mathematics

MDPI

Article

A Hierarchical Bayesian Model for Inferring and Decision Making in Multi-Dimensional Volatile Binary Environments

Changbo Zhu [1,2,3], Ke Zhou [4], Fengzhen Tang [1,2,3], Yandong Tang [1,2,3], Xiaoli Li [5] and Bailu Si [6,7]*

[1] State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[2] Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China
[3] University of Chinese Academy of Sciences, Beijing 100049, China
[4] Beijing Key Laboratory of Applied Experimental Psychology, School of Psychology, Beijing Normal University, Beijing 100875, China
[5] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China
[6] School of Systems Science, Beijing Normal University, Beijing 100875, China
[7] Chinese Institute for Brain Research, Beijing 102206, China
* Correspondence: bailusi@bnu.edu.cn

Abstract: The ability to track the changes of the surrounding environment is critical for humans and animals to adapt their behaviors. In high-dimensional environments, the interactions between each dimension need to be estimated for better perception and decision making, for example in volatile or social cognition tasks. We develop a hierarchical Bayesian model for inferring and decision making in multi-dimensional volatile environments. The hierarchical Bayesian model is composed of a hierarchical perceptual model and a response model. Using the variational Bayes method, we derived closed-form update rules. These update rules also constitute a complete predictive coding scheme. To validate the effectiveness of the model in multi-dimensional volatile environments, we defined a probabilistic gambling task modified from a two-armed bandit. Simulation results demonstrated that an agent endowed with the proposed hierarchical Bayesian model is able to infer and to update its internal belief on the tendency and volatility of the sensory inputs. Based on the internal belief of the sensory inputs, the agent yielded near-optimal behavior following its response model. Our results pointed this model a viable framework to explain the temporal dynamics of human decision behavior in complex and high dimensional environments.

Keywords: Bayesian inference; filtering; free energy; decision making; predictive coding; volatility

MSC: 62C10; 62C12; 62M45; 68T07

Citation: Zhu, C.; Zhou, K.; Tang, F.; Tang, Y.; Li, X.; Si, B. A Hierarchical Bayesian Model for Inferring and Decision Making in Multi-Dimensional Volatile Binary Environments. *Mathematics* **2022**, *10*, 4775. https://doi.org/10.3390/math10244775

Academic Editor: James Liou

Received: 15 November 2022
Accepted: 5 December 2022
Published: 15 December 2022

1. Introduction

Natural environments are volatile, with ever changing sensory distributions and reward contingencies [1]. In a volatile environment, a biological agent must maintain stable internal states and be able to efficiently capture effective sensory information at the same time [2,3]. These seemingly contradictory requirements are unified into Bayesian inference [4], an optimal probability inference process.

Neuroscience research has shown that Bayesian inference underlies brain functions, such as perception, memory and decision-making, and resulting adaptive animal behaviors [3,5–11]. Adaptive behaviors are rooted in perceptual inferences and adaptive behavioral responses [12–16]. To understand the mechanisms of adaptive behaviors, one basic approach is to employ a generative model to infer the probabilistic distribution of sensory information and reproduce the temporal dynamics of human perception and decision-making in dynamic environments [17,18]. In this view, a Bayesian agent with a generative

model is able to transform sensory inputs into behavioral responses [19]. With appropriate choice of parameters, a Bayesian agent could account for human decision behaviors.

"Observing the observer" is a meta Bayesian framework to simulate the perception processes of humans [20,21]. Perceptual and response models are two key components of this framework. According to this theoretic framework, inversion of the perceptual and response models can map from sensory inputs to response actions based on variational free energy principle [15,22,23] or Bayes' rule.

To deal with volatile environments, volatility models, such as Hierarchical Gaussian Filtering [23], are developed to deliver an estimation of the changes of the environment. Accumulating evidences from the research on human learning and perception have shown that volatile Bayesian models (e.g., Hierarchical Gaussian Filtering) well explain human behaviors, especially, in changing environments. For example, saccadic response speed can be modulated by prediction precision of the belief on sensory inputs [24]. The volatility of the sensory environment and changes in sensory inputs are overestimated by adults with autism spectrum disorders [14]. This overestimation of volatility leads to the reduced precision of prior belief on sensory inputs. In human social learning, hierarchical prediction errors are encoded by midbrain and septum activity [25]. These evidences manifest that hierarchical Bayesian inference provides an optimal scheme to diminish surprise and reduce uncertainty in a volatility world.

To gain theoretical understandings of decision making under uncertainty with finite resource, the multi-armed bandit problem has been formulated as an abstraction [26–29]. The goal of the multi-armed bandit problem is to maximize the overall rewards through a series of choices. In neuroscience, multi-armed bandit problem is widely used to investigate economic decision making, contingent learning and human social behavior [30–35]. Animals and humans often have to make perceptual inference and settle on a series of decisions in a complex volatility environment. In general, the state space of decision making is high-dimensional. For example, in social interactions, the behaviors of multi-agents play important roles in the decisions of each individual. In a particular situation, agents employ internal models to observe other agents' behaviors and to simulate their belief about actions [34,35]. The interactions between agents result in complex and correlated behaviors such as competition, cooperation, prediction and judgment. To describe multi-agents' behaviors in social tasks, models that are able to capture dynamic information and noisy correlation in multi-dimensional state space need to be developed [36].

Bayesian networks are widely used for the inference of features from observed data [37–40]. In recent years, hierarchical Bayesian networks are developed to model the compositional nature of complex features for recognition tasks [41,42]. To solve perceptual inference and decision making problems in high-dimensional volatile binary environments, in this paper, we develop a hierarchical Bayesian model to infer time-varying hidden states of multi-armed bandits and maximize rewards given uncertain high-dimensional sensory inputs.

In summary, our model is promising to solve complex inference and decision making problems in realistic environments, which are intrinsically dynamic and high dimensional. In addition, our model could be applied to reveal computational mechanisms underlying human cognition and behaviors [43,44].

The rest of this paper is structured as follows. Section 2 introduces the hierarchical Bayesian perceptual model in high-dimensional volatile binary environment. Section 3 derives a set of closed form update equations for perceptual inference. Section 4 develops a response model for reward maximization in volatile multi-armed bandits as a typical example. Experimental results are given in Section 5. Finally, the paper is concluded with discussions.

2. Hierarchical Bayesian Perceptual Model

2.1. Beyond Independency

As a classic task in neuroscience and reinforcement learning, a multi-armed bandit challenges the agent with uncertain reward distribution, revealing rewards probabilis-

tically. Since the agent has to estimate both the mean reward (for exploitation) and the precision of mean reward (for exploration), the multi-armed bandit captures the exploration–exploitation tradeoff dilemma in reward maximization under uncertainty [26,45].

Put simply, a one-armed bandit can be considered as a random binary number generator described by a Bernoulli distribution

$$\mathrm{Bern}(x_0; \mu_0) = \mu_0^{x_0}(1-\mu_0)^{1-x_0}, \tag{1}$$

where $x_0 \in \{0,1\}$, the state of the one-armed bandit, represents "reward" ($x_0 = 1$) or "no reward" ($x_0 = 0$). $\mu_0 \in [0,1]$ is the probability of being in the reward state. For a multi-armed bandit, $x_0^{(i)}$ and $\mu_0^{(i)}$ denote the observation and expectation of reward in the i-th arm. The binary vector $\boldsymbol{x}_0 = [x_0^{(1)}, x_0^{(2)}, \cdots, x_0^{(d_0)}]^T$ constitutes a binary pattern corresponding to the state of rewarding or non-rewarding of the arms at time t, with d_0 being the total number of the arms. Throughout the paper we use the notation (i) in the superscript to indicate the i-th element of a vector.

Assuming independence between the reward distributions of the arms, the joint probability of being in state $\boldsymbol{x}_0$ is given by the product of reward probabilities of the arms, equivalent to

$$p(\boldsymbol{x}_0) = \exp(\sum_{i=1}^{d_0}[x_0^{(i)} \ln \mu_0^{(i)} + (1-x_0^{(i)}) \ln(1-\mu_0^{(i)})]). \tag{2}$$

However, this independent model is not able to capture possible interaction structure of the arms.

In volatile environments, the reward distributions are non-stationary and often evolve dependently on each other, showing time-variant interaction strength. To quantitatively describe the interactions among the arms of a multi-armed bandit, we introduce the concept that there are low-order interactions among the natural parameters of the underlying multivariate Bernoulli distribution. Denoted by $\boldsymbol{x}_1$, the natural parameter vector is mapped to a point $\boldsymbol{\mu}_0$ in the probability space through a multivariate element-wise sigmoid function $\boldsymbol{s}$

$$\boldsymbol{\mu}_0(t) = \boldsymbol{s}(\boldsymbol{x_1}(t), \boldsymbol{\zeta}_1), \tag{3}$$

with the i-th element of $\boldsymbol{\mu}_0(t)$ being

$$\begin{aligned} \mu_0^{(i)}(t) =& s(x_1^{(i)}(t), \zeta_1^{(i)}) \\ =& \frac{1}{1+\exp(-\zeta_1^{(i)} x_1^{(i)}(t))}, \\ & i \in \{1, 2, \cdots, d_1\}. \end{aligned} \tag{4}$$

$d_1 = d_0$ is the dimension of $\boldsymbol{x}_1$. $\boldsymbol{s}(\boldsymbol{x_1}, \boldsymbol{\zeta}_1)$ is a vector-valued function defined by

$$[s(x_1^{(1)}, \zeta_1^{(1)}), s(x_1^{(2)}, \zeta_1^{(2)}), \cdots, s(x_1^{(d_1)}, \zeta_1^{(d_1)})]^T.$$

The parameter vector

$$\boldsymbol{\zeta}_1 = [\zeta_1^{(1)}, \zeta_1^{(2)}, \cdots, \zeta_1^{(d_1)}]^T$$

is the inverse temperature, with positive elements ($\zeta_1^{(i)} > 0$).

2.2. Perceiving Tendency and Volatility

In volatile environments, variables of interest, such as reward, are subject to changes. The changes of a variable are again subject to changes, and so forth. The nested nature of volatility is a hallmark of collective phenomena as observed in animal swarms, the financial market and social behavior. To quantitatively describe volatility and pairwise correlations of high dimensional variables, we have developed a hierarchical volatility

model, called General Hierarchical Brownian Filter (GHBF), based on the idea of nested Brownian motions [46]. Following this framework, we develop here a hierarchical perceptual model to estimate both the tendency and volatility in the states of a multi-armed bandit (Figure 1). More specifically, the natural parameters $\boldsymbol{x_1}$ of the underlying multivariate Bernoulli distribution is modeled by a general Brownian motion with pervasion matrix $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{d_1 \times d_1}$

$$\boldsymbol{x}_1 = \mathcal{B}(t; \boldsymbol{\Sigma}_1). \tag{5}$$

This Brownian motion captures the tendency of the learned parameter vector $\boldsymbol{x}_1$. The volatility (i.e., uncertainties and pairwise correlations) in $\boldsymbol{x}_1$ is given by $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{d_1 \times d_1}$, which is a symmetric positive definite matrix by definition.

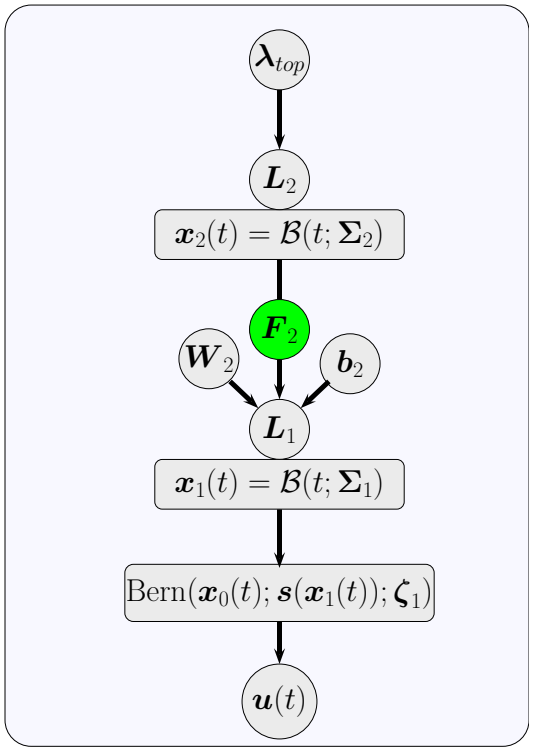

Figure 1. Overview of the hierarchical perceptual model.

Considering the fact that the pervasion intensity $\boldsymbol{\Sigma}_1$ is a symmetric positive definite matrix, it could be uniquely represented by a lower triangular matrix $\boldsymbol{L}_1 \in \mathbb{R}^{d_1 \times d_1}$ according to Cholesky decomposition

$$\boldsymbol{\Sigma_1} = \boldsymbol{L_1 L_1}^T.$$

To further evaluate the volatility $\boldsymbol{\Sigma}_1$ (i.e., uncertainties and pairwise correlations) in $\boldsymbol{x}_1$, we assume that its decomposition $\boldsymbol{L_1}$ is modeled by a general Brownian motion in its parameterized space. To be exact, the elements of $\boldsymbol{L_1}$ is parametrized by a $d_2 = d_1(d_1+1)/2$ dimensional vector $\boldsymbol{y_2}$, which results from concatenating the lower triangle elements of $\boldsymbol{L_1}$ in a column-wise fashion. The element in i-th row and j-th column of $\boldsymbol{L_1}$ is given by

$$\boldsymbol{L_1}^{(i,j)} = l_1^{(i,j)} = \begin{cases} 2\sinh(y_2^{(\frac{(2d_1-j+2)(j-1)}{2}+i-j+1)}), & 1 \le j < i \le d_1 \\ \exp(y_2^{(\frac{(2d_1-i+2)(i-1)}{2}+1)}), & j = i \end{cases} \tag{6}$$

where $\sinh(\cdot)$ denotes a hyperbolic sine function. The vector $\boldsymbol{y}_2$ gives the logarithm of volatility in the second level

$$\boldsymbol{y}_2 = \boldsymbol{W_2 x_2} + \boldsymbol{b_2}. \tag{7}$$

The coefficient matrix $\boldsymbol{W}_2$ is a d_2-by-d_2 diagonal matrix and represents the coupling strength from level two to level one. Here, $\boldsymbol{W}_2$ can simply take the form of a diagonal matrix spanned from a column vector $\boldsymbol{w}_2$ with all positive elements

$$\boldsymbol{W_2}^{(i,i)} = \boldsymbol{w}_2^{(i)}.$$

$\boldsymbol{b_2}$ and $\boldsymbol{x_2} \in \mathbb{R}^{d_2}$ represents trend and time-varying fluctuation in log-volatility of the natural parameter respectively. We may further assume that $\boldsymbol{x_2}$ evolves as a general Brownian motion with pervasion matrix $\boldsymbol{\Sigma}_2 \in \mathbb{R}^{d_2 \times d_2}$

$$\boldsymbol{x}_2 = \mathcal{B}(t; \Sigma_2). \tag{8}$$

We can rewrite the coupling (Equations (6) and (7)) as

$$\boldsymbol{L}_1 = \boldsymbol{F}_2(\boldsymbol{x}_2; \boldsymbol{w}_2, \boldsymbol{b}_2). \tag{9}$$

In the second level, the pervasion matrix $\boldsymbol{\Sigma}_2$ is chosen as a diagonal matrix. Let $\boldsymbol{L}_2 \in \mathbb{R}^{d_2 \times d_2}$ be the unique Cholesky decomposition of $\boldsymbol{\Sigma}_2$. We simply assume that $\boldsymbol{L}_2$ is a constant diagonal matrix spanned by vector $\boldsymbol{\lambda}_{top} \in \mathbb{R}^{d_2}$ with all positive components.

Figure 1 shows an overview of the hierarchical perceptual model. With this model, a Bayesian agent receives a series of sensory inputs or observations $\boldsymbol{u}_s = \{\boldsymbol{u}(t_1), \boldsymbol{u}(t_2), \cdots, \boldsymbol{u}(t_K)\}$. K is the total number of trials. At time t_k, the sensory input $\boldsymbol{u}(t_k)$ to the agent is determined by the state $\boldsymbol{x}_0(t_k)$ of the bandit deterministically, i.e., with a delta distribution $\delta(\cdot)$

$$P(\boldsymbol{u}(t_k) \mid \boldsymbol{x}_0(t_k)) = \delta(\boldsymbol{u}(t_k) = \boldsymbol{x}_0(t_k)). \tag{10}$$

In summary, the hierarchical perceptual model constitutes a generative model for sensory observations $\boldsymbol{u}(t)$ based on hidden representations of the tendency ($\boldsymbol{x}_1$) and the volatility ($\boldsymbol{x}_2$) of the observations.

3. Perceptual Inference Approximated by Variational Approximation

The aforementioned hierarchical perceptual model is constructed based on general continuous Brownian motions. It remains to develop update rules to estimate the posterior distributions for the hidden representations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. In order to derive a family of analytical and efficient updates, we discretize the continuous Brownian motions by applying Eulerian method. Sampling interval (SI) $\epsilon(t_k) = t_k - t_{k-1}$ is defined by the time that elapses between the arrival of consecutive sensory inputs $\boldsymbol{u}(t_{k-1})$ and $\boldsymbol{u}(t_k)$.

We use the variational Bayesian method [15,20,22,23] to reach an approximation to the posterior distributions of $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ given the sensory input $\boldsymbol{u}(t)$ (i.e., observation). To this end, we maximize the negative free energy, which is the lower bound of log-model evidence, to yield a variational approximation posterior (cf. Appendices A and B)

$$q(\boldsymbol{x}_h(t_k)) = \frac{1}{\mathcal{Z}_h} \exp(V_h(\boldsymbol{x}_h(t_k))), h = 1, 2,$$

where $\mathcal{Z}_h$ is a normalization constant. $V_h(\boldsymbol{x}_h(t_k)$ is the variational energy given by

$$V_h(\boldsymbol{x}_h(t_k) = E_{q(\boldsymbol{x}_s \setminus \{\boldsymbol{x}_h\}(t_k))}[\ln p(\boldsymbol{x}_s(t_k), \boldsymbol{u}(t_k) \mid \boldsymbol{\psi}_s, \epsilon(t_k))]. \tag{11}$$

Here we introduced the notation $\boldsymbol{x}_s = \{\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2\}$ to denote the set of all hidden states, $\boldsymbol{\psi}_s = \{\boldsymbol{w}_2, \boldsymbol{b}_2, \boldsymbol{\lambda}_{top}, \boldsymbol{\zeta}_1\}$ for the hyperparameters of the model, $\boldsymbol{x}_s \setminus \{\boldsymbol{x}_h\}$ for excluding $\boldsymbol{x}_h$ from the set $\boldsymbol{x}_s$, $E_{q(\boldsymbol{x})}(v)$ for the expectation of v under the distribution $q(\boldsymbol{x})$.

In order to complete the derivations, Gaussian quadratic form approximation is used as in [46]. In general, the variational energy $V_h(\boldsymbol{x}_h(t_k))$ will deviate from a Gaussian quadratic form. We have to use a Gaussian quadratic form

$$\bar{V}_h(\boldsymbol{x}_h(t_k)) = -\frac{1}{2}(\boldsymbol{x}_h(t_k) - \boldsymbol{\mu}_h(t_k))^T \boldsymbol{P}_h(t_k)(\boldsymbol{x}_h(t_k) - \boldsymbol{\mu}_h(t_k))$$

as an efficient approximation of $V_h(\boldsymbol{x}_h(t_k))$. $\boldsymbol{P}_h(t_k)$ is given by the inverse of the Hessian matrix at the last state $\boldsymbol{\mu}_h(t_{k-1})$, $\boldsymbol{P}_h(t_k) = (\boldsymbol{C}_h(t_k))^{-1} = -\nabla^2 V(\boldsymbol{\mu}_h(t_{k-1}))$, and then a local maximum point $\boldsymbol{\mu}_h(t_k)$ is found as the mode of the posterior Gaussian distribution. This approximation is made by neglecting higher order terms of the logarithm of $q(\boldsymbol{x}_h(t_k))$, and assuming Gaussian quadratic forms

$$\begin{aligned} &\boldsymbol{x}_h(t_k) \mid \boldsymbol{u}(t_k), \boldsymbol{\psi}_s \sim \mathcal{N}(\boldsymbol{\mu}_h(t_k), \boldsymbol{C}_h(t_k)). \\ &h = 1, 2 \end{aligned} \tag{12}$$

Under this approximation, the inference of the posterior distributions of $\boldsymbol{x}_h$ is reduced to the estimation of the mean $\boldsymbol{\mu}_h(t_k)$ and the covariance matrix $\boldsymbol{C}_h(t_k)$, or equivalently the precision matrix $\boldsymbol{P}_h(t_k) \equiv (\boldsymbol{C}_h(t_k))^{-1}$. Following [46], the update rules for the posterior distributions of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are derived.

At the bottom (zeroth) level of the hierarchical perceptual model, we can directly determine multivariate Bernoulli distribution $q(\boldsymbol{x}_0(t_k))$ with expectation:

$$\boldsymbol{\mu}_0(t_k) = \boldsymbol{u}(t_k). \tag{13}$$

At the first level, following Equation (11), $V_1(\boldsymbol{x}_1)$ can be calculated as

$$\begin{aligned} &V_1(\boldsymbol{x}_1(t_k)) = E_{q(\boldsymbol{x}_s \setminus \{\boldsymbol{x}_1\}(t_k))}[\ln p(\boldsymbol{x}_s(t_k), \boldsymbol{u}(t_k) \mid \boldsymbol{\psi}_s, \epsilon(t_k))] \\ =& \ln p(\boldsymbol{u}(t_k) \mid \boldsymbol{x}_0(t_k)) + E_{q(\boldsymbol{x}_0(t_k))}[\ln p(\boldsymbol{x}_0(t_k) \mid \boldsymbol{x}_1(t_k))] \\ &+ E_{q(\boldsymbol{x}_1(t_k), \boldsymbol{x}_2(t_k))}[\ln p(\boldsymbol{x}_1(t_k) \mid \boldsymbol{x}_2(t_k), \boldsymbol{W}_2, \boldsymbol{b}_2, \epsilon(t_k))] \\ \approx& \boldsymbol{\mu}_0^T(t_k) \ln \boldsymbol{s}(\boldsymbol{x}_1(t_k); \boldsymbol{\zeta}_1) + (\boldsymbol{1} - \boldsymbol{\mu}_0(t_k))^T \ln(\boldsymbol{1} - \boldsymbol{s}(\boldsymbol{x}_1(t_k); \boldsymbol{\zeta}_1)) \\ &- \frac{1}{2}(\boldsymbol{x}_1(t_k) - \boldsymbol{\mu}_1(t_{k-1}))^T \left(\epsilon(t_k)\hat{\boldsymbol{\Sigma}}_1(t_k) + \boldsymbol{C}_1(t_{k-1})\right)^{-1} (\boldsymbol{x}_1(t_k) - \boldsymbol{\mu}_1(t_{k-1})). \end{aligned}$$

where $\boldsymbol{1}$ is a d_0 dimensional column vector in which all elements are 1. Here we use the approximation

$$(\epsilon(t_k)\boldsymbol{\Sigma}_1(t_k) + \boldsymbol{C}_1(t_{k-1}))^{-1} \approx \left(\epsilon(t_k)\hat{\boldsymbol{\Sigma}}_1(t_k) + \boldsymbol{C}_1(t_{k-1})\right)^{-1}, \tag{14}$$

with $\hat{\boldsymbol{\Sigma}}_1(t_k)$ computed from the second level

$$\begin{aligned} &\hat{\boldsymbol{\Sigma}}_1(t_k) = \hat{\boldsymbol{L}}_1(t_k)\hat{\boldsymbol{L}}_1^T(t_k) \\ &\hat{\boldsymbol{L}}_1(t_k) = \boldsymbol{F}_2(\boldsymbol{\mu}_2(t_{k-1}); \boldsymbol{w}_2, \boldsymbol{b}_2). \end{aligned} \tag{15}$$

The variational energy $V_1(\boldsymbol{x}_1(t_k))$ is not a standard Gaussian quadratic form, so we have to employ a Gaussian quadratic form $\bar{V}_1(\boldsymbol{x}_1(t_k))$ to approximate it. To obtain this approximation form, we give the gradient and Hessian matrix of $V_1(\boldsymbol{x}_1(t_k))$ as follows:

$$\begin{aligned} &\nabla V_1(\boldsymbol{x}_1(t_k)) = \boldsymbol{\mu}_0(t_k) - \boldsymbol{s}(\boldsymbol{x}_1(t_k); \boldsymbol{\zeta}_1) \\ &- \frac{1}{2}\left(\epsilon(t_k)\hat{\boldsymbol{\Sigma}}_1(t_k) + \boldsymbol{C}_1(t_{k-1})\right)^{-1} (\boldsymbol{x}_1(t_k) - \boldsymbol{\mu}_1(t_{k-1})) \end{aligned} \tag{16}$$

and

$$\begin{aligned}\nabla^2 V_1(\boldsymbol{x}_1(t_k)) = &-\text{diag}(\boldsymbol{s}(\boldsymbol{x}_1(t_k);\boldsymbol{\zeta}_1)\odot(1-\boldsymbol{s}(\boldsymbol{x}_1(t_k);\boldsymbol{\zeta}_1)))\\ &-\frac{1}{2}\left(\epsilon(t_k)\hat{\boldsymbol{\Sigma}}_1(t_k)+\boldsymbol{C}_1(t_{k-1})\right)^{-1}\end{aligned} \tag{17}$$

where the operator $\odot$ is the Hadamard product. The operation $\text{diag}(\boldsymbol{v})$ is to transform a vector $\boldsymbol{v}$ into a diagonal square matrix with the elements of $\boldsymbol{v}$ on the principal diagonal.

Under the Gaussian quadratic form approximation which is based on a single-step Newton method, the tendency of $\boldsymbol{x}_0(t_k)$ is captured by

$$\boldsymbol{\mu}_1(t_k) = \boldsymbol{\mu}_1(t_{k-1}) + \text{diag}(\boldsymbol{\zeta}_1)\boldsymbol{C}_1(t_k)\boldsymbol{PE}_0(t_k) \tag{18}$$

where $\boldsymbol{PE}_0(t_k)$ is the prediction error

$$\boldsymbol{PE}_0(t_k) = \boldsymbol{\mu}_0(t_k) - \hat{\boldsymbol{\mu}}_0(t_k), \tag{19}$$

where $\hat{\boldsymbol{\mu}}_0(t_k) \equiv [\hat{\mu}_0^{(1)}(t_k), \hat{\mu}_0^{(2)}(t_k), \cdots, \hat{\mu}_0^{(d_0)}(t_k)]^T$ is the prediction according to Equation (3)

$$\hat{\boldsymbol{\mu}}_0(t_k) = \boldsymbol{s}(\boldsymbol{\mu}_1(t_{k-1}), \boldsymbol{\zeta}_1). \tag{20}$$

In Equation (18), the prediction error is scaled by the covariance matrix $\boldsymbol{C}_1(t_k)$ of the approximate Gaussian distribution, which is converted from the precision matrix

$$\begin{aligned}\boldsymbol{C}_1(t_k) \equiv &(\boldsymbol{P}_1(t_k))^{-1}\\ \boldsymbol{P}_1(t_k) = &\hat{\boldsymbol{\Pi}}_1(t_k) + \text{diag}(\boldsymbol{\zeta}_1)^2\hat{\boldsymbol{C}}_0(t_k).\end{aligned} \tag{21}$$

Here $\hat{\boldsymbol{C}}_0(t_k) = \text{diag}(\hat{\boldsymbol{\sigma}}_0(t_k))$ is the diagonal square matrix containing the observed variance

$$\hat{\boldsymbol{\sigma}}_0(t_k) = \begin{bmatrix} \hat{\mu}_0^{(1)}(t_k)(1-\hat{\mu}_0^{(1)}(t_k)) \\ \hat{\mu}_0^{(2)}(t_k)(1-\hat{\mu}_0^{(2)}(t_k)) \\ \vdots \\ \hat{\mu}_0^{(d_0)}(t_k)(1-\hat{\mu}_0^{(d_0)}(t_k)) \end{bmatrix}. \tag{22}$$

Prediction precision $\hat{\boldsymbol{\Pi}}_1(t_k)$ is given by

$$\hat{\boldsymbol{\Pi}}_1(t_k) = (\epsilon(t_k)\hat{\boldsymbol{\Sigma}}_1(t_k) + \boldsymbol{C}_1(t_{k-1}))^{-1}. \tag{23}$$

At the second level, the volatility, consisting of the uncertainties and pairwise correlations in natural parameters, is inferred by similar variational approximation method [46]. The mean is updated by

$$\begin{aligned}\boldsymbol{\mu}_2(t_k) = &\boldsymbol{\mu}_2(t_{k-1}) + \epsilon(t_k)\boldsymbol{C}_2(t_k)\boldsymbol{W}_2^T\\ &\cdot \hat{\boldsymbol{L}}_{g1}(t_k)(\boldsymbol{\Omega}_1(t_k)\otimes\boldsymbol{I}_{d_1})\mathbf{vec}\left(\boldsymbol{\Delta}_1^T(t_k)\right).\end{aligned} \tag{24}$$

Here the function $\mathbf{vec}(\boldsymbol{M}_{m\times n})$ is the vectorization of a matrix $\boldsymbol{M}$, a linear operation, to obtain a column vector of length $m \times n$ by concatenating the columns of the matrix $\boldsymbol{M}$ consecutively from column 1 to column n. The operator $\otimes$ is Kronecker product. $\boldsymbol{\Delta}_1(t_k)$ is given by

$$\boldsymbol{\Delta}_1(t_k) = \left[\boldsymbol{C}_1(t_k) + \boldsymbol{PE}_1(t_k)\boldsymbol{PE}_1^T(t_k)\right]\hat{\boldsymbol{\Pi}}_1(t_k) - \boldsymbol{I}_{d_1}. \tag{25}$$

The constant matrix $\boldsymbol{I}_d$ is a d-by-d unit square matrix. $\boldsymbol{PE}_1(t_k)$ is the prediction error on the hidden state $\boldsymbol{x}_1$

$$\boldsymbol{PE}_1(t_k) = \boldsymbol{\mu}_1(t_k) - \boldsymbol{\mu}_1(t_{k-1}). \tag{26}$$

$\hat{\boldsymbol{L}}_{g1}(t_k)$ is given by

$$\begin{aligned}
&\hat{\boldsymbol{L}}_{g1}(t_k) \\
&= \begin{bmatrix}
\exp\Big((\boldsymbol{W}_2^{(1)})^T\boldsymbol{\mu}_2(t_{k-1}) + b_2^{(1)}\Big)\boldsymbol{e}_2^T(1) \\
2\cosh\Big((\boldsymbol{W}_2^{(2)})^T\boldsymbol{\mu}_2(t_{k-1}) + b_2^{(2)}\Big)\boldsymbol{e}_2^T(2) \\
\exp\Big((\boldsymbol{W}_2^{(3)})^T\boldsymbol{\mu}_2(t_{k-1}) + b_2^{(3)}\Big)\boldsymbol{e}_2^T(3) \\
2\cosh\Big((\boldsymbol{W}_2^{(4)})^T\boldsymbol{\mu}_2(t_{k-1}) + b_2^{(4)}\Big)\boldsymbol{e}_2^T(4) \\
\vdots \\
\exp\Big((\boldsymbol{W}_2^{(d_2)})^T\boldsymbol{\mu}_2(t_{k-1}) + b_2^{(d_2)}\Big)\boldsymbol{e}_2^T(d_2)
\end{bmatrix},
\end{aligned} \tag{27}$$

where the constant vector $\boldsymbol{e}_2(d_2)$ is a d_1^2-dimension column vector. The j-th component in $\boldsymbol{e}_2^T(d_2)$ is 1 if $j = i$ or 0 if $j \neq i$. The column vector $\boldsymbol{W}_2^{(i)}$ is the i-th row in the coefficient matrix $\boldsymbol{W}_2$. $\boldsymbol{\Omega}_1(t_k)$ is

$$\boldsymbol{\Omega}_1(t_k) = \hat{\boldsymbol{L}}_1^T(t_k)\hat{\boldsymbol{\Pi}}_1(t_k). \tag{28}$$

The precision matrix is updated by

$$\begin{aligned}
\boldsymbol{P}_2(t_k) &= \hat{\boldsymbol{\Pi}}_2(t_k) + \boldsymbol{W}_2^T\hat{\boldsymbol{L}}_{g1}(t_k)\{ \\
&\epsilon(t_k)^2\boldsymbol{K}_{d_1d_1}\Big[\ \boldsymbol{\Omega}_1^T(t_k) \otimes [\boldsymbol{\Omega}_1(t_k)\boldsymbol{\Delta}_1(t_k)] \\
&\qquad + [\boldsymbol{\Delta}_1^T(t_k)\boldsymbol{\Omega}_1^T(t_k)] \otimes \boldsymbol{\Omega}_1(t_k) \\
&\qquad + \boldsymbol{\Omega}_1^T(t_k) \otimes \boldsymbol{\Omega}_1(t_k)\ \Big] \\
&+\epsilon(t_k)^2\Big[\ [\boldsymbol{L}_1^T(t_k)\boldsymbol{\Delta}_1^T(t_k)\boldsymbol{\Omega}_1^T(t_k)] \otimes \hat{\boldsymbol{\Pi}}_1(t_k) \\
&\qquad + [\boldsymbol{L}_1^T(t_k)\boldsymbol{\Omega}_1^T(t_k)] \otimes [\hat{\boldsymbol{\Pi}}_1(t_k)\boldsymbol{\Delta}_1(t_k)] \\
&\qquad + [\boldsymbol{L}_1^T(t_k)\boldsymbol{\Omega}_1^T(t_k)] \otimes \hat{\boldsymbol{\Pi}}_1(t_k)\ \Big] \\
&-\epsilon(t_k)\big[\boldsymbol{I}_{d_1} \otimes [\hat{\boldsymbol{\Pi}}_1(t_k)\boldsymbol{\Delta}_1(t_k)]\big] \\
&\}\hat{\boldsymbol{L}}_{g1}^T(t_k)\boldsymbol{W}_2 \\
&-\boldsymbol{W}_2^T\mathbf{diag}(\mathbf{lvec}(\delta_1(t_k)))\boldsymbol{W}_2
\end{aligned} \tag{29}$$

where the function $\mathbf{lvec}(\boldsymbol{L})$ is to transform a lower triangular matrix $\boldsymbol{L}$ into a column vector obtained by column stacking except all constant zero elements in the upper triangle part of the matrix. The prediction precision matrix $\hat{\boldsymbol{\Pi}}_2$ at the second level is given by

$$\hat{\boldsymbol{\Pi}}_2(t_k) = (\epsilon(t_k)\boldsymbol{\Sigma}_2 + \boldsymbol{C}_2(t_{k-1}))^{-1}. \tag{30}$$

The notation $\boldsymbol{K}_{mn}$ denotes a mn-by-mn commutation matrix. $\delta_1(t_k)$ is defined as

$$\delta_1(t_k) = \epsilon(t_k)[\boldsymbol{\Delta}_1^T(t_k)\boldsymbol{\Omega}_1^T(t_k)] \odot \hat{\boldsymbol{L}}_1(t_k). \tag{31}$$

4. Decision Making in Volatile Multi-Armed Bandits

To illustrate decision making on the basis of perceptual inference in volatile environments, we introduce, as a toy example, a two-armed bandit problem, which is a complex variant of a one-armed bandit gambling task in [30,47]. In this task, a cautious gambler is asked to bet on the outcomes of a two-armed bandit, and to maximize its overall score (Figure 2). We use upper-case letters A and B to denote the two arms of the bandit, and the

notations $x_0^{(1)}$ and $x_0^{(2)}$ for the states of arm A and B respectively. On each trail, the states of the two arms, i.e., the binary vector $\boldsymbol{x}_0 = [x_0^{(1)}, x_0^{(2)}]^T$, will be revealed to the gambler at the same time after the gambler makes a choice. There are two options available for the gambler to choose from. The first option "Same" represents the congruent states of the two arms, i.e., $[0,0]^T$ or $[1,1]^T$. The second option "Different" represents incongruent states of the two arms, i.e., $[1,0]^T$ or $[0,1]^T$. Once the gambler makes a decision (to choose "Same" or "Different"), the two arms would randomly generate their states by employing two univariate Bernoulli distributions (Equation (1)). To model a volatile environment, the state distributions of the arms are time-varying (Figure 3).

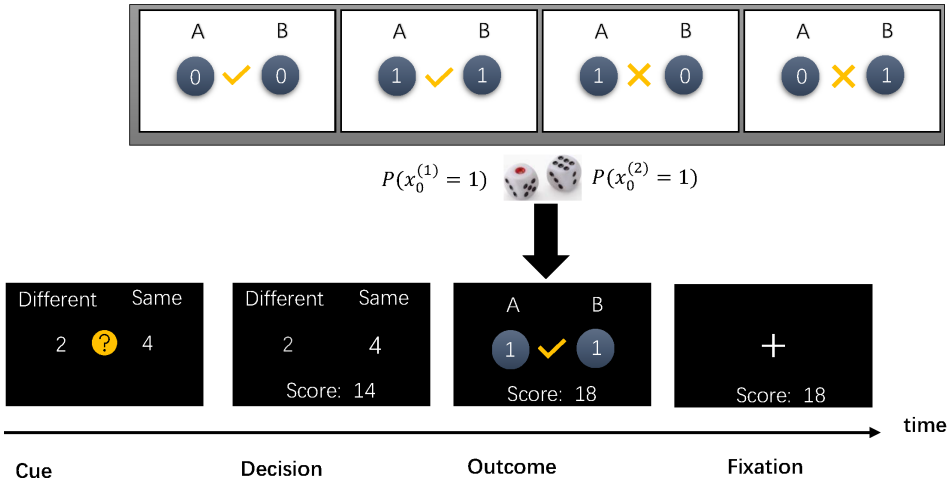

Figure 2. A gambling task. Cautious gamblers participated in a simple decision-making task in a volatile environment. There were four phases in a trial. (1) **Cue**: Two options and their rewards were presented; (2) **Decision**. Once the gambler had made a choice, the choice was displayed bigger and was highlighted; (3) **Outcome**. Once the two arms (denoted by letters A,B) had randomly generated their states, the outcome of the choice was output and then made an increment of the score only if the choice was right. (4) **Fixation**. This phase was the interval between trials. The screen only presented the score until the beginning of the next trial.

The gambler's response a is encoded as:

$$a = \begin{cases} 0, & \text{for choice 'Different'} \\ 1, & \text{for choice 'Same'} \end{cases}. \tag{32}$$

The gambler is rewarded if its choice matches the outcome of the bandit. To include volatility also in rewards, the magnitude of reward is varied from trial to trial. The reward is sampled from a reward set $\mathbb{S}_r = \{1, 2, 3, \cdots, N_r\}$, with equal probability of each reward being chosen $P(k) = 1/N_r, \forall k \in \mathbb{S}_r$.

The gambler starts the experiment with zero score. On each trial, once the chosen option turns out to be correct, the corresponding reward associated to the choice will be added to its overall score.

To maximize reward, a response model has to be defined. To this end, we first denote the rewards obtained for the correct choice of "Different" and "Same" as r_0 and r_1, respectively, can construct a reward table for each trial (Table 1).

Then we write a reward (utility) function $r(\boldsymbol{x}_0, a)$ on a trial basis according to the reward table

$$\begin{aligned} r(\boldsymbol{x}_0, a) = {} & (1 - (x_0^{(1)} - x_0^{(2)})^2)[a - (x_0^{(1)} - x_0^{(2)})^2]^2 r_1 \\ & + (x_0^{(1)} - x_0^{(2)})^2 [a - (x_0^{(1)} - x_0^{(2)})^2]^2 r_0. \end{aligned} \tag{33}$$

Table 1. Reward table.

		a	
		0	**1**
x_0	(0,0)	0	r_1
	(1,1)	0	r_1
	(1,0)	r_0	0
	(0,1)	r_0	0

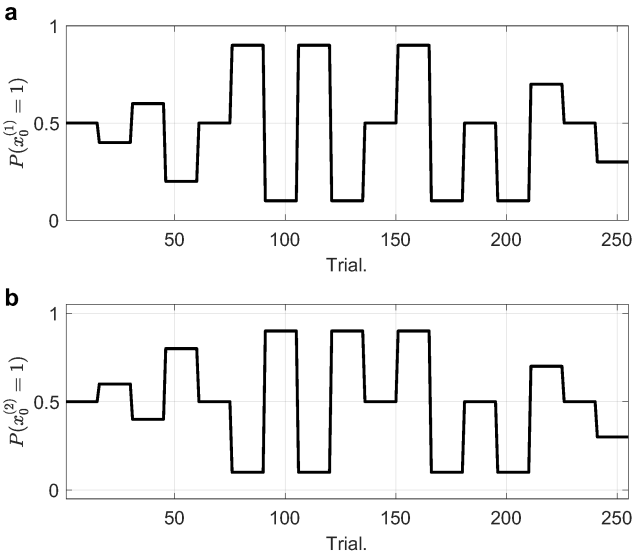

Figure 3. Expected states of the two-armed bandit. (**a**) Expected states of arm A. The mean of the state of arm A changes over time in a block fashion (black line). (**b**) Expected states of arm B. The mean of the state of arm B evolves over time (black line), showing variable correlations with that of arm A. By manipulating the expectations of the states of the two arms, we constructed a volatile environment. There were 17 blocks in the experiment. Each block consists of 15 trials.

Given the predicted state $\hat{\boldsymbol{\mu}}_0$ (Equation (20)), the expected reward of decision a under the corresponding predicted distribution $q(\boldsymbol{x}_0, \hat{\boldsymbol{\mu}}_0)$ is given by the value function

$$\begin{aligned} Q(a, \hat{\boldsymbol{\mu}}_0) &= \sum_{\boldsymbol{x}_0} r(\boldsymbol{x}_0, a)\mathrm{Bern}(\boldsymbol{x}_0; \hat{\boldsymbol{\mu}}_0) \\ &= \sum_{\boldsymbol{x}_0} r(\boldsymbol{x}_0, a)\mathrm{Bern}(x_0^{(1)}; \hat{\mu}_0^{(1)})\mathrm{Bern}(x_0^{(2)}; \hat{\mu}_0^{(2)}) \\ &= a^2 r_1[(1-\hat{\mu}_0^{(1)})(1-\hat{\mu}_0^{(2)}) + \hat{\mu}_0^{(1)}\hat{\mu}_0^{(2)}] \\ &\quad + (a-1)^2 r_0[(1-\hat{\mu}_0^{(1)})\hat{\mu}_0^{(2)} + \hat{\mu}_0^{(1)}(1-\hat{\mu}_0^{(2)})] \\ &= \begin{cases} r_1[(1-\hat{\mu}_0^{(1)})(1-\hat{\mu}_0^{(2)}) + \hat{\mu}_0^{(1)}\hat{\mu}_0^{(2)}], & a=1 \\ r_0[(1-\hat{\mu}_0^{(1)})\hat{\mu}_0^{(2)} + \hat{\mu}_0^{(1)}(1-\hat{\mu}_0^{(2)})], & a=0 \end{cases} \end{aligned} \tag{34}$$

The agent makes decisions according to a Boltzmann distribution constructed from the value function. The probability of choosing action a is defined by

$$P_a = \frac{\exp(Q(a, \hat{\boldsymbol{\mu}}_0))}{\sum_b \exp(Q(b, \hat{\boldsymbol{\mu}}_0))}. \tag{35}$$

For the binary decision-making task considered here, the probability of choosing action $a = 1$ is reduced to a sigmoid function

$$\begin{aligned} P_1 =& \frac{1}{1 + \exp(-(Q(1, \hat{\boldsymbol{\mu}}_0) - Q(0, \hat{\boldsymbol{\mu}}_0)))} \\ =& s(Q(1, \hat{\boldsymbol{\mu}}_0) - Q(0, \hat{\boldsymbol{\mu}}_0), 1), \end{aligned} \tag{36}$$

where $s(\cdot, \cdot)$ the sigmoid function defined in Equation (4).

In fact, a biological agent maximizes long-term rewards, instead of immediate rewards, using decision noise as a mechanician to tradeoff exploration and exploitation. We introduce a probability weighting function [47,48] with a noise parameter $\zeta_a > 0$ to include decision noise. The probability of choosing action $a = 1$ is

$$P(a = 1 \mid \hat{\boldsymbol{\mu}}_0, \zeta_a) = \frac{P_1^{\zeta_a}}{P_1^{\zeta_a} + (1 - P_1)^{\zeta_a}}. \tag{37}$$

Up to now, we have defined a response model (Equations (33)–(37)) based on Bayesian decision theory to maximize expected rewards. The response model is a function of the decision evidence $(Q(1, \hat{\boldsymbol{\mu}}_0) - Q(0, \hat{\boldsymbol{\mu}}_0))$, i.e., the difference between expected rewards for the two options ("Different", "Same"). If the decision evidence is positive, the probability of choosing "Same" exceeds 0.5, and the optimal action is to choose "Same" or $a = 1$. If the decision evidence is a negative number, the probability of choosing "Different" exceeds 0.5 and the optimal action is the option "Different" or $a = 0$.

5. Simulation Results

The combination of the perceptual model (Equations (5)–(10)) and the response model (Equations (33)–(37)) constitute a Bayesian model (denoted by $\mathcal{M}_1$) for decision making in volatile multi-armed bandits. To assess the model's ability to adapt to volatility, we simulated a gambler with the proposed Bayesian decision model to solve the two-armed bandit task (Figure 2). In the simulation, trials are organized into seventeen blocks, each of which contains 15 trials (Figure 3). The state expectations of the bandit change across blocks, resulting in volatility in sensory inputs. The reward set is specified as $\mathbb{S}_r = \{1, 2, 3, 4\}$.

For an ideal observer, it has the access to the actual state $\boldsymbol{u}(t) = [u^{(1)}(t), u^{(2)}(t)]^T$ generated by the bandit at each time t (Figure 3). Given this ideal information, the ideal observer could make the ideal actions $a_{ideal}(t)$

$$a_{ideal}(t) = \begin{cases} 0, & \text{if } u^{(1)}(t) \neq u^{(2)}(t) \\ 1, & \text{if } u^{(1)}(t) = u^{(2)}(t) \end{cases}. \tag{38}$$

Based on this series of ideal actions, the cumulative reward obtained by the ideal observer could be computed.

To measure the performance of decision making behavior in the above gambling task, we define a probabilistically optimal reference for comparison. For this purpose, we consider an informed agent, who is given the expectation of the states of the volatile bandit $[P(x_0^{(1)}(t) = 1), P(x_0^{(1)}(t) = 1)]^T$. The informed agent needs not learn the states of the bandit, and it uses the same action selection mechanism (Equations (34)–(37)) of the response model π_r to obtain the probabilistically optimal expectation of response action $a(t)$, denoted by $P^*(a(t) = 1)$. For a decision-making agent, only if the agent fully understands the volatile environments, the expectation of its action $a(t)$ can completely coincide with the probabilistically optimal expectation of response action $P^*(a(t) = 1)$. Overestimating or underestimating the environmental states will lead to the expectation of the agent's action $a(t)$ to deviate from the optimal behavior. Therefore, $P^*(a(t) = 1)$ constitutes the optimal decision making behavior of a learning agent could reach. The deviation of a

learning agent with sensory inputs $\boldsymbol{u}_s$, actions a_s and rewards $\boldsymbol{r}_s$ from the informed agent in decision making behavior is measured by regret $R(P(a(t_k)=1)|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)$, defined by

$$R(P(a(t_k)=1)|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) = \sum_{k=1}^{K} |P(a(t_k)=1) - P^*(a(t_k)=1)|, \tag{39}$$

where $P(a(t_k)=1)$ is generated by the learning agent.

5.1. Dynamics of Bayesian Decision Making

We employed a Bayesian agent $\mathcal{M}_1$, which is endowed with the proposed hierarchical perceptual model and binary response model (Figure 4), to perform the above gambling task (Figure 2). All free parameters of our Bayesian agent $\mathcal{M}_1$ is defined in Appendix C, and forms a random variable vector denoted by $\boldsymbol{\xi}_1$. Their initial sufficient statistics of all parameters are listed in Table 2. In details, the optimization of the free parameters was carried out in three steps as follows before the model was used for the gambling task.

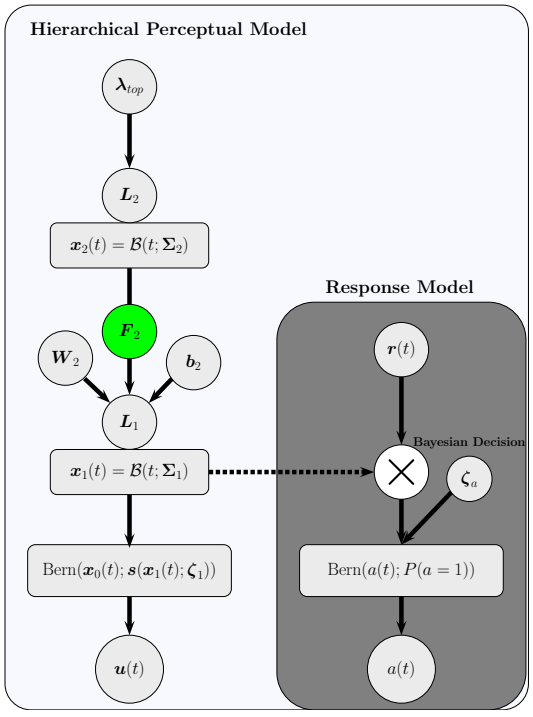

Figure 4. A Bayesian agent consists of the proposed hierarchical perceptual model and a binary response model based on Bayesian decision theory. The reward $\boldsymbol{r}(t) = [r_0(t), r_1(t)]^T$ is drawn uniformly from a set on each trial.

(1) **Generating synthetic data.** According to the expected states of the arms (Figure 3), we randomly generated a sequence of multivariate binary inputs

$$\boldsymbol{u}_s = \{\boldsymbol{u}(t_1), \boldsymbol{u}(t_2), \boldsymbol{u}(t_3), \cdots, \boldsymbol{u}(t_K)\}, (K = 255).$$

Then the series of ideal actions $a_s = \{a_{ideal}(t_1), a_{ideal}(t_2), \cdots, a_{ideal}(t_K)\}$ is computed by Equation (38).The random reward sequence $\boldsymbol{r}_s = \{\boldsymbol{r}(t_1), \boldsymbol{r}(t_2), \boldsymbol{r}(t_3), \cdots, \boldsymbol{r}(t_K)\}$ is generated from uniform distribution $\mathcal{U}(1,4)$ based on the reward set $\mathbb{S}_r = \{1,2,3,4\}$.

(2) **Initializing sufficient statistics of all random parameters**. To allow our model to work well for sensory inputs, we choose particular initial sufficient statistics of the random parameter vector $\boldsymbol{\xi}_1$, and determined the prior distribution of $\boldsymbol{\xi}_1$. The configuration for the parameters of the Bayesian agent (Figure 4) is shown in Table 2.

(3) **Maximizing negative free energy**. To obtain the optimal prior parameters $\left(\boldsymbol{\mu}_{\xi_1}^*, \boldsymbol{C}_{\xi_1}^*\right)$ of the parameter $\boldsymbol{\xi}_1$, we maximize negative free energy (Equations (A19)–(A21)) by using the quasi-Newton Broyden–Fletcher–Goldfarb–Shanno method based on a line search framework [49].

Table 2. Parameters of our hierarchical Bayesian model. Parameters labeled by 'Free' are optimized by the inversion of the model. Fixed parameters are constant and not optimized. The notation **1** is a constant column vector with all components being 1. The notation **0** is a zero vector. The matrix $\boldsymbol{O}_d$ is a d by d constant matrix in which all elements are 0. The notation $\text{logit}(\cdot)$ denotes a logistic function $\text{logit}(x) = \ln(\frac{x}{1-x})$. Given all initial priors, we search for the optimal priors on all optimized parameters $\boldsymbol{\mu}_{\tilde{\xi}}$ according to the free energy principle (Equations (A19) and (A21)).

Name	Description	Initial Value	Fixed or Free
	Parameters of our Bayesian perceptual model		
$d_0 = d_{\boldsymbol{u}}$	Dimension of sensory input $\boldsymbol{u}$	2	constant
d_1	Dimension of $\boldsymbol{x}_1$	2	constant
d_2	Dimension of $\boldsymbol{x}_2$	3	constant
$\epsilon(t_k)$	Sampling interval $\epsilon(t_k)$	1	constant
$\boldsymbol{\alpha}_{\boldsymbol{\lambda}_{top}}$	Upper bound on $\boldsymbol{\lambda}_{top}$	$\sqrt{0.1} \cdot \mathbf{1}$	constant
$\boldsymbol{\lambda}_{top}$	Volatility of $\boldsymbol{x}_2$		Free
$\boldsymbol{\mu}_{\boldsymbol{\lambda}_{top}^G}$	Mean of $\boldsymbol{\lambda}_{top}^G$	$\text{logit}(0.1) \cdot \mathbf{1}$	
$\boldsymbol{C}_{\boldsymbol{\lambda}_{top}^G}$	Covariance of $\boldsymbol{\lambda}_{top}^G$	$1 \times 10^{-2} \boldsymbol{I}_{d_2}$	
$\boldsymbol{\alpha}_{\boldsymbol{w}_2}$	Upper bound on $\boldsymbol{w}_2$	$1 \cdot \mathbf{1}$	constant
$\boldsymbol{w}_2$	Coupling strength		Free
$\boldsymbol{\mu}_{\boldsymbol{w}_2^G}$	Mean of $\boldsymbol{w}_2^G$	$\text{logit}(0.25) \cdot \mathbf{1}$	
$\boldsymbol{C}_{\boldsymbol{w}_2^G}$	Covariance of $\boldsymbol{w}_2^G$	$1 \times 10^{-2} \cdot \boldsymbol{I}_{d_2}$	
$\boldsymbol{b}_2$	Coupling bias	$\mathbf{0}$	Fixed
$\boldsymbol{\mu}_{\boldsymbol{b}_2}$	Mean of $\boldsymbol{b}_2$	$\mathbf{0}$	
$\boldsymbol{C}_{\boldsymbol{b}_2}$	Covariance of $\boldsymbol{b}_2$	$\boldsymbol{O}_3$	
$\boldsymbol{\mu}_2(t_0)$	Prior mean of $\boldsymbol{x}_2$		Free
$\boldsymbol{\mu}_{\boldsymbol{\mu}_2(t_0)}$	Mean of $\boldsymbol{\mu}_2(t_0)$	$\ln(0.16) \cdot \mathbf{1}$	
$\boldsymbol{C}_{\boldsymbol{\mu}_2(t_0)}$	Covariance of $\boldsymbol{\mu}_2(t_0)$	$1 \times 10^{-2} \cdot \boldsymbol{I}_3$	
$\boldsymbol{C}_2(t_0)$	Prior covariance of $\boldsymbol{x}_2$		Free
$\boldsymbol{\mu}_{\boldsymbol{c}_2^G}$	Mean of $\boldsymbol{c}_2^G$	$\ln(1)$	
$\boldsymbol{C}_{\boldsymbol{c}_2^G}$	Covariance of $\boldsymbol{c}_2^G$	$\boldsymbol{I}_{d_2}$	
$\boldsymbol{\mu}_1(t_0)$	Prior mean of $\boldsymbol{x}_1$		Free
$\boldsymbol{\mu}_{\boldsymbol{\mu}_1(t_0)}$	Mean of $\boldsymbol{\mu}_1(t_0)$	$\mathbf{0}$	
$\boldsymbol{C}_{\boldsymbol{\mu}_1(t_0)}$	Covariance of $\boldsymbol{\sigma}_1^G$	$\boldsymbol{O}_{d_2}$	
$\boldsymbol{C}_1(t_0)$	Prior covariance of $\boldsymbol{x}_1$		Free
$\boldsymbol{\mu}_{\boldsymbol{\sigma}_1^G}$	Mean of $\boldsymbol{\sigma}_1^G$	$\ln(0.16) \cdot \mathbf{1}$	
$\boldsymbol{C}_{\boldsymbol{\sigma}_1^G}$	Covariance of $\boldsymbol{\sigma}_1^G$	$0.1\boldsymbol{I}_{d_1}$	
$\boldsymbol{\zeta}_1$	Coefficient		Fixed
$\boldsymbol{\mu}_{\boldsymbol{\zeta}_1^G}$	Mean of $\boldsymbol{\zeta}_1^G$	$\mathbf{0}$	
$\boldsymbol{C}_{\boldsymbol{\zeta}_1^G}$	Covariance of $\boldsymbol{\zeta}_1^G$	$\boldsymbol{O}_2$	
	Parameters of our response model		
d_a	Dimension of a	1	Fixed
$\boldsymbol{\zeta}_a$	Coefficient		Fixed
$\mu_{\zeta_a^G}$	Mean of ζ_a^G	$\ln(2)$	
$C_{\zeta_a^G}$	Covariance of ζ_a^G	0	

In order to reveal the dynamic interaction between the natural parameters of the two-armed bandit, we show one example gambling process performed by the Bayesian agent characterized by the optimal parameters. The Bayesian agent tracked online the tendency $\boldsymbol{\mu}_1$ of the natural parameters $\boldsymbol{x}_1$ associated to the bandit (Figure 5), so that it is able to make decisions based on the estimated decision evidence. The evolution of $\boldsymbol{\mu}_1$ follows well of the trend of expected states of the bandit (Figure 3), generating good prediction of the states (Figure 6b,c).

After the observation of the tentative rewards (Figure 6a) during the cue phase of a trial, the Bayesian agent makes a choice in the decision phase according to the perceptual model and the response model. In the first block of the simulation (trial 1 to 15, Figure 3), the two arms have the same expected states of maximal uncertainty, i.e., $P(x_0^{(1)} = 1) = P(x_0^{(2)} = 1) = 0.5$, and the binary state patterns of the two-armed bandit are equal probable. Both of the belief states $\mu_0^{(1)}, \mu_0^{(2)}$ of two arms fluctuates around 0.5 (Figure 6). During this block, the prediction correlation $\hat{\rho}_1$ fluctuates and decreases slightly towards zero, reflecting the fact that the states of the two arms are uncorrelated. From the second block to the tenth block (trial 16 to 150, Figure 3), the expected states of the two arms are incongruent. Therefore, the changes in the prediction tendency $\hat{\mu}_1^{(1)}$ (Figure 5a, as well as in the predicted mean $\hat{\mu}_0^{(1)}$ of arm A in Figure 6b) are on average in opposite directions as the changes in the prediction tendency $\hat{\mu}_1^{(2)}$ (Figure 5b, as well as in the predicted mean $\hat{\mu}_0^{(2)}$ of arm B in Figure 6c). Meanwhile, the prediction correlation $\hat{\rho}_1$ continues to decrease during this stage (Figure 6d), manifesting the incongruency of the two arms. From the eleventh block to the seventeenth block (trial 151 to 255, Figure 3), the changes in $\hat{\mu}_1^{(1)}$ and $\hat{\mu}_1^{(2)}$ share the same trend (Figure 5), so do the changes in $\hat{\mu}_0^{(1)}$ and $\hat{\mu}_0^{(2)}$ (Figure 6b,c), due to the fact that the two arms have the same expected states. Consequently, the prediction correlation $\hat{\rho}_1$ continues to increase during this stage (Figure 6d).

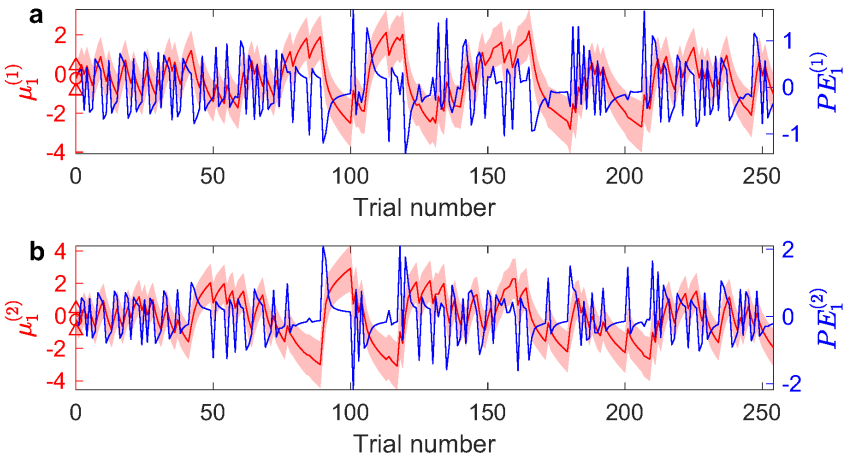

Figure 5. Temporal dynamics of the tendency $\boldsymbol{\mu}_1$ of the natural parameter at the first level. (**a**) The evolution of $\mu_1^{(1)}$, the first component of $\boldsymbol{\mu}_1$, is shown in red. The time-varying trajectory of the prediction error $PE_1^{(1)}$ is shown in blue. (**b**) The evolution of $\mu_1^{(2)}$, the second component of $\boldsymbol{\mu}_1$, is shown in red. The time-varying trajectory of the prediction error $PE_1^{(2)}$ is shown in blue. Light-red shaded area represents the uncertainty of each quantity (i.e., $\mu_1^{(i)}(t) \pm \sqrt{C_1^{(i,i)}(t)}, i \in \{1,2\}$). The red markers $\triangle, \circ$ represent the priors on the standard deviation and mean of each quantity.

The log-volatility in the natural parameters ($\mu_2^{(1)}$ and $\mu_2^{(3)}$, i.e., internal representation of the expected states) of the two arms has notable changes from the third block to the fourteenth block (trial 31 to 210, Figure 7a,c). The changes are more evident from the sixth block to the fourteenth block, during which volatility is more vigorous. From the second to the tenth blocks (trial 31 to 150), the expected states of the two arms are not equal.

Instead, they become incongruent (Figure 3). During this period, the log-volatility state $\mu_2^{(2)}$, corresponding to the prediction correlation $\hat{\rho}_1$, decreased and kept a descending trend (Figure 7b). This is consistent with the fact that the two arms are incongruent at the time. As a contrast, from the eleventh block to the seventeenth block, the expected states of the two arms are equal (trial 151 to 255, Figure 3), therefore, the Bayesian learner discovered an increasing log-volatility state $\mu_2^{(2)}$ during this stage (Figure 7b).

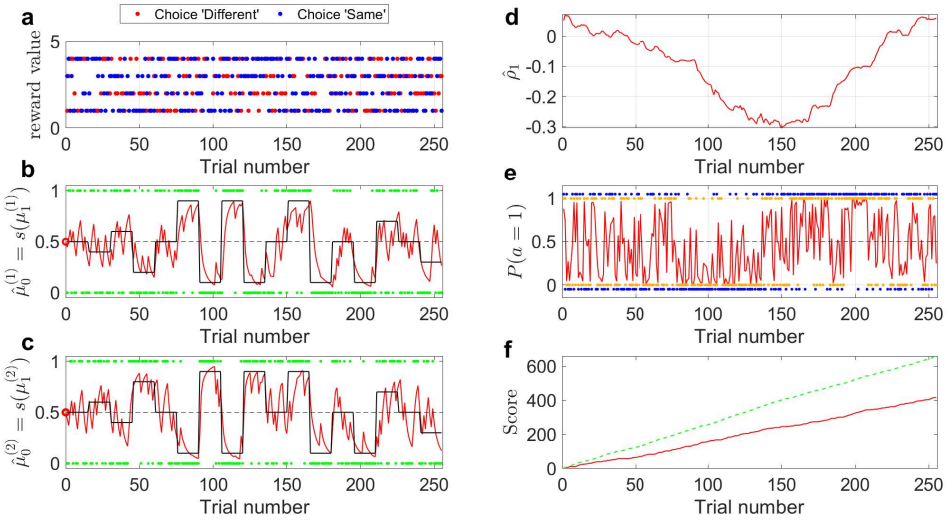

Figure 6. Temporal dynamics of the posterior states in a gambling task. (**a**) Rewards for two choices "Different' and "Same" were randomly generated by a discrete uniform distribution $\mathcal{U}(1,4)$. Blue dots represent the reward value for option "Same" on each trial, and red dots for option "Different". (**b**) The green dots are the sensory inputs of $u^{(1)}$ (i.e., states of arm A). The red line represents the estimated probability $\hat{\mu}_0^{(1)}(t_k) = s(\mu_1^{(1)}(t_{k-1}), \zeta_1^{(1)})$. (**c**) The green dots are the sensory inputs of $u^{(2)}$ (i.e., states of arm B). The red line represents the estimated probability $\hat{\mu}_0^{(2)}(t_k) = s(\mu_1^{(2)}(t_{k-1}), \zeta_1^{(2)})$. (**d**) Prediction correlation $\hat{\rho}_1(t)$ is extracted from the inverse prediction precision $\hat{\mathbf{\Pi}}_1(t)$ generated by the second (log-volatility) level. (**e**) Blue dots denote the optimal choice a_{ideal} on each trial. The red line is the trajectory of the expectation probability that the states of two arms of the bandit are the same (i.e., $P(a=1)$). The orange dots are the response action a generated by the agent on each trial. (**f**) The green dashed line is the cumulative reward of the ideal observer taking the ideal actions a_{ideal}. The red line shows the cumulative reward obtained by the Bayesian agent.

In Figure 8, our Bayesian agent $\mathcal{M}_1$ is compared with the informed agent. In our Bayesian decision model, the evidence for decision-making is quantified by the probability $P(a(t)=1)$ (red solid line in Figure 8a). It is close to the probabilistically optimal expectation of response action $P^*(a(t)=1)$ given by the informed agent (blue dashed line in Figure 8a). The action selection behavior of our Bayesian agent was similar to the optimal probability decision pattern. In this experiment, the regret $R(P(a(t_k)=1)|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)$ is worked out by substituting $P(a(t)=1)$ generated by our Bayesian agent into Equation (39), and is 27.3588.

Since the Bayesian agent made decisions based on the estimated decision evidence, it may be distracted by high rewards associated to the wrong action. In the third block (trial 31 to 45), the rewards of the option "Same" were higher than the option "Different", the Bayesian agent were biased towards choosing "Same", reflexing the fact that less likely but highly rewarded actions are worth to be tried (Figure 6e). This phenomenon was also evident in the beginning of the eleventh block (trial 151 to 155), where high rewards were more often assigned to the option "Different". The Bayesian agent from time to time reduced the probability of choosing the option "Same", leading to select "Different" more often (Figure 6e). The cumulative reward obtained by the Bayesian agent maintains a linear increasing trend irrespective of the volatility (red line in Figures 6f and 8b), keeping close to the reward gained by the informed agent (blue line in Figure 8b).

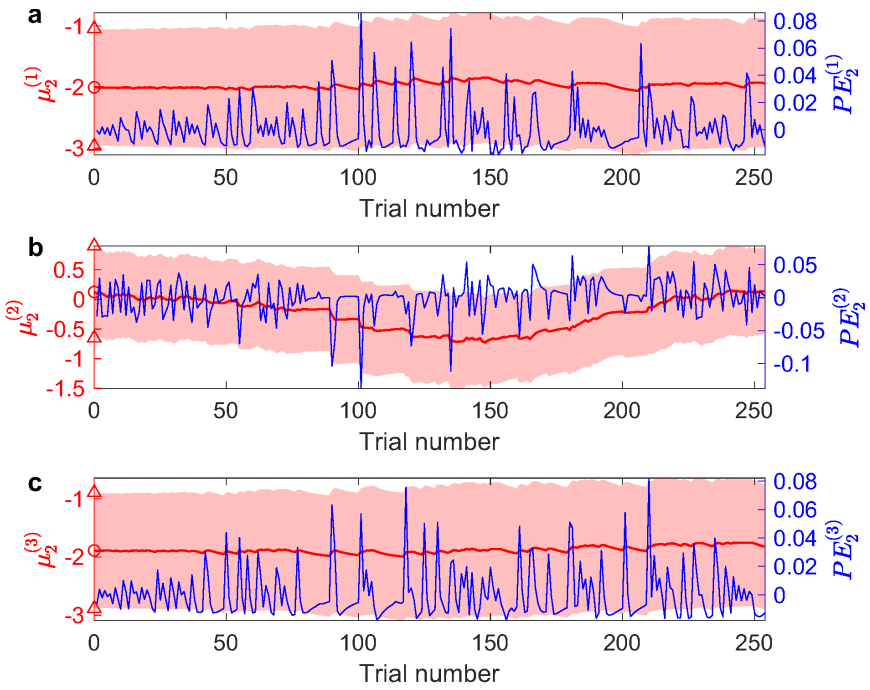

Figure 7. Temporal dynamics of the expectation of the logarithm of volatility $\boldsymbol{\mu}_2$ in the natural parameter $\boldsymbol{x}_1$ at the second level. Each panel shows the evolution of one element of $\boldsymbol{\mu}_2$ in red and the corresponding element of $\boldsymbol{PE}_2$ in blue. Light-red shaded area represents the uncertainty of each quantity (i.e., $\mu_2^{(i)}(t) \pm \sqrt{C_2^{(i,i)}(t)}, i \in \{1,2,3\}$). The red markers $\triangle, \circ$ represent the priors of the standard deviation and mean of each quantity.

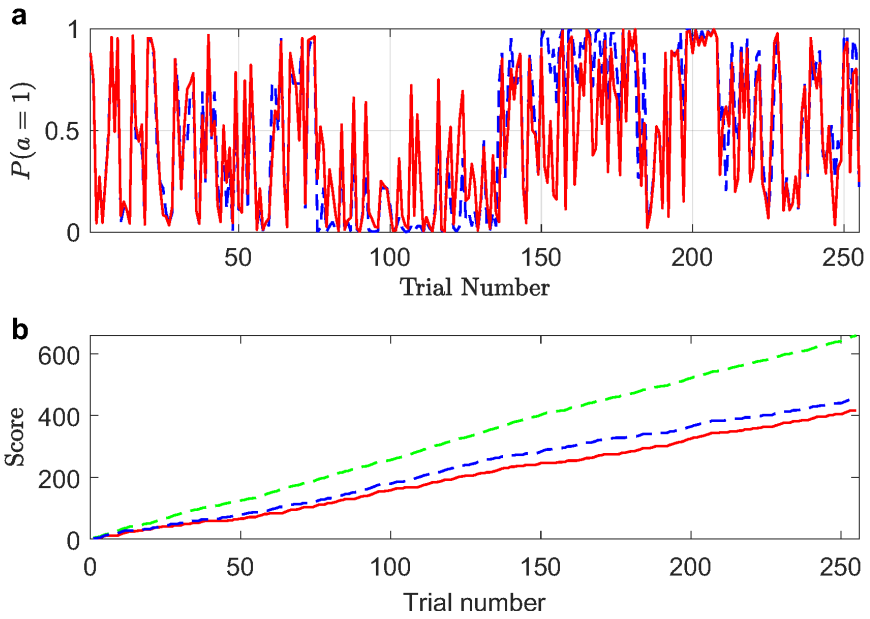

Figure 8. Temporal dynamics of the expectation of response action $P(a(t) = 1)$. (**a**) The expectation of response action $P(a(t) = 1)$ generated by our Bayesian agent $\mathcal{M}_1$ (red solid line) match closely to the probabilistically optimal expectation of response action $P^*(a(t) = 1)$ (blue dashed line). (**b**) The cumulative reward obtained by our Bayesian agent (red solid line) tightly follows the cumulative reward of the probabilistically optimal expectation of response action $P^*(a(t) = 1)$ given by the informed agent (blue dashed line). The ideal observer, who knows the actual outcomes of the bandit in advance, has the highest cumulative reward of the task.

5.2. Bayesian Model Selection

In order to evaluate the proposed hierarchical Bayesian model for inferring and decision making, we adopt the Bayesian model selection methodology [50]. It is a general principle to favor a model that achieves balanced tradeoff between complexity and flexibility. The proposed hierarchical Bayesian model has the sophisticated complexity to capture volatility in a multiscale fashion. We compare it with a well-known baseline model in psychology and Reinforcement Learning (RL), namely the Rescorla–Wagner (RW) model [26,51]. As a special case of the Temporal-Difference Learning method, the Rescorla–Wagner model updates value estimations based on prediction errors [26].

To perform fair comparisons, we construct a variant of the RW model using the same response model as the proposed hierarchical Bayesian model (cf. Appendix G). The agent with the RW model and the above response model is denoted by $\mathcal{M}_2$. Under the same variational Bayesian learning scheme, we search the optimal parameters for $\mathcal{M}_2$ on each sequence of sensory inputs (Appendix D).

We conducted a Bayesian model selection experiment to compare the proposed Bayesian agent $\mathcal{M}_1$ based on a variant of GHBF with the agent $\mathcal{M}_2$ based on the RW model. The detailed simulation was performed as the following steps.

(1) **Generating synthetic dataset** $\mathcal{D}$. According to Figure 3, we randomly generated 100 sequences of multivariate binary inputs $\boldsymbol{u}_s = \{\boldsymbol{u}(t_1), \boldsymbol{u}(t_2), \boldsymbol{u}(t_3), \cdots, \boldsymbol{u}(t_K)\}$ ($K = 255$). Then the series of ideal actions $a_s = \{a_{ideal}(t_1), a_{ideal}(t_2), \cdots, a_{ideal}(t_K)\}$ are computed according to Equation (38). Random reward sequences $\boldsymbol{r}_s = \{\boldsymbol{r}(t_1), \boldsymbol{r}(t_2), \boldsymbol{r}(t_3), \cdots, \boldsymbol{r}(t_K)\}$ are generated from uniform distribution $\mathcal{U}(1,4)$ based on the reward set $\mathbb{S}_r$. Here we used the notation $\mathcal{D}$ to denote the set of sensory and action sequences

$$\mathcal{D} = \{\boldsymbol{u}_s, \boldsymbol{r}_s, a_s | \boldsymbol{u}_s \text{ and } \boldsymbol{r}_s \text{ are repeatedly generated}\}.$$

(2) **Initializing sufficient statistics of all random parameters in our Bayesian agent** $\mathcal{M}_1$. We choose particular initial sufficient statistics of a parameter vector $\boldsymbol{\xi}_1$ to allow the Bayesian agent $\mathcal{M}_1$ to work well on all sequences of sensory inputs. Then we determined the prior distribution of $\boldsymbol{\xi}_1$. All configurations for parameters of the agent based on GHBF (Figure 4) are shown in Table 2 .

(3) **Initializing sufficient statistics of all random parameters in the RW-agent** $\mathcal{M}_2$. We determined a particular initial value of a parameter vector $\boldsymbol{\xi}_2$ (Table A2) for the agent $\mathcal{M}_2$. All configurations for parameters of the agent based on Rescorla–Wagner model were shown in Table A2. The response model of the RW model uses the same parameter configuration as in the Bayesian agent in step 2.

(4) **Maximizing negative free energy**. On each sequence of sensory inputs, we performed an optimization method to obtain the optimal prior parameters $\left(\boldsymbol{\mu}^*_{\xi_1}, \boldsymbol{C}^*_{\xi_1}\right)$ of the parameter $\boldsymbol{\xi}_1$ for the agent $\mathcal{M}_1$ and the optimal prior parameters $\left(\boldsymbol{\mu}^*_{\xi_2}, \boldsymbol{C}^*_{\xi_2}\right)$ of the parameter $\boldsymbol{\xi}_2$ for the agent $\mathcal{M}_2$ according to Equation (A21) respectively. In this paper, we implemented the quasi-Newton Broyden–Fletcher–Goldfarb–Shanno method based on a line search framework [49] to obtain negative free energy maximization (Equations (A19)–(A21)).

(5) **Evaluating negative free energy**. On each sequence of sensory inputs, we can evaluate the maximum negative free energies $\mathcal{F}^*_{\xi_1}$ for the agent $\mathcal{M}_1$ and $\mathcal{F}^*_{\xi_2}$ for the agent $\mathcal{M}_2$ according to Equation (A22). Then Bayesian Factors are evaluated by Equations (A33) and (A34).

In each gambling task, two agents $\mathcal{M}_1, \mathcal{M}_2$ generate their time-courses of the predicted states $\boldsymbol{\mu}_0(t)$ on the expectation of the states of the two-armed bandit. The predicted states are recorded into a $d_0 \times K$ matrix $\boldsymbol{T}$

$$\boldsymbol{T} = [\boldsymbol{\mu}_0(t_1), \boldsymbol{\mu}_0(t_2), \cdots, \boldsymbol{\mu}_0(t_K)] \in \mathbb{R}^{d_0 \times K}.$$

Given any pair of sequences (randomly generated in step 1) $(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) \in \mathcal{D}$, our Bayesian agent $\mathcal{M}_1$ and the RW agent dynamically infer the states, and form inference trajectories $\boldsymbol{T}_1(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)$ and $\boldsymbol{T}_2(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)$ respectively. The mean dynamic inference trajectory $\bar{T}_i$ and the standard deviation $\sigma_{\boldsymbol{T}_i}$ are computed as

$$\begin{aligned} \bar{\boldsymbol{T}}_i &= \frac{1}{|\mathcal{D}|} \sum_{(\boldsymbol{u}_s, a_s, \boldsymbol{u}_s) \in \mathcal{D}} \boldsymbol{T}_i(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) \\ \sigma_{\boldsymbol{T}_i} &= \sqrt{\frac{1}{|\mathcal{D}| - 1} \sum_{(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) \in \mathcal{D}} [T_i(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) - \bar{T}_i]^2}, \end{aligned} \tag{40}$$

where the notation $|\mathcal{D}|$ is the number of elements in the dataset $\mathcal{D}$ (e.g., $|\mathcal{D}| = 100$).

Figure 9 shows that both the Bayesian agent (red lines) and the RW agent (blue dashed lines) are able to track the ground truth or real probabilities well, showing quick jumps at the points where the ground truth undertakes remodeling. However, the RW agent produces more variable predictions, as shown by larger standard deviation (blue shaded areas vs. red shaded areas). The RW agent often overshoots its estimation (blue dashed lines vs. black lines). These results demonstrate that the RW agent overfits observations.

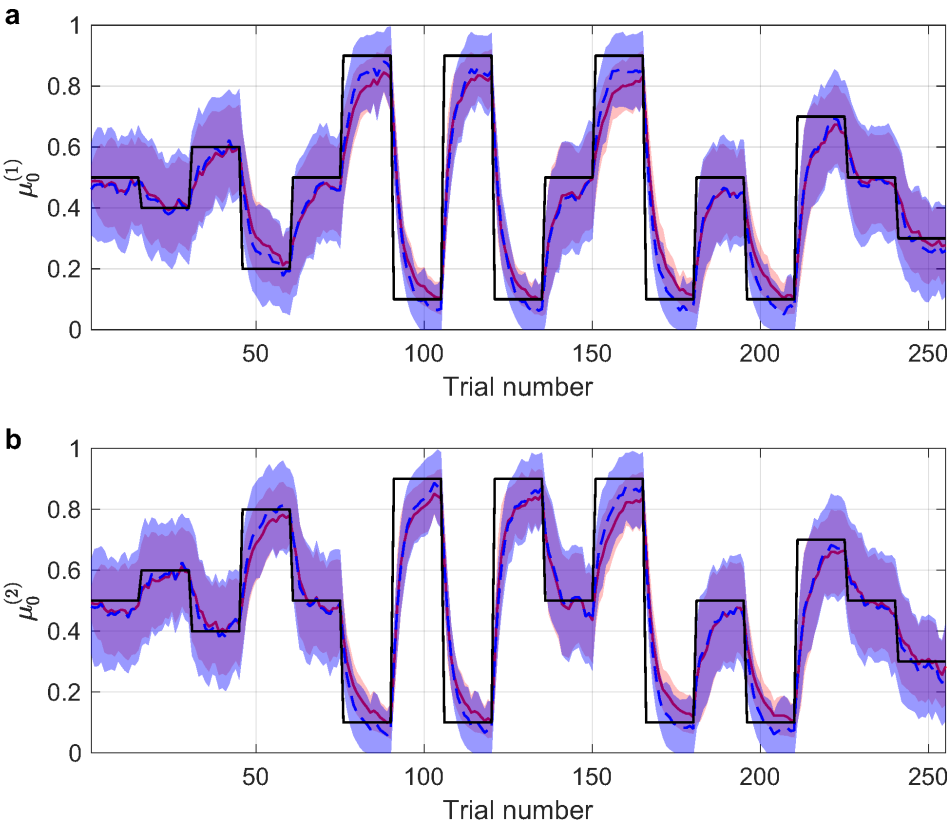

Figure 9. The mean inference trajectory $\bar{T}_i (i = 1, 2)$ of the predicted states of the two-armed bandit x_0. (**a**) The evolution of mean inference trajectories corresponding to $\mu_0^{(1)}$ over the dataset $\mathcal{D}$. (**b**) The evolution of mean inference trajectories corresponding to $\mu_0^{(2)}$ over the dataset $\mathcal{D}$. In both panels, the groundtruth is shown by black lines. The mean inference trajectories given by the Bayesian agent and the RW agent are in red and blue, respectively. The shaded areas correspond to the standard deviations $\pm\sigma_{\boldsymbol{T}_i}$.

In each gambling task, given the sensory inputs $\boldsymbol{u}_s$, actions a_s and rewards $\boldsymbol{r}_s$, two agents $\mathcal{M}_1, \mathcal{M}_2$ generate their series of the expectations of the response action $P(a(t) = 1)$, denoted by $\{P_1(a(t_k) = 1) | k = 1, 2, \cdots, K\}$ and $\{P_2(a(t_k) = 1) | k = 1, 2, \cdots, K\}$ respectively.

Their regrets can be evaluated by substituting $P_i(a(t_k) = 1)$ into Equation (39). The mean regret $\bar{R}_i$ and the standard deviation σ_{T_i} on the synthetic dataset $\mathcal{D}$ are computed as

$$\begin{aligned} \bar{R}_i =& \frac{1}{|\mathcal{D}|} \sum_{(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) \in \mathcal{D}} R(P_i(a(t_k) = 1) | \boldsymbol{u}_s, a_s, \boldsymbol{r}_s) \\ \sigma_{\boldsymbol{R}_i} =& \sqrt{\frac{1}{|\mathcal{D}| - 1} \sum_{(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) \in \mathcal{D}} [R(P_i(a(t_k) = 1) | \boldsymbol{u}_s, a_s, \boldsymbol{r}_s) - \bar{R}_i]^2}. \end{aligned} \tag{41}$$

The mean $\bar{R}_1$ and standard deviation $\sigma_{\boldsymbol{R}_1}$ of our Bayesian agent (based on GHBF) $\mathcal{M}_1$ are smaller than the mean $\bar{R}_2$ and standard deviation $\sigma_{\boldsymbol{R}_2}$ of the RW-agent (based on the RW model) $\mathcal{M}_2$ (Figure 10).

To evaluate the two models more formally, given the three sequences $(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s)$, we computed Bayesian Factor BF without Bayesian Information Criterion (BIC)

$$\begin{aligned} BF :=& \frac{p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s | \mathcal{M}_1)}{p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s | \mathcal{M}_2)} \\ \approx& \exp(\mathcal{F}^*_{\xi_1} - \mathcal{F}^*_{\xi_2}), \end{aligned} \tag{42}$$

and Bayesian Factor BF_{BIC} with BIC (cf. Appendices E and F)

$$BF_{BIC} := \exp\left(\mathcal{F}^*_{\xi_1} - \mathcal{F}^*_{\xi_2} - \frac{d_{\xi_1} - d_{\xi_2}}{2} \ln(K)\right), \tag{43}$$

where d_{ξ_i} is the number of free parameters estimated by the model. The notations $\mathcal{F}^*_{\xi_1}, \mathcal{F}^*_{\xi_2}$ are respectively the maximal negative free energies of the two agents $\mathcal{M}_1, \mathcal{M}_2$ on the given pair of the sequences $(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s)$. Under both measures, Bayesian Factors on the observation dataset $\mathcal{D}$ are concentrated on the range larger than 100 (i.e., $BF > 100$, $BF_{BIC} > 100$) (Figure 11a,b), meaning decisive evidence for the Bayesian agent outperforming the RW agent according to Table A1.

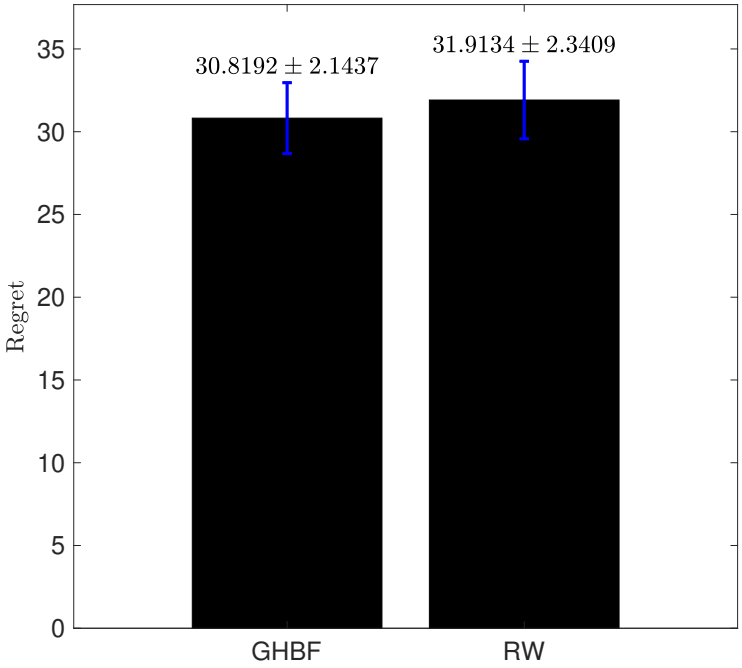

Figure 10. The statistic mean $\bar{R}_i$ and standard deviation $\sigma_{\boldsymbol{R}_i}$ of the regrets on the synthetic dataset $\mathcal{D}$.

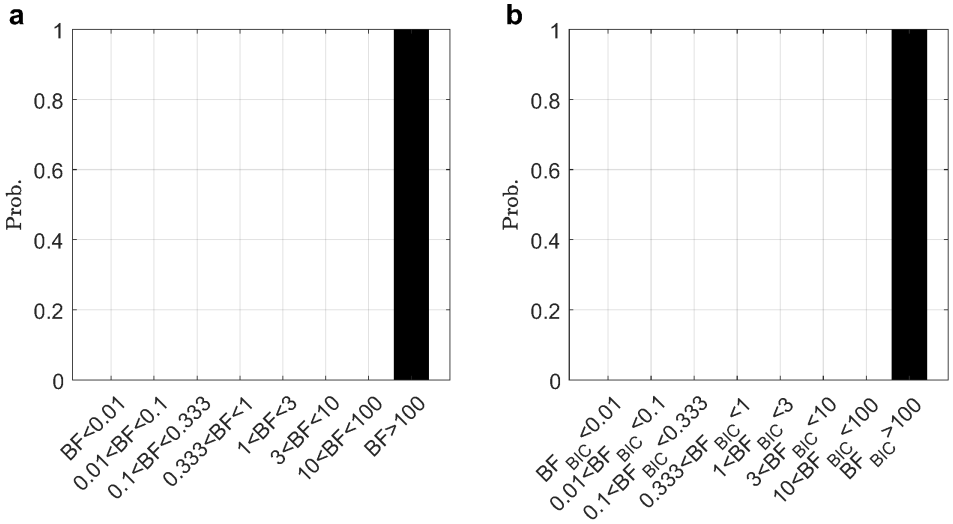

Figure 11. Histogram of Bayesian Factor. (**a**) Bayesian Factor without the Bayesian Information Criterion BF. (**b**) Bayesian Factor with the Bayesian Information Criterion BF_{BIC}.

6. Discussion

6.1. Contributions of This Work

In this article, we have introduced a hierarchical Bayesian model that describes how to infer volatility (i.e., environmental uncertainty and correlations) in a multi-dimensional space. In this model, the bottom level is to learn the state expectation of a multi-armed bandit, which is described by a multivariate Bernoulli distribution. The natural parameter $\boldsymbol{x}_1$ of the Bernoulli distribution is learned by the first level. Under the Brownian and Gaussian assumption on $\boldsymbol{x}_1$, volatility can be strictly determined by the Cholesky Decomposition of pervasion intensity of Brownian motion $\boldsymbol{x}_1$ [46]. Therefore, we can define the volatility in $\boldsymbol{x}_1$ as the Cholesky Decomposition of pervasion intensity of $\boldsymbol{x}_1$. Next, the volatility in $\boldsymbol{x}_1$ can be represented by $\boldsymbol{x}_2$, with evolves as a Brownian motion. The low-order interactions between the dimensions of the Bernoulli distribution and the environmental uncertainties are captured in the second level, corresponding to $\boldsymbol{x}_2$.

The hierarchical Bayesian model assumes that the tendency of a binary pattern evolves as a general Brownian motion at the first level. The tendency can be updated by Equation (18), where prediction error $\boldsymbol{PE}_0(t_k)$ is the information gap between the agent's belief and sensory input. This quantity is a target that the agent should learn to diminish. The parameter vector $\boldsymbol{\zeta}_1$ functions as weighting vector to weight prediction error $\boldsymbol{PE}_0(t_k)$. The covariance $\boldsymbol{C}_1(t_k)$ plays the role of complex adaptive learning rate in Equation (18).

In principle, the proposed model could be easily generalized to a Bayesian framework for decision making in high-dimensional multinary environments, by defining appropriate forms of perceptual models and response models. In this study, the input space was assumed to be binary. For multinary environments, the representations of the tendency of the inputs could be defined accordingly to form a hierarchical perceptual model. Here we derived a response model from Bayesian decision theory with the goal of maximizing expected rewards or minimizing expected risk or loss [47,52]. For other problems of interest, it is sufficient to construct a compatible response model addressing the particular optimization criteria of the question. For example, recognition and navigation tasks could be formulated in the proposed Bayesian framework to cope with the interactions between multimodal information [53,54].

In summary, the main contributions of this work are twofold. First, the model captures the correlations between the dimensions of the sensory space, and is able to make decisions contingent on the structure of the sensory inputs. Simulations show that our model is applicable to complex inference and decision making tasks that could not be tackled by methods with independence assumptions of the high dimensional input features. Second, the model represents the tendency and volatility of the sensory inputs in a hierarchical

manner based on the idea of nested Brownian motions. The resulting hierarchical computational framework naturally allows the interactions between layers, and is able to track the dynamics of the environment.

6.2. Related Works

The proposed hierarchical Bayesian model is most related to the Rescorla–Wagner model [35,51,55,56]. Equation (18) has the form of a generalized form of the Rescorla–Wagner equation in reinforcement learning [26]

$$\boldsymbol{\mu}_1(t_k) = \boldsymbol{\mu}_1(t_{k-1}) + \boldsymbol{\Gamma}\Delta\boldsymbol{\mu}_1(t_k),$$

where $\Delta\boldsymbol{\mu}_1(t_k)$ is an error signal (or learned target) at time t_k. In the cognitive neuroscience field, some variants of the RW model have been introduced for the behavioral paradigm of multi-armed bandits. However, due to the limitation of the RW model, it is difficult to capture the volatility of the signal. More importantly, since the learning rate in RW model is constant, it is difficult to interpret the subject's dynamic process of capturing effective information during the experiment. As an example, given the reward $R(t_{k-1})$ at time t_k, the standard RW model estimates the value state variable V by

$$\begin{aligned} V(t_k) &= V(t_{k-1}) + \alpha PE(t_k) \\ PE(t_k) &= R(t_{k-1}) - V(t_{k-1}), \end{aligned} \tag{44}$$

which is simplified to

$$V(t_k) = (1-\alpha)V(t_{k-1}) + \alpha R(t_{k-1}). \tag{45}$$

It is clear to see that the learning rate α plays a role of a moving average, weighting initial value $V(t_0)$ and a reward sequence $R(t_1), R(t_2), \cdots, R(t_k)$. This is an inflexible filtering method to cope with volatility. For small learning rate α, the RW model prefers to predict based on the input history, a good scenario for slow changing signals. The RW model with large learning rate prefers to rely on most recent rewards, a good scenario for fast changing signals. However, the RW model did not unify the two learning processes (i.e., the learning rate is not able be adapted depending on the environment and agent state). In this sense, our hierarchical Bayesian model provides a theoretically justified mechanism to adapt learning rate dynamically according to the volatility of the environment and the states of the agent.

For a single-step update, the time complexity of our hierarchical Bayesian model is $O(d_0^4)$ (Equations (13)–(31)), while the time complexity of the RW model is $O(d_0)$. In our model, capturing volatility to form adaptive learning rate leads to a higher computational cost. Experiments show that this computational cost is necessary for the model to flexibly adapt to volatile environments. On the synthetic dataset $\mathcal{D}$, the trajectories of the state estimation formed by our hierarchical Bayesian model (light red shadow area in Figure 9) are distributed narrower than those of the RW model (light blue shadow area in Figure 9), indicating the stability and robustness of the proposed model.

6.3. Strengths and Limitations

Our hierarchical Bayesian model is general enough to be easily applied in high-dimensional environments. The number of parameters of the model scales quadratically with respect to the dimension of the input space. Given the number of dimensions d_0, corresponding to the dimension of $\boldsymbol{x}_0(t)$ at the bottom level (i.e., sensory input $\boldsymbol{u}(t)$), the dimension of the parameter $\boldsymbol{\xi}_1$ of our perceptual model is $d_0 + 2d_1 + 5d_2 = \frac{d_0(d_0+5)}{2}$ (cf. Appendix C). In the Bayesian learning process, the optimization algorithm (i.e., quasi-Newton Broyden–Fletcher–Goldfarb–Shanno method) needs to numerically evaluate the gradient of the negative free energy with respect to each component of the model parameter $\boldsymbol{\xi}_1$. For a large number of dimensions d_0, parallel computing framework based

on CPU and GPU need to be developed in order to improve the evaluation efficiency of numerical gradients.

6.4. Future Work

In this paper, we construct a hierarchical Bayesian model for inferring and decision-making in multivariate volatile binary environment, and test and validate it on a synthetic dataset. We plan to use this model to explain human decision-making behaviors and brain activities. To this end, we need to collect behavior and neuroimaging data while human subjects are performing the same task of multi-armed bandit as defined in this paper. For theoretical interest, the mechanism of the adaptive learning rate and correlation among natural parameters are worthy for further clarification in our hierarchical model, and we look forward to analyzing these critical mechanisms in future investigation.

7. Conclusions

We have introduced a hierarchical Bayesian model for decision making in high-dimensional volatile environments, and derived a family of interpretable closed form update rules. Based on this framework, we define a Bayesian agent endowed with the proposed hierarchical Bayesian model, as a mentalizing model of a biological agent, to perform an abstract multi-armed bandit task. Simulations show that our model is applicable to complex tasks that could not be tackled by models with independency assumptions. Crucially, the proposed model contains a hierarchical perceptual model that is able to capture different covariances (e.g., prediction covariance, posterior covariance, likelihood covariance). As an important indicator of mental process, prediction correlation is dynamically estimated in the second level of the hierarchical perceptual model. Prediction correlation describes quantitatively (weak) pairwise interactions among different perception quantities (e.g., natural parameters of multi-armed bandits). In conclusion, the proposed hierarchical Bayesian model provides a powerful tool to solve complex perception and decision making problems in high-dimensional volatile environments [57], as well as to quantify complex phenomena such such as perceptual decision making, spatial navigation, social interactions and exploratory behaviors [58–63].

Author Contributions: C.Z. and B.S. conceived the model. C.Z. conducted the simulations, with inputs from B.S., K.Z., F.T., Y.T. and X.L. C.Z. and B.S. analyzed the results. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science and Technology Innovation 2030 Major Program of China grant number 2022ZD0205000.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Custom code is written by the authors based on MATLAB. Raw data and MATLAB code to analyze results can be accessed at https://github.com/changbozhu/GHBF-mvBern-simulation.git (accessed on 8th December 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this paper:

GHBF	General Hierarchical Brownian Filter
SI	Sampling interval
RL	Reinforcement Learning
RW	Rescorla–Wagner
BF	Bayesian Factor
BIC	Bayesian Information Criterion

Appendix A. Bayesian Agent

A Bayesian agent summarizes its past experience of perceptions and decisions to adapt to the external environment. After observing sensory input $\boldsymbol{u}$, the agent integrates internal salability priors obtained from past experience with the current information provided by sensory input. Then it yields inferences and predictions of about the external environment. Based on the estimated states of the external environment, the agent makes a decision to manipulate the external environment. Bayes' rule subserves optimal probability inferences and calculus for representing beliefs and acting in external environment in an efficient and consistent manner [22,23,64,65].

More specifically, a Bayesian agent $\mathcal{M}$ is defined by the likelihood $p(\boldsymbol{u}|\boldsymbol{x},\mathcal{M})$ and a priori $p(\boldsymbol{x}|\mathcal{M})$ on the hidden state $\boldsymbol{x}$. After recieving sensory input $\boldsymbol{u}$, the agent infers a posterior distribution $p(\boldsymbol{x}|\boldsymbol{u},\mathcal{M})$ according to Bayes' rule for future perception and action

$$p(\boldsymbol{x}|\boldsymbol{u},\mathcal{M}) = \frac{p(\boldsymbol{x}|\mathcal{M})p(\boldsymbol{u}|\boldsymbol{x},\mathcal{M})}{\int p(\boldsymbol{x}|\mathcal{M})p(\boldsymbol{u}|\boldsymbol{x},\mathcal{M})d\boldsymbol{x}}. \tag{A1}$$

To act in the external environment, the agent selects an action $\boldsymbol{a}^*$ from an action set $\mathcal{A}$ according to the prediction or posterior distribution of hidden states $p(\boldsymbol{x}|\boldsymbol{u},\mathcal{M}) \approx q(\boldsymbol{x};\boldsymbol{\chi})$, with $\boldsymbol{\chi}$ being the sufficient statistics of posterior hidden states $\boldsymbol{x}$. $q(\boldsymbol{x};\boldsymbol{\chi})$ is an approximation for the true posterior $p(\boldsymbol{x}|\boldsymbol{u},\mathcal{M})$. In general, a response model π_r is defined to map hidden states into actions, which can be a deterministic or stochastic mapping.

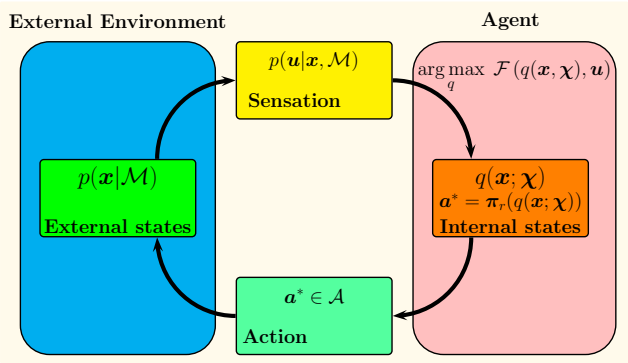

Figure A1. Interaction between an agent and the external environment.

Unfortunately, the integral in Equation (A1) is intractable to compute. To calculate the above posterior $p(\boldsymbol{x}|\boldsymbol{u},\mathcal{M})$, we resort to variational Bayesian methods [22] to approximate Bayesian inference efficiently. This is done by finding a lower bound on the logarithm of model evidence $\ln p(\boldsymbol{u}|\mathcal{M})$, called negative free energy $\mathcal{F}(q(\boldsymbol{x};\boldsymbol{\chi}),\boldsymbol{u})$

$$\begin{aligned}
&\ln p(\boldsymbol{u}|\mathcal{M}) \\
=&\ln \int q(\boldsymbol{x};\boldsymbol{\chi})\frac{p(\boldsymbol{u},\boldsymbol{x}|\mathcal{M})}{q(\boldsymbol{x};\boldsymbol{\chi})}d\boldsymbol{x} \\
\geq& \int q(\boldsymbol{x};\boldsymbol{\chi})\ln\frac{p(\boldsymbol{u},\boldsymbol{x}|\mathcal{M})}{q(\boldsymbol{x};\boldsymbol{\chi})}d\boldsymbol{x} \\
=&\int q(\boldsymbol{x};\boldsymbol{\chi})\ln p(\boldsymbol{u},\boldsymbol{x}|\mathcal{M})d\boldsymbol{x} \\
&-\int q(\boldsymbol{x};\boldsymbol{\chi})\ln q(\boldsymbol{x};\boldsymbol{\chi})d\boldsymbol{x} \\
=&H(\boldsymbol{x};\boldsymbol{\chi}) - U(\boldsymbol{x};\boldsymbol{\chi}) \\
=&\ln p(\boldsymbol{u}|\mathcal{M}) - D_{KL}[q(\boldsymbol{x};\boldsymbol{\chi})||p(\boldsymbol{x}|\boldsymbol{u},\mathcal{M})] \\
=&\mathcal{F}(q(\boldsymbol{x};\boldsymbol{\chi}),\boldsymbol{u})
\end{aligned}, \tag{A2}$$

where $H(\boldsymbol{x};\boldsymbol{\chi}) = -\int q(\boldsymbol{x};\boldsymbol{\chi}) \ln q(\boldsymbol{x};\boldsymbol{\chi}) d\boldsymbol{x}$ is the entropy, and

$$U(\boldsymbol{x};\boldsymbol{\chi}) = -\int q(\boldsymbol{x};\boldsymbol{\chi}) \ln p(\boldsymbol{u},\boldsymbol{x}|\mathcal{M}) d\boldsymbol{x}$$

is the internal energy. Equation (A2) tells that the lower bound is negative free energy, i.e., the entopy $H(\boldsymbol{x};\boldsymbol{\chi})$ minus the internal energy $U(\boldsymbol{x};\boldsymbol{\chi})$.The Kullback–Leibler divergence $D_{KL}[q(\boldsymbol{x};\boldsymbol{\chi})||p(\boldsymbol{x}|\boldsymbol{u},\mathcal{M})] \geq 0$ measures the difference between the approximation and the true posterior. The better the approximation $q(\boldsymbol{x};\boldsymbol{\chi})$ is, the smaller the divergence is. The minimal divergence 0 occurs when the ideal approximation $q(\boldsymbol{x};\boldsymbol{\chi})$ is equal to $p(\boldsymbol{x}|\boldsymbol{u},\mathcal{M})$. The agent therefore could obtain the optimal approximation posterior $q(\boldsymbol{x};\boldsymbol{\chi})$ by maximizing negative free energy $\mathcal{F}(q(\boldsymbol{x};\boldsymbol{\chi}),\boldsymbol{u})$

$$q(\boldsymbol{x};\boldsymbol{\chi}_0) = \arg\max_q \mathcal{F}(q(\boldsymbol{x};\boldsymbol{\chi}),\boldsymbol{u}). \tag{A3}$$

We use the Lagrange method to solve this maximization problem. The Lagrangian functional is defined as

$$\begin{aligned} &\bar{\mathcal{F}}(q(\boldsymbol{x};\boldsymbol{\chi}),\boldsymbol{u}) \\ &= \mathcal{F}(q(\boldsymbol{x};\boldsymbol{\chi}),\boldsymbol{u}) + v[\int q(\boldsymbol{x};\boldsymbol{\chi}) d\boldsymbol{x} - 1]' \end{aligned} \tag{A4}$$

where v is a Lagrange multiplier. The solution of the optimal problem (Equation (A3)) is also the solution of the variational equation (Equation (A5))

$$\frac{\delta \bar{F}(q(\boldsymbol{x};\boldsymbol{\chi}),\boldsymbol{u})}{\delta q} = 0. \tag{A5}$$

Appendix B. Variational Bayesian Inference

Given a Bayesian perceptual model $p(\boldsymbol{x}_s,\boldsymbol{u}|\boldsymbol{\psi},\epsilon)$, where $\boldsymbol{\psi}$ and ϵ are parameters, the model evidence $p(\boldsymbol{u}|\boldsymbol{\psi},\epsilon)$ is often analytically intractable. Therefore, exact Bayesian posteriors could not be analytically calculated. We apply variational Bayesian methods to transform the calculation of exact Bayesian posteriors $p(\boldsymbol{x}_s|\boldsymbol{u},\boldsymbol{\psi},\epsilon)$ into finding the optimal variational posteriors $q(\boldsymbol{x}_s)$ (cf. Equations (A2)–(A5) in Appendix A). The lower bound on the logarithm of the model evidence $p(\boldsymbol{u}|\boldsymbol{\psi},\epsilon)$ is given by

$$\begin{aligned} &\ln p(\boldsymbol{u}|\boldsymbol{\psi},\epsilon) \\ =&\ln \int q(\boldsymbol{x}_s) \frac{p(\boldsymbol{u},\boldsymbol{x}_s|\boldsymbol{\psi},\epsilon)}{q(\boldsymbol{x}_s)} d\boldsymbol{x}_s \\ \geq& \int q(\boldsymbol{x}_s) \ln \frac{p(\boldsymbol{u},\boldsymbol{x}_s|\boldsymbol{\psi},\epsilon)}{q(\boldsymbol{x}_s)} d\boldsymbol{x}_s \\ =& \int q(\boldsymbol{x}_s) \ln p(\boldsymbol{u},\boldsymbol{x}_s|\boldsymbol{\psi}_s,\epsilon) d\boldsymbol{x}_s \\ &- \int q(\boldsymbol{x}_s) \ln q(\boldsymbol{x}_s) d\boldsymbol{x}_s \\ =& -U(\boldsymbol{x}_s) + H(\boldsymbol{x}_s) \\ =&\mathcal{F}(q(\boldsymbol{x}_s)). \end{aligned} \tag{A6}$$

Then we use an important assumption that marginal variational posteriors over latent variables are independent, i.e., the joint variational posterior distribution factorizes with respect to all marginal posteriors

$$q(\boldsymbol{x}_s) = \prod_{\boldsymbol{x}_h \in \boldsymbol{x}_s} q(\boldsymbol{x}_h), \tag{A7}$$

where $\boldsymbol{x}_h$ is one element of $\boldsymbol{x}_s$. The factorized form in Equation (A7) corresponds to the so-called *mean field approximation*, an approximation scheme developed in statistical mechanics.

It should be noted that we now wish to maximize the negative free energy $\mathcal{F}(q(\boldsymbol{x}_s))$ with respect to each approximation posterior $q(\boldsymbol{x}_h)$ under the constraint of normalized probability $\int q(\boldsymbol{x}_h)d\boldsymbol{x}_h = 1, \forall h$. The Lagrangian functional $\bar{F}(q(\boldsymbol{x}_s))$ is defined as

$$\begin{aligned}
&\bar{\mathcal{F}}(q(\boldsymbol{x}_s)) = \bar{\mathcal{F}}(q(\boldsymbol{x}_{s\backslash h}), q(\boldsymbol{x}_h)) \\
\triangleq &\mathcal{F}(q(\boldsymbol{x}_s)) + \sum_{h=1}^{H} \kappa_h \left(\int q(\boldsymbol{x}_h)d\boldsymbol{x}_h - 1 \right) \\
= &- \int q(\boldsymbol{x}_s) \ln p(\boldsymbol{x}_s, \boldsymbol{u}|\boldsymbol{\psi}_s, \epsilon)d\boldsymbol{x}_s \\
&+ \int q(\boldsymbol{x}_s) \ln q(\boldsymbol{x}_s)d\boldsymbol{x}_s \\
&+ \sum_{h=1}^{H} \kappa_h \left(\int q(\boldsymbol{x}_h)d\boldsymbol{x}_h - 1 \right) \\
= &- \int q(\boldsymbol{x}_{s\backslash h})q(\boldsymbol{x}_h) \ln p(\boldsymbol{x}_s, \boldsymbol{u}|\boldsymbol{\psi}_s, \epsilon)d\boldsymbol{x}_{s\backslash h}d\boldsymbol{x}_h \\
&+ \int q(\boldsymbol{x}_{s\backslash h})q(\boldsymbol{x}_h) \ln q(\boldsymbol{x}_{s\backslash h})q(\boldsymbol{x}_h)d\boldsymbol{x}_{s\backslash h}d\boldsymbol{x}_h \\
&+ \sum_{i\in\mathbb{H}\backslash\{h\}} \kappa_i \left(\int q(\boldsymbol{x}_i)d\boldsymbol{x}_i - 1 \right) \\
&+ \kappa_h \left(\int q(\boldsymbol{x}_h)d\boldsymbol{x}_h - 1 \right),
\end{aligned} \tag{A8}$$

where κ_h is a Larangian multiplier. We use $\boldsymbol{x}_{s\backslash h}$ to denote the set defined by the subtraction of two sets $\boldsymbol{x}_s - \{\boldsymbol{x}_h\}$, the notation $\mathbb{H}$ for an index set $\{1, 2, 3, \ldots, H\}$, and the notation $\mathbb{H}\backslash\{h\}$ for the subtraction of two sets $\mathbb{H} - \{h\}$. The variation of Equation (A8) with respect to $q(\boldsymbol{x}_h)$ is

$$\begin{aligned}
&\frac{\delta\bar{\mathcal{F}}(q(\boldsymbol{x}_s))}{\delta q(\boldsymbol{x}_h)} = \frac{\delta\bar{\mathcal{F}}(q(\boldsymbol{x}_{s\backslash h}), q(\boldsymbol{x}_h))}{\delta q(\boldsymbol{x}_h)} \\
= &- \int q(\boldsymbol{x}_{s\backslash h}) \ln p(\boldsymbol{x}_s, \boldsymbol{u}|\boldsymbol{\psi}_s, \epsilon)d\boldsymbol{x}_{s\backslash h} \\
&+ \int q(\boldsymbol{x}_{s\backslash h}) \ln q(\boldsymbol{x}_{s\backslash h})d\boldsymbol{x}_{s\backslash h} \\
&+ \ln q(\boldsymbol{x}_h) + \kappa_h + 1 \\
= &0.
\end{aligned} \tag{A9}$$

The optimal variational approximation posterior $q(\boldsymbol{x}_h)$ is of the form of Boltzmann distribution

$$\begin{aligned}
q(\boldsymbol{x}_h) = &\frac{1}{\mathcal{Z}_h} \exp(V_h(\boldsymbol{x}_h)) \\
V_h(\boldsymbol{x}_h) = &\int q(\boldsymbol{x}_{s\backslash h}) \ln p(\boldsymbol{x}_s, \boldsymbol{u}|\boldsymbol{\psi}_s, \epsilon)d\boldsymbol{x}_{s\backslash h} \\
H_e(\boldsymbol{x}_{s\backslash h}) = &- \int q(\boldsymbol{x}_{s\backslash h}) \ln q(\boldsymbol{x}_{s\backslash h})d\boldsymbol{x}_{s\backslash h} \\
\mathcal{Z}_h = &\exp(-H_e(\boldsymbol{x}_{s\backslash h}) + \kappa_h + 1),
\end{aligned} \tag{A10}$$

where $V_h(\boldsymbol{x}_h) = \int q(\boldsymbol{x}_{s\backslash h}) \ln p(\boldsymbol{x}_s, \boldsymbol{u}|\boldsymbol{\psi}_s, \epsilon)d\boldsymbol{x}_{s\backslash h}$ corresponds to negative internal energy over the hidden variable $\boldsymbol{x}_h$. The quantity $V_h(\boldsymbol{x}_h)$ is often called variational energy.

Appendix C. Probabilistic Representation of Parameters

In a Bayesian model, an unknown parameter can be treated as a random variable. Probability models could be employed to determine the parameters. Put simply, the probability density function of each random parameter is modeled by a delta-function at each time, and their values follow various multivariate Gaussian distributions [22,23]. In addition, different parameters may have different constraints, therefore we introduce parameterizations to represent these constrained parameters [46].

The coupling mapping $\boldsymbol{F}_2$ contains bias $\boldsymbol{b}_2$ and coupling strength $\boldsymbol{w}_2$ as parameters (Equation (9)). We make an assumption on $\boldsymbol{b}_2$ that it is a multivariate Gaussian distribution with the mean $\boldsymbol{\mu}_{\boldsymbol{b}_2}$ and the variance $\boldsymbol{C}_{\boldsymbol{b}_2}$

$$q(\boldsymbol{b}_2) = \mathcal{N}(\boldsymbol{b}_2; \boldsymbol{\mu}_{\boldsymbol{b}_2}, \boldsymbol{C}_{\boldsymbol{b}_2}). \tag{A11}$$

In principle, there should not be any constraints on the coupling strength $\boldsymbol{w}_2$. However, there is no reason to choose $\boldsymbol{w}_2$ to be a negative element, since the negativity in $\boldsymbol{w}_2$ could be counterbalanced by the negativity in $\boldsymbol{x}_2$. Therefore, the lower bound on each component $\boldsymbol{w}_2^{(i)}$ is chosen to be 0. In addition, considering the fact that $\boldsymbol{w}_2$ is involved in the update of the positive definite precision matrix $\boldsymbol{P}_2$ (Equation (29)), each component $\boldsymbol{w}_2^{(i)} (i = 1, 2, \ldots, d_2)$ should have a upper bound. If the value of $\boldsymbol{w}_2^{(i)}$ is too large, $\boldsymbol{P}_2$ would be degenerated. To avoid such violations, we set the upper bound of $\boldsymbol{w}_2^{(i)}$ to be a constant value $\boldsymbol{\alpha}_{\boldsymbol{w}_2}^{(i)} > 0$, i.e., the i-th component of a constant column vector $\boldsymbol{\alpha}_{\boldsymbol{w}_2}$. We use a sigmoid function to map a multivariate Gaussian variable into a bounded variable $\boldsymbol{w}_2$. This transformation and the priors on $\boldsymbol{w}_2$ are given as

$$\begin{aligned}
&\boldsymbol{w}_2^{(i)} = \boldsymbol{W}_2^{(i,i)} = \frac{\boldsymbol{\alpha}_{\boldsymbol{w}_2}^{(i)}}{1 + \exp(-\boldsymbol{w}_2^{(i)G})}, \forall i \in \{1, 2, \cdots, d_2\} \\
&\boldsymbol{w}_2 = \boldsymbol{\alpha}_{\boldsymbol{w}_2} \odot \boldsymbol{s}(\boldsymbol{w}_2^G, \boldsymbol{1}) \\
&q(\boldsymbol{w}_2) = q(\boldsymbol{w}_2^G) = \mathcal{N}(\boldsymbol{w}_2^G; \boldsymbol{\mu}_{\boldsymbol{w}_2^G}, \boldsymbol{C}_{\boldsymbol{w}_2^G}).
\end{aligned} \tag{A12}$$

The parameter $\boldsymbol{\lambda}_{top}$ naturally has a lower bound $\boldsymbol{0}$ constrained by variances, but if $\boldsymbol{\lambda}_{top}$ is not bounded from above, it may cause some violations: for a large $\boldsymbol{\lambda}_{top}$, it yields small prediction precision $\hat{\boldsymbol{\Pi}}_2$ where all variances are close to 0 and causes the posterior precision $\boldsymbol{P}_2$ not to be a positive definite matrix. That is to say, an unbounded vector $\boldsymbol{\lambda}_{top}$ violates the conditions of the update equations, yielding an improbable perceptual inference. Therefore, we set an upper bound $\boldsymbol{\alpha}_{\boldsymbol{\lambda}_{top}}$ on $\boldsymbol{\lambda}_{top}$, through a bounded sigmoid function similar as in Equation (A12)

$$\begin{aligned}
&\boldsymbol{\lambda}_{top}^{(i)} = \frac{\boldsymbol{\alpha}_{\boldsymbol{\lambda}_{top}}^{(i)}}{1 + \exp(-\boldsymbol{\lambda}_{top}^{(i)G})}, \forall i \in \{1, 2, \cdots, d_2\} \\
&\boldsymbol{\lambda}_{top} = \boldsymbol{\alpha}_{\boldsymbol{\lambda}_{top}} \odot \boldsymbol{s}(\boldsymbol{\lambda}_{top}^G, \boldsymbol{1}) \\
&q(\boldsymbol{\lambda}_{top}) = q(\boldsymbol{\lambda}_{top}^G) = \mathcal{N}(\boldsymbol{\lambda}_{top}^G; \boldsymbol{\mu}_{\boldsymbol{\lambda}_{top}^G}, \boldsymbol{C}_{\boldsymbol{\lambda}_{top}^G}).
\end{aligned} \tag{A13}$$

In the hierarchical model, we have introduced sensory noise parameters $\boldsymbol{\zeta}_1$, with all positive components. We represent these parameters in logarithmic space to preserve nonnegativity. More specifically, $\boldsymbol{\zeta}_1$ is expressed in its log-space by a Gaussian random vector $\boldsymbol{\zeta}_1^G$

$$\begin{aligned}
&\zeta_1^{(i)} = \exp(\zeta_1^{(i)G}), \zeta_1^{(i)G} \in \mathbb{R} \\
&\boldsymbol{\zeta}_1 = \mathbf{exp}(\boldsymbol{\zeta}_1^G) \\
&q(\boldsymbol{\zeta}_1) = q(\boldsymbol{\zeta}_1^G) = \mathcal{N}(\boldsymbol{\zeta}_1^G; \boldsymbol{\mu}_{\boldsymbol{\zeta}_1^G}, \boldsymbol{C}_{\boldsymbol{\zeta}_1^G}).
\end{aligned} \tag{A14}$$

Here, we employ an element-wise exponential function $\mathbf{exp}(\cdot)$ to map a multivariate Gaussian random variable $\boldsymbol{\zeta}_1^G$ into $\boldsymbol{\zeta}_1$.

Aside from these structural parameters, the initial priors on all hidden states are also determined following similar way. In details, we use a Gaussian random variable to express the initial mean $\boldsymbol{\mu}_h(t_0)$

$$\begin{aligned}
&q(\boldsymbol{\mu}_h(t_0)) = \mathcal{N}(\boldsymbol{\mu}_h(t_0); \boldsymbol{\mu}_{\boldsymbol{\mu}_h(t_0)}, \boldsymbol{C}_{\boldsymbol{\mu}_h(t_0)}) \\
&\forall h \in \{1, 2\}.
\end{aligned} \tag{A15}$$

Each of the initial prior covariances $\{\boldsymbol{C}_h(t_0)|h = 1, 2\}$ is restricted to a principal diagonal and positive definite matrix. All principal diagonal elements in $\boldsymbol{C}_h(t_0)$ form a column vector $\boldsymbol{c_h}$. Since the components in $\boldsymbol{c_h}$ are positive, they are represented by multivariate Gaussian random variables in log-space

$$\begin{aligned}
&\boldsymbol{c}_h = \mathbf{exp}(\boldsymbol{c}_h^G) \\
&q(\boldsymbol{C}_h) = q(\boldsymbol{c}_h^G) = \mathcal{N}(\boldsymbol{c}_h^G; \boldsymbol{\mu}_{\boldsymbol{c}_h^G}, \boldsymbol{C}_{\boldsymbol{c}_h^G}) \\
&\forall h \in \{1, 2\}.
\end{aligned} \tag{A16}$$

For the response model expressed by Equations (34)–(37), there is only one inverse temperature parameter ζ_a, which is also restricted to be positive. We can use the same representation method as $\boldsymbol{\zeta}_1$ to express ζ_a.

$$\begin{aligned}
&\zeta_a = \exp(\zeta_a^G) \\
&q(\zeta_a) = q(\zeta_a^G) = \mathcal{N}(\zeta_a^G; \mu_{\zeta_a^G}, C_{\zeta_a^G})
\end{aligned} \tag{A17}$$

where $\mu_{\zeta_a^G}, C_{\zeta_a^G}$ are the mean and variance of a Gaussian random variable ζ_a^G respectively.

Appendix D. Variational Bayesian Learning

A Bayesian agent receives and encodes sensory input $\boldsymbol{u}(t)$, and then makes a perceptual decision (i.e., action) $a(t) \in \mathcal{A}$ based on random reward $\boldsymbol{r}(t)$ and perceptual evidence. These two successive processes correspond to the two main functional models of an agent: a perceptual model to encode sensory inputs and a response model to make perceptual decisions [20,21,23]. Here, we employ a GHBF as the perceptual model $\mathcal{M}_p$ with perceptual parameter vector $\boldsymbol{\psi}$ and a simple response model defined by Equations (34)–(37) as the response model π_r with the response parameter vector $\boldsymbol{\psi}_r$. The combined model is denoted by $\mathcal{M} = (\mathcal{M}_p, \pi_r)$. All its parameters are denoted by $\boldsymbol{\xi}$.

We introduce the following mean field approximation to fit the parameters of the combined model with the sensory inputs $\boldsymbol{u}_s = \{\boldsymbol{u}(t_1), \boldsymbol{u}(t_2), \cdots, \boldsymbol{u}(t_K)\}$, actions $a_s = \{a(t_1), a(t_2), \cdots, a(t_K)\}$ and random rewards $\boldsymbol{r}_s = \{\boldsymbol{r}(t_1), \boldsymbol{r}(t_2), \cdots, \boldsymbol{r}(t_K)\}$

$$\begin{aligned}
&q(\boldsymbol{\xi}) \approx q(\boldsymbol{\psi}) q(\boldsymbol{\psi}_r) \\
=&q(\boldsymbol{\lambda}_{top}) q(\boldsymbol{\sigma}_{\boldsymbol{u}}) \left(\prod_{h=2}^{H} q(\boldsymbol{w}_h) p(\boldsymbol{b}_h) \right) \left(\prod_{h=1}^{H} q(\boldsymbol{\mu}_h(t_0)) q(\boldsymbol{C}_h(t_0)) \right).
\end{aligned} \tag{A18}$$

Then

$$\begin{aligned}
&\ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s|\mathcal{M}) = \ln \int p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \boldsymbol{\xi}|\mathcal{M}) d\boldsymbol{\xi} \\
&= \ln \int \frac{p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \boldsymbol{\xi}|\mathcal{M}) q(\boldsymbol{\xi})}{q(\boldsymbol{\xi})} d\boldsymbol{\xi} \\
&\geq \int q(\boldsymbol{\xi}) \ln(\frac{p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \boldsymbol{\xi}|\mathcal{M})}{q(\boldsymbol{\xi})}) d\boldsymbol{\xi} \\
&= \int q(\boldsymbol{\xi}) \ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \boldsymbol{\xi}|\mathcal{M}) - q(\boldsymbol{\xi}) \ln q(\boldsymbol{\xi}) d\boldsymbol{\xi} \\
&\triangleq \mathcal{F}(q(\boldsymbol{\xi}))
\end{aligned} \tag{A19}$$

We use the Lagrange multiplier method to work out the optimal variational posterior as below

$$\begin{aligned}
q(\boldsymbol{\xi}) &= \frac{1}{\mathcal{Z}_{\boldsymbol{\xi}}} \exp(V(\boldsymbol{\xi})) \\
\mathcal{V}(\boldsymbol{\xi}) &= \ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \boldsymbol{\xi}|\mathcal{M}).
\end{aligned} \tag{A20}$$

Then we execute Laplace's approximation to determine a Gaussian approximation of the variational posterior solution (Equation (A21))

$$\begin{aligned}
\boldsymbol{\xi}^* &= \arg\max_{\boldsymbol{\xi}} \mathcal{V}(\boldsymbol{\xi}) = \arg\max_{\boldsymbol{\xi}} \ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \boldsymbol{\xi}|\mathcal{M}) \\
&= \arg\max_{\boldsymbol{\xi}} \ln p(\boldsymbol{\xi}, a_s|\boldsymbol{u}_s, \boldsymbol{r}_s, \mathcal{M}) p(\boldsymbol{u_s}) \\
&= \arg\max_{\boldsymbol{\xi}} \ln p(\boldsymbol{\xi}, a_s|\boldsymbol{u}_s, \boldsymbol{r}_s, \mathcal{M}) \\
&= \arg\max_{\boldsymbol{\xi}} \ln p(a_s|\boldsymbol{\xi}, \boldsymbol{u}_s, \boldsymbol{r}_s, \mathcal{M}) + \ln p(\boldsymbol{\xi}) \\
&= \arg\max_{\boldsymbol{\xi}} \sum_{k=1}^{K} \ln p(a(t_k)|\boldsymbol{u}(t_k), \boldsymbol{r}(t_k), \boldsymbol{\xi}, \mathcal{M}) + \ln p(\boldsymbol{\xi}) \\
&= \arg\max_{\boldsymbol{\xi}} \sum_{k=1}^{K} \ln p\big(a(t_k)|\boldsymbol{r}(t_k), \boldsymbol{\chi}_s(t_k) = \mathcal{M}_p(\boldsymbol{u}(t_k), \boldsymbol{\psi}), \boldsymbol{\psi}_r\big) \\
&\qquad + \ln p(\boldsymbol{\psi}) \\
\boldsymbol{\mu}_{\boldsymbol{\xi}}^* &= \boldsymbol{\xi}^* \\
\boldsymbol{C}_{\boldsymbol{\xi}}^* &= -\frac{\partial^2 \mathcal{V}(\boldsymbol{\xi}^*)}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^T},
\end{aligned} \tag{A21}$$

where $p\big(a(t_k)|\boldsymbol{\chi}_s(t_k) = \mathcal{M}_p(\boldsymbol{u}(t_k), \boldsymbol{\psi}), \boldsymbol{\psi}_r\big)$ is given by a particular response model. $\boldsymbol{\chi}_s(t_k)$ is the set of sufficient statistics of posterior hidden states in our hierarchical Bayesian perceptual model at time t_k

$$\boldsymbol{\chi}_s(t_k) = \{\boldsymbol{\mu}_1(t_k), \boldsymbol{C}_1(t_k), \boldsymbol{\mu}_2(t_k), \boldsymbol{C}_2(t_k)\}.$$

Finally, the maximum value $\mathcal{F}_{\boldsymbol{\xi}}^*$ of the negative free energy $\mathcal{F}_{\boldsymbol{\xi}}$ is given by

$$\mathcal{F}_{\boldsymbol{\xi}} \leq \mathcal{F}_{\boldsymbol{\xi}}^* = \mathcal{V}(\boldsymbol{\mu}_{\boldsymbol{\xi}}^*) + \frac{d_{\boldsymbol{\xi}}}{2} \ln 2\pi e + \frac{1}{2} \ln \det(\boldsymbol{C}_{\boldsymbol{\xi}}^*). \tag{A22}$$

Appendix E. Evaluating Negative Free Energy

For a Bayesian agent $\mathcal{M}$ with parameters $\boldsymbol{\xi}$, the posterior $p(\boldsymbol{\xi}|\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \mathcal{M})$ on parameters $\boldsymbol{\xi}$ is approximated by a multivariate Gaussian distribution $q(\boldsymbol{\xi})$ under the *Laplacian approximation*

$$p(\boldsymbol{\xi}|\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \mathcal{M}) \approx q(\boldsymbol{\xi}) = \mathcal{N}(\boldsymbol{\xi}; \boldsymbol{\mu}_{\boldsymbol{\xi}}, \boldsymbol{C}_{\boldsymbol{\xi}}),$$

where $\boldsymbol{C}_{\boldsymbol{\xi}}$ is a covariance matrix. The mean $\boldsymbol{\mu}_{\boldsymbol{\xi}}$ is determined by maximizing the quantity $p(\boldsymbol{\xi}|\boldsymbol{u}_s, a_s, \mathcal{M})$

$$\begin{aligned} \boldsymbol{\mu}^*_{\boldsymbol{\xi}} &= \arg\max_{\boldsymbol{\xi}} p(\boldsymbol{\xi}|\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \mathcal{M}) \\ &= \arg\max_{\boldsymbol{\xi}} \frac{p(\boldsymbol{\xi}, \boldsymbol{u}_s, \boldsymbol{r}_s, \boldsymbol{r}_s, a_s|\mathcal{M})}{p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s|\mathcal{M})} \\ &= \arg\max_{\boldsymbol{\xi}} p(\boldsymbol{\xi}, \boldsymbol{u}_s, \boldsymbol{r}_s, a_s|\mathcal{M}). \end{aligned} \tag{A23}$$

The optimal $q(\boldsymbol{\xi})$ is determined by maximizing the negative free energy $\mathcal{F}_{\boldsymbol{\xi}}$

$$\begin{aligned} &\max_{\boldsymbol{\xi}} \ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s|\boldsymbol{\xi}, \mathcal{M}) \\ \geq &\max_{q(\boldsymbol{\xi})} \mathcal{F}_{\boldsymbol{\xi}} = \max_{q(\boldsymbol{\xi})} \int q(\boldsymbol{\xi}) \ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \boldsymbol{\xi}|\mathcal{M}) - q(\boldsymbol{\xi}) \ln q(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \end{aligned} \tag{A24}$$

We use the notation $\mathcal{V}(\boldsymbol{\xi})$ to denote the quantity $\ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s, \boldsymbol{\xi}|\mathcal{M})$ and then use Taylor's theorem to expand $\mathcal{V}(\boldsymbol{\xi})$ at the point $\boldsymbol{\mu}^*_{\boldsymbol{\xi}}$

$$\begin{aligned} \mathcal{V}(\boldsymbol{\xi}) \approx &\mathcal{V}(\boldsymbol{\mu}^*_{\boldsymbol{\xi}}) \\ &+ \frac{\partial \mathcal{V}(\boldsymbol{\mu}^*_{\boldsymbol{\xi}})}{\partial \boldsymbol{\xi}} (\boldsymbol{\xi} - \boldsymbol{\mu}^*_{\boldsymbol{\xi}}) \\ &+ \frac{1}{2} (\boldsymbol{\xi} - \boldsymbol{\mu}^*_{\boldsymbol{\xi}})^T \frac{\partial^2 \mathcal{V}(\boldsymbol{\mu}^*_{\boldsymbol{\xi}})}{\partial^2 \boldsymbol{\xi}} (\boldsymbol{\xi} - \boldsymbol{\mu}^*_{\boldsymbol{\xi}}). \end{aligned} \tag{A25}$$

The first term $\int q(\boldsymbol{\xi}) \mathcal{V}(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}$ in the negative free energy $\mathcal{F}_{\boldsymbol{\xi}}$ is evaluated by

$$\begin{aligned} &\int q(\boldsymbol{\xi}) \mathcal{V}(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \\ \approx &\mathcal{V}(\boldsymbol{\mu}^*_{\boldsymbol{\xi}}) + \frac{\partial \mathcal{V}(\boldsymbol{\mu}^*_{\boldsymbol{\xi}})}{\partial \boldsymbol{\xi}} E_{q(\boldsymbol{\xi}|\boldsymbol{\mu}^*_{\boldsymbol{\xi}}, \boldsymbol{C}_{\boldsymbol{\xi}})} [\boldsymbol{\xi} - \boldsymbol{\mu}^*_{\boldsymbol{\xi}}] \\ &+ \frac{1}{2} E_{q(\boldsymbol{\xi}|\boldsymbol{\mu}^*_{\boldsymbol{\xi}}, \boldsymbol{C}_{\boldsymbol{\xi}})} [(\boldsymbol{\xi} - \boldsymbol{\mu}^*_{\boldsymbol{\xi}})^T \frac{\partial^2 \mathcal{V}(\boldsymbol{\mu}^*_{\boldsymbol{\xi}})}{\partial^2 \boldsymbol{\xi}} (\boldsymbol{\xi} - \boldsymbol{\mu}^*_{\boldsymbol{\xi}})] \\ = &\mathcal{V}(\boldsymbol{\mu}^*_{\boldsymbol{\xi}}) + \frac{1}{2} \mathrm{tr} \left(\boldsymbol{C}_{\boldsymbol{\xi}} \frac{\partial^2 \mathcal{V}(\boldsymbol{\mu}^*_{\boldsymbol{\xi}})}{\partial^2 \boldsymbol{\xi}} \right) \end{aligned} \tag{A26}$$

The last term $H_e(\boldsymbol{\xi}) = -\int q(\boldsymbol{\xi}) \ln q(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}$ is given by

$$\begin{aligned} H_e(\boldsymbol{\xi}) &= -\int q(\boldsymbol{\xi}) \ln q(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \\ &= -E_{q(\boldsymbol{\xi}|\boldsymbol{\mu}_{\boldsymbol{\xi}}^*, \boldsymbol{C}_{\boldsymbol{\xi}})}\left[\ln q(\boldsymbol{\xi}|\boldsymbol{\mu}_{\boldsymbol{\xi}}^*, \boldsymbol{C}_{\boldsymbol{\xi}})\right] \\ &= -E_{q(\boldsymbol{\xi}|\boldsymbol{\mu}_{\boldsymbol{\xi}}^*, \boldsymbol{C}_{\boldsymbol{\xi}})}[\\ &\quad -\frac{d_{\boldsymbol{\xi}}}{2}\ln 2\pi - \frac{1}{2}\ln\det(\boldsymbol{C}_{\boldsymbol{\xi}}) \\ &\quad -\frac{1}{2}(\boldsymbol{\xi}-\boldsymbol{\mu}_{\boldsymbol{\xi}}^*)^T \boldsymbol{C}_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\xi}-\boldsymbol{\mu}_{\boldsymbol{\xi}}^*) \\ &\quad] \\ &= \frac{d_{\boldsymbol{\xi}}}{2}\ln 2\pi + \frac{1}{2}\ln\det(\boldsymbol{C}_{\boldsymbol{\xi}}) + \frac{1}{2}\mathrm{tr}(I_{d_{\boldsymbol{\xi}}}) \\ &= \frac{d_{\boldsymbol{\xi}}}{2}\ln 2\pi e + \frac{1}{2}\ln\det(\boldsymbol{C}_{\boldsymbol{\xi}}) \end{aligned} \tag{A27}$$

Therefore, the negative free energy $\mathcal{F}_{\boldsymbol{\xi}}$ is calculated as

$$\begin{aligned} \mathcal{F}_{\boldsymbol{\xi}} &= E_{q(\boldsymbol{\xi})}[\mathcal{V}(\boldsymbol{\mu}_{\boldsymbol{\xi}}^*)] + H_e(\boldsymbol{\xi}) \\ &= \mathcal{V}(\boldsymbol{\mu}_{\boldsymbol{\xi}}^*) + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{C}_{\boldsymbol{\xi}}\frac{\partial^2\mathcal{V}(\boldsymbol{\mu}_{\boldsymbol{\xi}}^*)}{\partial^2\boldsymbol{\xi}}\right) \\ &\quad + \frac{d_{\boldsymbol{\xi}}}{2}\ln 2\pi e + \frac{1}{2}\ln\det(\boldsymbol{C}_{\boldsymbol{\xi}}) \end{aligned} \tag{A28}$$

$\mathcal{F}_{\boldsymbol{\xi}}$ is a scalar function of the covariance $\boldsymbol{C}_{\boldsymbol{\xi}}$. The optimal point or a stationary point $\boldsymbol{C}_{\boldsymbol{\xi}}^*$ is found where $\mathcal{F}_{\boldsymbol{\xi}}$ reaches the maximum. This is done by making the partial derivative $\frac{\partial\mathcal{F}_{\boldsymbol{\xi}}}{\partial\boldsymbol{C}_{\boldsymbol{\xi}}}$ to be a zero matrix $\boldsymbol{O}$.

$$\begin{aligned} \frac{\partial\mathcal{F}_{\boldsymbol{\xi}}}{\partial\boldsymbol{C}_{\boldsymbol{\xi}}} &= \frac{1}{2}\frac{\partial^2\mathcal{V}(\boldsymbol{\mu}_{\boldsymbol{\xi}}^*)}{\partial^2\boldsymbol{\xi}} + \frac{1}{2}\boldsymbol{C}_{\boldsymbol{\xi}}^{-1} = \boldsymbol{O} \\ \Longrightarrow \boldsymbol{C}_{\boldsymbol{\xi}}^* &= -\left(\frac{\partial^2\mathcal{V}(\boldsymbol{\mu}_{\boldsymbol{\xi}}^*)}{\partial^2\boldsymbol{\xi}}\right)^{-1} \end{aligned} \tag{A29}$$

At the optimal point $\boldsymbol{C}_{\boldsymbol{\xi}}^*$, the maximal value of $\mathcal{F}_{\boldsymbol{\xi}}$ is

$$\begin{aligned} \mathcal{F}_{\boldsymbol{\xi}}^* &= \mathcal{V}(\boldsymbol{\mu}_{\boldsymbol{\xi}}^*) + \frac{d_{\boldsymbol{\xi}}}{2}\ln 2\pi e + \frac{1}{2}\ln\det(\boldsymbol{C}_{\boldsymbol{\xi}}^*) \\ &\approx \max_{\boldsymbol{\xi}} \ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s|\boldsymbol{\xi}, \mathcal{M}) \\ &= \ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s|\boldsymbol{\mu}_{\boldsymbol{\xi}}^*, \mathcal{M}). \end{aligned} \tag{A30}$$

Appendix F. Bayesian Model Selection

Grounded on probability theory, Bayesian model selection is to evaluate different models based on the observed data, favoring the model with balanced tradeoff between complexity and flexibility. Given a series of sensory inputs $\boldsymbol{u}_s = \{\boldsymbol{u}(t_1), \boldsymbol{u}(t_2), \cdots, \boldsymbol{u}(t_K)\}$, a series of actions $a_s = \{a(t_1), a(t_2), \cdots, a(t_K)\}$ and a series of random rewards $\boldsymbol{r}_s = \{\boldsymbol{r}(t_1), \boldsymbol{r}(t_2), \cdots, \boldsymbol{r}(t_K)\}$, Bayesian model selection is to select the optimal agent $\mathcal{M}^*$ to best interpret sensory inputs and actions

$$\mathcal{M}^* = \arg\max_{M} p(\mathcal{M}|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s). \tag{A31}$$

Taking two different agents $\mathcal{M}_2, \mathcal{M}_1$ into account, we can define Bayesian Factor as

$$\begin{aligned} p(\mathcal{M}_2|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) &= \frac{p(\mathcal{M}_2)p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_2)}{p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)} \\ p(\mathcal{M}_1|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s) &= \frac{p(\mathcal{M}_1)p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_1)}{p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)} \\ \frac{p(\mathcal{M}_1|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)}{p(\mathcal{M}_2|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)} &= BF\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)} \\ BF &= \frac{p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_1)}{p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_2)}, \end{aligned} \tag{A32}$$

where $p(\mathcal{M}_i)$ is the prior distribution of $\mathcal{M}_i$. Here, we make a general assumption that the prior distribution of an agent is a non-informative prior. Under the assumption of non-informative priors, the prior distribution is equivalent to a uniform distribution $\frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)} = 1$. Then the ratio of the posterior distributions $\frac{p(\mathcal{M}_1|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)}{p(\mathcal{M}_2|\boldsymbol{u}_s, a_s, \boldsymbol{r}_s)}$ is simply given by the Bayesian Factor.

Bayesian model selection problem is reduced to selecting an agent with maximal model evidence $p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_i)$. In the Bayesian learning framework, log-model evidence $\ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M})$ can be approximated by the optimal negative free energy

$$\mathcal{F}^*_{\boldsymbol{\xi}} \approx \ln p(\boldsymbol{u}_s, \boldsymbol{r}_s, a_s|\boldsymbol{\mu}^*_{\boldsymbol{\xi}}, \mathcal{M})$$

defined in Equation (A30). By computing the negative free energies of two different agents $\mathcal{F}^*_{\boldsymbol{\xi}_1}, \mathcal{F}^*_{\boldsymbol{\xi}_2}$, Bayesian Factor is given by

$$\begin{aligned} BF =& \frac{p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_1)}{p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_2)} \\ =& \exp(\ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_1) - \ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_2)) \\ \approx& \exp(\mathcal{F}^*_{\boldsymbol{\xi}_1} - \mathcal{F}^*_{\boldsymbol{\xi}_2}). \end{aligned} \tag{A33}$$

For the ease of using Bayesian Factor, Harold Jeffreys gave a scale for the interpretation of Bayesian Factor (Table A1) [66]. If $BF > 1$, the agent $\mathcal{M}_1$ is more strongly supported by the observed date, and vice versa (if $BF < 1$, the agent $\mathcal{M}_2$ is more strongly supported).

Table A1. Bayes Factors and interpretations.

Bayesian Factor BF	Interpretations
$0 < BF < \frac{1}{100}$	Decisive evidence for $\mathcal{M}_2$
$\frac{1}{100} < BF < \frac{1}{10}$	Strong evidence for $\mathcal{M}_2$
$\frac{1}{10} < BF < \frac{1}{3}$	Moderate evidence for $\mathcal{M}_2$
$\frac{1}{3} < BF < 1$	Weak evidence for $\mathcal{M}_2$
$1 < BF < 3$	Weak evidence for $\mathcal{M}_1$
$3 < BF < 10$	Moderate evidence for $\mathcal{M}_1$
$10 < BF < 100$	Strong evidence for $\mathcal{M}_1$
$BF > 100$	Decisive evidence for $\mathcal{M}_1$

According to the Bayesian Information Criterion (BIC) [67], log-model evidence $\ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s, |\mathcal{M}_i)$ can be approximated by

$$\begin{aligned} \ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_i) &\approx \ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\boldsymbol{\mu}^*_{\boldsymbol{\xi}_i}, \mathcal{M}_i) - \frac{d_{\boldsymbol{\xi}_i}}{2}\ln(K) \\ \Rightarrow \ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_i) &= \mathcal{F}^*_{\boldsymbol{\xi}_i} - \frac{d_{\boldsymbol{\xi}_i}}{2}\ln(K), \end{aligned} \tag{A34}$$

where K is the number of sensory inputs in $\boldsymbol{u}_s$. d_{ξ_i} is the number of free parameters estimated by the model. Therefore, Bayesian Factor is modified by

$$\begin{aligned} BF_{BIC} =& \frac{p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_1)}{p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_2)} \\ =& \exp(\ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_1) - \ln p(\boldsymbol{u}_s, a_s, \boldsymbol{r}_s|\mathcal{M}_2)) \\ \approx& \exp\left(\mathcal{F}^*_{\xi_1} - \mathcal{F}^*_{\xi_2} - \frac{d_{\xi_1} - d_{\xi_2}}{2}\ln(K)\right). \end{aligned} \tag{A35}$$

Appendix G. Rescorla–Wagner Model

The Rescorla–Wagner (RW) model is a basic model in reinforcement learning (RL) field and cognitive neuroscience field [26,51]. As a baseline model for comparison, we construct a two dimensional RW model to capture the dynamic expectation $\boldsymbol{\mu}_0 = E[\boldsymbol{x}_0]$ of the two armed bandits in the above gambling task

$$\begin{aligned} &\mu_0^{(i)}(t_k) = \mu_0^{(i)}(t_{k-1}) + \alpha\Delta\mu_0^{(i)}(t_k) \\ &\Delta\mu_0^{(i)}(t_k) = u^{(i)}(t_k) - \mu_0^{(i)}(t_k) \\ &\forall t_k, \mu_0^{(i)}(t_k) \in [0,1] \\ &i = 1, 2, \end{aligned} \tag{A36}$$

where $\alpha \in (0,1)$ is a positive learning rate. To yield a prediction $\hat{\boldsymbol{\mu}}_0(t_k)$ on $\boldsymbol{x}_0(t_k)$ before receiving the actual sensory input $\boldsymbol{u}(t_k)$ at time t_k, the RW model uses its most recent state i.e., $\boldsymbol{\mu}_0(t_{k-1})$ as the prediction

$$\hat{\boldsymbol{\mu}}_0(t_k) := \boldsymbol{\mu}_0(t_{k-1}).$$

To produce an action based on the predicted state $\hat{\boldsymbol{\mu}}_0 = [\hat{\mu}_0^{(1)}, \hat{\mu}_0^{(2)}]^T$, the RW model needs a response model to work with. We use the same response model based on Bayesian decision theory (Section 4) for fair comparison.

To perform the variational Bayesian learning scheme (cf. Appendix D), we assume that all parameters of the RW model are random variables. Following similar treatmeants as in Appendix C, the initial prior state $\boldsymbol{\mu}_0(0) = [\mu_0^{(1)}(0), \mu_0^{(2)}(0)]^T$ is represented in logit-space $\boldsymbol{\mu}_{\boldsymbol{\mu}_0^G(0)}$

$$\boldsymbol{\mu}_0(0) = \boldsymbol{s}(\boldsymbol{\mu}_0^G(0), \mathbf{1}), \tag{A37}$$

where $\boldsymbol{\mu}_{\boldsymbol{\mu}_0^G(0)}$ is a two-dimensional Gaussian distribution with mean $\boldsymbol{\mu}_{\boldsymbol{\mu}_0^G(0)}$

$$p(\boldsymbol{\mu}_0^G(0)) = \mathcal{N}(\boldsymbol{\mu}_0^G(0); \boldsymbol{\mu}_{\boldsymbol{\mu}_0^G(0)}, \boldsymbol{C}_{\boldsymbol{\mu}_0^G(0)}). \tag{A38}$$

Since the learning rate α is a value between 0 and 1, it is represented by a random variable α^G in the logit-space. We further assume that α^G is a Gaussian random variable with mean μ_{α^G} and variance C_{α^G}

$$\alpha = s(\alpha^G, 1) \tag{A39}$$

$$p(\alpha^G) = \mathcal{N}(\alpha^G; \mu_{\alpha^G}, C_{\alpha^G}). \tag{A40}$$

In this paper, all parameter configurations for the RW model are listed in Table A2.

Table A2. Parameters of the Rescorla–Wagner model. Parameters labeled by 'Free' are optimized by the inversion of the model. Fixed parameters are constant and not optimized. The notation **0** is a zero vector. Given all initial priors, we search the optimal priors on the free parameters $\boldsymbol{\mu}_{\tilde{\xi}}$ according to the free energy principle (Equations (A19) and (A21)).

Name	Description	Initial Value	Fixed or Free
	Parameters of Rescorla–Wagner model		
$d_{\boldsymbol{u}}$	Dimension of $\boldsymbol{u}$	2	constant
d_0	Dimension of $\boldsymbol{\mu}_0$	2	constant
$\boldsymbol{\mu}_0(t_0)$	Prior initial state		Fixed
$\boldsymbol{\mu}_{\boldsymbol{\mu}_0^G(t_0)}$	Mean of $\boldsymbol{\mu}_0^G(t_0)$	**0**	
$\boldsymbol{C}_{\boldsymbol{\mu}_0^G(t_0)}$	Covariance of $\boldsymbol{\mu}_0^G(t_0)$	$\boldsymbol{I}_{d_0}$	
α	Learning rate α		Free
μ_{α^G}	Mean of α^G	0	
C_{α^G}	Covariance of α^G	0.01	

References

1. Cisek, P.; Puskas, G.A.; El-Murr, S. Decisions in changing conditions: The urgency-gating model. *J. Neurosci.* **2009**, *29*, 11560–11571. [CrossRef] [PubMed]
2. Weiss, A.; Chambon, V.; Lee, J.K.; Drugowitsch, J.; Wyart, V. Interacting with volatile environments stabilizes hidden-state inference and its brain signatures. *Nat. Commun.* **2021**, *12*, 2228. [CrossRef] [PubMed]
3. Vargas, D.V.; Lauwereyns, J. Setting the space for deliberation in decision-making. *Cogn. Neurodyn.* **2021**, *15*, 743–755. [CrossRef] [PubMed]
4. Knill, D.C.; Richards, W. *Perception as Bayesian Inference*; Cambridge University Press: Cambridge, UK, 1996.
5. Ernst, M.O.; Banks, M.S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **2002**, *415*, 429–433. [CrossRef]
6. Weilnhammer, V.A.; Stuke, H.; Sterzer, P.; Schmack, K. The neural correlates of hierarchical predictions for perceptual decisions. *J. Neurosci.* **2018**, *38*, 5008–5021. [CrossRef] [PubMed]
7. Zhang, W.; Wu, S.; Doiron, B.; Lee, T.S. A Normative Theory for Causal Inference and Bayes Factor Computation in Neural Circuits. In *Advances in Neural Information Processing Systems*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2019; Volume 32.
8. Friston, K.; FitzGerald, T.; Rigoli, F.; Schwartenbeck, P.; Pezzulo, G. Active inference and learning. *Neurosci. Biobehav. Rev.* **2016**, *68*, 862–879. [CrossRef] [PubMed]
9. Shikauchi, Y.; Miyakoshi, M.; Makeig, S.; Iversen, J.R. Bayesian models of human navigation behaviour in an augmented reality audiomaze. *Eur. J. Neurosci.* **2021**, *54*, 8308–8317. [CrossRef]
10. Zhang, J.; Gu, Y.; Chen, A.; Yu, Y. Unveiling Dynamic System Strategies for Multisensory Processing: From Neuronal Fixed-Criterion Integration to Population Bayesian Inference. *Research* **2022**, *2022*, 9787040. [CrossRef]
11. Zhou, L.; Gu, Y. Cortical Mechanisms of Multisensory Linear Self-motion Perception. *Neurosci. Bull.* **2022**, 1–13. [CrossRef]
12. Chikkerur, S.; Serre, T.; Tan, C.; Poggio, T. Attention as a Bayesian inference process. In *Human Vision and Electronic Imaging XVI*; Rogowitz, B.E., Pappas, T.N., Eds.; Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series; SPIE Press: California, USA, 2011; Volume 7865, p. 786511.
13. Vossel, S.; Mathys, C.; Stephan, K.E.; Friston, K.J. Cortical coupling reflects Bayesian belief updating in the deployment of spatial attention. *J. Neurosci.* **2015**, *35*, 11532–11542. [CrossRef]
14. Lawson, R.P.; Mathys, C.; Rees, G. Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* **2017**, *20*, 1293–1299. [CrossRef] [PubMed]
15. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138. [CrossRef] [PubMed]
16. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* **2005**, *360*, 815–836. [CrossRef] [PubMed]
17. Stefanics, G.; Heinzle, J.; Horváth, A.A.; Stephan, K.E. Visual mismatch and predictive coding: A computational single-trial ERP study. *J. Neurosci.* **2018**, *38*, 4020–4030. [CrossRef]
18. Wang, B.A.; Schlaffke, L.; Pleger, B. Modulations of insular projections by prior belief mediate the precision of prediction error during tactile learning. *J. Neurosci.* **2020**, *40*, 3827–3837. [CrossRef]
19. Sun, Y.; Gomez, F.; Schmidhuber, J. Planning to Be Surprised: Optimal Bayesian Exploration in Dynamic Environments. In *Artificial General Intelligence*; Schmidhuber, J., Thórisson, K.R., Looks, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 41–51.
20. Daunizeau, J.; den Ouden, H.E.M.; Pessiglione, M.; Kiebel, S.J.; Stephan, K.E.; Friston, K.J. Observing the Observer (I): Meta-Bayesian Models of Learning and Decision-Making. *PLoS ONE* **2010**, *5*, e15554. [CrossRef]

21. Daunizeau, J.; Den Ouden, H.E.; Pessiglione, M.; Kiebel, S.J.; Friston, K.J.; Stephan, K.E. Observing the observer (II): Deciding when to decide. *PLoS ONE* **2010**, *5*, e15555. [CrossRef]
22. Beal, M.J. Variational Algorithms for Approximate Bayesian Inference. Ph.D. Thesis, University College London (UCL), London, UK, 2003.
23. Mathys, C.D.; Daunizeau, J.; Friston, K.J.; Stephan, K.E. A Bayesian Foundation for Individual Learning Under Uncertainty. *Front. Hum. Neurosci.* **2011**, *5*, 39. [CrossRef]
24. Vossel, S.; Mathys, C.; Daunizeau, J.; Bauer, M.; Driver, J.; Friston, K.; Stephan, K. Spatial Attention, Precision, and Bayesian Inference: A Study of Saccadic Response Speed. *Cereb. Cortex* **2013**, *24*, 1436–1450. [CrossRef]
25. Diaconescu, A.O.; Mathys, C.; Weber, L.A.; Kasper, L.; Mauer, J.; Stephan, K.E. Hierarchical prediction errors in midbrain and septum during social learning. *Soc. Cogn. Affect. Neurosci.* **2017**, *12*, 618–634. [CrossRef]
26. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
27. Si, B.; Herrmann, J.M.; Pawelzik, K. Gain-based Exploration: From Multi-armed Bandits to Partially Observable Environments. In Proceedings of the International Conference on Natural Computation, Haikou, China, 24–27 August 2007; pp. 177–182.
28. Atan, O.; Tekin, C.; van der Schaar, M. Global bandits. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 5798–5811. [CrossRef]
29. Xu, X.; Xie, H.; Lui, J.C.S. Generalized Contextual Bandits with Latent Features: Algorithms and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–13. [CrossRef] [PubMed]
30. Behrens, T.E.J.; Woolrich, M.W.; Walton, M.E.; Rushworth, M.F.S. Learning the value of information in an uncertain world. *Nat. Neurosci.* **2007**, *10*, 1214–1221. [CrossRef] [PubMed]
31. Walton, M.E.; Behrens, T.E.; Buckley, M.J.; Rudebeck, P.H.; Rushworth, M.F. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* **2010**, *65*, 927–939. [CrossRef] [PubMed]
32. Costa, V.D.; Mitz, A.R.; Averbeck, B.B. Subcortical substrates of explore-exploit decisions in primates. *Neuron* **2019**, *103*, 533–545. [CrossRef] [PubMed]
33. Hampton, A.N.; Bossaerts, P.; O'Doherty, J.P. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6741–6746. [CrossRef] [PubMed]
34. Heuer, L.; Orland, A. Cooperation in the Prisoner's Dilemma: An experimental comparison between pure and mixed strategies. *R. Soc. Open Sci.* **2019**, *6*, 182142. [CrossRef]
35. Hill, C.A.; Suzuki, S.; Polania, R.; Moisa, M.; O'doherty, J.P.; Ruff, C.C. A causal account of the brain network computations underlying strategic social behavior. *Nat. Neurosci.* **2017**, *20*, 1142–1149. [CrossRef]
36. Bolis, D.; Balsters, J.; Wenderoth, N.; Becchio, C.; Schilbach, L. Beyond autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity. *Psychopathology* **2017**, *50*, 355–372. [CrossRef]
37. Konishi, T.; Kubo, T.; Watanabe, K.; Ikeda, K. Variational Bayesian Inference Algorithms for Infinite Relational Model of Network Data. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 2176–2181. [CrossRef]
38. Chien, J.T.; Ku, Y.C. Bayesian Recurrent Neural Network for Language Modeling. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 361–374. [CrossRef] [PubMed]
39. Qi, Y.; Liu, B.; Wang, Y.; Pan, G. Dynamic ensemble modeling approach to nonstationary neural decoding in Brain-computer interfaces. In Proceedings of the Advances in Neural Information Processing Systems 32 (Nips 2019), Vancouver, BC, Canada, 8–14 December 2019.
40. Li, H.; Barnaghi, P.; Enshaeifar, S.; Ganz, F. Continual Learning Using Bayesian Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4243–4252. [CrossRef] [PubMed]
41. Wang, H.; Yeung, D.Y. Towards Bayesian deep learning: A framework and some existing methods. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 3395–3408. [CrossRef]
42. Du, C.; Zhu, J.; Zhang, B. Learning Deep Generative Models With Doubly Stochastic Gradient MCMC. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 3084–3096. [CrossRef] [PubMed]
43. Mirza, M.B.; Adams, R.A.; Mathys, C.; Friston, K.J. Human visual exploration reduces uncertainty about the sensed world. *PLoS ONE* **2018**, *13*, e0190429. [CrossRef] [PubMed]
44. Adolphs, R. Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* **2003**, *4*, 165–178. [CrossRef] [PubMed]
45. Pezzulo, G.; Friston, K.J. The value of uncertainty: An active inference perspective. *Behav. Brain Sci.* **2019**, *42*, e47. [CrossRef] [PubMed]
46. Zhu, C.; Zhou, K.; Han, Z.; Tang, Y.; Tang, F.; Si, B. General hierarchical Brownian filter in multi-dimensional volatile environments. 2022, *submitted*.
47. Mathys, C.D.; Lomakina, E.I.; Daunizeau, J.; Iglesias, S.; Brodersen, K.H.; Friston, K.J.; Stephan, K.E. Uncertainty in perception and the Hierarchical Gaussian Filter. *Front. Hum. Neurosci.* **2014**, *8*, 825. [CrossRef]
48. Al-Nowaihi, A.; Dhami, S. *Probability Weighting Functions*; University of Leicester: Leicester, UK, 2010.
49. Nocedal J.;Wright S. J. *Numerical Optimization*; Spinger: New York, NY, USA, 2006.
50. Ando, T. *Bayesian Model Selection and Statistical Modeling*; CRC Press: Cleveland, OH, USA, 2010.
51. Zhang, L.; Gläscher, J. A brain network supporting social influences in human decision-making. *Sci. Adv.* **2020**, *6*, eabb4159. [CrossRef]
52. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*; Springer Inc.: New York, NY, USA, 2013.
53. Zeng, T.; Si, B. A brain-inspired compact cognitive mapping system. *Cogn. Neurodyn.* **2021**, *15*, 91–101. [CrossRef]

54. Chen, S.; Tang, J.; Zhu, L.; Kong, W. A multi-stage dynamical fusion network for multimodal emotion recognition. *Cogn. Neurodyn.* **2022**, 1–10. [CrossRef]
55. Walkenbach, J.; Haddad, N.F. The Rescorla-Wagner theory of conditioning: A review of the literature. *Psychol. Rec.* **1980**, *30*, 497–509. [CrossRef]
56. Zhang, L.; Lengersdorff, L.; Mikus, N.; Gläscher, J.; Lamm, C. Using reinforcement learning models in social neuroscience: frameworks, pitfalls and suggestions of best practices. *Soc. Cogn. Affect. Neurosci.* **2020**, *15*, 695–707. [CrossRef] [PubMed]
57. Zheng, N.; Ding, J.; Chai, T. DMGAN: Adversarial Learning-Based Decision Making for Human-Level Plant-Wide Operation of Process Industries Under Uncertainties. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 985–998. [CrossRef] [PubMed]
58. Chen, X.; Yang, T. A neural network model of basal ganglia's decision-making circuitry. *Cogn. Neurodyn.* **2021**, *15*, 17–26. [CrossRef] [PubMed]
59. Mao, D. Neural Correlates of Spatial Navigation in Primate Hippocampus. *Neurosci. Bull.* **2022**, 1–13.. [CrossRef] [PubMed]
60. Zheng, L.; Liu, W.; Long, Y.; Zhai, Y.; Zhao, H.; Bai, X.; Zhou, S.; Li, K.; Zhang, H.; Liu, L.; et al. Affiliative bonding between teachers and students through interpersonal synchronisation in brain activity. *Soc. Cogn. Affect. Neurosci.* **2020**, *15*, 97–109. [CrossRef]
61. Wang, Y.; Yang, X.; Tang, Z.; Xiao, S.; Hewig, J. Hierarchical neural prediction of interpersonal trust. *Neurosci. Bull.* **2021**, *37*, 511–522. [CrossRef]
62. Wang, W.; Fu, C.; Kong, X.; Osinsky, R.; Hewig, J.; Wang, Y. Neuro-behavioral dynamic prediction of interpersonal cooperation and aggression. *Neurosci. Bull.* **2022**, *38*, 275–289. [CrossRef]
63. Dong, W.; Chen, H.; Sit, T.; Han, Y.; Song, F.; Vyssotski, A.L.; Gross, C.T.; Si, B.; Zhan, Y. Characterization of exploratory patterns and hippocampal–prefrontal network oscillations during the emergence of free exploration. *Sci. Bull.* **2021**, *66*, 2238–2250. [CrossRef]
64. Friston, K.; FitzGerald, T.; Rigoli, F.; Schwartenbeck, P.; Pezzulo, G. Active Inference: A Process Theory. *Neural Comput.* **2017**, *29*, 1–49. [CrossRef] [PubMed]
65. Friston, K.; Mattout, J.; Trujillo-Barreto, N.; Ashburner, J.; Penny, W. Variational free energy and the Laplace approximation. *NeuroImage* **2007**, *34*, 220–234. [CrossRef] [PubMed]
66. Harold Jeffreys, S. *Theory of Probability*; Clarendon Press: Oxford, UK, 1961.
67. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]