodward, 2005; Howes and Kapur, 2009; Moutoussis et al., 2011). In our scheme, reducing prior precision leads to earlier decisions (Figure 5), which suggests that impaired dopaminergic neuromodulation may underlie the increased impulsivity observed in patients with schizophrenia. To our knowledge this is the first hypothesis which provides a mechanistic explanation for how impulsivity emerges out of a likely pathophysiology in schizophrenia. We intend to explore this in future empirical work, but note that explaining increased impulsivity in terms of decreased precision fits nicely with previous work by Moutoussis et al. These authors suggested that increased impulsivity in schizophrenic patients arises not from increased sensitivity to the costs of gathering new information, but rather from increased stochasticity (what they describe as ‘cognitive noise’) in planning and choice behaviour (Moutoussis et al., 2011). This is exactly what our scheme predicts. Our findings also bear a striking resemblance to previous work suggesting that behavioural differences across three different tasks between impulsive and non-impulsive Parkinson’s disease patients can be explained by increasing uncertainty about obtaining future rewards (Averbeck et al., 2013). In one sense, the main contribution of the work presented here is to provide a principled framework within which to understand findings of this sort – and to derive plausible and testable hypotheses about the pathophysiology that may underlie them.

An interesting phenomenon highlighted by our simulations is the degeneracy of reflection impulsivity, when this is simply taken to mean hasty decision-making. Since the amount of information sampled prior to making a decision can be altered by several parameters in the agent’s generative model (Figure 5), there is no one-to-one mapping between reflection impulsivity as a behavioural trait and the parameters of the model. This suggests that understanding the mechanisms that give rise even to an apparently simple trait such as reflection impulsivity necessarily requires careful model-based analysis of both behaviour and its neural correlates. Such model-based analysis is likely to become increasingly important both for characterising the mechanisms that give rise to inter-individual differences in behaviour, and for improving diagnostic criteria in psychiatry (Montague et al., 2012).

It is interesting to consider whether the precise patterns of behaviour underlying reflection impulsivity are similar across different patient groups (for example, in psychosis (Moutoussis et al., 2011), Parkinson’s disease (Djamshidian et al., 2012; Averbeck et al., 2013) and other disorders (Lunt et al., 2012)). To the extent that these disorders are characterised by altered dopaminergic function, our model predicts similar behavioural changes. However, it is also possible that specific disease features play a role, particularly given the multiplicity of potential routes to impulsivity discussed above corresponding to the

differences, rather than similarities, between the panel rows of Fig. 5. In fact, the particular strength in inverting individual (generative) models of a task lies in the potential identification of characteristic parameter configurations of a specific patient group that may induce the same overt behaviour. More specifically, the underlying causes of impulsivity in patients with Parkinson's disease, addiction or psychosis are likely to differ, even though the behavioural consequences in a task may be the same. In the particular setting of this model, this translates into the fact that on the one hand, impulsive and hasty behaviour can have various causes such as decreased prior precision or negative expectations about transition probabilities and, on the other hand, aberrant precision can have different effects depending on the context, such as on action selection, perception or policy. Addressing this requires fitting the model to data collected from different patient groups, which we hope to carry out in the near future. A similar comment applies to the relationship between delusions and impulsivity (Fear and Healy, 1997; Moritz and Woodward, 2005; Menon et al., 2008), given that delusions are common in schizophrenia patients, but less common in other groups that also show increased impulsivity. Delusions might depend upon the dopaminergic abnormalities also found in other disorders, but emerge only in the particular context of a schizophrenia pathology. Alternatively, there may be specific mechanisms underlying delusions that relate to (reflection) impulsivity in ways that are indirectly related to dopaminergic function. For example, reduced precision at the relevant cognitive level might contribute to down-grading evidence against a delusion which otherwise has great personal significance.

In general, the active inference framework considered here explains stochastic action selection as resulting from agents seeking to minimise surprise (variational free energy), rather than maximise the value of their actions. This means that the precision term which governs the sensitivity of action to the utility of different policies must itself be optimised, based on an agent's confidence in attaining desirable states (an extremely large precision term would lead the agent to deterministically select the action with greatest utility) (Friston et al., 2013). This provides a principled explanation for stochastic choice behaviour which does not need recourse to noise or temperature parameters in the decision rule governing choice (Daw et al., 2006; FitzGerald et al., 2009). In the context of the urn task, free energy minimisation explains why agents do not always choose to sample until the last possible opportunity, even when the cost of acquiring more evidence is set to zero (Figure 5).

One limitation of our approach is that it only considers processes that occur in discrete time, whereas the brain, as a dynamical system, functions in continuous time. This criticism can be mitigated by the observation that the solutions that variational updates converge to the same solutions that would be found through gradient descent methods. These gradient descent schemes can be implemented as (biologically plausible) Bayesian filtering or predictive coding (Mumford, 1992; Friston, 2005; Friston et al., 2007, 2008). This means that while our discrete time models are further abstracted from brain function than models we have considered elsewhere (for example (Adams et al., 2012; Friston et al., 2012)) they are still capable of embodying mechanistic hypotheses about cognition, which is essential if we want to relate model parameters to neurobiology. Discrete versus continuous time formulations of evidence accumulation are important because evidence accumulation is a general problem in inference – that goes beyond the discrete trial, binomial task considered

here. In future work, we will consider how evidence accumulation might be performed through free energy minimisation when the agent is presented with a stream of continuous, graded evidence as in the random-dot motion paradigm (Gold and Shadlen, 2007).

In summary, we have considered evidence accumulation on the urn task in terms of active inference using a generic variational scheme. By specifying a generative model of the agent's beliefs about the causal structure of the environment – together with its beliefs about which states it expects to end up in – we were able to simulate plausible decision-making behaviour. In addition, the partitioning of model parameters required by variational Bayes naturally motivates hypotheses about which brain regions or systems might encode particular parameters. Expected precision, which we hypothesise is encoded by dopaminergic firing, plays a key role in modulating impulsivity, leading one to hypothesise that dopaminergic dysfunction may underlie the behavioural changes observed in schizophrenia, and perhaps other disorders. Our simulations demonstrate the power and flexibility of the active inference framework and generate plausible hypotheses for future investigation.

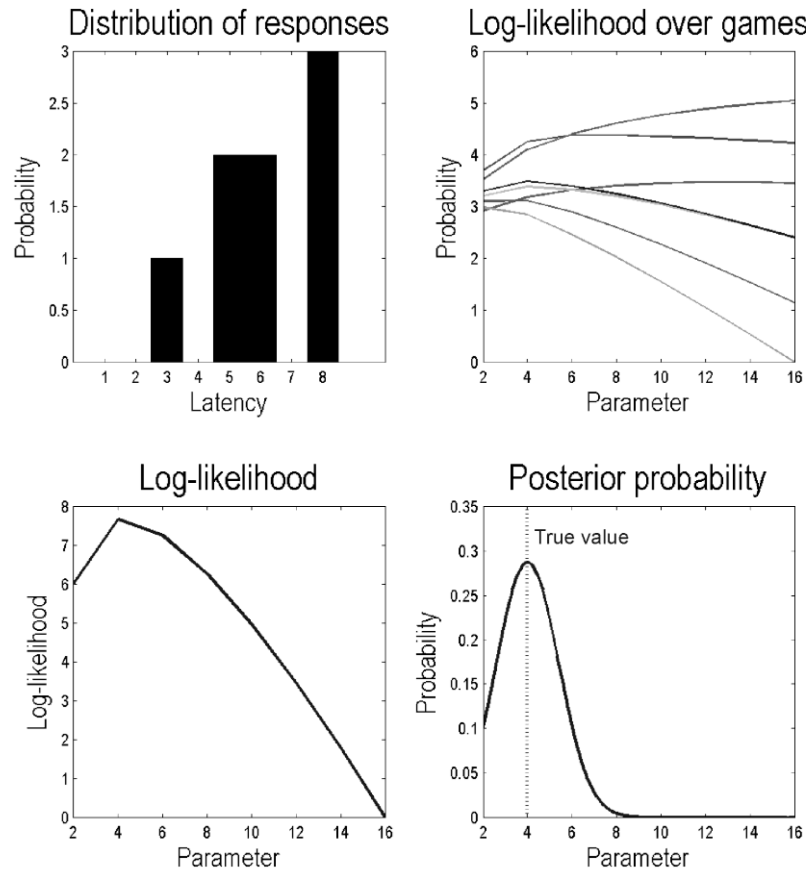
## Acknowledgements

This work was supported by Wellcome Trust Senior Investigator Awards to KJF [088130/Z/09/Z] and RJD [098362/Z/12/Z]; The Wellcome Trust Centre for Neuroimaging is supported by core funding from Wellcome Trust Grant 091593/Z/10/Z.

## Appendix

An important application of the Markovian scheme described in the main text is to parameterise or phenotype individual subjects in terms of key model parameters that have both a computational and putative neurobiological meaning. This computational phenotyping rests on estimating model parameters from observed behaviour. This is relatively simple to do with Markovian models because they provide the probability of any allowable action at any point in time. This means the probability of observed choices can be computed for any given model parameter, using simulations of the sort described above. By searching over ranges of parameters the ensuing log likelihood functions serve to estimate the model parameters that best explain a particular subject's behaviour.

Figure A1 shows the results of a simulated estimation of model parameters. In this example, eight games were simulated and the outcomes (and choices) were used to evaluate the probability of choice behaviour under different values of prior precision  $\alpha = (2, \dots, 16)$ . The observed distribution of latencies over the eight games is shown on the upper left, while the corresponding log likelihood of obtaining the choice behaviour is shown on the lower left – as a function of the prior precision, for each of the eight trials. Because the outcomes from the eight games were conditionally independent, their log likelihood is the sum of log likelihoods from each game. Under the (Laplace) assumption that the posterior distribution over prior precision is Gaussian, one can now use the log likelihood to estimate the posterior expectation and precision of the model parameter. In this case, the scale parameter of the priors over precision and had a true value of four (vertical line) – which was estimated accurately.



**Figure A1.**

Results of a simulated estimation of model parameters. In this example, eight games were simulated and the outcomes (and choices) were used to evaluate the decision probability under different values of prior precision. *Upper left*: the observed distribution of latencies over the eight games. *Upper right*: the log likelihood of obtaining the choice behaviour under different levels of the prior precision (from 2 to 16), with one function for each games. *Lower left*: because the choices from the eight games were conditionally independent, the total log likelihood is the sum of log likelihoods from each game. Under the (Laplace) assumption that the posterior distribution over prior precision is Gaussian, one can use the log likelihood to estimate the posterior expectation and precision of the model parameter. In this case, the scale parameter of the priors over precision had a true value of eight (vertical line).

## References

- Adams RA, Perrinet LU, Friston K. Smooth pursuit and visual occlusion: active inference and oculomotor control in schizophrenia. *PLoS One*. 2012; 7:e47502. [PubMed: 23110076]
- Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. *Front psychiatry*. 2013; 4:47. [PubMed: 23750138]
- Attias H. A variational Bayesian framework for graphical models. *Adv Neural Inf Process Syst*. 2000; 12:209–215.

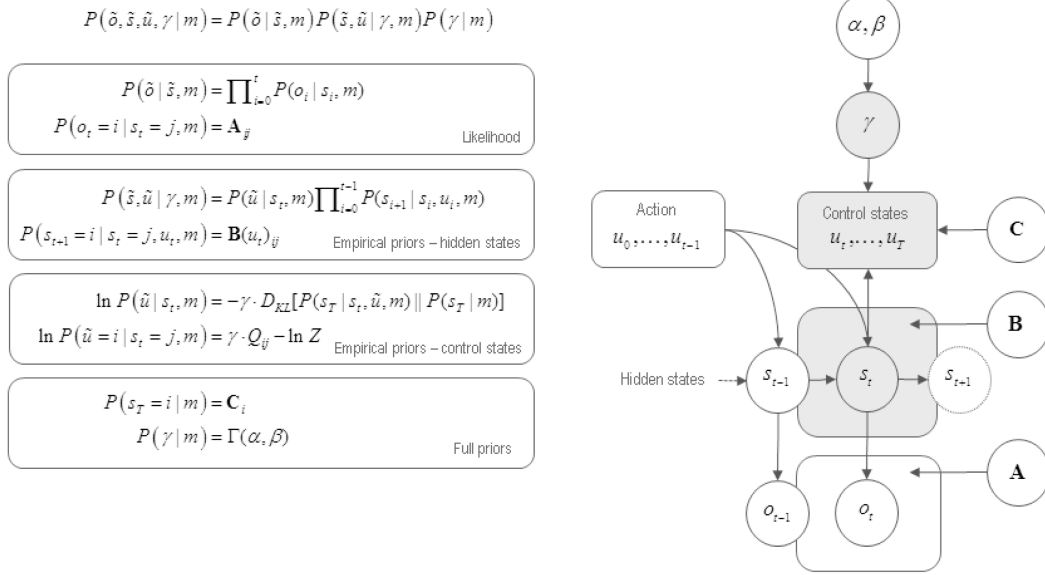
- Averbeck BB, Djamshidian A, O'Sullivan SS, Housden CR, Roiser JP, Lees AJ. Uncertainty about mapping future actions into rewards may underlie performance on multiple measures of impulsivity in behavioral addiction: evidence from Parkinson's disease. *Behav Neurosci.* 2013; 127:245–255. [PubMed: 23565936]
- Averbeck BB, Evans S, Chouhan V, Bristow E, Shergill SS. Probabilistic learning and inference in schizophrenia. *Schizophr Res.* 2011; 127:115–122. [PubMed: 20810252]
- Beal, MJ. Variational algorithms for approximate Bayesian inference. 2003.
- Brown H, Adams RA, Parees I, Edwards M, Friston K. Active inference, sensory attenuation and illusions. *Cogn Process.* 2013; 14:411–427. [PubMed: 23744445]
- Clark L, Robbins TW, Ersche KD, Sahakian BJ. Reflection Impulsivity in Current and Former Substance Users. *Biol Psychiatry.* 2006; 60:515–522. [PubMed: 16448627]
- Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature.* 2012; 482:85–88. [PubMed: 22258508]
- Corcoran R, Rowse G, Moore R, Blackwood N, Kinderman P, Howard R, Cummins S, Bentall RP. A transdiagnostic investigation of “theory of mind” and “jumping to conclusions” in patients with persecutory delusions. *Psychol Med.* 2008; 38:1577–1583. [PubMed: 18005499]
- Dagher A, Robbins TW. Personality, Addiction, Dopamine: Insights from Parkinson's Disease. *Neuron.* 2009; 61:502–510. [PubMed: 19249271]
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. *Nature.* 2006; 441:876–879. [PubMed: 16778890]
- Deneve S. Making decisions with unknown sensory reliability. *Front Neurosci.* 2012; 6:75. [PubMed: 22679418]
- Djamshidian A, O'Sullivan SS, Sanotsky Y, Sharman S, Matviyenko Y, Foltynie T, Michalczuk R, Aviles-Olmos I, Fedoryshyn L, Doherty KM, Filts Y, Selikhova M, Bowden-Jones H, Joyce E, Lees AJ, Averbeck BB. Decision making, impulsivity, and addictions: do Parkinson's disease patients jump to conclusions? *Mov Disord.* 2012; 27:1137–1145. [PubMed: 22821557]
- Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A. The cost of accumulating evidence in perceptual decision making. *J Neurosci.* 2012; 32:3612–3628. [PubMed: 22423085]
- Esslinger C, Braun U, Schirrmbeck F, Santos A, Meyer-Lindenberg A, Zink M, Kirsch P. Activation of midbrain and ventral striatal regions implicates salience processing during a modified beads task. *PLoS One.* 2013; 8:e58536. [PubMed: 23484034]
- Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci.* 2005; 8:1481–1489. [PubMed: 16251991]
- Fear CF, Healy D. Probabilistic reasoning in obsessive-compulsive and delusional disorders. *Psychol Med.* 1997; 27:199–208. [PubMed: 9122300]
- FitzGerald THB, Seymour B, Dolan RJ. The role of human orbitofrontal cortex in value comparison for incommensurable objects. *J Neurosci.* 2009; 29:8388–8395. [PubMed: 19571129]
- Frank MJ, Seeberger LC, O'reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science (80- ).* 2004; 306:1940–1943.
- Friston K. Functional integration and inference in the brain. *Prog Neurobiol.* 2002; 68:113–143. [PubMed: 12450490]
- Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci.* 2005; 360:815–836. [PubMed: 15937014]
- Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free energy and the Laplace approximation. *Neuroimage.* 2007; 34:220–234. [PubMed: 17055746]
- Friston K, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ. The anatomy of choice: active inference and agency. *Front Hum Neurosci.* 2013; 7:598. [PubMed: 24093015]
- Friston K, Trujillo-Barreto N, Daunizeau J. DEM: a variational treatment of dynamic systems. *Neuroimage.* 2008; 41:849–885. [PubMed: 18434205]
- Friston KJ. The free-energy principle: a unified brain theory? *Nat Rev Neurosci.* 2010; 11:127–138. [PubMed: 20068583]
- Friston KJ, Schwartenbeck P, FitzGerald THB, Moutoussis M, Behrens TEJ, Dolan RJ. The anatomy of choice: a free energy treatment of decision-making. *Philos Trans R Soc Lond B Biol Sci.* 2014

- Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan KE, Bestmann S. Dopamine, affordance and active inference. *PLoS Comput Biol.* 2012; 8:e1002327. [PubMed: 22241972]
- Furl N, Averbeck BB. Parietal cortex and insula relate to evidence seeking relevant to reward-related decisions. *J Neurosci.* 2011; 31:17572–17582. [PubMed: 22131418]
- Garety PA, Freeman D. Cognitive approaches to delusions: A critical review of theories and evidence. *Br J Clin Psychol.* 1999; 38:113–154. [PubMed: 10389596]
- Garety PA, Hemsley DR, Wessely S. Reasoning in deluded schizophrenic and paranoid patients. Biases in performance on a probabilistic inference task. *J Nerv Ment Dis.* 1991; 179:194–201. [PubMed: 2007889]
- Gold JI, Shadlen MN. The neural basis of decision making. *Annu Rev Neurosci.* 2007; 30:535–574. [PubMed: 17600525]
- Howes OD, Kapur S. The dopamine hypothesis of schizophrenia: version III--the final common pathway. *Schizophr Bull.* 2009; 35:549–562. [PubMed: 19325164]
- Huang Y, Rao RPN. Reward optimization in the primate brain: a probabilistic model of decision making under uncertainty. *PLoS One.* 2013; 8:e53344. [PubMed: 23349707]
- Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MFS, Behrens TEJ. Mechanisms underlying cortical activity during value-guided choice. *Nat Neurosci.* 2012; 15:470–6. S1–3. [PubMed: 22231429]
- Huq SF, Garety PA, Hemsley DR. Probabilistic judgements in deluded and non-deluded subjects. *Q J Exp Psychol A.* 1988; 40:801–812. [PubMed: 3212213]
- Kagan J. Reflection-impulsivity: The generality and dynamics of conceptual tempo. *J Abnorm Psychol.* 1966; 71:17–24. [PubMed: 5902550]
- Krimer LS, Jakab RL, Goldman-Rakic PS. Quantitative three-dimensional analysis of the catecholaminergic innervation of identified neurons in the macaque prefrontal cortex. *J Neurosci.* 1997; 17:7450–7461. [PubMed: 9295391]
- Laruelle, M. Persecutory Delusions: Assessment, Theory and Treatment. In: Freeman, D.; Bentall, R.; Garety, P., editors. *Persecutory delusions.* Oxford University Press; Oxford: 2008. p. 239–266.
- Lunt L, Bramham J, Morris RG, Bullock PR, Selway RP, Xenitidis K, David AS. Prefrontal cortex dysfunction and “Jumping to Conclusions”: bias or deficit? *J Neuropsychol.* 2012; 6:65–78. [PubMed: 22257612]
- Mathys C, Daunizeau J, Friston KJ, Stephan KE. A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci.* 2011; 5:39. [PubMed: 21629826]
- Menon M, Mizrahi R, Kapur S. “Jumping to conclusions” and delusions in psychosis: relationship and response to treatment. *Schizophr Res.* 2008; 98:225–231. [PubMed: 17897811]
- Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci.* 2012; 16:72–80. [PubMed: 22177032]
- Moritz S, Woodward TS. Jumping to conclusions in delusional and non-delusional schizophrenic patients. *Br J Clin Psychol.* 2005; 44:193–207. [PubMed: 16004654]
- Moutoussis M, Bentall RP, El-Deredy W, Dayan P. Bayesian modelling of Jumping-to-Conclusions bias in delusional patients. *Cogn Neuropsychiatry.* 2011; 16:422–447. [PubMed: 21480015]
- Moutoussis M, Fearon P, El-Deredy W, Dolan RJ, Friston KJ. Bayesian inferences about the Self (and Others): A Review. *Conscious Cogn.* 2014; 25:67–76. [PubMed: 24583455]
- Mumford D. On the computational architecture of the neocortex. *Biol Cybern.* 1992; 66:241–251. [PubMed: 1540675]
- O’Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science.* 2004; 304:452–454. [PubMed: 15087550]
- Schultz W, Dayan P, Montague PR. A Neural Substrate of Prediction and Reward. *Science (80- ).* 1997; 275:1593–1599.
- Schwartenbeck P, FitzGerald T, Dolan RJ, Friston K. Exploration, novelty, surprise, and free energy minimization. *Front Psychol.* 2013; 4:710. [PubMed: 24109469]
- Schwartenbeck P, FitzGerald THB, Mathys C, Dolan RJ, Friston KJ. The dopaminergic midbrain encodes the confidence that goals can be reached. *Cereb Cortex.* 2014

- Seamans J. Dopamine anatomy. *Scholarpedia*. 2007; 2:3737.
- Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P, Dolan RJ. Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. *Brain*. 2012; 135:1871–1883. [PubMed: 22508958]
- Smiley JF, Levey AI, Ciliax BJ, Goldman-Rakic PS. D1 dopamine receptor immunoreactivity in human and monkey cerebral cortex: predominant and extrasynaptic localization in dendritic spines. *Proc Natl Acad Sci U S A*. 1994; 91:5720–5724. [PubMed: 7911245]
- Speechley WJ, Whitman JC, Woodward TS. The contribution of hypersalience to the “jumping to conclusions” bias associated with delusions in schizophrenia. *J Psychiatry Neurosci*. 2010; 35:7–17. [PubMed: 20040242]
- Steeves TDL, Ko JH, Kideckel DM, Rusjan P, Houle S, Sandor P, Lang AE, Strafella AP. Extrastriatal dopaminergic dysfunction in tourette syndrome. *Ann Neurol*. 2010; 67:170–181. [PubMed: 20225192]
- Vossel S, Mathys C, Daunizeau J, Bauer M, Driver J, Friston KJ, Stephan KE. Spatial Attention, Precision, and Bayesian Inference: A Study of Saccadic Response Speed. *Cereb Cortex*. 2013 doi: 10.1093.
- Waltz JA, Frank MJ, Robinson BM, Gold JM. Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biol Psychiatry*. 2007; 62:756–764. [PubMed: 17300757]
- Yang T, Shadlen MN. Probabilistic reasoning by neurons. *Nature*. 2007; 447:1075–1080. [PubMed: 17546027]
- Yao W-D, Spealman RD, Zhang J. Dopaminergic signaling in dendritic spines. *Biochem Pharmacol*. 2008; 75:2055–2069. [PubMed: 18353279]
- Zeki S, Watson J, Lueck C. A direct demonstration of functional specialization in human visual cortex. *J Neurosci*. 1991; 11:641–649. [PubMed: 2002358]

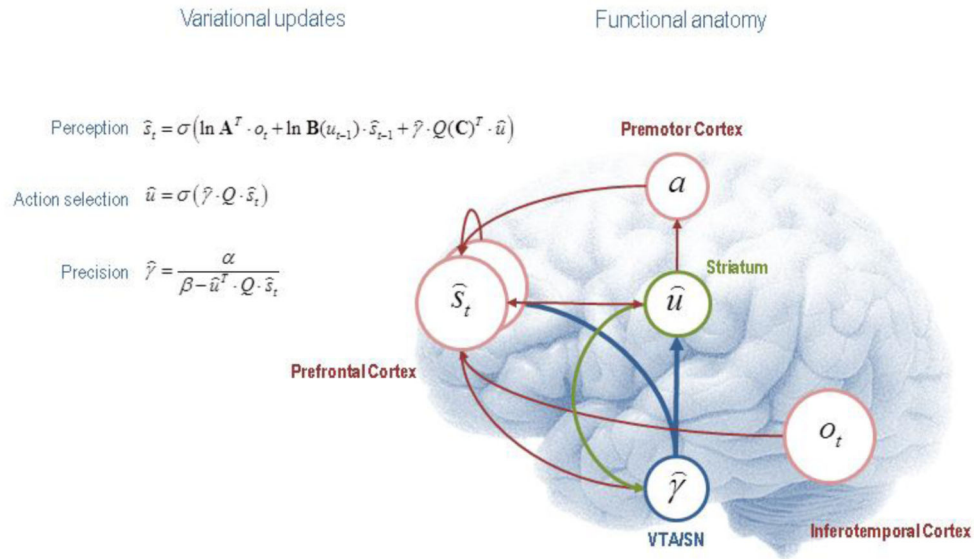


The generative model

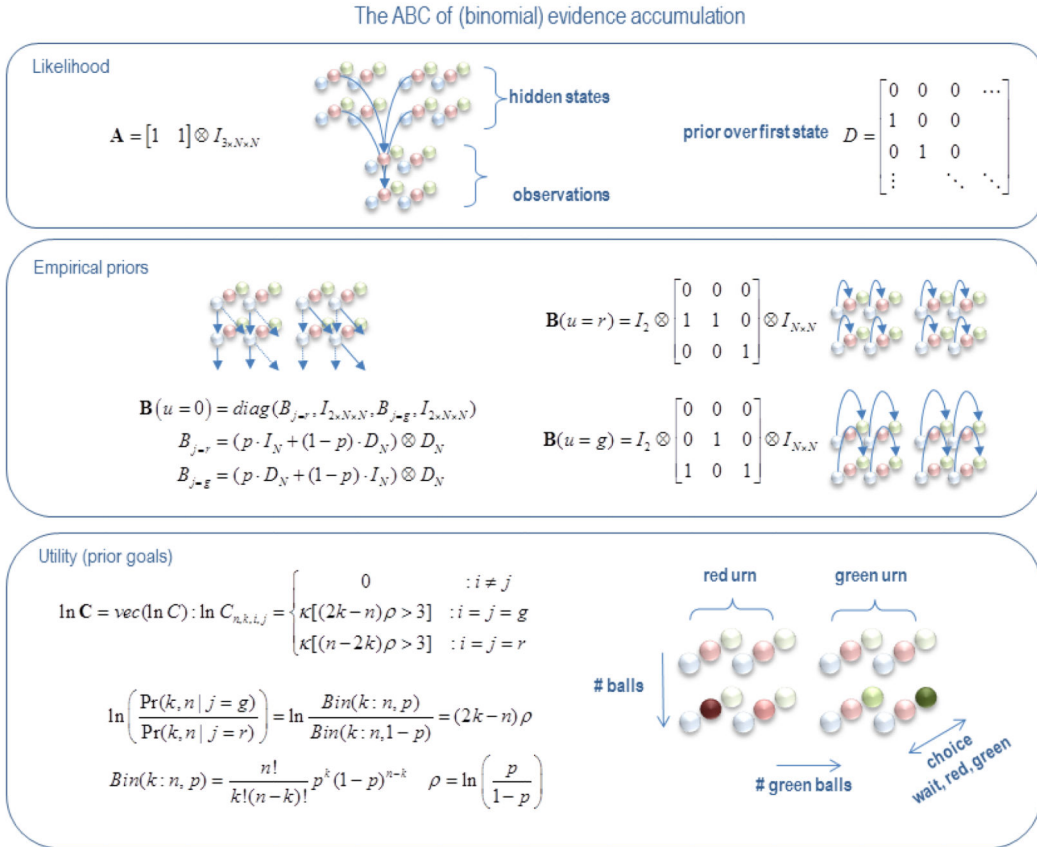


**Figure 1.**

This figure illustrates the dependencies among hidden states and control states of a generative model of choice behaviour. *Left panel:* these equations specify the generative model in terms of the joint probability over observations, the states causing these observations and the precision of prior beliefs about those states. The form of these equations rests upon Markovian or independence assumptions about controlled state transitions. *Right panel:* The dependencies among these variables are depicted on the right. This Bayesian graph illustrates the dependencies among successive hidden states and how they depend upon past and future control states. Note that future control states (policies) depend upon the current state, because policies depend upon relative entropy or divergence between distributions over the final state that are, and are not, conditioned on the current state. The resulting choices depend upon the precision of beliefs about control states, which in turn depend upon the parameters of the model. Observed outcomes depend on, and only on, the hidden states at any given time. Given this generative model, an agent can make inferences about observed outcomes using standard (variational Bayes) schemes for approximate Bayesian inference (see next figure). See (Friston et al., 2013) for further details.

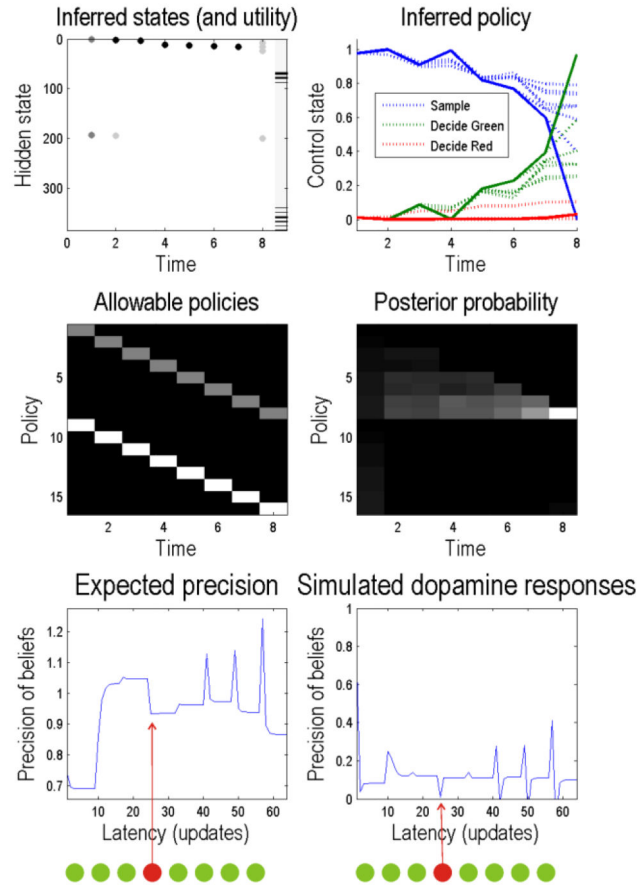
**Figure 2.**

This figure illustrates the cognitive and functional anatomy implied by the variational Bayesian inversion of the generative model in the previous figure. Here, we have associated the Bayesian updates of expected states with perception, of future control states (policies) within action selection and, finally, expected precision with evaluation. The forms of these updates show that sufficient statistics from each set of variables are passed among each other until convergence to an internally consistent (Bayes optimal) solution. In terms of neuronal implementation, this might be likened to the exchange of neuronal signals via extrinsic connections among functionally specialised brain systems. In this schematic, we have associated perception (inference about the current state of the world) with the prefrontal cortex, while assigning action selection to the basal ganglia. Crucially, precision has been associated with dopaminergic projections from ventral tegmental area and substantia nigra that, necessarily, project to both cortical (perceptual) and subcortical (action selection) systems. See (Friston et al., 2013) for further details. The precise form of this model, and the anatomical mappings implied in the figure, are not meant to be definitive, but we present them to illustrate the underlying plausibility of our simple scheme.



**Figure 3.**

This figure illustrates the construction of the model for the urn task. As described in the main text, this has three main components. *Upper panel:* the likelihood maps from the hidden states to observations. Here, we assume that all attributes of hidden states are available for observation apart from the state of the urn, which can contain a preponderance of red or green balls. This means that for every observed state there are two hidden states – one for a red urn and one for the green urn. *Middle panel:* for every control state there is a probability transition matrix. Here there are three control states wait, or choose red and green. If the agent waits, the number of drawn balls increases by one and the number of green balls increases (or not) probabilistically, depending upon whether the hidden state corresponds to a red or green urn. If the agent makes a decision, then the hidden state jumps from an undecided to declaring red or green, having witnessed the previous outcomes. *Lower panel:* prior beliefs about policies or sequences of control states are based upon goals encoded by the utility. Here, the utility reflects a confident decision – defined as a log odds ratio of greater than three, when comparing the accumulated evidence under the appropriate hidden urn state, relative to the other. Here,  $\otimes$  denotes a Kronecker tensor product. The graphics depict a subset of hidden states encoding the number of drawn balls, the number of green balls, the choice (undecided, red or green) and the hidden state of the urn (red or green).

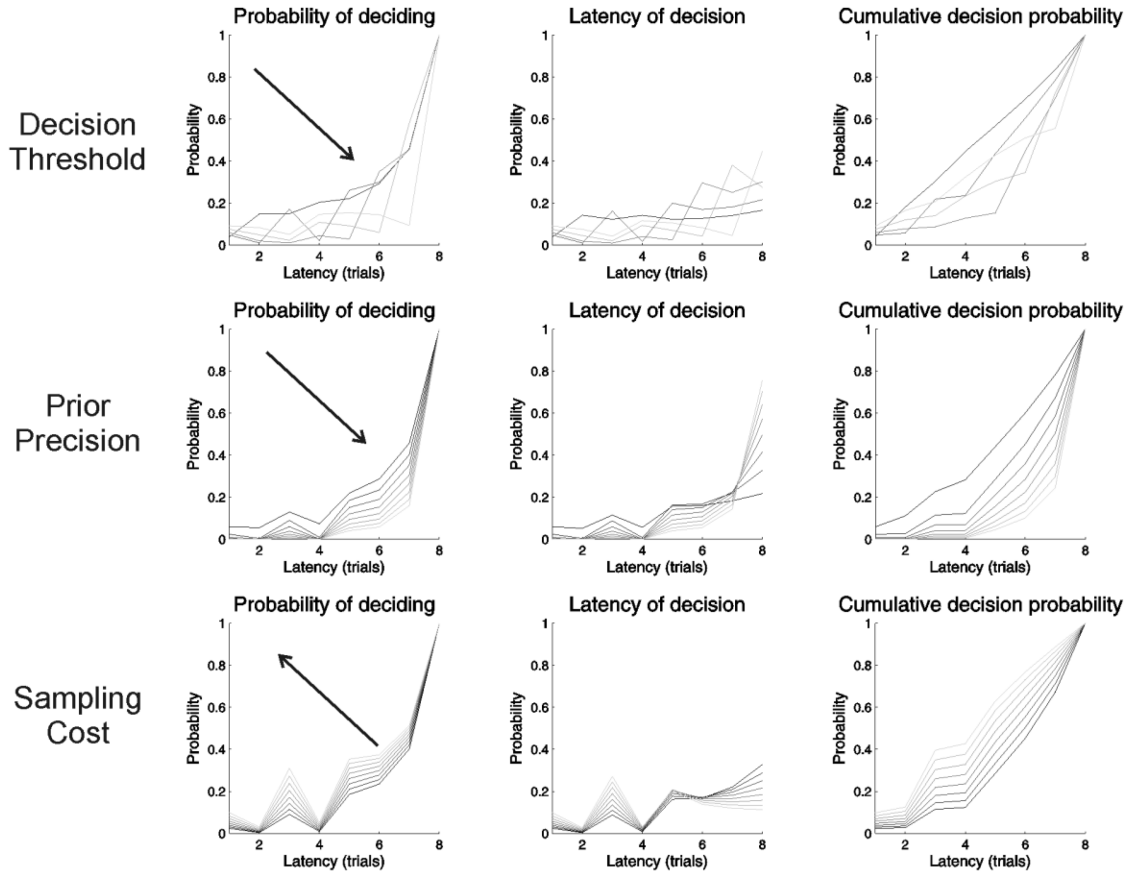


**Figure 4.**

This figure shows the results of a simulation of an eight draw game. *Upper left*: these show the inferred hidden states as a function of trial or draws. In this example, all the drawings returned green balls, apart from the fourth ball which was red. The image scale on the right of the graph indicates those final states that have a high utility or prior probability. Here, the agent was prevented from deciding until the last draw, where it chooses green. We delayed the decision to reveal the posterior beliefs about policies shown on the *upper right*: the solid lines correspond to the posterior expectation of waiting (blue) or deciding (red or green respectively). The dotted lines show the equivalent beliefs at earlier times. These probabilities are based upon the posterior expectations about policies shown in the middle two panels. *Middle left*: these are the allowable policies which, in this example, allow a decision at each of the eight trials. The corresponding posterior expectations are shown in the middle right panel in image format. It shows a progressive increase in the eighth policy that corresponds to a green choice on the final trial. The lower panels show the expected precision over trials. *Lower left* shows the precision ( $\gamma$ ) as a function of variational updates (eight per trial). Simulated dopamine responses are shown on the lower right. These are the updates after deconvolution with an exponentially decaying kernel of eight updates. Both expected precision and simulated dopamine responses show increases after sampling information reinforcing currently held beliefs (here drawing a green ball after the first trial),

and a decrease when sampling conflicting information (when a red ball is drawn on trial four).

Cognitive phenotypes



**Figure 5.** these simulated results illustrate the effects of increasing the threshold criterion for decision (top row), prior precision (middle row) and sampling cost (bottom row) on the propensity to make an early decision. *Upper panels:* these show the probability of deciding as a function of the number of draws (latency) for increasing the decision threshold (from 0 to 8). The left panel shows the effects in terms of the probability of deciding, given no previous decision, the middle panel shows the corresponding distribution over latencies to a choice, and the right panel shows the cumulative probability of making a decision. (Lower thresholds are indicated by darker grey lines). Increasing the threshold or evidence required for a confident decision effectively delays the decision. *Middle panels:* these show the equivalent results when increasing the scale parameter of the prior beliefs about precision (from 2 to 16). (Lower prior precision values are indicated by darker grey lines) As confidence in policy selection increases, there is again a tendency to wait for more evidence. *Bottom panels:* these show the results when increasing the cost to sample  $w$  (from 0 to  $1/8$ ). (Lower

sampling costs are indicated by darker grey lines) As sampling costs increase, the agent tends to decide at earlier time points.