Retrospective inference as a form of bounded rationality, and its beneficial influence on learning

Thomas HB FitzGerald[1,2,3],Timothée Devaux[2,4], Will D. Penny[1,2], Heidi M Bonnici[1], Rick A Adams[2,3,5]

1. *School of Psychology, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, UK*
2. *The Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London, WC1N 3BG, UK*
3. *Max Planck-UCL Centre for Computational Psychiatry and Ageing Research, Russell Square House, London, WC1B 5EH, UK*
4. *École normale supérieure, 45 rue d'Ulm, F-75230 Paris cedex 05, France*
5. *Department of Computer Science, University College London, Malet Place, London, WC1E 6BT, UK*

Correspondence: Thomas FitzGerald

Department of Psychology, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, UK

E-mail: t.fitzgerald@uea.ac.uk.

Telephone: +44 1603 59 2270

Abstract

Probabilistic models of cognition typically assume that agents make inferences about current states by combining new sensory information with fixed beliefs about the past, an approach known as Bayesian filtering. This is computationally parsimonious, but, in general, leads to suboptimal beliefs about past states, since it ignores the fact that new observations typically contain information about the past as well as the present. This is disadvantageous both because knowledge of past states may be intrinsically valuable, and because it impairs learning about fixed or slowly changing parameters of the environment. For these reasons, in offline data analysis it is usual to infer on every set of states using the entire time series of observations, an approach known as (fixed-interval) Bayesian smoothing. Unfortunately, however, this is impractical for real agents, since it requires the maintenance and updating of beliefs about an ever-growing set of states. We propose an intermediate approach, finite retrospective inference (FRI), in which agents perform update beliefs about a limited number of past states. (Formally, this represents online fixed-lag smoothing with a sliding window.) This can be seen as a form of bounded rationality in which agents seek to optimise the accuracy of their beliefs subject to computational and other resource costs. We show through simulation that this approach has the capacity to significantly increase the accuracy of both inference and learning, using a simple variational scheme applied to both randomly generated hidden Markov models (HMMs), and a specific application of the HMM, in the form of the widely used probabilistic reversal task. We then demonstrate, using behavioural data from 96 subjects performing a reversal task, clear evidence that subjects do indeed perform retrospective inference, as revealed through its effect on parameter learning. Our proposal thus constitutes a theoretical contribution to normative accounts of bounded rationality, which makes testable empirical predictions, of which we provide a preliminary validation.

Introduction

To behave adaptively, agents need to continuously update their beliefs about present states of the world using both existing knowledge and incoming sensory information, a process that can be formalised according to the principles of probabilistic inference [1,2]. This simple insight has generated a large field of inquiry than spans most areas of the mind and brain sciences and seeks to build probabilistic accounts of cognition [3–8].

In this paper, we take this framework for granted, and consider an important and related problem, that of using new sensory information to update beliefs about the past. This is important because, under conditions of uncertainty, new observations can contain significant information about past states as well as present ones.

In offline cognition or data analysis (in which agents are dealing with complete data sets, and are not required to respond to them in real time), it is possible to make inferences about all time points simultaneously (Figure 1).

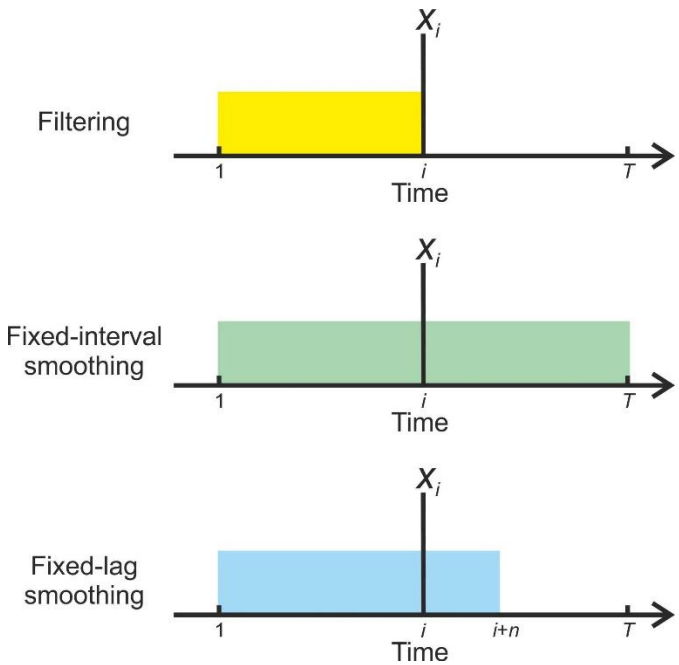

Figure 1 Illustration of different strategies for inference. In filtering (top, yellow) beliefs about state x at time i are based solely on observations made up to and including that time ($\mathbf{o}_{1:i}$), and are not revised in the light of subsequent information. In fixed-interval smoothing (middle, green), observations from the whole data set ($\mathbf{o}_{1:T}$) are used to inform each set of beliefs. In fixed-lag smoothing (bottom, blue), beliefs are retrospectively updated up to some fixed lag n , so $\mathbf{o}_{1:i+n}$ are used. (See Methods for a more formal description, and explanation of the notation) Fixed-lag smoothing allows an agent to perform finite retrospective inference, which constitutes a principled trade-off between the reduced inferential accuracy resulting from filtering and the potentially severe computational costs of retrospection to an indefinite temporal depth.

In other words, one uses every observation to inform every belief about hidden states. This option is unavailable to real, embodied agents because they need to perceive and act in time (online). They thus need to perform retrospective inference to increase the accuracy of their beliefs about the past. To perform retrospective inference optimally (or, equivalently in this context, to be strictly rational) it

is necessary for an agent to update beliefs about a sequence of states stretching backwards to the beginning of the current task or context, or perhaps even to the beginning of its existence. This sequence is both indefinitely long and constantly growing, and representing and updating these beliefs will thus, in many situations, place intolerable demands on any real organism.

We propose an alternative approach, finite retrospective inference (FRI), in which agents update beliefs about states falling within a limited temporal window stretching into the past [9]. Selecting the size of this window, and thus the depth of retrospective belief updating constitutes a form of bounded rationality [10–12], since it trades off inferential accuracy against resource costs. The depth of updating performed by an agent in a particular context might be selected using a form of 'metareasoning' in response to environmental demands [13,14]. Alternatively, it might be phenotypically specified (and thus, presumably, selected for during species evolution). In either case, a bounded-rational approach to retrospection has the potential to explain and quantify how humans and other organisms approach but do not attain optimal performance on a number of cognitive tasks.

FRI differs from existing probabilistic accounts of online cognition [6,8,15,16], which typically only consider inferences about present states, an approach known as 'Bayesian filtering' (though see [17,18]). It thus constitutes a novel hypothesis about cognitive function that extends probabilistic models to subsume a broader range of problems. Importantly, as we will illustrate in the simulations described below, FRI makes testable predictions about behaviour and brain activity in real agents that can be tested in future experimental studies.

Methods

Approximating normative inference

Consider the situation in which an agent seeks to infer on a series of T time-varying hidden states $\mathbf{x}_{1:T} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ given a set of time-invariant parameters $\boldsymbol{\theta}$ that are known with certainty, a series of observations $\mathbf{o}_{1:T} = \{\mathbf{o}_1, \ldots, \mathbf{o}_T\}$, and an initial distribution $\mathbf{x}_0$. To simplify our discussion, in what follows we will assume that all the processes under consideration share the structure shown in Figure 2, and hence have the same conditional independence properties. In particular, we assume that

$$p\left(\mathbf{x}_i \mid \mathbf{x}_{1:i-1}, \mathbf{x}_{i+1:T}\right) = p\left(\mathbf{x}_i \mid \mathbf{x}_{i-1}\right), \tag{1}$$

meaning that states at time i depend only upon the states at the immediately preceding time (this is the Markov property) and that

$$p\left(\mathbf{o}_i \mid \mathbf{x}_{1:T}, \mathbf{o}_{1:i-1}, \mathbf{o}_{i+1:T}\right) = p\left(\mathbf{o}_i \mid \mathbf{x}_i\right), \tag{2}$$

meaning that observations depend only on the current states, and not previous states or observations. However, the general principles presented in this paper apply equally in cases where many, if not all, of these properties are relaxed (for example in processes with a higher-order temporal structure).

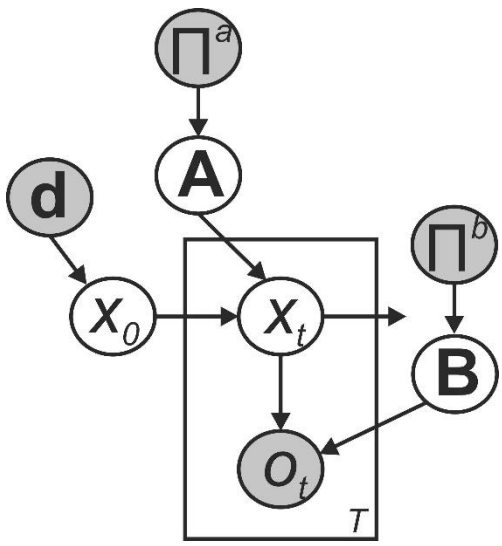

Figure 2: Bayesian graph illustrating the structure of the Hidden Markov model described in the text. (Shaded circles indicate variables with known values, unshaded circles indicate hidden variables) Transitions between hidden states x_0 to x_T are governed by the transition matrix $\mathbf{A}$, and are first-order Markovian. Observations o_1 to o_T depend only on the current hidden state and the emission matrix $\mathbf{B}$. Where the parameters of $\mathbf{A}$ and $\mathbf{B}$ need to be learnt, as depicted here we include appropriate sets of Dirichlet priors, parameterised by the matrices $\mathbf{\Pi}^a$ and $\mathbf{\Pi}^b$ respectively. Beliefs about the initial hidden state x_0 are governed by the parameter vector $\mathbf{d}$.

By the chain rule of probability, the joint distribution over all states is given by

$$p\left(\mathbf{x}_{1:T} \mid \mathbf{o}_{1:T}, \mathbf{x}_0, \boldsymbol{\theta}\right) = \prod_{i=1:T} p\left(\mathbf{x}_i \mid \mathbf{x}_{i-1}, \mathbf{o}_{1:T}, \boldsymbol{\theta}\right). \tag{3}$$

However, inferring on the joint distribution rapidly becomes computationally intractable, and is often unnecessary. Thus instead of inferring on the joint distribution we can instead infer on the marginal distributions over states at each time point (infer on the sequence of most likely states rather than the most likely sequence of states), an approach known as fixed-interval Bayesian smoothing [19]:

$$p\left(\mathbf{x}_{1:T} \mid \mathbf{o}_{1:T}, \mathbf{x}_0, \boldsymbol{\theta}\right) \approx \prod_{i=1:T} p\left(\mathbf{x}_i \mid \mathbf{o}_{1:T}, \mathbf{x}_0, \boldsymbol{\theta}\right). \tag{4}$$

This provides a powerful approach for analysing sequential data, and is widely used in offline data analysis. However, it presents serious practical difficulties for agents performing online inference of the kind that is mandatory for real, embodied agents. (That is, where agents have to make inferences, and very likely take actions, whilst the process is unfolding). Specifically, it requires the agent to store and update an ever-growing set of beliefs about the past, resulting in a set of calculations that will rapidly overwhelm the cognitive capacities of plausible embodied agents. This means that 'true' rationality, defined here as cognition that accords precisely with the principles of optimal probabilistic inference, is impossible for real agents, who must instead seek a feasible approximation.

In probabilistic models of online cognition, this is typically achieved by conditioning inference only on past and current observations, an approach known as Bayesian filtering [19]. This means that agents make inferences of the form

$$\prod_{i=1:T} p\left(\mathbf{x}_i \mid \mathbf{o}_{1:T}, \mathbf{x}_0, \boldsymbol{\theta}\right) \approx \prod_{i=1:T} p\left(\mathbf{x}_i \mid \mathbf{o}_{1:i}, \mathbf{x}_0, \boldsymbol{\theta}\right). \tag{5}$$

This can be implemented in a straightforward fashion by recursive application of

$$p\left(\mathbf{x}_i \mid \mathbf{o}_{1:i}, \mathbf{x}_0, \boldsymbol{\theta}\right) = \frac{p\left(\mathbf{o}_i \mid \mathbf{x}_i, \boldsymbol{\theta}\right) p\left(\mathbf{x}_i \mid \mathbf{x}_{i-1}, \boldsymbol{\theta}\right)}{p\left(\mathbf{o}_i \mid \mathbf{x}_0, \boldsymbol{\theta}\right)} p\left(\mathbf{x}_{i-1} \mid \mathbf{o}_{1:i-1}, \mathbf{x}_0, \boldsymbol{\theta}\right). \tag{6}$$

Filtering is thus computationally parsimonious, since it requires only a single set of calculations at each time step and only requires an agent to store fixed beliefs about the past (in the case of first-order processes, only about the immediately preceding time step). However, this parsimony comes at a cost, since it reduces the accuracy of an agent's beliefs about the past, and consequently, as will discussed later, impairs learning.

To remedy this, an agent that is performing Bayesian filtering, can implement smoothing recursively by performing an additional 'backwards pass' through the data:

$$p\left(\mathbf{x}_i \mid \mathbf{o}_{1:T}, \mathbf{x}_0, \boldsymbol{\theta}\right) = p\left(\mathbf{x}_i \mid \mathbf{o}_{1:i}, \mathbf{x}_0, \boldsymbol{\theta}\right) \int \left[\frac{p\left(\mathbf{x}_{i+1} \mid \mathbf{x}_i\right) p\left(\mathbf{x}_{i+1} \mid \mathbf{o}_{1:T}, \mathbf{x}_0, \boldsymbol{\theta}\right)}{p\left(\mathbf{x}_{i+1} \mid \mathbf{o}_{1:i}, \mathbf{x}_0, \boldsymbol{\theta}\right)} \right] d\mathbf{x}_{i+1}. \tag{7}$$

Here $p\left(\mathbf{x}_i \mid \mathbf{o}_{1:i}, \mathbf{x}_0, \boldsymbol{\theta}\right)$ is the state estimate derived from filtering, $p\left(\mathbf{x}_{i+1} \mid \mathbf{x}_i\right)$ is the dynamic model governing transitions between states, $p\left(\mathbf{x}_{i+1} \mid \mathbf{o}_{1:T}, \mathbf{x}_0, \boldsymbol{\theta}\right)$ is the smoothed state estimate at $i+1$, and $p\left(\mathbf{x}_{i+1} \mid \mathbf{o}_{1:i}, \mathbf{x}_0, \boldsymbol{\theta}\right)$ is the predicted distribution at $i+1$ given by

$$p\left(\mathbf{x}_{i+1} \mid \mathbf{o}_{1:i}, \mathbf{x}_0, \boldsymbol{\theta}\right) = \int p\left(\mathbf{x}_{i+1} \mid \mathbf{x}_i\right) p\left(\mathbf{x}_i \mid \mathbf{o}_{1:i}, \mathbf{x}_0, \boldsymbol{\theta}\right) d\mathbf{x}_i. \tag{8}$$

Thus (fixed-interval) smoothing can be carried out in a straightforward manner, beginning with the current state estimate derived from filtering, and working iteratively backwards. Nonetheless, it requires the agent to perform a set of calculations that grows linearly with the time series, and store a similarly growing set of beliefs about past states, and thus introduces significant extra costs for an agent over and above filtering, which are likely to become unsustainable for real agents in ecological contexts. We thus propose that agents make use of an intermediate strategy, finite retrospective inference, in which they perform retrospective belief updating to a limited degree, in a manner that reflects both the desirability of accurate inference and the need to limit resource (and other) costs.

Finite retrospective inference

To implement FRI, we propose that agents perform fixed-lag smoothing, an approach that is intermediate between full (fixed-interval) smoothing and filtering. In fixed-lag smoothing, agents update beliefs about all states within a fixed-length time window that includes the present time but stretches a set distance into the past (Figure 1) [9]. This window moves forward in time at the same rate that observations are gathered, meaning that cognition occurs within a sliding window. (In principle the sliding window approach can also be used to infer on the joint distribution of short sequences of states [9], but we focus on smoothing in this paper for the sake of simplicity) We are unaware of a precedent for this approach in treatments of cognition, however it has been employed in other contexts [19–22]. This means that, for a window of length n (one that stretches $n-1$ timesteps into the past) considered at time t, agents approximate the true marginal distribution as follows:

$$\prod_{i=1:T} p\left(\mathbf{x}_i \mid \mathbf{o}_{1:T}, \mathbf{x}_0, \boldsymbol{\theta}\right) \approx \prod_{i=1:T} p\left(\mathbf{x}_i \mid \mathbf{o}_{1:i+n-1}, \mathbf{x}_0, \boldsymbol{\theta}\right). \tag{9}$$

As can be seen by comparing equations (5) and (9), filtering is a special case of fixed-lag smoothing in which $n=1$. Smoothing can thus be performed by iteratively evaluating

$$\prod_{i=t-n+1}^{t} p\left(\mathbf{x}_i \mid \mathbf{o}_{1:t}, \mathbf{d}, \boldsymbol{\theta}\right) = \int p\left(\mathbf{x}_{t-n} \mid \mathbf{o}_{1:t-n}, \mathbf{d}, \boldsymbol{\theta}\right) \prod_{i=t-n+1}^{t} p\left(\mathbf{x}_i \mid \mathbf{o}_{t-n+1:t}, \mathbf{x}_{t-n}, \mathbf{d}, \boldsymbol{\theta}\right) d\mathbf{x}_{t-n}. \tag{10}$$

This simply requires the agent to track $p\left(\mathbf{x}_{t-n} \mid \mathbf{o}_{1:t-n}, \mathbf{d}, \boldsymbol{\theta}\right)$, the filtered estimate of the states that obtain at the timestep immediately preceding the current window. Practically, fixed-lag smoothing can be implemented using equation (7), with the proviso that backward recursion is only performed $n-1$ times. In other words, rather than propagating new information right the way back through a time series as is typical in offline applications, it is only propagated to a fixed depth ($n-1$), limiting the computational cost to the agent. This allows agents to adopt a bounded rational strategy in which they trade off inferential accuracy and computational (and potentially other) costs to select an appropriate depth of processing.

Parameter learning using retrospective inference

We next consider the more general situation in which there is uncertainty about both states and parameters, and agents must therefore perform learning as well as inference. This is often referred to as a 'dual estimation' problem [23,24], and is characteristic of many real-world situations. To do so, we condition beliefs about parameters on a set of hyperparameters $\boldsymbol{\lambda}$, such that states and observations are independent of the hyperparameters when conditioned on the parameters, meaning that

$$p\left(\mathbf{x}_{1:T}, \boldsymbol{\theta} \mid \mathbf{o}_{1:T}, \boldsymbol{\lambda}\right) = p\left(\mathbf{x}_{1:T} \mid \mathbf{o}_{1:T}, \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta} \mid \mathbf{o}_{1:T}, \boldsymbol{\lambda}\right). \tag{11}$$

Learning and inference are inextricably related to one another, since beliefs about states depend on beliefs about parameters, and vice versa. Since parameters are fixed, accurately estimating them involves accumulating evidence across entire time series, and thus beliefs about multiple sets of states. This means that increasing the accuracy of beliefs about the past, through retrospective belief updating, also increases the accuracy of parameter estimation. Crucially, since improved parameter estimates will also result in more accurate beliefs about the present and better predictions about the future. Thus, in the context of uncertainty about model parameters, retrospective belief updating is advantageous even for an agent that has no intrinsic interest in the past. This is a very important point, since it argues for the wide importance of retrospective belief updating across a variety of situations and agents.

At a practical level, learning using FRI is very similar to offline learning. We treat each window as a time series in its own right, with $\boldsymbol{\lambda}$ is replaced by $\tilde{\boldsymbol{\lambda}}$, the summary statistics of $p\left(\boldsymbol{\theta} \mid \mathbf{o}_{1:t-n}, \boldsymbol{\lambda}\right)$, which is the posterior distribution over the parameters conditioned on all observations preceding the current window, and perform learning and inference as normal. The use of a sliding window does, however introduce a small additional complexity, since successive windows overlap and thus share data points. Thus, if we treated each window as a separate time series we would count each observation multiple times and, as a result, overweight them. To avoid this, when updating $\tilde{\boldsymbol{\lambda}}$ we only use information about states at the first time-point in the window (i.e. $p\left(\mathbf{x}_{t-n+1} \mid \mathbf{o}_{1:t}, \boldsymbol{\theta}\right)$). (A more specific example of this is provided for the HMM below) This also means that only the best available estimate of each set of

states (in other words, the estimate that will not be revised in light of future evidence) contributes to stored beliefs about the parameters of the model.

Retrospective inference in Hidden Markov models

To illustrate the utility of bounded-rational retrospective inference for an agent, we applied the principles described above to Hidden Markov models (Figure 2). (In principle though, they equally to a broad range of models with alternative properties such as continuous state spaces and higher-order temporal structure) In a hidden Markov model (HMM), the system moves though a series of T time-varying hidden states, each of which is drawn from a discrete state space of dimension K. Hidden states $x_{1:T}$ are not observed directly, but instead must be inferred from observed variables. Thus at time t (where $t \in \mathbb{N} : t = [1, T]$), $x_t \in \mathbb{N} : x_t = [1, K]$ and $o_t \in \mathbb{N} : o_t = [1, M]$. Initial state probabilities are encoded in a row vector $\mathbf{d}$, which we will hereafter assume to encode a uniform distribution. Transitions between states are first-order Markovian, and the transition probabilities are encoded in a $K \times K$ matrix $\mathbf{A}$, such that

$$A_{jk} \equiv p\left(x_{t,k} = 1 \mid x_{t-1,j} = 1\right), \tag{12}$$

where $A_{jk} \in [0,1]$ and $\sum_k A_{jk=1}$. This means that each row $\mathbf{A}_{j\bullet}$ encodes the transition probabilities from state j to the entire state space. From this it follows [25] that

$$p\left(x_t \mid x_{t-1}, \mathbf{A}\right) = \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{x_{t-1,j} x_{tk}}. \tag{13}$$

Similarly, the $M \times K$ matrix $\mathbf{B}$ encodes the emission probabilities such that

$$B_{jk} \equiv p\left(o_{t,k} = 1 \mid x_{t,j} = 1\right), \tag{14}$$

where $B_{jk} \in [0,1]$ and $\sum_k B_{jk=1}$. Thus that each row $\mathbf{B}_{j\bullet}$ encodes the probability of each observed variable when in state j, and

$$p\left(o_t \mid x_t, \mathbf{B}\right) = \prod_{k=1}^{M} \prod_{j=1}^{K} B_{jk}^{x_{tj} o_{tk}}. \tag{15}$$

Pure inference in HMMs

To calculate the smoothed marginal posterior $\gamma\left(x_t\right)$ in an HMM, we can make use of the forward-backward algorithm [26]. This involves recursive forward and backward sweeps, that calculates two quantities $\alpha\left(x_t\right)$ and $\beta\left(x_t\right)$ for each time point [25] such that

$$
\begin{aligned}
\gamma(x_t) &= p(x_t \mid o_{1:T}, \mathbf{A}, \mathbf{B}), \\
&= \frac{p(o_{1:T}, x_t \mid \mathbf{A}, \mathbf{B})\, p(o_{i+1:T} \mid x_i, \mathbf{A}, \mathbf{B})}{p(o_{1:T} \mid \mathbf{A}, \mathbf{B})}, \\
&= \frac{\alpha(x_i)\beta(x_i)}{\sum \alpha(x_i)\beta(x_i)}, \\
\alpha(x_t) &\equiv p(o_{1:t}, x_t \mid \mathbf{A}, \mathbf{B}), \\
\beta(x_t) &\equiv p(o_{t+1:T} \mid x_t, \mathbf{A}, \mathbf{B}).
\end{aligned} \tag{16}
$$

$\alpha(x_i)$ thus corresponds to the unnormalised filtered posterior, and is given by

$$
\begin{aligned}
\alpha(x_1) &= p(o_1 \mid x_1, \mathbf{B})\, p(x_1 \mid \mathbf{d}), \\
&= (\mathbf{B}\hat{o}_t) \circ \mathbf{d},
\end{aligned} \tag{17}
$$

for the first state, and

$$
\begin{aligned}
\alpha(x_i) &= p(o_i \mid x_i, \mathbf{B})\, \alpha(x_{i-1})\, p(x_i \mid x_{i-1}, \mathbf{A}), \\
&= (\mathbf{B}\hat{o}_t) \circ \left(\mathbf{A}^{\mathrm{T}} \alpha(x_{i-1})\right),
\end{aligned} \tag{18}
$$

for all subsequent states. Here $\hat{o}_t$ is a column vector encoding the observed variable at time t and $\circ$ denotes the Hadamard or element-wise product. $\beta(x_i)$ is given by

$$
\begin{aligned}
\beta(x_i) &= \beta(x_{i+1})\, p(o_{i+1} \mid x_{i+1}, \mathbf{B})\, p(x_{i+1} \mid x_i, \mathbf{A}), \\
&= \mathbf{A}\left(\beta(x_{i+1}) \circ (\mathbf{B}\hat{o}_{t+1})\right).
\end{aligned} \tag{19}
$$

To apply the sliding window approach to this model, at each timestep we simply evaluate the filtered posterior using (18) and then perform backward inference a fixed number of steps using (19).

Dual estimation in HMMs

To learn the transition probabilities of an HMM we first need to define an additional quantity, the dual-slice marginal $\xi(x_t, x_{t-1})$, which corresponds to the joint probability distribution $p(x_t, x_{t-1} \mid o_{1:T}, \theta)$ [25,27]. It is simple to show that

$$
\begin{aligned}
\xi(x_t, x_{t-1}) &= p(x_t, x_{t-1} \mid o_{1:T}, \theta, \mathbf{d}), \\
&= \frac{\alpha(x_{t-1})\, p(o_t \mid x_t, \theta)\, p(x_t \mid x_{t-1}, \theta)\, \beta(x_t)}{p(o_{1:T} \mid \theta, \mathbf{d})}, \\
&\propto \mathbf{A} \circ \left(\alpha(x_{t-1})\left((\mathbf{B}\hat{o}_t) \circ \beta(x_t)\right)^{\mathrm{T}}\right).
\end{aligned} \tag{20}
$$

(For a more detailed exposition of this see [25].)

Introducing learning renders exact inference impossible, which necessitates the use of an approximation. Broadly speaking, such approximations fall into two categories: sampling approaches [28], which are computationally expensive but asymptotically exact, and variational approaches which

are more computationally efficient but require the introduction of a tractable approximate distribution [29]. We focus here on implementing model inversion using variational Bayes [30], which we believe has some neurobiological plausibility [31]. This is not a strong claim, however, about the actual mechanisms used by human observers (or indeed any other agent), and similar results could be derived under any appropriate scheme. (See Appendix for further description of the variational methods employed here)

In the offline case, this model has been described in [30,32], and the reader is referred to these sources for more detailed expositions. Briefly, we start by placing Dirichlet priors over each row of the transition matrix $\mathbf{A}$ and the observation matrix $\mathbf{B}$ such that

$$\begin{aligned}
p\left(\mathbf{A}_{j\bullet}\right) &= Dir\left(\mathbf{\Pi}^{a}_{j\bullet}\right), \\
\mathbb{E}\left[p\left(a_{ij}\right)\right] &= \frac{\pi^{a}_{ij}}{\sum \mathbf{\Pi}^{a}_{j\bullet}}, \\
\mathbb{E}\left[\ln p\left(a_{ij}\right)\right] &= \psi\left(\pi^{a}_{ij}\right) - \psi\left(\sum \mathbf{\Pi}^{a}_{j\bullet}\right). \\
p\left(\mathbf{B}_{j\bullet}\right) &= Dir\left(\mathbf{\Pi}^{b}_{j\bullet}\right), \\
\mathbb{E}\left[p\left(b_{ij}\right)\right] &= \frac{\pi^{b}_{ij}}{\sum \mathbf{\Pi}^{b}_{j\bullet}}, \\
\mathbb{E}\left[\ln p\left(b_{ij}\right)\right] &= \psi\left(\pi^{b}_{ij}\right) - \psi\left(\sum \mathbf{\Pi}^{b}_{j\bullet}\right).
\end{aligned} \tag{21}$$

where $\mathbf{\Pi}^{a}$ and $\mathbf{\Pi}^{b}$ are matrices encoding the concentration parameters of the Dirichlet distributions, and ψ is the digamma function. Since the Dirichlet distribution is the conjugate prior for a multinomial likelihood, this enables us to carry out parameter learning using a set of simple update equations as described below.

The log joint probability distribution for the model thus becomes

$$\begin{aligned}
\ln p\left(o_{1:T}, x_{1:T}, \mathbf{A}, \mathbf{B} \mid \mathbf{d}, \mathbf{\Pi}^{a}, \mathbf{\Pi}^{b}\right) = &\sum_{i=1:T} \ln p\left(o_i \mid x_i, \mathbf{B}\right) + \sum_{i=2:T} \ln p\left(x_i \mid x_{i-1}, \mathbf{A}\right) \\
&+ \ln p\left(\mathbf{B} \mid \mathbf{\Pi}^{b}\right) + \ln p\left(\mathbf{A} \mid \mathbf{\Pi}^{a}\right) + \ln p\left(x_1 \mid \mathbf{d}\right),
\end{aligned} \tag{22}$$

and model inversion can be performed by iteratively evaluating the following update equations for the states and parameters (see Appendix for a full derivation).

$$\begin{aligned}
\hat{x}_j &= \mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\gamma\left(x_j\right)\right], \\
\hat{\mathbf{M}}_j &= \mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\xi\left(x_{j-1}, x_j\right)\right], \\
\ln \hat{a}_{ij} &= \psi\left(\hat{\pi}^{a}_{ij}\right) - \psi\left(\sum \hat{\mathbf{\Pi}}^{a}_{j\bullet}\right) \\
\hat{\mathbf{\Pi}}^{a} &= \mathbf{\Pi}^{a} + \sum_{t=2:T} \hat{\mathbf{M}}_t, \\
\ln \hat{b}_{ij} &= \psi\left(\hat{\pi}^{b}_{ij}\right) - \psi\left(\sum \hat{\mathbf{\Pi}}^{b}_{j\bullet}\right), \\
\hat{\mathbf{\Pi}}^{b} &= \mathbf{\Pi}^{b} + \sum_{t=1:T} \hat{\mathbf{x}}_t \hat{\mathbf{o}}_t^{\mathbf{T}}.
\end{aligned} \tag{23}$$

These have intuitive interpretations. The update equations for the transition matrix $\mathbf{A}$ correspond to accumulating evidence about the number of times each state transition occurs, whilst those for the observation matrix $\mathbf{B}$ correspond to a similar evidence accumulation process, this time about the number of times that a particular observation was made whilst occupying a particular state.

To carry out online learning and inference, we simply apply the sliding window approach described earlier to this model. This means that we only evaluate $\hat{\mathbf{x}}_j$ and $\widehat{\mathbf{M}}_j$ for timepoints that fall within the current window, and parameter learning is performed by updating the concentration parameters using the following equations:

$$\begin{aligned}
\tilde{\mathbf{\Pi}}^a &= \mathbf{\Pi}^a + \widehat{\mathbf{M}}_{t-n+1}, \\
\widehat{\mathbf{\Pi}}^a &= \tilde{\mathbf{\Pi}}^a + \sum_{j=t-n+2}^{t} \widehat{\mathbf{M}}_j, \\
\tilde{\mathbf{\Pi}}^b &= \mathbf{\Pi}^b + \hat{\mathbf{x}}_{t-n+1}\hat{\mathbf{o}}_{t-n+1}^{\mathbf{T}}, \\
\widehat{\mathbf{\Pi}}^b &= \tilde{\mathbf{\Pi}}^b + \sum_{j=t-n+2}^{t} \hat{\mathbf{x}}_j\hat{\mathbf{o}}_j^{\mathbf{T}}.
\end{aligned} \tag{24}$$

Here we $\tilde{\mathbf{\Pi}}^a$ and $\tilde{\mathbf{\Pi}}^b$ denote the fixed-lag parameters that are incremented across time steps, $\mathbf{\Pi}^a$ and $\mathbf{\Pi}^b$ denote the values of the fixed-lag concentration parameters from the previous time step (in other words, the evidence that has been accumulated prior to the current window), and $\widehat{\mathbf{\Pi}}^a$ and $\widehat{\mathbf{\Pi}}^b$ denote the full estimates of the concentration parameters based on timesteps 1 to t.

The probabilistic reversal task as a special case of the HMM

To illustrate the utility of FLR even for relatively straightforward tasks, we simulated inference and learning on a probabilistic reversal paradigm [33]. Briefly, subjects are required to track an underlying hidden state that occasionally switches between one of two possible values, based on probabilistic feedback (in other words, feedback that is only, for example, 85% reliable) This paradigm is both simple and widely used, and the small state space makes illustrating results in graphical form relatively straightforward. In addition, the fact that the paradigm is widely used makes it an appealing tool for exploring to what extent human subjects actually employ FLR when solving this sort of task. The task can be modelled as an HMM, in which there are only two hidden states $x_t \in \{1,2\}$, which probabilistically generate one of two possible observations $o_t \in \{1,2\}$ [9,33–35]. The parameter *r* encodes the probability of a reversal between trials, and *v* encodes the reliability of observations. Thus:

$$\begin{aligned}
\mathbf{A} &= \begin{bmatrix} 1-r & r \\ r & 1-r \end{bmatrix}, \\
\mathbf{B} &= \begin{bmatrix} v & 1-v \\ 1-v & v \end{bmatrix}.
\end{aligned} \tag{25}$$

(Introducing learning requires a slight modification of the standard HMM parameter update equations to reflect the symmetry of the $\mathbf{A}$ and $\mathbf{B}$ matrices, as described in the Appendix.)

Simulations

Probabilistic reversal task

To illustrate the effects of retrospective inference on an agent's beliefs whilst doing the probabilistic reversal task, we simulated 1000 instantiations of a 256 trial task session, with parameters set as $r = 0.1$ and $v = 0.85$, plausible values for real versions of the task (e.g. [9]). For the 'pure inference' agent, we set extremely strong (and accurate) prior beliefs about $\mathbf{A}$ and $\mathbf{B}$ by setting initial values of $\mathbf{\Pi}^{(a)} = 10^6 \mathbf{A}$ and $\mathbf{\Pi}^{(b)} = 10^6 \mathbf{B}$. For the 'dual estimation' agent, we kept the prior beliefs about $\mathbf{B}$ identical, but set weak priors on the transition matrix of

$$\mathbf{\Pi}^{(a)} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}. \tag{26}$$

Window lengths for retrospective inference were set at 1, 2, 4, 8, 16, 32, 64, 128 and 256 trials, and we also simulated an agent performing offline (fixed interval) smoothing for comparison. To assess the accuracy of inference and learning, we calculates the log likelihood assigned to the true sequence of hidden states, or true value of r, and averaged this across simulations.

Random HMMs

To show that the effects that we illustrate are not due to some specific feature of the probabilistic reversal paradigms, we performed similar simulations, this time using HMMs with three possible hidden states, three possible observations, and randomly generated transition probabilities. We generated 10 such HMMs, and simulated 100 instantiations of each, whilst varying the diagonal terms of the emission matrix $\mathbf{B}$ at intervals of 0.05 between 0.65 and 0.95 (and setting the off-diagonal terms to be equal). (This corresponds to varying the degree of perceptual uncertainty) Prior beliefs for the pure inference and dual estimation agents were set as described for the reversal task, and accuracy was assessed in the same way.

Experimental paradigm and modelling

We tested 96 healthy subjects with no self-declared history of neurological or psychiatric disorders on a probabilistic reversal task known as the 'beads task' [36] using a ratio of 85:15 beads per jar. In this task, a subject is told a sequence of beads will be drawn from either a jar containing (for example) an 85:15 mixture of red:blue beads or from another jar with the opposite distribution. The subject is then asked to estimate the probability of the jar's identity after each bead. There were 32 male and 64 female subjects, of mean age 24.5 years (sd = 5.5). The subjects were drawn from a UCL subject pool (to which the study was advertised) – composed of undergraduates (the majority) and the general population – and they were given small financial compensation for their participation. They gave written informed consent, and the study was approved by UCL's Ethics Committee (3450/002). Subjects were told that the jar "might change" during the task; they viewed 150 draws (with replacement) of a bead and rated their belief about the identity of the jar on a sliding scale after each draw. The sliding scale went from certainty that the jar was red (i.e., 85:15 red:blue) via full uncertainty to certainty that the jar was blue (i.e., 85:15 blue:red). Previous draws were not displayed. Each subject saw the same sequence (generated by a computer sampling from two 85:15 populations), in which the jar changed after 25, 50, 60, 70, 80, 90, 100 and 125 draws.

Prior to model fitting we first scaled the range of each subjects' rating data by 0.99, and then transformed the scaled data into logit space to generate of vector of behavioural responses $y_{1:T}$.

These steps were necessary to allow us to carry out model fitting using the Variational Laplace algorithm, as described below.

To explore between-subject differences in strategy, we used a version of the variational model described above, which included as free parameters r and v, as well as two additional parameters $l^{(a)}$ and $l^{(b)}$ that parameterise the strength of prior beliefs such that $\mathbf{\Pi}^{(a)} = l^{(a)}\mathbf{A}$ and $\mathbf{\Pi}^{(b)} = l^{(b)}\mathbf{B}$. Window lengths were set at 1, 2, 4, 8, 16, 32, 64 and 128 trials. Additionally, we fitted models in which no learning occurred by setting $l^{(a)} = l^{(b)} = 10^6$. (Note that since, in the absence of learning, predicted behaviour is identical whatever window length is used, we only fitted a single model, with window length 1)

To map the agent's beliefs onto observed responses, we used a Gaussian observation model,

$$p\left(y_t \mid x_t, y_{t-1}, s, w, h\right) = N\left(\left(1-w\right)\text{logit}\left(x_{t2}\right)s + wy_{t-1}, \frac{1}{e^h}\right). \tag{27}$$

Where $\text{logit}\left(p\right) = \ln\left(p\right) - \ln(1-p)$, and $N\left(\mu, \sigma^2\right)$ is the normal distribution with mean μ and variance σ^2. Here the mean of the distribution is given by a logit transform of x_{t2}, the agent's belief that state 2 currently obtains, combined with two additional parameters that we included to account for suboptimalities in behaviour. The first (s) governs the scaling between beliefs and responses (in other words, the degree to which subjects tended to over or under-report the strength of their beliefs). The second (w) parameterises the degree of inertia or 'choice stickiness' shown by subjects in their responses. (Where this is high, subjects show a bias to making the same response as on the previous trial over and above effects predicted by the probabilistic model. Where it is low, this bias is weak or absent). Including these extra parameters was strongly favoured by our model comparison results, as described below. Parameter h represents the log precision of the Gaussian observation noise, and thus accounts for individual differences in the noisiness of responses.

Prior distributions over appropriately transformed parameters were specified as:

$$\begin{aligned}
&\text{logit}\left(r\right) \sim N\left(\text{logit}(0.1), 3\right),\\
&\text{logit}\left(v\right) \sim N\left(\text{logit}\left(0.85\right), 3\right),\\
&\ln\left(l^a\right) \sim N\left(4, 5\right),\\
&\ln\left(l^b\right) \sim N\left(4, 5\right),\\
&\ln\left(s\right) \sim N\left(0, 5\right),\\
&\text{logit}\left(w\right) \sim N\left(\text{logit}\left(0.1\right), 5\right),\\
&\ln\left(h\right) = N\left(2.5, 5\right),
\end{aligned} \tag{28}$$

and model fitting was performed using Variational Laplace [37,38]. These priors were selected to weakly favour plausible parameter values. Off-diagonal terms in the prior covariance matrices were set to zero. Model comparison was performed using the negative variational free energy (L) as an approximation to the log model evidence.

To examine the parameter estimates derived from our modelling, we created a weighted average set of estimates for each subject by multiplying the maximum a posteriori (MAP) estimate derived from each model by the posterior probability of that model for that subject, and summing. This yielded a

point estimate of each parameter for each subject, and group-level results of these are reported below.

Results

Simulations results

To explore the properties of fixed-lag retrospection in pure inference problems (in other words, ones where no learning is necessary), we simulated behaviour on both the probabilistic reversal task and on random HMMs. As expected, in both cases, FLR considerably improved the accuracy of agents' final (offline) beliefs about past hidden states. (Online estimates of current states are identical under all approaches.) Strikingly, in both cases, this improvement occurred even when agents only retrospected over short windows (Figure 3), suggesting that, in certain problems at least, a limited capacity for retrospection can yield significantly improves inference.

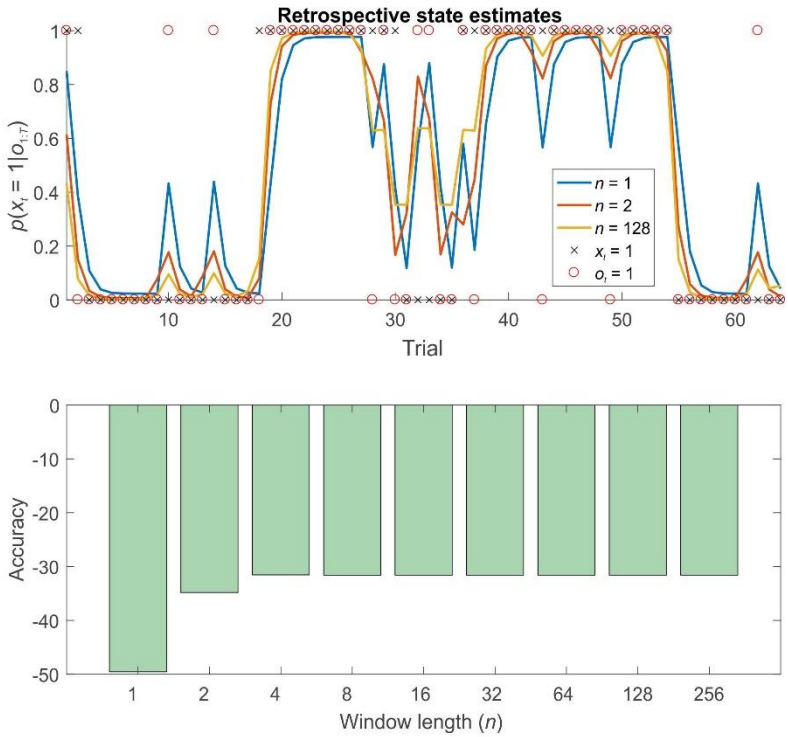

Figure 3: Retrospective belief updating improves state estimates during pure inference on a probabilistic reversal task. Figure 3a: illustration of the first 64 trials of a 256 trial session of the reversal task using different strategies. The final (retrospective) posteriors are shown in blue (*n*=1, filtering), orange (*n*=2) and gold (*n*=16). Black crosses show the true hidden state, and red circles the observations made on each trial. Retrospective belief updating allows agents to infer the true underlying states more accurately. Figure 3b: relative log accuracy of models of different window lengths, averaged across simulated time series (see main text for details). (Accuracy is quantified as the log likelihood assigned to the true sequence of states by the agent, averaged across simulations) This illustrates that, in this context at least, even the use of a very short window leads to significantly more accurate beliefs, but that this benefit saturates relatively rapidly (by about *n*=8). Thus for a bounded rational agent performing pure inference, the optimal window length may be surprisingly low, depending on the relevant computational costs.

Simulations of dual estimation problems in which there was uncertainty about r clearly illustrated that retrospective inference increases the accuracy of both retrospective *and* online state estimation, as a result of increased accuracy in parameter learning (Figures 4 and 5). One important feature to note is that even when the maximum possible depth of retrospection is employed ($n = 256$), the accuracy of online state estimation always falls significantly short of offline estimation. This indicates the fact that, however great the representational and computational sophistication of an agent, there is always a cost to performing inference online, rather than with a complete data set. If sufficiently high, this cost provides an incentive to perform additional (subsequent) offline processing, perhaps during sleep, and it is conceivable that this might be linked to the extended process of memory consolidation. (Simulations using randomly generated HMMs show similar properties, but these are not reported for reasons of brevity)

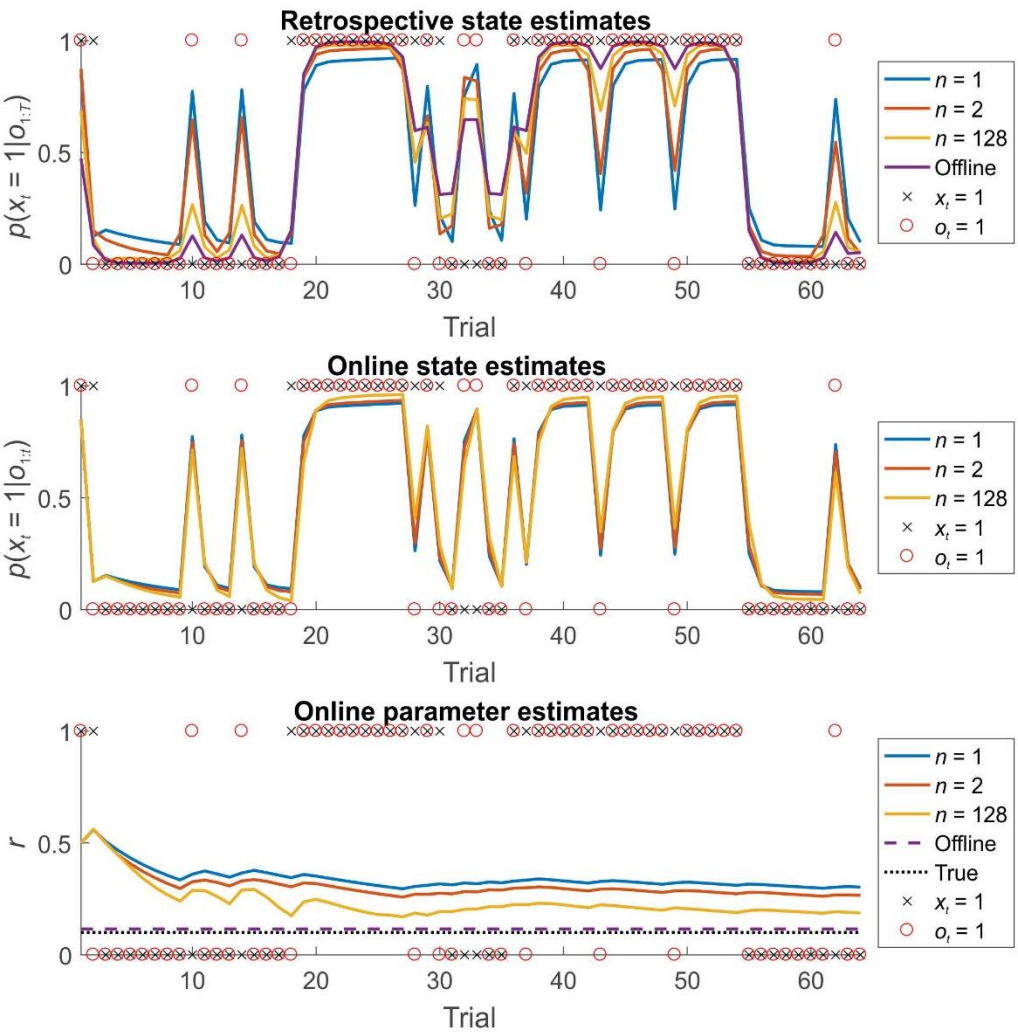

Figure 4: State and parameter estimation for agents performing dual estimation on the first 64 trials of a 256 trial session of the reversal task using different strategies. Figure 4a: the accuracy of retrospective belief estimates increases with greater window lengths, but still falls well short of the performance of an offline agents, who has access to the entire time series simultaneously. Figure

4b: the accuracy of online (filtered) beliefs about the current state subtly but consistently increases with greater window length. Note that this effect is entirely due to the beneficial effects of greater window lengths on parameter learning. Figure 4c: the effect of window length on parameter learning. With greater window lengths, parameter estimates converge more rapidly on the true value. (Estimates from agents performing retrospective inference with windows of length 1, 2 and 128 time steps are shown in blue, orange and gold respectively. Estimates from an agent performing offline inference is shown in purple. True hidden states are indicated with black crosses, whilst observations are indicated with red circles. The true value of parameter r is indicated with a dotted black line in 5c)

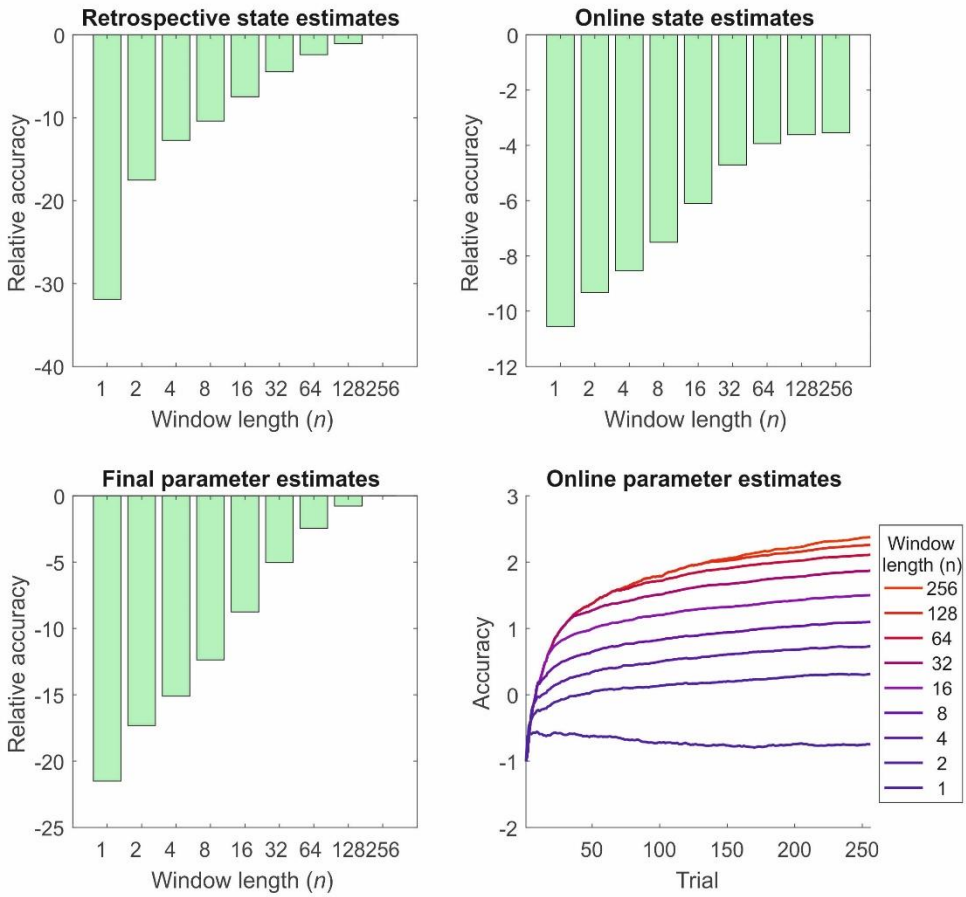

Figure 5: Accuracy of inference and learning performed by agents using different window length, averaged across 1000 simulations (see main text for more details). Accuracy is quantified as the log likelihood assigned to the true sequence of states or the true parameter value by the agent, averaged across simulations. Figure 5a: accuracy of retrospective state estimation relative to the performance of an offline agent. Accuracy increases with window length, becoming identical for online and offline agents with the same effective window length (256 trials). Figure 5b: accuracy of online (filtered) state estimates relative to the filtered state estimates of an offline agent. Accuracy increases with window length, but never becomes equivalent to that of an offline agent. This difference reflects the fact that the parameter estimates of the online agent are based only on observations made up to the present time, rather than on the entire data set. (In other words, $p(\theta \mid o_{1:t}, \lambda)$ at trial t rather than $p(\theta \mid o_{1:T}, \lambda)$). This can be thought of as a cost of online inference. Figure 5c: accuracy of final parameter estimates relative to the performance of an offline

agent. Accuracy progressively increases with window length, becoming equivalent for online and offline agents with the same effective window length. Figure 5d: average accuracy of parameter estimates across trials. Accuracy of parameter estimation increases with window length, and these differences progressively appear as the session goes on.

Behavioural modelling results

We performed two preliminary analyses to reduce the size of our model space, using fixed-effects comparison over model families [39]. We first compared families that included a response-inertia parameter w, a response-scaling parameter s, both, or neither. The family of models including both parameters was strongly favoured by our data (posterior probability > 0.999). Second, we grouped models into families according to whether they included parameter learning or not. The family allowing for learning was favoured with a posterior probability of >0.999. This provides strong evidence that subjects did indeed learn the parameters of the task, rather than using a fixed set of parameter estimates, as is commonly assumed in reversal learning models [9,33–35].

As a result of these preliminary analyses, we confined further analysis to models that included both scaling and inertia parameters, and allowed for parameter learning. We first confirmed that these models fitted the data well, as indexed by the mean proportion of variance explained, which in all cases was above 0.85 (Table 1). We then used random effects Bayesian model selection (as described by [40]) to characterise the depth of retrospective inference performed by subjects on our task. This suggested that subjects tended to perform retrospection to a considerable degree, with the greatest probability masses assigned to the three longest window lengths (Table 1) (Figure 6). To further test the hypothesis that subjects carried out retrospective inference, we performed an additional comparison between the pure filtering model ($n=1$) on the one hand, and the whole family of smoothing models on the other. This provided overwhelming evidence in favour of the family of smoothing models (posterior probability > 0.999), strongly supporting the hypothesis that subjects performed retrospective inference.

	Group-level Analysis				**Canonical Subject Analysis**		
Depth of retrospection (*n*)	***L* (summed)**	**Posterior Probability**	**Exceedance Probability**	**Mean Proportion Variance Explained**	***L***	**Posterior Probability**	**Proportion Variance Explained**
1	-19326	0.0464	<0.0001	0.8554	36.86	0.0004	0.9890
2	-19297	0.0727	<0.0001	0.8570	39.39	0.0053	0.9891
4	-19254	0.0412	<0.0001	0.8575	39.26	0.0047	0.9892
8	-19242	0.1432	0.0049	0.8579	42.31	0.0988	0.9897
16	-19249	0.0370	<0.0001	0.8584	41.82	0.0604	0.9896
32	-19204	0.1762	0.0262	0.8599	42.47	0.1151	0.9898
64	-19204	0.1808	0.0319	0.8607	43.18	0.2346	0.9899
128	-19213	0.3024	0.9370	0.8611	43.90	0.4807	0.9900

Table 1: Summary of results from our main analyses. Both the group-level analysis, where we fitted our models to data from individual subjects, and the canonical subject analysis, in which we averaged behavioural responses across subjects prior to model fitting, clearly favoured models with a greater depth of retrospective inference (larger values of n). All models predicted both individual and canonical subject data with a high degree of accuracy.

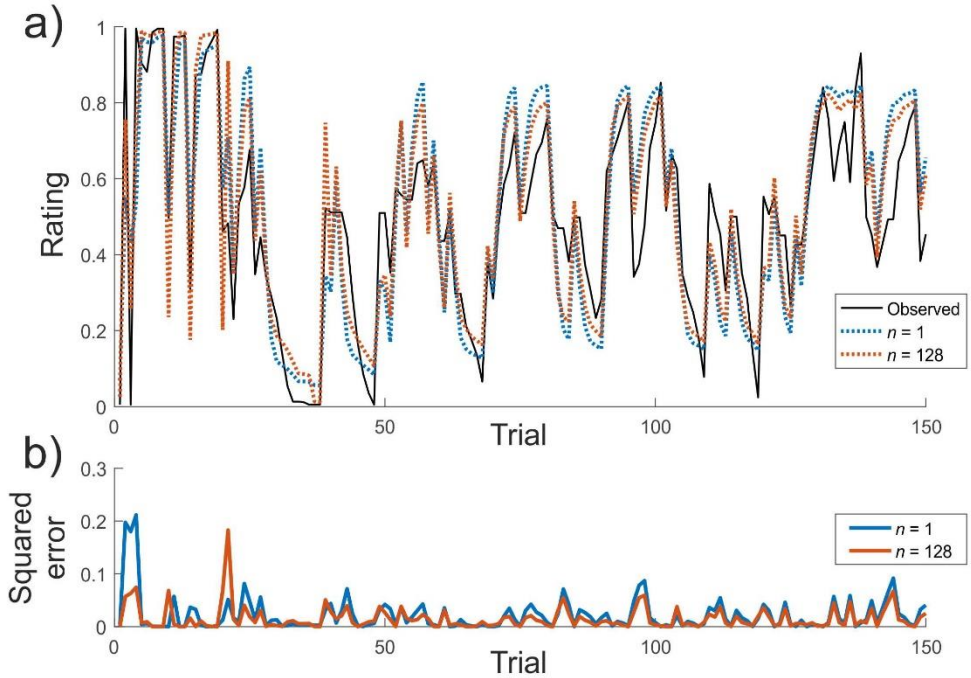

Figure 6: Illustrations of filtering ($n = 1$, blue) and best smoothing ($n = 128$, orange) model fits to example subject data (subject 3). Figure 6a: both models provide good fits to the observed data (black), but the smoothing model gives more accurate predictions, as can be seen by inspecting the trial-by-trial squared error, as shown in Figure 6b. This pattern was borne out in our group-level analysis (Table 1).

To test whether greater depth of retrospection increased performance on the task, we first derived an accuracy score for each subject, reflecting the mean probability they assigned to the 'true' task state at each trial. We then performed a median split on subjects, based on the window length of the model that had the greatest posterior probability for their data, as calculated during our mixture model analysis. Accuracy score were greater in the long window length group (mean: 0.713, sd: 0.065) than the short window length group (mean: 0.689, sd: 0.056), and this difference approached conventional significance levels using a Wilcoxon rank sum test (p = 0.068). These results are consistent with a link between increased depth of retrospection and task performance, but do not provide compelling evidence for it, perhaps because out task was not specifically designed to probe this.

Analysis of the group-level parameter estimates (Table 2) indicates that, at the start of the session, subjects tended to overestimate the probability of a reversal, as indicated by the relatively high value for r , but started with fairly accurate beliefs about the cue validity v , and that these prior beliefs were reasonably strong (as indicated by the values of $l^{(a)}$ and $l^{(b)}$. At a group level, subjects tended to overrate the strengths of their beliefs, as indicated by a mean value of the scaling parameter s that is greater than 1, and showed a significant degree of behavioural inertia, as indicated by a mean value of w that is well above zero.

Parameter	Group Mean and Standard Deviation
r	0.27 (0.23)
v	0.80 (0.15)

$l^{(a)}$	129 (249)
$l^{(b)}$	148 (337)
s	1.25 (0.77)
w	0.48 (0.21)
h	3.19 (6.14)

Table 2: Weighted average of parameter estimates across subjects and models. Maximum a posteriori (MAP) estimates for each parameter in each subject were multiplied by the posterior probability of that model for that subject, and summed to give single subject parameter estimates.(Standard deviation across subject given in brackets). These indicate that, at the start of the session, subjects tended to overestimate the probability of a reversal, as indicated by the relatively high value for r, but started with fairly accurate beliefs about the cue validity v, and that these prior beliefs were reasonably strong (as indicated by the values of $l^{(a)}$ and $l^{(b)}$. At a group level, subjects tended to overrate the strengths of their beliefs, as indicated by a mean value of the scaling parameter s that is greater than 1, and showed a significant degree of behavioural inertia, as indicated by a mean value of w that is well above zero.

Since behavioural data from individual subjects can be assumed to contain a significant degree of noise, we also performed a complementary set of analyses in which we created a 'canonical subject' by averaging responses across all subjects, and fitted our reduced model space to these data (Figure 7). (This can be compared to fitting a model to grand mean evoked responses in neuroimaging, [41].) Our models were able to predict these canonical responses with a very high degree of accuracy (in each case, with a proportion of variance explained of over 0.98, see Table 1). Reassuringly, when we examined the variational free energy of each of the fitted models, an extremely similar pattern to the group-level results emerged, with clear evidence in favour of models with greater depth of retrospective processing (Table 1).

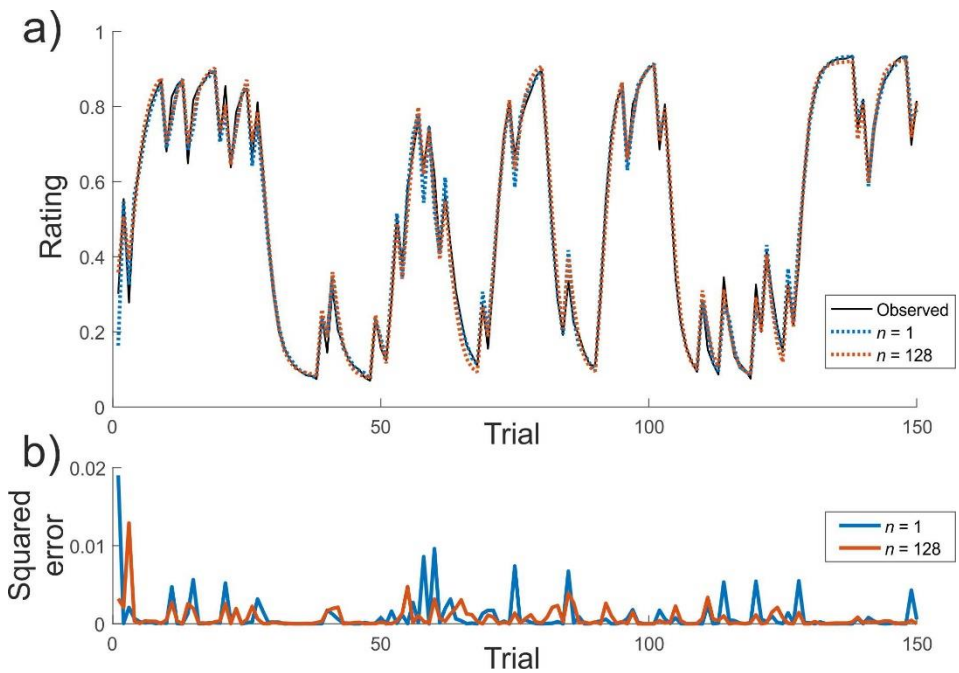

Figure 7: Illustrations of filtering ($n=1$, blue) and best smoothing ($n=128$, orange) model fits to canonical subject data. Figure 7a: both models provide extremely good fits to the observed data

(black), but the smoothing model yields more accurate predictions, as can be seen by inspecting the trial-by-trial squared error, as shown in Figure 7b. (See Table 1 for more information)

Discussion

In this paper, we consider the problem of accurately updating beliefs about the past from the perspective of probabilistic cognition. Specifically, we propose that humans and other agents use finite retrospective inference, in which beliefs about past states are modifiable within a certain temporal window, but are fixed thereafter. We show, using simulations of inference and learning in the context of a probabilistic reversal task, that even a fairly limited degree of retrospection results in significantly improved accuracy of beliefs about both states and parameters. The hypothesis that agents perform retrospective inference makes clear predictions about behaviour on appropriate tasks that are quantitatively dissociable from those made under the hypothesis that agents use pure filtering, which we test using data collected from 96 subjects who performed such a reversal task. Our results provide clear evidence in favour of the hypothesis that subjects carried out retrospective inference whilst performing our task. Implementing retrospective inference also makes specific predictions about brain function, since it requires beliefs about past states to be explicitly represented and updated. Our work thus provides testable hypotheses that can be explored in future behavioural and neurobiological studies.

Perhaps the most significant feature of our simulations is the demonstration that, where there is uncertainty about time-invariant model parameters, finite retrospective inference significantly improves the accuracy of learning. This is important both because these parameters may be of intrinsic interest, and because better learning will result in more accurate beliefs about present and future states. Even if an agent has no intrinsic interest in past events, it still has a clear incentive to perform retrospective inference, since this will allow it to act better in the future. This provides a new twist on the often-advanced hypothesis that the primary function of memory in general, and episodic memory in particular, is to improve predictions about the future [42]. Here, in addition to playing a role in constructing imagined future scenarios [43], the explicit representation of events or episodes in the past may be essential for updating beliefs about current and future states or learning time-invariant properties of an agent's environment [17].

A similar point may be made about the potential importance of retrospective inference for the generation and selection of appropriate cognitive models, a process known as structure learning [44–46]. In this paper, we confine ourselves to considering inference about hidden states and learning about fixed model parameters, but structure learning is an equally important process, and one that is likely to be strongly affected by the depth of retrospective processing employed by an individual. In future work, we plan to address this explicitly, both through simulations and experimental work.

The specific retrospective inference model we describe here differs importantly from previous approaches to modelling probabilistic reversal tasks [33] and change point detection more generally [47] in two key ways, first through the fact that we allow for parameter learning, and second, because we simulate agents that are able to update beliefs about past states. Both these processes are important for normative behaviour, and it will be important to establish how closely human performance across a number of domains reflects this. Our approach also importantly differs from models such as the hierarchical Gaussian filter [48], which use higher-level variables operating at longer time scales to provide an implicit time window, but do not make postdictive inferences of the sort discussed here. A closer analogy can be drawn with generalised filtering approaches [49], which

infer both on the current state and its derivatives (rate of change, acceleration and so on), and require a finite window of data to perform updates. This is something we intend to return to in future work.

Retrospective inference provides a natural explanation for a number of 'postdictive' phenomena in perception, in which perception of an event is influenced by other events that only occur afterwards [50,51]. A classic example of this is the colour phi phenomenon [52]. Here, two differently coloured dots are briefly displayed to the subject at different spatial locations. If the interval between the flashes is sufficiently short, subjects report perceiving a single moving dot, rather than two separate dots. Critically, they also perceive the colour of the dot as changing during motion, meaning that they perceive the second colour as occurring before it is presented on screen. This means that information about the colour of the second dot has somehow been propagated backwards in (perceptual) time. That such postdictive phenomena might be explained by smoothing has previously been pointed out by [18], but our proposal builds on this by suggesting a limited window of updating, as well as highlighting the importance of such belief updating for learning.

The existence of postdictive perceptual phenomena (among other considerations) have led to what is often called the 'multiple drafts' account of consciousness [53], in which the contents of conscious are subject to continual revision in the light of new information (at short timescales, at least), and what subjects report is critically dependent upon when they are asked. For example, in the colour phi experiment, subjects' reported perceptual experience would differ if they were asked to report it before the second dot is shown, as opposed to when they are asked to report it afterwards. This accords extremely well with our proposal (at least if we make the further supposition that the contents of consciousness can, in some sense, be identified with the outcome of optimal perceptual inference). Under FRI (unlike filtering), reported perceptual experience will be critically dependent on when the report is made, since online retrospective inference makes beliefs time-dependent. In other words, my belief about what happened at time t may be different depending on whether you ask me for it at time $t+1$ or time $t+10$. This means that FRI has the potential to provide the computational underpinning of a 'probabilistic multiple drafts' model of perceptual experience.

One intriguing possibility raised by FRI is that different individuals might perform retrospective belief updating to different extents, either on particular tasks or in general, and that this might partially explain between-subject differences in performance on particular tasks (see [9] for evidence of this). Such differences might even help explain facets of psychopathology [54]. For example, impaired learning due to reduced or absent retrospection might lead to the tendency to form delusional beliefs [55–57]. For example, say someone looked at you in an unusual way – making you feel they were spying on you – but then subsequently ignored you: if you could not use the latter information to revise your initial suspicion, you would be more likely to become paranoid about that person. This idea is supported by the finding of altered neuronal responses in subjects with delusions (as compared with healthy controls) during performance of a retrospective belief updating task [58], and is something we intend to return to in future.

Implementing retrospective inference also has important implications for neurobiology. In particular, since agents need to be able to dynamically update beliefs about past states, they are required to store explicit, ordered representations of the past, and it should be possible to find evidence of this in appropriate neuronal structures [59]. (For some evidence of this, see [57].) Intriguingly, this fits extremely well with an extensive literature on hippocampal function [60–64], a finding supported by the results of our previous study, which found a relation between depth of retrospective processing and grey matter density in the hippocampus [9]. On the further supposition that retrospective inference is implemented using filtering and smoothing as described above, this leads to the hypothesis that forward and backward sweeps through recently encountered states, as are known to

occur in the hippocampus [63,65–67] may play a key role in retrospective belief updating. What is less clear, at present, is how to implement retrospective inference within established, neurobiologically-grounded accounts of probabilistic inference in the brain [6,16,68] – though see [31] for related suggestions. This is an extremely important question, and one we intend to return to in future work.

Probabilistic models of cognition are an enormously exciting tool for understanding the complex workings of the mind and brain [3,5,6,69]. The ideas we propose represent a development of such approaches to encompass inference about states in the past, as well as the present. On the further hypothesis that the depth of processing employed is flexible and tailored to the demands of a particular problem or environment, such retrospective processing can also be linked to broader notions of bounded rationality [10–12]. We show, through simulations of simple environments, that even a limited degree of retrospection can yield significantly more accurate beliefs about both time-varying states and time-invariant parameters, and thus has the potential to support more adaptive, successful behaviour to justify its extra resource costs. This makes it a plausible strategy for real, biological agents to employ. We support this hypothesis using data from a classic probabilistic reversal task, which strongly indicate that real subjects indeed perform a significant degree of retrospective inference whilst performing the task. FRI makes both behavioural and neuronal predictions in a number of contexts and thus naturally suggests further avenues for exploration in future work.

Acknowledgements

THBF is supported by a European Research Council (ERC) Starting Grant under the Horizon 2020 program (Grant Agreement 804701)

References

1. Gregory RL. Perceptions as hypotheses. Philos Trans R Soc Lond B Biol Sci. 1980;290: 181–97.
2. von Helmholtz H. Handbuch der Physiologischen Optik. Leipzig: Leopold Voss; 1867.
3. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 2012;36: 181–204. doi: http://dx.doi.org/10.1017/S0140525X12000477
4. Friston KJ. The free-energy principle: a unified brain theory? Nat Rev Neurosci. Nature Publishing Group; 2010;11: 127–38. doi:10.1038/nrn2787
5. Pouget A, Beck JM, Ma WJ, Latham PE. Probabilistic brains: knowns and unknowns. Nat Neurosci. 2013;16: 1170–8. doi:10.1038/nn.3495
6. Aitchison L, Lengyel M. The Hamiltonian Brain: Efficient Probabilistic Inference with Excitatory-Inhibitory Neural Circuit Dynamics. PLoS Comput Biol. 2016;12. doi:10.1371/journal.pcbi.1005186
7. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. How to grow a mind: statistics, structure, and abstraction. Science. 2011;331: 1279–85. doi:10.1126/science.1192788
8. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci. 1999;2: 79–87. doi:10.1038/4580

9. FitzGerald THB, Hämmerer D, Friston KJ, Li S-C, Dolan RJ. Sequential inference as a mode of cognition and its correlates in fronto-parietal and hippocampal brain regions. Daunizeau J, editor. PLOS Comput Biol. 2017;13: e1005418. doi:10.1371/journal.pcbi.1005418

10. Simon HH. Theories of Bounded Rationality. In: McGuire C, Radner R, editors. Decision and organization. North-Holland; 1972. pp. 161–176.

11. Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev. 1996;103: 650–69.

12. Ortega PA, Braun DA, Dyer J, Kim K-E, Tishby N. Information-Theoretic Bounded Rationality. arXiv. 2015;1512.06789.

13. Lieder F, Griffiths TL. Strategy selection as rational metareasoning. Psychol Rev. 2017;124: 762–794. doi:10.1037/rev0000075

14. Russell S, Wefald E. Principles of metareasoning. Artif Intell. Elsevier; 1991;49: 361–395. doi:10.1016/0004-3702(91)90015-C

15. Friston K, Kiebel S. Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci. 2009;364: 1211–21. doi:10.1098/rstb.2008.0300

16. Ma WJ, Beck JM, Latham PE, Pouget A. Bayesian inference with probabilistic population codes. Nat Neurosci. 2006;9: 1432–1438. doi:10.1038/nn1790

17. Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nat Hum Behav. Nature Publishing Group; 2017;1: 0064. doi:10.1038/s41562-017-0064

18. Rao RPN, Eagleman DM, Sejnowski TJ. Optimal Smoothing in Visual Motion Perception. Neural Comput. 2001;13: 1243–1253. doi:10.1162/08997660152002843

19. Sarkka S. Bayesian filtering and smoothing. Cambridge: Cambridge University Press; 2013.

20. Chen B, Tugnait JK. Tracking of multiple maneuvering targets in clutter using IMM/JPDA filtering and fixed-lag smoothing. Automatica. Pergamon; 2001;37: 239–249. doi:10.1016/S0005-1098(00)00158-8

21. Moore JB. Discrete-time fixed-lag smoothing algorithms. Automatica. Pergamon; 1973;9: 163–173. doi:10.1016/0005-1098(73)90071-X

22. Cohn SE, Sivakumaran NS, Todling R, Cohn SE, Sivakumaran NS, Todling R. A Fixed-Lag Kalman Smoother for Retrospective Data Assimilation. Mon Weather Rev. 1994;122: 2838–2867. doi:10.1175/1520-0493(1994)122<2838:AFLKSF>2.0.CO;2

23. Friston K, Trujillo-Barreto N, Daunizeau J. DEM: a variational treatment of dynamic systems. Neuroimage. 2008;41: 849–885.

24. Wan EA, Van Der Merwe R, Nelson AT. Dual estimation and the unscented transformation. Proceedings of the 12th International Conference on Neural Information Processing Systems. MIT Press; 1999. pp. 666–672.

25. Bishop CM. Pattern Recognition and Machine Learning. New York: Springer; 2006.

26. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77: 257–286. doi:10.1109/5.18626

27. Baum LE, Petrie T, Soules G, Weiss N. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. Ann Math Stat. 1970;41: 164–171.

doi:10.1214/aoms/1177697196

28. Andrieu C, de Freitas N, Doucet A, Jordan MI. An Introduction to MCMC for Machine Learning. Mach Learn. Kluwer Academic Publishers; 2003;50: 5–43. doi:10.1023/A:1020281327116

29. Blei DM, Kucukelbir A, McAuliffe JD. Variational Inference: A Review for Statisticians. J Am Stat Assoc. Taylor & Francis; 2017; 0–0. doi:10.1080/01621459.2017.1285773

30. Beal MJ. Variational algorithms for approximate Bayesian inference. University of London. 2003.

31. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active Inference: A Process Theory. Neural Comput. 2017;29: 1–49. doi:10.1162/NECO_a_00912

32. Mackay DJC. Ensemble learning for hidden Markov models. Technical report, Cavendish Laboratory, University of Cambridge. 1997.

33. Hampton AN, Bossaerts P, O'Doherty JP. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci. 2006;26: 8360–7. doi:10.1523/JNEUROSCI.1010-06.2006

34. Schlagenhauf F, Huys QJM, Deserno L, Rapp MA, Beck A, Heinze HJ, et al. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. Neuroimage. 2014;89: 171–180. doi:10.1016/j.neuroimage.2013.11.034

35. Costa VD, Tran VL, Turchi J, Averbeck BB. Reversal learning and dopamine: a bayesian perspective. J Neurosci. 2015;35: 2407–16. doi:10.1523/JNEUROSCI.1989-14.2015

36. 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.

37. Daunizeau J. The variational Laplace approach to approximate Bayesian inference. arXiv. 2017;1703.02089.

38. Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W. Variational free energy and the Laplace approximation. NeuroImage. 2007 Jan. doi:10.1016/j.neuroimage.2006.08.035

39. Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, et al. Comparing families of dynamic causal models. PLoS Comput Biol. Public Library of Science; 2010;6: e1000709. doi:10.1371/journal.pcbi.1000709

40. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. Neuroimage. 2009;46: 1004–17. doi:10.1016/j.neuroimage.2009.03.025

41. FitzGerald THB, Moran RJ, Friston KJ, Dolan RJ. Precision and neuronal dynamics in the human posterior parietal cortex during evidence accumulation. Neuroimage. 2015;107: 219–228. doi:10.1016/j.neuroimage.2014.12.015

42. Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK. The Future of Memory: Remembering, Imagining, and the Brain. Neuron. 2012;76: 677–694. doi:10.1016/j.neuron.2012.11.001

43. Hassabis D, Kumaran D, Maguire EA. Using Imagination to Understand the Neural Basis of Episodic Memory. J Neurosci. 2007;27: 14365–14374. doi:10.1523/JNEUROSCI.4549-07.2007

44. Acuña DE, Schrater P. Structure learning in human sequential decision-making. PLoS Comput

Biol. 2010;6: e1001003. doi:10.1371/journal.pcbi.1001003

45. Braun DA, Mehring C, Wolpert DDM. Structure learning in action. Behav Brain Res. 2010;206: 157–65. doi:10.1016/j.bbr.2009.08.031

46. Tervo DGR, Tenenbaum JB, Gershman SJ. Toward the neural implementation of structure learning. Current Opinion in Neurobiology. 2016. pp. 99–105. doi:10.1016/j.conb.2016.01.014

47. Wilson RC, Nassar MR, Gold JI. Bayesian Online Learning of the Hazard Rate in Change-Point Problems. Neural Comput. 2010;22: 2452–2476. doi:10.1162/NECO_a_00007

48. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A bayesian foundation for individual learning under uncertainty. Front Hum Neurosci. 2011;5: 39. doi:10.3389/fnhum.2011.00039

49. Friston K. Variational filtering. Neuroimage. 2008;41: 747–66.

50. Shimojo S. Postdiction: Its implications on visual awareness, hindsight, and sense of agency. Frontiers in Psychology. 2014. doi:10.3389/fpsyg.2014.00196

51. Eagleman DM, Sejnowski TJ. Motion Integration and Postdiction in Visual Awareness. Science (80-). 2000;287: 2036–2038.

52. Kolers PA, von Grünau M. Shape and color in apparent motion. Vision Res. 1976;16: 329–335. doi:10.1016/0042-6989(76)90192-9

53. Dennett DC, Kinsbourne M. Time and the observer: The where and when of consciousness in the brain. Behav Brain Sci. Cambridge University Press; 1992;15: 183–201. doi:10.1017/S0140525X00068229

54. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends Cogn Sci. 2012;16: 72–80. doi:10.1016/j.tics.2011.11.018

55. Hemsley DR, Garety PA. The formation of maintenance of delusions: a Bayesian analysis. Br J Psychiatry. 1986;149: 51–6.

56. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. Front psychiatry. 2013;4: 47. doi:10.3389/fpsyt.2013.00047

57. Corlett PR, Aitken MRF, Dickinson A, Shanks DR, Honey GD, Honey RAE, et al. Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. Neuron. 2004;44: 877–888. doi:10.1016/j.neuron.2004.11.022

58. Corlett PR, Murray GK, Honey GD, Aitken MRF, Shanks DR, Robbins W, et al. Disrupted prediction-error signal in psychosis : evidence for an associative account of delusions. Brain. 2007; 2387–2400. doi:10.1093/brain/awm173

59. Pezzulo G, van der Meer MAA, Lansink CS, Pennartz CMA. Internally generated sequences in learning and executing goal-directed behavior. Trends Cogn Sci. 2014;18: 647–57. doi:10.1016/j.tics.2014.06.011

60. Jensen O, Lisman JE. Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. Trends Neurosci. 2005;28: 67–72. doi:10.1016/j.tins.2004.12.001

61. Fortin NJ, Agster KL, Eichenbaum HB. Critical role of the hippocampus in memory for sequences of events. Nat Neurosci. 2002;5: 458–62. doi:10.1038/nn834

62. Lehn H, Steffenach H-A, van Strien NM, Veltman DJ, Witter MP, Håberg AK. A specific role of

the human hippocampus in recall of temporal sequences. J Neurosci. 2009;29: 3475–84. doi:10.1523/JNEUROSCI.5370-08.2009

63. Pastalkova E, Itskov V, Amarasingham A, Buzsaki G. Internally Generated Cell Assembly Sequences in the Rat Hippocampus. Science (80-). 2008;321: 1322–1327. doi:10.1126/science.1159775

64. Penny WD, Zeidman P, Burgess N. Forward and Backward Inference in Spatial Cognition. PLoS Comput Biol. 2013;9: e1003383. doi:10.1371/journal.pcbi.1003383

65. Davidson TJ, Kloosterman F, Wilson MA. Hippocampal replay of extended experience. Neuron. 2009;63: 497–507. doi:10.1016/j.neuron.2009.07.027

66. Wikenheiser AM, Redish AD. The balance of forward and backward hippocampal sequences shifts across behavioral states. Hippocampus. 2013;23: 22–29. doi:10.1002/hipo.22049

67. Diba K, Buzsáki G. Forward and reverse hippocampal place-cell sequences during ripples. Nat Neurosci. 2007;10: 1241–1242. doi:10.1038/nn1961

68. Friston K. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci. 2005;360: 815–36. doi:10.1098/rstb.2005.1622

69. 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. doi:10.3389/fnhum.2013.00598

Appendix: Derivation of the update equations for the Bayesian HMM

The log joint for the Bayesian HMM we describe is:

$$\begin{aligned}\ln p\left(o_{1:T}, x_{1:T}, \mathbf{A}, \mathbf{B} \mid \mathbf{d}, \mathbf{\Pi}^a, \mathbf{\Pi}^b\right) = &\sum_{i=1:T} \ln p\left(o_i \mid x_i, \mathbf{B}\right) + \sum_{i=2:T} \ln p\left(x_i \mid x_{i-1}, \mathbf{A}\right) \\ &+ \ln p\left(\mathbf{B} \mid \mathbf{\Pi}^b\right) + \ln p\left(\mathbf{A} \mid \mathbf{\Pi}^a\right) + \ln p\left(x_1 \mid \mathbf{d}\right),\end{aligned} \tag{29}$$

with individual factors given by:

$$\begin{aligned}p\left(o_i \mid x_i, \mathbf{B}\right) &= \prod_{k=1}^{M} \prod_{j=1}^{K} B_{j,k}^{x_{i,j} o_{i,k}}, \\ p\left(\mathbf{A}_{j,\bullet}\right) &= Dir\left(\mathbf{A}_{j,\bullet} \mid \mathbf{\Pi}_{j,\bullet}^a\right) = C\left(\mathbf{A}_{j,\bullet}\right) \prod_{k=1}^{K} A_{j,k}^{\Pi_{j,k}^a - 1}, \\ p\left(\mathbf{B}_{j,\bullet}\right) &= Dir\left(\mathbf{B}_{j,\bullet} \mid \mathbf{\Pi}_{j,\bullet}^b\right) = C\left(\mathbf{B}_{j,\bullet}\right) \prod_{k=1}^{K} B_{j,k}^{\Pi_{j,k}^b - 1}, \\ p\left(x_1 \mid \mathbf{d}\right) &= \prod_{j=1}^{K} d_j^{x_{1,j}}.\end{aligned} \tag{30}$$

where

$$C\left(\mathbf{A}_{j\bullet}\right)=\frac{\Gamma\left(\sum_{k=1}^{K}A_{j,k}\right)}{\prod_{k=1}^{K}\Gamma\left(A_{j,k}\right)}, \tag{31}$$

and Γ denotes the gamma function.

Since exact inference in this model is intractable, we define a tractable approximating distribution by factorising such that

$$q\left(x_{1:T},\mathbf{A},\mathbf{B}\right)=q\left(x_{1:T}\right)q\left(\mathbf{A}\right)q\left(\mathbf{B}\right) \tag{32}$$

The optimal solution for $q^{*}\left(x_{1:T}\right)$ is given by:

$$\begin{aligned}
\ln q^{*}\left(x_{1:T}\right)&=\mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\ln p\left(o_{1:T},x_{1:T},\mathbf{A},\mathbf{B}\mid x_{0},\mathbf{\Pi}^{a},\mathbf{\Pi}^{b}\right)\right]+\text{const},\\
&=\mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\sum_{i=1}^{T}\ln p\left(x_{i}\mid x_{i-1},\mathbf{A}\right)+\sum_{i=1}^{T}\ln p\left(o_{i}\mid x_{i},\mathbf{B}\right)\right]+\text{const},\\
&=\mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\sum_{i=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K}x_{i-1,j}x_{i,k}\ln A_{jk}+\sum_{i=1}^{T}\sum_{k=1}^{M}\sum_{j=1}^{K}x_{i,j}o_{i,k}\ln B_{j,k}\right]+\text{const},\\
&\propto\sum_{i=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K}x_{i-1,j}x_{i,k}\ln\widehat{A}_{j,k}+\sum_{i=1}^{T}\sum_{k=1}^{M}\sum_{j=1}^{K}x_{i,j}o_{t,k}\ln\widehat{B}_{j,k}\\
\ln\widehat{A}_{jk}&\equiv\mathbb{E}\left[\ln A_{jk}\right]=\psi\left(\Pi_{jk}^{a}\right)-\psi\left(\sum\mathbf{\Pi}_{j,\bullet}^{a}\right),\\
\ln\widehat{B}_{jk}&\equiv\mathbb{E}\left[\ln B_{jk}\right]=\psi\left(\Pi_{jk}^{b}\right)-\psi\left(\sum\mathbf{\Pi}_{j,\bullet}^{b}\right).
\end{aligned} \tag{33}$$

This involves evaluating the entire joint probability distribution over $x_{1:T}$, which as discussed in the main text rapidly becomes computationally infeasible, and is often unnecessary. However, to infer on the optimal marginal distribution over states, we simply need to perform smoothing using the matrices $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$. (These are not true transition and observation matrices since the rows do not sum to one. However, the relevant normalisation constants are estimated and applied automatically during smoothing) We can then use the results of smoothing to calculate the joint distributions $\xi\left(x_{i-1},x_{i}\right)$. If we now define a set of summary statistics:

$$\begin{aligned}
\ln\widehat{\mathbf{x}}_{i}&\equiv\mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\ln p\left(x_{i}\mid o_{1:T},\mathbf{d},\mathbf{A},\mathbf{B}\right)\right]=\ln\mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\gamma\left(x_{i}\right)\right],\\
\ln\widehat{\mathbf{M}}_{i}&\equiv\mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\ln p\left(x_{i-1},x_{i}\mid o_{1:T},\mathbf{d},\mathbf{A},\mathbf{B}\right)\right]=\ln\mathbb{E}_{\mathbf{A},\mathbf{B}}\left[\xi\left(x_{i-1},x_{i}\right)\right].
\end{aligned} \tag{34}$$

it is easy to find the optimal solutions for $q^{*}\left(\mathbf{A}\right)$ and $q^{*}\left(\mathbf{B}\right)$.

$$\begin{aligned}
\ln q^{*}\left(\mathbf{A}\right)&=\mathbb{E}_{x_{1:T},\mathbf{B}}\left[\ln p\left(o_{1:T},x_{1:T},\mathbf{A},\mathbf{B}\mid\mathbf{d},\mathbf{\Pi}^{a},\mathbf{\Pi}^{b}\right)\right]+\text{const},\\
&=\mathbb{E}_{x_{1:T},\mathbf{B}}\left[\sum_{i=2}^{T}\ln p\left(x_{i}\mid x_{i-1},\mathbf{A}\right)+\ln p\left(\mathbf{A}\mid\mathbf{\Pi}^{a}\right)\right]+\text{const},\\
&\propto\sum_{i=2}^{T}\sum_{k=1}^{K}\sum_{j=1}^{K}\widehat{M}_{i,j,k}\ln A_{j,k}+\sum_{k=1}^{K}\sum_{j=1}^{K}\left(\Pi_{j,k}^{a0}-1\right)\ln A_{j,k}.
\end{aligned} \tag{35}$$

Thus by inspection the optimised factor is:

$$
\begin{aligned}
q^*(\mathbf{A}) &= \prod_{j=1}^{K} \mathrm{Dir}\left(\mathbf{A}_{j,\bullet} \mid \mathbf{\Pi}^a_{j,\bullet}\right), \\
\mathbf{\Pi}^a &= \mathbf{\Pi}^{a0} + \sum_{i=2}^{T} \mathbf{M}_i .
\end{aligned}
\tag{36}
$$

Similarly

$$
\begin{aligned}
\ln q^*(\mathbf{B}) &= \mathbb{E}_{x_{1:T},\mathbf{A}}\left[\ln p\left(o_{1:T}, x_{1:T}, \mathbf{A}, \mathbf{B} \mid \mathbf{d}, \mathbf{\Pi}^a, \mathbf{\Pi}^b\right)\right] + \text{const}, \\
&= \mathbb{E}_{x_{1:T},\mathbf{A}}\left[\sum_{i=1}^{T} \ln p\left(o_i \mid x_i, \mathbf{B}\right) + \ln p\left(\mathbf{B} \mid \mathbf{\Pi}^b\right)\right] + \text{const}, \\
&\propto \sum_{i=1}^{T}\sum_{k=1}^{M}\sum_{j=1}^{K} \widehat{x}_{i,j} o_{i,k} \ln B_{j,k} + \sum_{k=1}^{M}\sum_{j=1}^{K}\left(\Pi^{b0}_{j,k} - 1\right)\ln B_{j,k}, \\
q^*(\mathbf{B}) &= \prod_{j=1}^{K} \mathrm{Dir}\left(\mathbf{B}_{j,\bullet} \mid \mathbf{\Pi}^b_{j,\bullet}\right), \\
\mathbf{\Pi}^b &= \mathbf{\Pi}^{b0} + \sum_{i=1}^{T} \widehat{\mathbf{x}}_{i,j} \mathbf{o}^{\mathbf{T}}_{i,k} .
\end{aligned}
\tag{37}
$$

Thus we can perform approximate inference simply by iteratively evaluating the update equations for each factor.

For the probabilistic reversal model, we enforce symmetry between the rows of $\mathbf{A}$, and make use of a single Dirichlet prior that has two concentration parameters (in other words, a beta distribution).

$$
\begin{aligned}
p\left(x_i \mid x_{i-1}, \mathbf{A}\right) &= \prod_{j=1}^{2} A_{1,1}^{x_{i-1,j} x_{i,3-j}} A_{1,2}^{x_{i-1,j} x_{ij}}, \\
p\left(\mathbf{A}_{1,\bullet} \mid \mathbf{\Pi}^a\right) &= \mathrm{Dir}\left(\mathbf{A}_{1,\bullet} \mid \mathbf{\Pi}^a\right), \\
p\left(\mathbf{A}_{1,j} \mid \mathbf{\Pi}^a\right) &= p\left(\mathbf{A}_{2,3-j} \mid \mathbf{\Pi}^a\right), \\
\mathbf{\Pi}^a &= \mathbf{\Pi}^{a0} + \left[\sum_{i=2}^{T}\sum_{j=1}^{2} \widehat{M}_{i,j,3-j} \quad \sum_{i=2}^{T}\sum_{j=1}^{2} \widehat{M}_{i,j,j}\right], \\
\mathbf{A}_{2,\bullet} &= \left[A_{1,2} \quad A_{1,1}\right].
\end{aligned}
\tag{38}
$$