THE NEURAL CORRELATES OF NOVELTY AND VARIABILITY IN HUMAN DECISION-MAKING UNDER AN ACTIVE INFERENCE FRAMEWORK

Shuo Zhang[1,2], **Yan Tian**[1], **Quanying Liu**[2,*], **Haiyan Wu**[1,*]
[1] Centre for Cognitive and Brain Sciences and Department of Psychology,
University of Macau, Macau, China
[2] Department of Biomedical Engineering,
Southern University of Science and Technology
Shenzhen, China
[*] Corresponding to
liuqy@sustech.edu.cn & haiyanwu@um.edu.mo

December 6, 2024

ABSTRACT

Active inference integrates perception, decision-making, and learning into a united theoretical framework, providing an efficient way to trade off exploration and exploitation by minimizing (expected) free energy. In this study, we asked how the brain represents values and uncertainties (novelty and variability), and resolves these uncertainties under the active inference framework in the exploration-exploitation trade-off. 25 participants performed a contextual two-armed bandit task, with electroencephalogram (EEG) recordings. By comparing the model evidence for active inference and reinforcement learning models of choice behavior, we show that active inference better explains human decision-making under novelty and variability, which entails exploration or information seeking. The EEG sensor-level results show that the activity in the frontal, central, and parietal regions is associated with novelty, while activity in the frontal and central brain regions is associated with variability. The EEG source-level results indicate that the expected free energy is encoded in the frontal pole and middle frontal gyrus and uncertainties are encoded in different brain regions but with overlap. Our study dissociates the expected free energy and uncertainties in active inference theory and their neural correlates, speaking to the construct validity of active inference in characterizing cognitive processes of human decisions. It provides behavioral and neural evidence of active inference in decision processes and insights into the neural mechanism of human decisions under uncertainties.

1 Introduction

Active inference from the free energy principle provides a powerful explanatory tool for understanding the dynamic relationship between an agent and its environment [1]. Free energy is a measure of an agent's uncertainty about the environment, which can be understood as the difference between the real environment state and the agent's estimated environment state [2]. In addition, expected free energy is the free energy about the future and can be used to guide the optimization process of decision-making. Under the active inference framework, perception, action, and learning are all driven by the minimization of free energy (Figure 1). By minimizing free energy, people can optimize decisions, which encompasses both the reduction of uncertainty about the environment (through *exploration*) and the maximization of rewards (through *exploitation*). Active inference [3] is a pragmatic implementation of the free energy principle to action, proposing that agents not only minimize free energy through perception but also through actions that enable them to reach preferable states. Briefly, in active inference, the agent has an internal cognitive model to approximate the hidden states of the environment (perception) and actively acts to enable oneself to reach preferable states (action)(see *Section 2.1*).

In recent years, the active inference framework has been applied to understanding cognitive processes and behavioral policies in human decisions. Many works provide support for the potential of the active inference framework to describe complex cognitive processes and give theoretical insights into behavioral dynamics [4–7]. For instance, it is theoretically deduced in the active inference framework on the exploration and exploitation trade-off [3, 8], which trade-off is essential to the functioning of cognitive agents in many decision contexts [9, 10]. Specifically, *exploration* is to take actions that offer extra information about the current environment, actions with higher uncertainty, while *exploitation* is to take actions to maximize immediate rewards given the current belief, actions with higher expected reward. The exploration-exploitation trade-off refers to an inherent tension between information (resolving uncertainty) and goal-seeking, particularly when the agent is confronted with incomplete information about the environment [11]. However, these theoretical studies have rarely been confirmed experimentally with lab empirical evidence from both behavioral and neural responses [1, 2]. We aimed to validate the active framework in a decision-making task with electroencephalogram (EEG) neural recordings.

The decision-making process frequently involves grappling with varying forms of uncertainty, such as novelty - the kind of uncertainty that can be reduced through sampling, and variability - the inherent uncertainty (variance) presented by a stable environment. Studies have investigated these different forms of uncertainty in decision-making, focusing on their neural correlates [12–15]. These studies utilized different forms of multi-armed bandit tasks, e.g the restless multi-armed bandit tasks [12, 16], risky/safe bandit tasks [15, 17, 18], contextual multi-armed bandit tasks [19–21]. However, these tasks only separate either variability from novelty in uncertainty or actions from states (perception). In our work, we develop a contextual multi-armed bandit task to enable participants to actively reduce novelty, avoid variability, and maximize rewards using various policies (see *Section 2.2* and Figure 4 (a)). Our task makes it possible to study whether the brain represents these different types of uncertainty distinctly [22] and whether the brain represents both the value of reducing uncertainty and the degree of uncertainty. The active inference framework presents a theoretical approach to investigate these questions. Within this framework, uncertainties can be reduced to novelty and variability. Novelty is represented by the uncertainty about model parameters associated with choosing a particular action, while variability is signified by the variance of the environment's hidden states. The value of reducing novelty, the value of reducing variability, and extrinsic value together constitute expected free energy (see *Section 2.1*).

Our study aims to utilize the active inference framework to investigate how the brain represents the decision-making process and how the brain disassociates the representations of novelty and variability (the degree of uncertainty and the value of reducing uncertainty). To achieve these aims, we utilize the active inference framework to examine the exploration-exploitation trade-off, with behavioral and EEG data (see *Methods*). Our study provides results of 1) how participants trade off the exploration and exploitation in the contextual two-armed bandit task (behavioral evidence)(see *Section 3.1*), 2) how brain signals differ under different levels of ambiguities and risks (sensor-level EEG evidence, see *Section 3.2*), 3) how our brain encodes the trade-off of exploration and exploitation, evaluates the values of reducing novelty and reducing variability during action selection, and 4) updates the information about the environment during belief update (source-level EEG evidence, see *Section 3.3*).

2 Methods

2.1 The free energy principle and active inference

The Free energy principle [1] is a theoretical framework that proposes that both biological and non-biological systems tend to minimize their (variational) free energy to maintain a non-equilibrium steady state. In the context of the brain, the free energy principle suggests that the brain functions as an "inference machine" that aims to minimize the difference between its internal cognitive model about the environment and the true causes (hidden states) of perceived sensory inputs. This minimization is achieved through active inference.

Active inference can be regarded as a form of planning as inference, in which an agent samples the environment to maximize the evidence for its internal cognitive model of how sensory samples are generated. This is sometimes known as self-evidencing [3]. Under the active inference framework, variational free energy can be viewed as the objective function that underwrites belief updating; namely, inference and learning. By minimizing the free energy expected following an action (i.e., expected free energy) we can optimize decisions and resolve uncertainty.

Mathematically, the minimization of free energy is formally related to Variational Bayesian methods [23]. Variational inference is used to estimate both hidden states of the environment and the parameters of the cognitive model. This process can be viewed as an optimization problem that seeks to find the best model parameters and action policy to maximize the sensory evidence. By minimizing variational free energy and expected free energy, optimal model parameters can be estimated and better decisions can be made [24]. Active inference bridges the sensory input, cognitive processes, and action output, enabling us to quantitatively describe the neural processes of learning about the environment. The brain receives sensory input o from the environment, and the cognitive model encoded by the brain $q(s)$

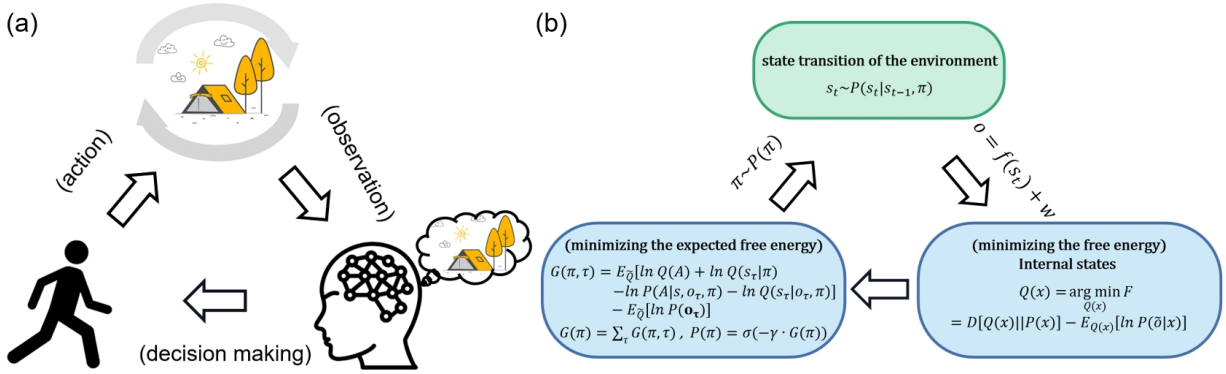

Figure 1: Active inference. (a) Qualitatively, agents receive observations from the environment and use these observations to optimize Bayesian beliefs under an internal cognitive (a.k.a., world or generative) model of the environment. Then agents actively sample the environment states by action, choosing actions that would make them in more favorable states. The environment changes its state according to agents' policies (action sequences) and transition functions. Then again, agents receive new observations from the environment. (b) From a quantitative perspective, agents optimize the Bayesian beliefs under an internal cognitive (a.k.a., world or generative) model of the environment by minimizing the variational free energy. Then agents select policies minimizing the expected free energy, namely, the surprise expected in the future under a particular policy.

makes an inference on the cause of sensory input $p(s|o)$ (a.k.a., the hidden state of the environment). In the free energy principle, minimizing free energy refers to minimizing the difference (*e.g.*, KL divergence) between the cognitive model encoded by the brain and the causes of the sensory input. Thus, free energy is an information-theoretic quantity that bounds the evidence for the data model. Free energy can be minimized by the following two means [25]:

- Minimize free energy through *perception*. Based on existing observations, by maximizing model evidence, the brain improves its internal cognitive model, reducing the gap between the true cause of the sensory input and the estimated distribution of the internal cognitive model.

- Minimize free energy through *action*. The agent actively samples the environment, making the sensory input more in line with the cognitive model by sampling preferred states (i.e., prior preferences over observations). Minimizing free energy through action is one of the generalizations afforded by the free energy principle over Bayesian formulations which only address perception.

Active inference formulates the necessary cognitive processing as a process of belief updating, where choices depend on agents' expected free energy. Expected free energy serves as a universal objective function, guiding both perception and action. In brief, expected free energy can be seen as the expected surprise following some policies. The expected surprise can be reduced by resolving uncertainty, and one can select policies with lower expected free energy which can encourage information-seeking and resolve uncertainty. Additionally, one can minimize expected surprise by avoiding surprising or aversive outcomes [26, 27]. This leads to goal-seeking behavior, where goals can be viewed as prior preferences or rewarding outcomes.

Technically, expected free energy can also be expressed as expected information gain plus expected value, where the value corresponds to (log) prior preferences. We will refer to both formulations in what follows. Resolving novelty, minimizing variability, and maximizing information gain have epistemic value while maximizing expected value has pragmatic or instrumental value. These two types of values can be referred to in terms of intrinsic and extrinsic value, respectively [8, 28].

2.1.1 The generative model

Active inference builds on partially observable Markov decision processes: (O, S, U, T, R, P, Q)(see Table 1).

In this model, the generative model P is parameterized as follows and the model parameters are $\eta = a, c, d, \beta$ [3]:

Table 1: Ingredients for computational modeling of active inference

Notations	Definition	Description
O	a finite set of observations (outcomes)	Sensory input that brains receive
S	a finite set of hidden states	The true hidden states of the environment that generate sensory inputs to brains
U	a finite set of actions	Agent performs actions that change the environment
T	a finite set of time-sensitive policies	A policy is an action sequence over time
R	a generative process $R(\widetilde{o}, \widetilde{s}, \widetilde{u})$	The generative process generates observations and next states (state transitions of the environment) based on current states and actions
P	a generative model $P(\widetilde{o}, \widetilde{s}, \pi, \eta)$	The generative model describes what the agent believes about the environment (how observations are generated)
Q	an approximate posterior	The Bayesian beliefs under the generative model that is optimized to minimize variational free energy. By definition, these beliefs correspond to approximate posteriors.

$$
\begin{aligned}
&P(\widetilde{o}, \widetilde{s}, \pi, \eta) = P(\pi)P(\eta)\prod_{t=1}^{T} P(o_t|s_t)P(s_t|s_{t-1}, \pi), \\
&P(o_t|s_t) = Cat(A), \\
&P(s_{t+1}|s_t, \pi) = Cat(B(\pi(t))), \\
&P(s_0) = d, \\
&P(\pi|\gamma) = \sigma(-\gamma \cdot G(\pi)), \\
&P(A) = Dir(a), \\
&P(\gamma) = \Gamma(1, \beta),
\end{aligned} \tag{1}
$$

where o is observations or sensory inputs ($\widetilde{o}$ is the history of observations), s is the hidden states of the environment ($\widetilde{s}$ is the history of hidden states), π is agent's policies, A is the likelihood matrix mapping from hidden states to observations, B is the transition function for hidden states under the policy in time t, d is the prior expectation of each state at the beginning of each trial, γ is the inverse temperature of beliefs about policies, β is the prior expectation of policies' temperature parameters, a is the concentration parameters of the likelihood matrix, σ is the softmax function, $Cat()$ is the categorical distribution, $Dir()$ is the Dirichlet distribution, and $\Gamma()$ is the Gamma distribution.

The posterior probability of the corresponding hidden states and parameters ($x = \widetilde{s}, \pi, A, B, \beta$) is as Eq.(2):

$$
Q(x) = Q(s_1|\pi)...Q(s_T|\pi)Q(\pi)Q(A)Q(B)Q(\gamma) = \arg\min_{Q(x)} F \approx P(x|\widetilde{o}) \tag{2}
$$

The generative model is a conceptual representation of how agents understand their environment. This model fundamentally posits that agents' observations are contingent upon states, and the transitions of these states inherently depend on both the state itself and the chosen policy. It is crucial to note that within this model, the policy is considered a stochastic variable requiring inference, thus considering planning as a form of inference. This inference process involves deducing the optimal policy from the agents' observations. All the conditional abilities rest on likelihood and state transition models that are parameterized using a Dirichlet distribution [29]. The Dirichlet distribution's sufficient statistic is its concentration parameter, which is equivalently interpreted as the cumulative frequency of previous occurrences. In essence, this means that the agents incorporate the frequency of past combinations of states and observations into the generative model. Therefore, the generative model plays a pivotal role in inferring the probabilities and uncertainties related to the hidden states and observations.

2.1.2 Variational free energy and expected free energy

Perception, decision-making, and learning in active inference are all achieved by minimizing the variational and expected free energy with respect to the model parameters and hidden states. The variational free energy can be expressed

in various forms with respect to the reduced posterior as Eq.(3):

$$\begin{aligned} F &= E_{Q(x)}[\ln Q(x) - \ln P(x, \widetilde{o})] \\ &= E_{Q(x)}[\ln Q(x) - \ln P(x|\widetilde{o}) - \ln P(\widetilde{o})] \\ &= E_{Q(x)}[\ln Q(x) - \ln P(\widetilde{o}|x) - \ln P(x)] \\ &= D_{KL}[Q(x)||P(x)] - E_{Q(x)}[\ln P(\widetilde{o})] \end{aligned} \tag{3}$$

Here, $x = \widetilde{s}, \pi, A, B, \beta$, including the hidden states and parameters. These forms of free energy are consistent with the variational inference in statistics. Minimizing free energy is equal to maximizing model evidence, that is, minimizing surprise. In addition, free energy can also be written in other forms as Eq.(4):

$$F = \underbrace{D_{KL}[Q(x)||P(x)]}_{complexity} - \underbrace{E_{Q(x)}[\ln P(\widetilde{o}|x)]}_{accuracy} \tag{4}$$

The initial term, denoted as $D_{KL}[Q(x)||P(x)]$, is conventionally referred to as “complexity". This term, reflecting the divergence between $Q(x)$ and $P(x)$, quantifies the volume of information intended to be encoded within $Q(x)$ that is not inherent in $P(x)$. The subsequent term, $E_Q[\ln P(\widetilde{o}|s)]$, designated as “accuracy", represents the likelihood of an observation expected under approximate posterior (Bayesian) beliefs about hidden states.

The minimization of variational free energy facilitates a progressive alignment between the approximate posterior distribution of hidden states, as encoded by the brain's cognitive function, and the actual posterior distribution of the environment. However, it is noteworthy that our policy beliefs are future-oriented. We want policies that possess the potential to effectively guide us toward achieving the future state that we desire. It follows that these policies should aim to minimize the free energy in the future, or in other words, expected free energy. Thus, expected free energy depends on future time points τ and policies π, and x can be replaced by the possible hidden state s_τ and the likelihood matrix A. The relationship between policy selection and expected free energy is inversely proportional: a lower expected free energy under a given policy heightens the probability of that policy's selection. Hence, expected free energy emerges as a crucial factor influencing policy choice.

$$\begin{aligned} P(\pi) &= \sigma(-\gamma \cdot G(\pi) \\ G(\pi) &= \sum_t G(\pi, \tau) \end{aligned} \tag{5}$$

Next, we can derive the expected free energy in the same way as the variational free energy:

$$G(\pi, \tau) = E_{\widetilde{Q}}[\ln Q(s_\tau, A|\pi) - \ln P(o_\tau, s_\tau, A|\pi)] \tag{6}$$

$$\begin{aligned} G(\pi, \tau) &= E_{\widetilde{Q}}[\ln Q(s_\tau, A|\pi) - \ln P(o_\tau, s_\tau, A|\pi)] \\ &= E_{\widetilde{Q}}[\ln Q(A) + \ln Q(s_\tau|\pi) - \ln P(A|s_\tau, o_\tau, \pi) - \ln P(s_\tau|o_\tau, \pi) - \ln P(o_\tau)] \\ &\approx E_{\widetilde{Q}}[\ln Q(A) + \ln Q(s_\tau|\pi) - \ln P(A|s_\tau, o_\tau, \pi) - \ln Q(s_\tau|o_\tau, \pi) - \ln P(o_\tau)] \end{aligned} \tag{7}$$

In Eq.(7), it is important to note that we anticipate observations that have not yet occurred. Consequently, we designate $\widetilde{Q} = Q(o_\tau, s_\tau, A|\pi)$. If we establish a relationship between $\ln P(o_\tau)$ and the prior preference, it enables us to express expected free energy in terms of epistemic value and extrinsic value. The implications of such a relationship offer a new lens to understand the interplay between cognitive processes and their environmental consequences, thereby enriching our understanding of decision-making under the active inference framework.

$$G(\pi, \tau) = \underbrace{E_{\widetilde{Q}}[\underset{novelty}{\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)} + \underset{salience}{\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)}]}_{Epistemic\ value} - \underbrace{E_{\widetilde{Q}}[\ln P(o_\tau)]}_{Extrinsic\ value} \tag{8}$$

In this context, extrinsic value aligns with the concept of expected utility. On the other hand, epistemic value corresponds to the anticipated information gain or the value of reducing uncertainty, encapsulating the exploration of both model parameters (novelty) and the hidden states (salience), which are to be illuminated by future observations. We

can add coefficients (AL, AI, and EX) before these three terms of Eq.(8) to better simulate the diverse exploration strategies of agents:

$$G(\pi,\tau) = \underbrace{AL \cdot E_{\widetilde{Q}}[\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)]}_{\text{Value of reducing novelty}} + \underbrace{AI \cdot E_{\widetilde{Q}}[\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)]}_{\text{Value of reducing variability}} - \underbrace{EX \cdot E_{\widetilde{Q}}[\ln P(o_\tau)]}_{\text{Extrinsic value}} \tag{9}$$

To align with different types of uncertainties and avoid conflicts with active inference terminology, the first two terms in Eq. (9) are referred to as the value of reducing novelty and variability, respectively, while the corresponding terms in Eq. (8) are termed novelty and variability.

Belief updates play a dual role by facilitating both inference and learning processes. The inference is here understood as the optimization of expectations about the hidden states. Learning, on the other hand, involves the optimization of model parameters. This optimization necessitates the finding of sufficient statistics of the approximate posterior that minimize the variational free energy. Active inference employs the technique of gradient descent to identify the optimal update method [3]. In the present work, our focus is primarily centered on the updated methodology related to the likelihood mapping A and the concentration parameter a (rows correspond to observations, and columns correspond to hidden states):

$$a_0 = \begin{bmatrix} 1 & 1 & prior & prior & 0 & 0 & 0 & 0 \\ 0 & 0 & prior & prior & 0 & 0 & 0 & 0 \\ 0 & 0 & prior & prior & 0 & 0 & 0 & 0 \\ 0 & 0 & prior & prior & 0 & 0 & 0 & 0 \\ 0 & 0 & prior & prior & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{10}$$

$$\begin{aligned} A &= Cat(a), \\ a_{t+1} &= a_t + \alpha(o_t \otimes s_t), \end{aligned} \tag{11}$$

where a_0 is the concentration parameter at the beginning of the experiment, $prior$ is the prior concentration of a_0, and α is the learning rate.

2.2 Contextual two-armed bandit task

In this study, we developed a “contextual two-armed bandit task" (Figure 2), which was based on the conventional multi-armed bandit task [8, 30]. Participants were instructed to explore two paths that offer rewards with the aim of maximizing cumulative rewards. One path provided constant rewards in each trial, labeled the “Safe" while the other, referred to as the “Risky", probabilistically offered varying amounts of rewards. The risky path had two different contexts, “Context 1" and “Context 2", each corresponding to different reward distributions. The risky path would give more rewards in “Context 1" and give fewer rewards in “Context 2". The context of the risky path changed randomly in each trial, and agents could only know the specific context of the current trial’s risky path by accessing the “Cue" option, although this comes with a cost. The actual reward distribution of the risky path in “Context 1" was $[+12\ (55\%), +9\ (25\%), +6\ (10\%), +3\ (5\%), +0\ (5\%)]$ and the actual reward distribution of the risky path in “Context 2" was $[+12\ (5\%), +9\ (5\%), +6\ (10\%), +3\ (25\%), +0\ (55\%)]$. For a comprehensive overview of the specific settings, please refer to Figure 2.

We ran some simulation experiments to demonstrate how active inference agents performed the “contextual two-armed bandit task" (Figure 3). Active inference agents with different parameter configurations could exhibit different decision-making policies, as demonstrated in the simulation experiment (see Figure 3). By adjusting parameters such as AL, AI, EX (Eq.(9)), $prior$ (Eq.(10)), and α (Eq.(11)), agents could operate under different policies. Agents with a low learning rate would initially incur a cost to access the cue, enabling them to thoroughly explore and understand the reward distributions of different contexts. Once sufficient environmental information was obtained, the agent would evaluate the actual values of various policies and select the optimal policy for exploitation. In the experimental setup, the optimal policy required selecting the risky path in a high-reward context and the safe path in a low-reward context after accessing the cue. However, in particularly difficult circumstances, an agent with a high learning rate might become trapped in a local optimum and consistently opt for the safe path, especially if the initial high-reward scenarios encountered yield minimal rewards.

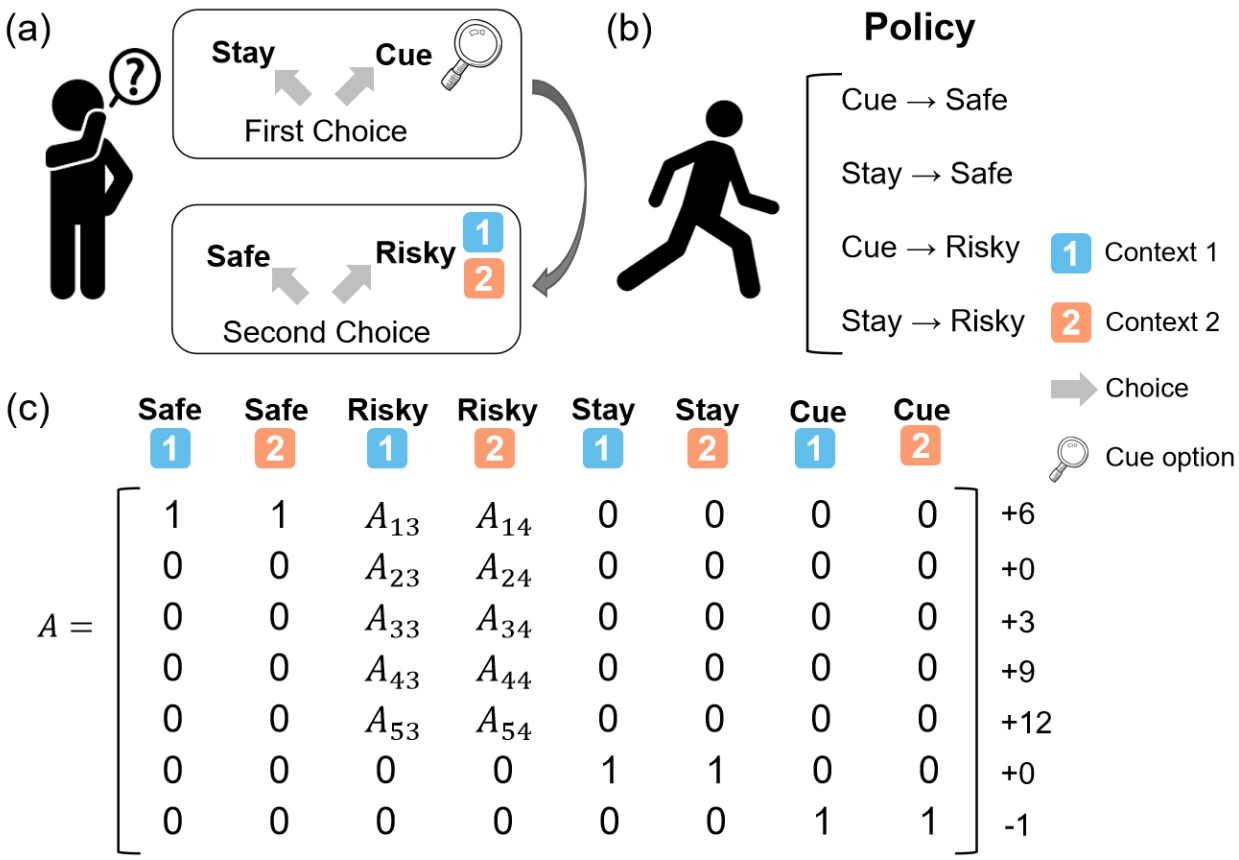

Figure 2: The contextual two-armed bandit task. (a) In this task, agents need to make two choices in each trial. The first choice is: “Stay" and “Cue". The “Stay" option gives you nothing while the “Cue" option gives you a -1 reward and the context information about the “Risky" option in the current trial. The second choice is: “Safe" and “Risky". The “Safe" option always gives you a +6 reward and the “Risky" option gives you a reward probabilistically, ranging from 0 to +12 depending on the current context (context 1 or context 2); (b) The four policies in this task are: “Cue" and “Safe", “Stay" and “Safe", “Cue" and “Risky", and “Stay" and “Risky"; (c) The likelihood matrix maps from 8 hidden states (columns) to 7 observations (rows).

Figure 3 shows how an active inference agent with $AI = AL = EX = 1$ performs our task. We can see the active inference agent exhibits human-like policies and efficiency in completing tasks. In the early stages of the simulation, the agent tended to prefer the “Cue" option, as it provided more information, reducing novelty and reducing variability. Similarly, in the second choice, the agent favored the “Risky" option, even though initially the expected rewards for the “Safe" and “Risky" options were the same, but the “Risky" option offered greater informational value and reduced novelty. In the latter half of the experiment, the agent again preferred the “Cue" option due to its higher expected reward. For the second choice, the agent made decisions based on specific contexts, opting for the “Risky" option in “Context 1" for a higher expected reward, and the “Safe" option in “Context 2" where the informational value of the “Risky" option was outweighed by the difference in expected rewards between the “Safe" option and the “Risky" option in “Context 2".

2.3 EEG collection and analysis

2.3.1 Participants

Participants were recruited via an online recruitment advertisement. We recruited 25 participants (male: 14, female: 11, mean age: 20.82 ± 2.12 years old), concurrently collecting electroencephalogram (EEG) and behavioral data. All participants signed an informed consent form before the experiments. This study was approved by the local ethics committee of the University of Macau (BSERE22-APP006-ICI).

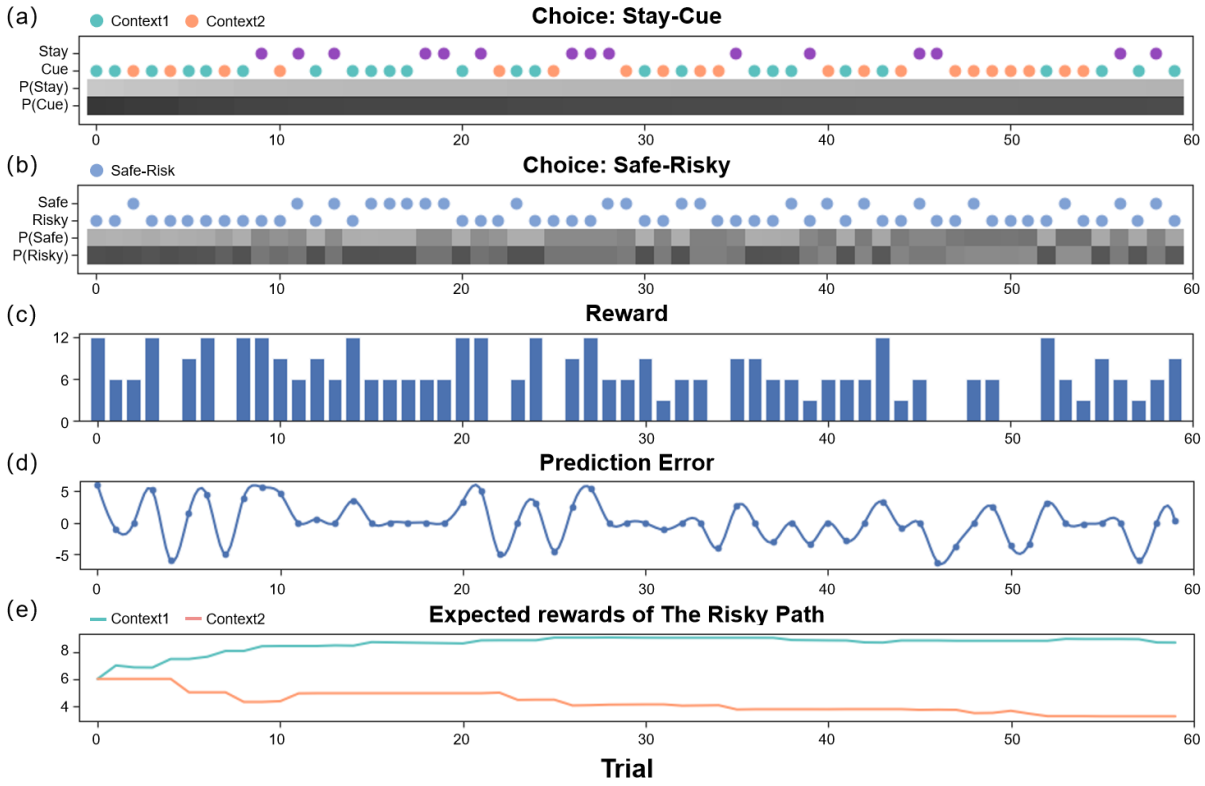

Figure 3: The simulation experiment results. This figure demonstrates how an agent selects actions and updates beliefs over 60 trials in the active inference framework. The first two panels (a-b) display the agent's policy and depict how the policy probabilities are updated (choosing between the stay or cue option in the first choice, and selecting between the safe or risky option in the second choice). The scatter plot indicates the agent's actions, with green representing the cue option when the context of the risky path is "Context 1" (high-reward context), orange representing the cue option when the context of the risky path is "Context 2" (low-reward context), purple representing the stay option when the agent is uncertain about the context of the risky path, and blue indicating the safe-risky choice. The shaded region represents the agent's confidence, with darker shaded regions indicating greater confidence. The third panel(c) displays the rewards obtained by the agent in each trial. The fourth panel(d) shows the prediction error of the agent in each trial, which decreases over time. Finally, the fifth panel(e) illustrates the expected rewards of the "Risky Path" in the two contexts of the agent.

2.3.2 Data collection

In the behavioral experiment, in order to enrich the behavioral data of participants, a "you can ask" stage was added at the beginning of each trial. When the participants see "you can ask", they know that they can choose whether to ask for cue information in the next stage; when the participants see "you can't ask", they know that they can't choose whether to ask and it defaults that participants choose the "Stay" option. Additionally, to make the experiment more realistic, we added a background story of "finding apples" to the experiment. Specifically, participants were presented with the following instructions: *"You are on a quest for apples in a forest, beginning with 5 apples. You encounter two paths: 1) The left path offers a fixed yield of 6 apples per excursion. 2) The right path offers a probabilistic reward of 0/3/6/9/12 apples, and it has two distinct contexts, labeled "Context 1" and "Context 2," each with a different reward distribution. Note that the context associated with the right path will randomly change in each trial. Before selecting a path, a ranger will provide information about the context of the right path ("Context 1" or "Context 2") in exchange for an apple. The more apples you collect, the greater your monetary reward will be."*

The participants were provided with the task instructions (i.e., prior beliefs) above and asked to press a space bar to proceed. They were told that the total number of apples collected would determine the monetary reward they would receive. For each trial, the experimental procedure is illustrated in Figure 4 (a), and comprises five stages:

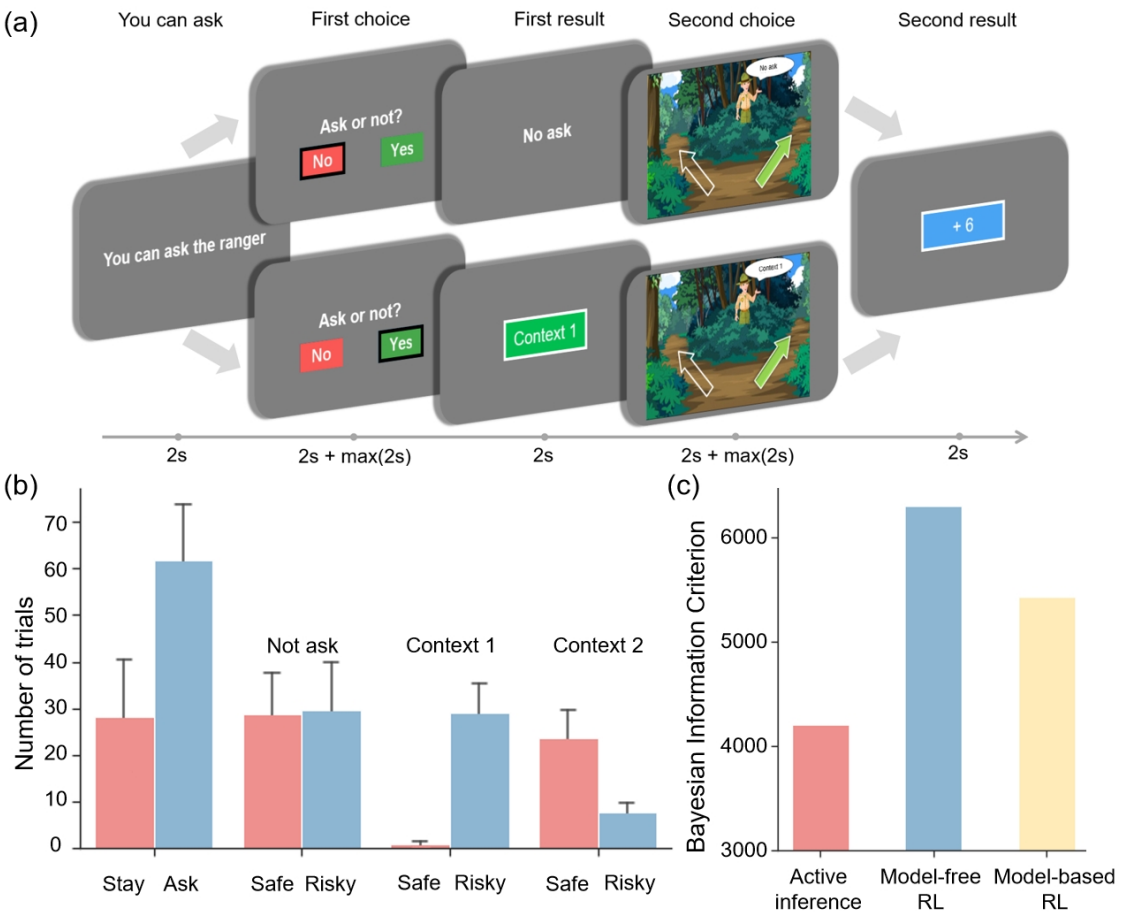

Figure 4: The experiment task and behavioral result. (a) The five stages of the experiment, which include the “You can ask" stage to prompt the participants to decide whether to request information from the Ranger, the “First choice" stage to decide whether to ask the ranger for information, the “First result" stage to display the result of the “First choice" stage, the “Second choice" stage to choose between left and right paths under different uncertainties and the “Second result" stage to show the result of the “Second choice" stage. (b) The number of times each option was selected. The error bar indicates the variance among participants. (c) The Bayesian Information Criterion of active inference, model-free reinforcement learning, and model-based reinforcement learning.

1. “You can ask" stage: Participants are informed if they can choose to ask in the “First choice" stage. If they can’t ask, it defaults that participants choose to not ask. This stage lasts for 2 seconds.

2. “First choice" stage: Participants decide whether to press the right or left button to ask the ranger for information, at the cost of an apple. In this stage, participants have two seconds to decide which option to choose, and they can not press buttons within these two seconds. Then, they need to respond by pressing a button within another two seconds. This stage corresponds to the action selection in active inference.

3. “First result" stage: Participants either receive information about the context of the right path for the current trial or gain no additional information based on their choices. This stage lasts for 2 seconds and corresponds to the belief update in active inference.

4. “Second choice" stage: Participants decide whether to select the RIGHT or LEFT key to choose the safe path or risky path. In this stage, participants have two seconds to decide which option to choose, and they can not press buttons within these two seconds. Then, they need to respond by pressing a button within another two seconds. This stage corresponds to the action selection in active inference.

5. “Second result" stage: Participants are informed about the number of apples rewarded in the current trial and their total apple count, which lasts for 2 seconds. This stage corresponds to the belief update in active inference.

Each stage was separated by a jitter ranging from 0.6 to 1.0 seconds. The entire experiment consisted of a single block with a total of 120 trials. The participants were required to use any two fingers of one hand to press the buttons (left arrow and right arrow on the keyboard).

2.3.3 EEG processing

The processing of EEG signals was conducted using the EEGLAB toolbox [31] in the Matlab and the MNE package [32]. The preprocessing of EEG data involved multiple steps, including data selection, downsampling, high- and low-pass filtering, and independent component analysis (ICA) decomposition. Two-second data segments were selected at various stages during each trial in Figure 4 (a). Subsequently, the data was downsampled to a frequency of 250Hz and subjected to high- and low filtering within the 1-30 Hz frequency range. In instances where channels exhibited abnormal data, these were resolved using interpolation and average values. Following this, ICA was applied to identify and discard components flagged as noise.

After obtaining the preprocessed data, our objective was to gain a more comprehensive understanding of the specific functions associated with each brain region, mapping EEG signals from the sensor level to the source level. To accomplish this, we employed the head model and source space available in the “fsaverage" of the MNE package. We utilized eLORETA [33] for mapping the EEG data to the source space and used the “aparc sub" parcellation for annotation [34].

We segmented the data into five intervals that corresponded to the five stages of the experiment. The first stage, known as the “You can ask" stage, informed participants whether they could ask the ranger. In the second stage, referred to as the “First choice" stage, participants decided whether to seek cues. The third stage, called the “First result" stage, revealed the results of participants’ first choices. The fourth stage, known as the “Second choice" stage, involved choosing between choosing the safe or risky path. Finally, the fifth stage, named the “Second result" stage, encompassed receiving rewards. The two seconds in the two choosing stages when participants were thinking about which option to choose and the two seconds in the two result stages when the results were presented were used for analysis. Each interval lasted two seconds, and this categorization allowed us to investigate brain activity responses to different phases of the decision-making process. Specifically, we examined the processes of action selection and belief update within the framework of active inference.

3 Results

3.1 Behavioral results

To assess the evidence for active inference over reinforcement learning, we fitted active inference (Eq.(9)), model-free reinforcement learning, and model-based reinforcement learning models to the behavioral data of each participant. This involved optimizing the free parameters of active inference and reinforcement learning models. The resulting likelihood was used to calculate the Bayesian Information Criterion (BIC) [35] as the evidence for each model. The free parameters for the active inference model (AL, AI, EX, $prior$ (Eq.(10)), and α (Eq.(11))) scaled the contribution of the three terms that constituted the expected free energy in Eq.(9). These coefficients could be regarded as precisions that characterized each participant’s prior beliefs about contingencies and rewards. For example, increasing α meant participants would update their beliefs about reward contingencies more quickly, increasing AL meant participants would like to reduce novelty more, and increasing AI meant participants would like to learn the hidden state of the environment and reduce variability more. The free parameters for the model-free reinforcement learning model were the learning rate α and the temperature parameter γ and the free parameters for the model-based were the learning rate α, the temperature parameter γ and $prior$ (the details for the model-free reinforcement learning model can be found in Eq.S1-11 and the details for the model-based reinforcement learning model can be found in Eq.S12-23 in the Supplementary Method). The parameter fitting for these three models was conducted using the ’BayesianOptimization’ package [36] in Python, first randomly sampling 1000 times and then iterating for an additional 1000 times.

The model comparison results demonstrated that active inference provided a better performance to fit participants’ behavioral data compared to the basic model-free reinforcement learning and model-based reinforcement learning (Figure 4 (c)). Notably, the active inference could better capture the participants’ exploratory inclinations [37, 38]. This was evident in our experimental observations (Figure 4 (b)) where participants significantly favored asking the ranger over opting to stay. Asking the ranger, which provided environmental information, emerged as a more beneficial policy within the context of this task.

Moreover, participants’ preferences for information gain (i.e., epistemic value) were found to vary depending on the context. When participants lacked information about the context and the risky path had the same average rewards as the safe path but with greater variability, they showed an equal preference for both options (Figure 4 (b), “Not ask").

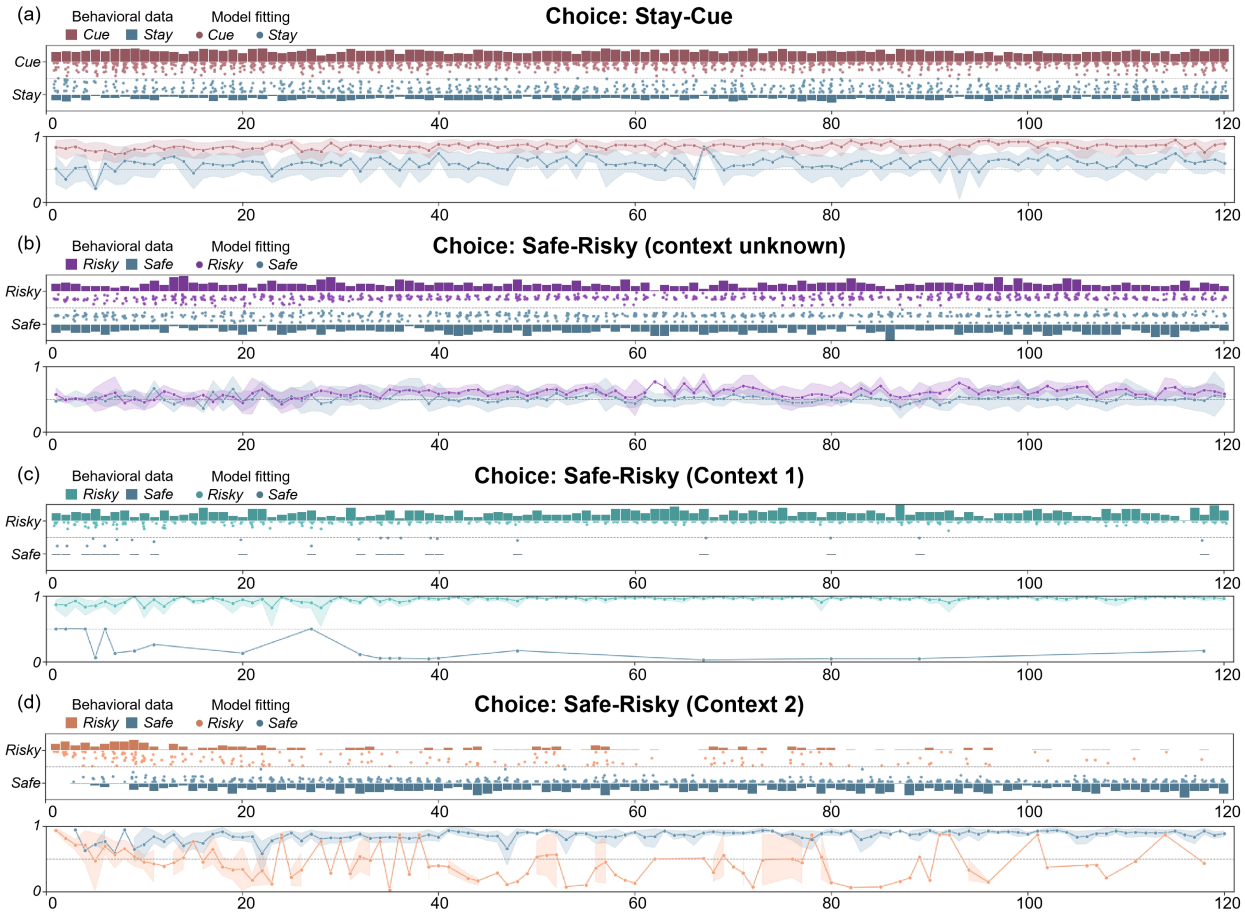

Figure 5: The comparison between the active inference model and the behavioral data in (a) the “First choice" stage, and the “Second choice" stage ((b) context unknown, (c) “Context 1", and (d) “Context 2"). The bar graphs show participants’ behavior data in each trial, and the height shows the proportion of participants who chose a certain option in each trial. The scatter plots show the model’s fitting results for the two choices of the participants. The closer the point is to the bar graph on both sides, the higher the fitting accuracy. The line graphs show the trend of the model fitting accuracies with the trials.

However, in “Context 1" (Figure 4 (b), high-reward context), where the risky path offered greater rewards than the safe path, participants strongly favored the risky path, which not only provided higher rewards but also had more epistemic value. In contrast, in “Context 2" (Figure 4 (b), low-reward context), where the risky path had fewer rewards than the safe path, participants mostly chose the safe path but occasionally opted for the risky path, recognizing that despite its fewer rewards, it offered epistemic value.

Figure 5 shows the comparison between the active inference model and the behavioral data, where we can see that the model can fit the participants behavioral strategies well. In the “Stay-Cue" choice, participants always tend to choose to ask the ranger and rarely choose not to ask. When the context was unknown, participants chose the “Safe" option or the “Risky" option very randomly, and they did not show any aversion to variability. When given “Context 1", where the “Risky" option gave participants a high average reward, participants almost exclusively chose the “Risky" option, which provided more information in the early trials and was found to provide more rewards in the later rounds. When given “Context 2", where the “Risky" option gave participants a low average reward, participants initially chose the “Risky" option and then tended to choose the “Safe" option. We can see that participants still occasionally chose the “Risky" option in the later trials of the experiment, which the model does not capture. This may be due to the influence of forgetting. Participants chose the “Risky" option again to establish an estimate of the reward distribution.

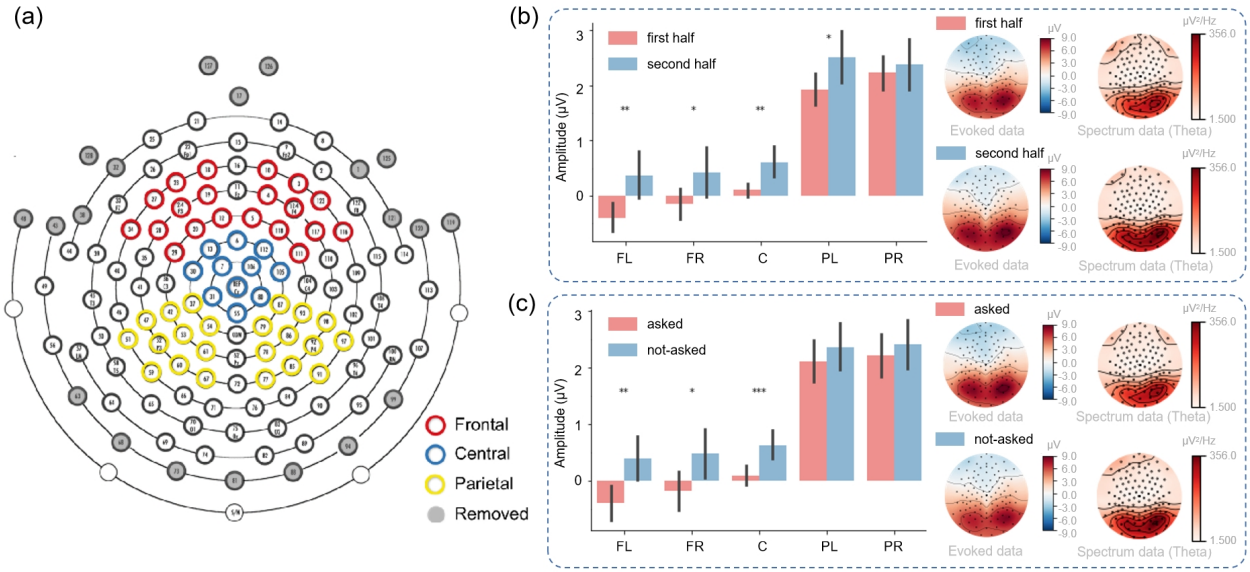

Figure 6: EEG results at the sensor level. (a) The electrode distribution. (b) The signal amplitude of different brain regions in the first and second half of the experiment in the “Second choice" stage. The error bar indicates the amplitude variance in each region. The right panel shows the visualization of the evoked data and spectrum data. (c) The signal amplitude of different brain areas in the “Second choice" stage where participants know the context or do not know the context of the right path. The error bar indicates the amplitude variance in each region. The right panel shows the visualization of the evoked data and spectrum data. FL: frontal-left; FR:frontal-right; C: central; PL: parietal-left; PR: parietal-right.

3.2 EEG results at sensor level

As depicted in Figure 6 (a), we divided electrodes into five clusters: left frontal, right frontal, central, left parietal, and right parietal. Within the “Second choice" stage, participants were required to make decisions under varying degrees of uncertainty (the uncertainty about the hidden states and the uncertainty about the model parameters). Thus, we investigated whether distinct brain regions exhibited different responses under such uncertainty.

In the first half of the experimental trials, participants would have greater uncertainty about model parameters compared to the latter half of the trials [8]. We thus analyzed data from the first half and latter half trials and identified statistically significant differences in the signal amplitude of the left frontal region ($p < 0.01$), the right frontal region ($p < 0.05$), the central region ($p < 0.01$), and the left parietal region ($p < 0.05$), suggesting a role for these areas in encoding the statistical structure of the environment (Figure 6 (b)). We postulated that when participants had constructed the statistical model of the environment during the second half of the trials, brains could effectively utilize the statistical model to make more confident decisions and exhibit greater neural responses.

To investigate whether distinct brain regions exhibited differential responses under the uncertainty about the hidden states, we divided all trials into two groups: the "asked" trials and the “not-asked" trials based on whether participants chose to ask in the “First choice" stage. In the *not-asked* (Figure 6 (c)), participants had greater uncertainty about the hidden states of the environment compared to the *asked trials*. We identified statistically significant differences in the signal amplitude of the left frontal region ($p < 0.01$), the right frontal region ($p < 0.05$), and the central region ($p < 0.001$), suggesting a role for these areas in encoding the hidden states of the environment. It might suggest that when participants knew the hidden states, they could effectively integrate the information with the environmental statistical structure to make more precise or confident decisions and exhibit greater neural response. The right panel of Figure 6 (c) revealed a higher signal in the theta band during not-asked trials, suggesting a correlation between theta band signal and uncertainty about the hidden states [39].

3.3 EEG results at source level

In the final analysis of the neural correlates of the decision-making process, as quantified by the epistemic and intrinsic values of expected free energy, we presented a series of linear regressions in source space. These analyses tested for

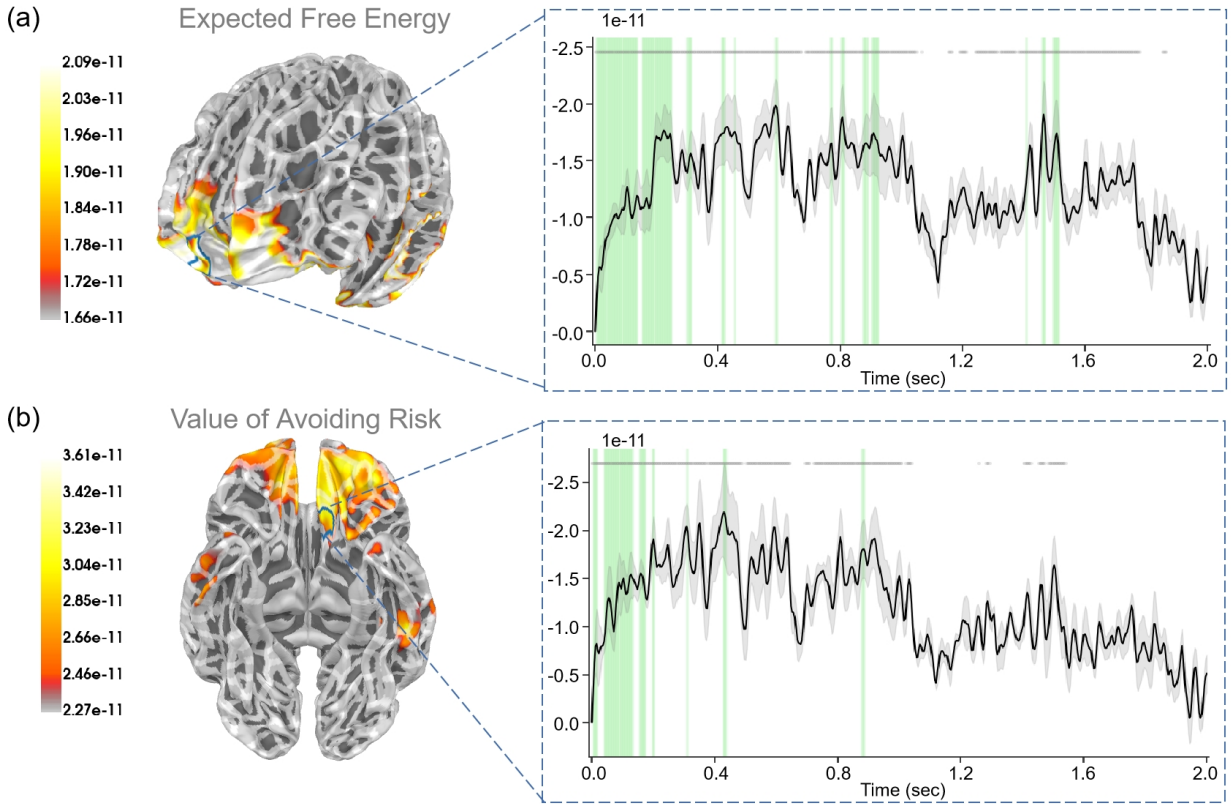

Figure 7: The source estimation results of expected free energy and active inference in the “First choice" stage. (A) The regression intensity (β) of expected free energy. The right panel indicates the regression intensity between the frontal pole (1, right half) and the expected free energy. The green-shaded regions indicate $p < 0.05$ after FDR correction (the average $t - value$ during these significant periods equals -3.228). (B) The regression intensity (β) of the value of reducing variability. The right panel indicates the regression intensity between the medial orbitofrontal cortex (5, left half) and the value of reducing variability. The green-shaded regions indicate $p < 0.05$ after FDR correction (the average $t - value$ during these significant periods equals -3.081). The black lines indicate the average intensities and the gray-shaded regions indicate the ranges of variations. The gray lines indicate $p < 0.05$ before FDR.

correlations over trials between constituent terms in expected free energy (the value of reducing variability, the value of reducing novelty, extrinsic value, and expected free energy itself) and neural responses in source space. Additionally, we also investigated the neural correlate of (the degree of) variability, (the degree of) novelty, and prediction error. Because we were dealing with a two-second time series, we were able to identify the periods of time during decision-making when the correlates were expressed. The linear regression was run by the “mne.stats.linear regression" function in the MNE package ($Activity \sim Regressor + Intercept$). $Activity$ is the activity amplitude of the EEG signal in the source space and $regressor$ is one of the regressors that we mentioned (e.g., expected free energy, the value of reducing novelty, etc.).

In these analyses, we focused on the induced power of neural activity at each time point, in the brain source space. To illustrate the functional specialization of these neural correlates, we presented whole-brain maps of correlation coefficients and picked out the brain region with the most significant correlation for reporting fluctuations in selected correlations over two-second periods. These analyses were presented in a descriptive fashion to highlight the nature and variety of the neural correlates, which we unpacked in relation to the existing EEG literature in the discussion. The significant results after false discovery rate (FDR) [40, 41] correction were shown in shaded regions. Additional regression results can be found in Supplementary Materials.

3.3.1 "First choice" stage – action selection

During the "First choice" stage, participants were presented with the choice of either choosing to stay or ask the ranger to get information regarding the present context of the risky path, the latter choice coming at a cost. Here, we examined "expected free energy" ($G(\pi, \tau)$, Eq.(9)), "value of reducing variability" ($AI \cdot E_{\widetilde{Q}}[\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)]$) and "extrinsic value" ($EX \cdot E_{\widetilde{Q}}[\ln P(o_\tau)$).

We found a robust correlation ($p < 0.05$) between the "expected free energy" regressor and the frontal pole (Figure 7 (a)). In addition, the superior temporal gyrus also displayed strong correlations with expected free energy. With respect to the "value of reducing variability" regressor, we identified a strong correlation ($p < 0.05$) with the medial orbitofrontal cortex (Figure 7 (b)). In addition, the postcentral gyrus and precentral gyrus also displayed strong correlations with the value of reducing variability. For the "extrinsic value" regressor, we observed a strong correlation ($p < 0.05$) with the middle temporal gyrus (see Figure S1(a)). In addition, the inferior temporal gyrus, and superior temporal gyrus also exhibited strong correlations with extrinsic value. Interestingly, we observed that during the "First choice" stage, expected free energy and extrinsic value regressors were both strongly correlated. However, expected free energy correlations appeared later than those of extrinsic value at the beginning, suggesting that the brain initially encoded reward values before integrating these values with information values for decision-making.

3.3.2 "First result" stage – belief update

During the "First result" stage, participants were presented with the outcome of their first choice, which informed them of the current context: either "Context 1" or "Context 2" for the risky path, or no additional information if they opted not to ask. This process correlated with the "reducing variability" regressor, as it corresponded to resolving uncertainties about hidden states. We assumed that the brain learning hidden states (reducing variability) corresponded to the value of reducing variability. Thus, the "reducing variability" regressor could be $AI \cdot (\ln Q(s_t|\pi) - \ln Q(s_t|o_t, \pi))$.

For "reducing variability", we observed a robust correlation ($p < 0.05$) within the medial orbitofrontal cortex (Figure 8 (a)). In addition, the rostral middle frontal gyrus, the lateral orbitofrontal cortex, and the superior temporal gyrus also displayed strong correlations with reducing variability.

3.3.3 "Second choice" stage – action selection

During the "Second choice" stage, participants chose between the risky path and the safe path based on the current information, with the aim of maximizing rewards. This required a balance between exploration and exploitation. Here, we examined "expected free energy" ($G(\pi, \tau)$, Eq.(9)), "value of reducing novelty" ($AL \cdot E_{\widetilde{Q}}[\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)]$), "extrinsic value" ($EX \cdot E_{\widetilde{Q}}[\ln P(o_\tau)]$), and "novelty" ($E_{\widetilde{Q}}[\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)]$).

For "expected free energy" (Figure 9(a), we identified strong correlations ($p < 0.001$) in the rostral middle frontal gyrus. In addition, the caudal middle frontal gyrus, middle temporal gyrus, pars triangularis, and superior temporal gyrus also displayed strong correlations with expected free energy. Regarding the "value of reducing novelty", we found that the rostral middle frontal gyrus showed strong correlations ($p < 0.05$). In addition, the superior frontal gyrus, insula, and lateral orbitofrontal cortex also displayed strong correlations with the value of reducing novelty. For "extrinsic value", strong correlations ($p < 0.001$) were evident in the rostral middle frontal gyrus (see Figure S1(b)). In addition, the middle temporal gyrus, pars opercularis, and precentral gyrus also displayed strong correlations with extrinsic values. In the "Second choice" stage, participants made choices under different degrees of novelty. For "the degree of novelty", we found no significant correlations after FDR correction (see Figure S3). Generally, the correlations between regressors and brain signals were more pronounced in the "Second choice" stage compared to the "First choice" stage.

3.3.4 "Second result" stage – belief update

During the "Second result" stage, participants obtained specific rewards based on their second choice: selecting the safe path yields a fixed reward, whereas choosing the risky path results in variable rewards, contingent upon the context. Here we examined "extrinsic value" (r_t), "prediction error" ($r_t - E_{\widetilde{Q}}[\ln P(o_\tau)]$), and "reducing novelty" ($\ln Q(A) - \ln P(A|s_t, o_t, \pi)$). Here, we also assumed that learning model parameters (reducing novelty) corresponded to the value of reducing novelty.

For "extrinsic value", we observed strong correlations ($p < 0.05$) in the lateral occipital gyrus, inferior parietal gyrus, and superior parietal gyrus (see Figure S2(a)). For "prediction error", we observed strong correlations ($p < 0.05$) in

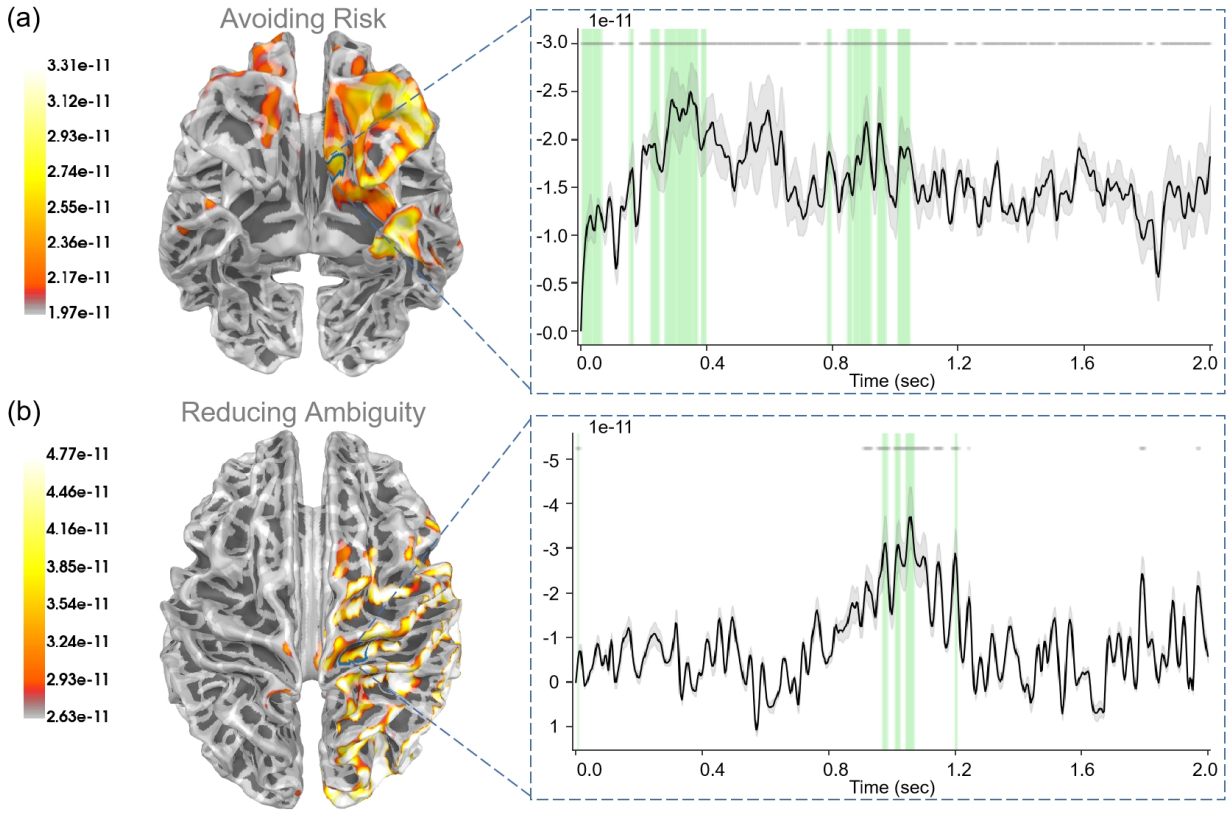

Figure 8: The source estimation results of reducing variability and reducing novelty in the two result stages. (A) The regression intensity (β) of reducing variability in the "First result" stage. The right panel indicates the regression intensity between the medial orbitofrontal cortex (5, left half) and reducing variability. The green-shaded regions indicate $p < 0.05$ after FDR correction (the average $t - value$ during these significant periods equals -3.001). (B) The regression intensity (β) of reducing novelty in the "Second result" stage. The right panel indicates the regression intensity between the precentral gyrus (15, right half) and reducing novelty. The green-shaded regions indicate $p < 0.05$ after FDR correction (the average $t - value$ during these significant periods equals -3.278). The black lines indicate the average intensities and the gray-shaded regions indicate the ranges of variations. The gray lines indicate $p < 0.05$ before FDR.

the bank of the superior temporal sulcus, inferior temporal gyrus, and lateral occipital gyrus(see Figure S2(b)). For "reducing novelty", we observed strong correlations ($p < 0.05$) in the precentral gyrus (see Figure 8).

4 Discussion

In this study, we utilized active inference to explore the neural correlates involved in the human decision-making process under novelty and variability. By employing a contextual two-bandit task, we demonstrated that the active inference framework effectively describes real-world decision-making. Our findings indicate that active inference not only provides explanations for decision-making under different kinds of uncertainty but also reveals the common and unique neural correlates associated with different types of uncertainties and decision-making policies. This was supported by evidence from both sensor-level and source-level EEG.

4.1 The varieties of human exploration strategies in active inference

In the diverse realm of human behavior, it has been observed that human exploration strategies vary significantly depending on the current situation. Such strategies can be viewed as a blend of *directed exploration*, where actions with higher levels of uncertainty are favored, and *random exploration*, where actions are chosen at random [42]. In the framework of active inference, the randomness in exploration is derived from the precision parameter employed

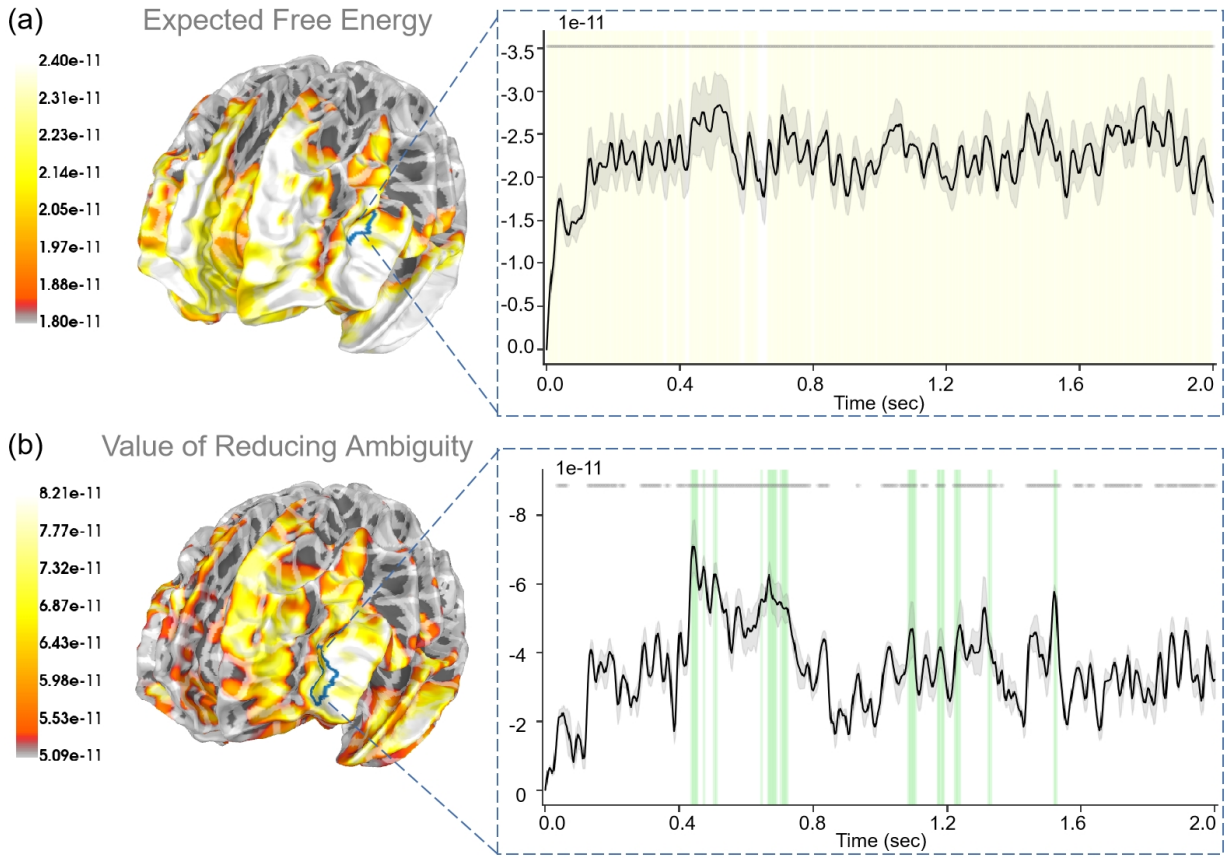

Figure 9: The source estimation results of expected free energy and the value of reducing novelty in the "Second choice" stage. (a) The regression intensity (β) of expected free energy. The right panel indicates the regression intensity between the rostral middle frontal gyrus (1, left half) and expected free energy, the black line indicates the average intensity of this region, and the gray-shaded region indicates the range of variation. The yellow-shaded regions indicate $p < 0.001$ after FDR (the average $t - value$ during these significant periods equals -4.819) and the gray lines indicate $p < 0.001$ before FDR. (b) The regression intensity (β) of the value of reducing novelty. The right panel indicates the regression intensity between the rostral middle frontal gyrus (6, left half) and the value of reducing novelty, the black line indicates the average intensity of this region, and the gray-shaded region indicates the range of variation. The green-shaded regions indicate $p < 0.05$ after FDR (the average $t - value$ during these significant periods equals -3.067) and the gray lines indicate $p < 0.05$ before FDR.

during policy selection. As the precision parameter increases, the randomness in agents' actions also increases. On the other hand, the directed exploration stems from the computation of expected free energy. Policies that lead to the exploration of more disambiguating options, hence yielding higher information gain, are assigned increased expected free energy by the model [3, 4, 11].

Our model-fitting results indicate that people show high variance in their exploration strategies (Figure 4 (b)). Exploration strategies, from a model-based perspective, incorporate a fusion of model-free learning and model-based learning. Intriguingly, these two learning ways exhibit both competition and cooperation within the human brain [43, 44]. The simplicity and effectiveness of model-free learning contrast with its inflexibility and data inefficiency. Conversely, model-based learning, although flexible and capable of forward planning, demands substantial cognitive resources. The active inference model tends to lean more toward model-based learning, as this model incorporates a cognitive model of the environment to guide the agent's actions. Our simulation results showed these model-based behaviors in which the agent constructs an environment model and uses the model to maximize rewards (Figure 3). Active inference can integrate model-free learning through adding a habitual term [3]. This allows the active inference agent to exploit the cognitive model (model-based) for planning in the initial task stages and utilize habits for increased accuracy and efficiency in later stages.

4.2 The strength of the active inference framework in decision-making

Active inference is a comprehensive framework elucidating neurocognitive processes (Figure 1). It unifies perception, decision-making, and learning within a single framework centered around the minimization of free energy. One of the primary strengths of the active inference model lies in its robust statistical [45] and neuroscientific underpinnings [46], allowing for a lucid understanding of an agent's interaction within its environment.

Active inference offers a superior exploration mechanism compared with basic model-free reinforcement learning (Figure 4 (c)). Since traditional reinforcement learning models determine their policies solely on the state, this setting leads to difficulty in extracting temporal information [47] and increases the likelihood of entrapment within local minima. In contrast, the policies in active inference are determined by both time and state. This dependence on time [48] enables policies to adapt efficiently, such as emphasizing exploration in the initial stages and exploitation later on. Moreover, this mechanism prompts more exploratory behavior in instances of state novelty. A further advantage of active inference lies in its adaptability to different task environments [4]. It can configure different generative models to address distinct tasks, and compute varied forms of free energy and expected free energy.

Despite these strengths, the active inference framework also has its limitations [49]. One notable limitation pertains to its computational complexity (Figure 2 (c)), resulting from its model-based architecture, restricting the traditional active inference model's application within continuous state-action spaces. Additionally, the model heavily relies on the selection of priors, meaning that poorly chosen priors could adversely affect decision-making, learning, and other processes [8]. However, sometimes it's just the opposite. As illustrated in the model comparison, priors can be a strength of Bayesian approaches. Under the complete class theorem [50, 51], any pair of behavioral data and reward functions can be described in terms of ideal Bayesian decision-making with particular priors. In other words, there always exists a description of behavioral data in terms of some priors. This means that one can, in principle, characterize any given behavioral data in terms of the priors that explain that behavior. In our example, these were effectively priors over the precision of various preferences or beliefs about contingencies that underwrite expected free energy.

4.3 Representing uncertainties at the sensor level

The employment of EEG signals in decision-making processes under uncertainty has largely concentrated on event-related potential (ERP) and spectral features at the sensor level [52–55]. In our study, the sensor level results reveal greater neural responses in multiple brain regions during the second half trials compared to the first half, and similarly, during not-asked trials as opposed to asked trials (Figure 6).

In our setting, after the first half of the trials, participants had learned some information about the environmental statistical structure, thus experiencing less novelty in the latter half of the trials. This increased understanding enabled them to better utilize the statistical structure for decision-making than they did in the first half of the trials. In contrast, during the not-asked trials, the lack of knowledge of the environment's hidden states led to higher-variability actions. This elevated variability was reflected in increased positive brain activities.

Novelty and variability, two pivotal factors in decision-making, are often misinterpreted and can vary in meaning depending on the context. Regarding the sensor level results, we find an overall greater neural response for the second half of the trials than the first half of the trials (Figure 6 (b)). It may indicate a generally greater neural response for the lower novelty trials, which may contrast with previous studies showing greater neural response for higher novelty trials in previous studies [55, 56]. For example, a late positive potential (LPP) was identified in their work, which differentiated levels of novelty, with the amplitude of the LPP serving as an index for perceptual novelty levels. However, the novelty in their task was defined as the perceptual difficulty of distinguishing, while our definition of novelty corresponds to the information gained from certain policies. Furthermore, Zheng et al. [57] used a wheel-of-fortune task to examine the ERP and oscillatory correlations of neural feedback processing under conditions of variability and novelty. Their findings suggest that risky gambling enhanced cognitive control signals, as evidenced by theta oscillation. In contrast, ambiguous gambling heightened affective and motivational salience during feedback processing, as indicated by positive activity and delta oscillation. Future work may focus on this oscillation level analysis and reveal more evidence on it.

4.4 Representation of decision-making process in human brain

In our experiment, each stage corresponded to distinct phases of the decision-making process. Participants made decisions to optimize cumulative rewards based on current information about the environment during the two choice stages while acquiring information about the environment during the two result stages.

During the “First choice" stage, participants had to decide whether to pay an additional cost in exchange for information regarding the environment’s hidden states. Here, the epistemic value stemmed from resolving the uncertainty about the hidden states and reducing variability. The frontal pole appears to play a critical role in this process by combining extrinsic value with epistemic value, expected free energy, to guide decision-making (Figure 7). Our results also showed the medial orbitofrontal cortex, postcentral gyrus, and precentral gyrus were correlated with the value of reducing variability. Previous study [58] demonstrated that the frontal pole was strongly activated in the risky condition and the ambiguous condition during decision-making. Another study also demonstrated that the frontal pole played an important role in the interaction between beliefs (variability and novelty) and payoffs (gains and losses) [59].

As for the “First result" stage, participants learned about the environment’s hidden states and avoided risks in the environment. Our results indicated that the medial orbitofrontal cortex, rostral middle frontal gyrus, and lateral orbitofrontal cortex played a crucial role in both valuing the uncertainty about hidden states and learning information about these hidden states (Figure 8 (a)). A previous study [60] found that both the medial and lateral orbitofrontal cortex encoded variability and reward probability while the lateral orbitofrontal cortex played a dominant role in coding experienced value. Another study [61] indicated that the medial orbitofrontal cortex was related to risk-taking and risk-taking was driven by specific orbitofrontal cortex reward systems. Throughout the “First result" stage, participants are processing the state information relevant to the current trial. The orbitofrontal cortex is postulated to play a key role in processing this state information and employing it to construct an environmental model.

In the “Second choice" stage, participants chose between a safe path and a risky path based on their current information. When knowing the environment’s hidden states, participants tended to resolve the uncertainty about model parameters by opting for the risky path. Conversely, without knowledge of the hidden states, Participants leaned towards variability reduction by choosing the safe path. Expected free energy is also correlated with brain signals but in different regions, such as the rostral middle frontal gyrus, caudal middle frontal gyrus, and middle temporal gyrus. Our results also highlighted the significance of the rostral middle frontal gyrus, superior frontal gyrus, insula, and lateral orbitofrontal cortex, in evaluating the value of reducing novelty. These results suggest that some brain regions may evaluate both the value of reducing novelty and the value of reducing variability [62].

For the “Second result" stage, participants got rewards according to their actions, constructing the value function and the state transition function. Our results highlighted the role of the precentral gyrus and superior parietal gyrus in learning the state transition function and reducing novelty (Figure 8 (b)). Participants made their decisions in different contexts and there were multiple studies emphasizing the role of the superior parietal gyrus in uncertain decision-making [63–65].

In the two “choice" stages, we observed stronger correlations for the expected free energy compared to the extrinsic value, suggesting that the expected free energy could better represent the actual value of the brain used to guide actions [66]. Compared with the “First choice" stage, the correlations in the “Second choice" stage were more significant. This may indicate that the brain is activated more when making decisions for rewards than when making decisions for information. We found neural correlates for the value of reducing variability and the value of reducing novelty, but not the degree of variability and novelty (after FDR correction). Future work should design a task that highlights different degrees of variability and ambiguities. In the two result stages, the regression results of the “Second result" stage were not very reliable. This may be due to our discrete reward structure. Participants may not be good at remembering specific probabilities, but only the mean reward.

It should be acknowledged that our EEG-based regression results are somewhat unstable, and the brain regions with significant regression are inconsistent before and after FDR correction. In future work, we should collect more precise neural data to reduce this instability.

5 Conclusion

In the current study, we introduce the active inference framework to investigate the neural mechanisms underlying an exploration and exploitation decision-making task. Compared to model-free reinforcement learning, active inference provides a superior exploration bonus and offers a better fit to the participants’ behavioral data. Given that the behavioral task in our study only involved variables from a limited number of states and rewards, future research should strive to apply the active inference framework to more complex tasks. Specific brain regions may play key roles in balancing exploration and exploitation. The frontal pole and middle frontal gyrus were primarily involved in action selection (expected free energy). The precentral gyrus was mainly engaged in evaluating the value of reducing variability and the rostral middle frontal cortex was also engaged in evaluating the value of reducing novelty. Furthermore, the medial orbitofrontal cortex participated in learning the hidden states of the environment (reducing variability) and the precentral gyrus participated in learning the model parameters of the environment (reducing novelty). In essence, our findings suggest that active inference is capable of investigating human behaviors in decision-making under uncer-

tainty. Overall, this research presents evidence from both behavioral and neural perspectives that support the concept of active inference in decision-making processes. We also offer insights into the neural mechanisms involved in human decision-making under various forms of uncertainty.

Data and Code availability

All experiment codes and analysis codes are available at GitHub: `https://github.com/andlab-um/FreeEnergyEEG`.

Acknowledgements

This work was mainly supported by the Science and Technology Development Fund (FDCT) of Macau[0127/2020/A3, 0041/2022/A], the Natural Science Foundation of Guangdong Province (2021A1515012509), Shenzhen-Hong Kong-Macao Science and Technology Innovation Project (Category C) (SGDX2020110309280100), MYRG of University of Macau (MYRG2022-00188-ICI), NSFC-FDCT Joint Program 0095/2022/AFJ, the SRG of University of Macau (SRG202000027-ICI), the National Key R&D Program of China (2021YFF1200804), National Natural Science Foundation of China (62001205), Shenzhen Science and Technology Innovation Committee (2022410129, KCXFZ2020122117340001), Guangdong Provincial Key Laboratory of Advanced Biomaterials (2022B1212010003).

Author contributions

S.Z, Q.L., and H.W. developed the study concept and designed the study; S.Z. and H.W. prepared experimental materials; Q.L. and H.W. supervised the experiments and analyses; S.Z. and Y. T. performed the data collection; S.Z. performed the data analyses; all authors drafted, revised and reviewed the manuscript and approved the final manuscript for submission.

Competing interests

The authors declare no competing interests.

References

[1] K. Friston, "The free-energy principle: a unified brain theory?" *Nature reviews neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[2] K. J. Friston and K. E. Stephan, "Free-energy and the brain," *Synthese*, vol. 159, pp. 417–458, 2007.

[3] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, G. Pezzulo *et al.*, "Active inference and learning," *Neuroscience & Biobehavioral Reviews*, vol. 68, pp. 862–879, 2016.

[4] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo, "Active inference: a process theory," *Neural computation*, vol. 29, no. 1, pp. 1–49, 2017.

[5] M. Kirchhoff, T. Parr, E. Palacios, K. Friston, and J. Kiverstein, "The markov blankets of life: autonomy, active inference and the free energy principle," *Journal of The royal society interface*, vol. 15, no. 138, p. 20170792, 2018.

[6] T. Parr, G. Pezzulo, and K. J. Friston, *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press, 2022.

[7] K. J. Friston, J. Daunizeau, and S. J. Kiebel, "Reinforcement learning or active inference?" *PloS one*, vol. 4, no. 7, p. e6421, 2009.

[8] P. Schwartenbeck, J. Passecker, T. U. Hauser, T. H. FitzGerald, M. Kronbichler, and K. J. Friston, "Computational mechanisms of curiosity and goal-directed exploration," *Elife*, vol. 8, 2019.

[9] C. A. O'Reilly III and M. L. Tushman, "Organizational ambidexterity in action: How managers explore and exploit," *California management review*, vol. 53, no. 4, pp. 5–22, 2011.

[10] R. C. Wilson, A. Geana, J. M. White, E. A. Ludvig, and J. D. Cohen, "Humans use directed and random exploration to solve the explore–exploit dilemma." *Journal of Experimental Psychology: General*, vol. 143, no. 6, p. 2074, 2014.

[11] S. J. Gershman, "Uncertainty and exploration." *Decision*, vol. 6, no. 3, p. 277, 2019.

[12] N. D. Daw, J. P. O'doherty, P. Dayan, B. Seymour, and R. J. Dolan, "Cortical substrates for exploratory decisions in humans," *Nature*, vol. 441, no. 7095, pp. 876–879, 2006.

[13] D. Badre, B. B. Doll, N. M. Long, and M. J. Frank, "Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration," *Neuron*, vol. 73, no. 3, pp. 595–607, 2012.

[14] J. F. Cavanagh, C. M. Figueroa, M. X. Cohen, and M. J. Frank, "Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation," *Cerebral cortex*, vol. 22, no. 11, pp. 2575–2586, 2012.

[15] E. Payzan-LeNestour, S. Dunne, P. Bossaerts, and J. P. ODoherty, "The neural representation of unexpected uncertainty during value-based decision making," *Neuron*, vol. 79, no. 1, pp. 191–201, 2013.

[16] S. Guha, K. Munagala, and P. Shi, "Approximation algorithms for restless bandit problems," *Journal of the ACM (JACM)*, vol. 58, no. 1, pp. 1–50, 2010.

[17] M. S. Tomov, V. Q. Truong, R. A. Hundia, and S. J. Gershman, "Dissociable neural correlates of uncertainty underlie different exploration strategies," *Nature Communications*, vol. 11, no. 1, 2020.

[18] H. Fan, S. J. Gershman, and E. A. Phelps, "Trait somatic anxiety is associated with reduced directed exploration and underestimation of uncertainty," *Nature Human Behaviour*, pp. 1–12, 2022.

[19] A. G. Collins and M. J. Frank, "Within-and across-trial dynamics of human eeg reveal cooperative interplay between reinforcement learning and working memory," *Proceedings of the National Academy of Sciences*, vol. 115, no. 10, pp. 2502–2507, 2018.

[20] E. Schulz, E. Konstantinidis, and M. Speekenbrink, "Exploration-exploitation in a contextual multi-armed bandit task," in *International conference on cognitive modeling*, 2015, pp. 118–123.

[21] A. G. Collins and M. J. Frank, "How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis," *European Journal of Neuroscience*, vol. 35, no. 7, pp. 1024–1035, 2012.

[22] I. Levy, J. Snell, A. J. Nelson, A. Rustichini, and P. W. Glimcher, "Neural representation of subjective value under risk and ambiguity," *Journal of neurophysiology*, vol. 103, no. 2, pp. 1036–1047, 2010.

[23] M. Galdo, G. Bahg, and B. M. Turner, "Variational bayesian methods for cognitive science." *Psychological methods*, vol. 25, no. 5, p. 535, 2020.

[24] K. Friston, "Active inference and free energy," *Behavioral and brain sciences*, vol. 36, no. 3, p. 212, 2013.

[25] C. L. Buckley, C. S. Kim, S. McGregor, and A. K. Seth, "The free energy principle for action and perception: A mathematical review," *Journal of Mathematical Psychology*, vol. 81, pp. 55–79, 2017.

[26] P.-Y. Oudeyer and F. Kaplan, "What is intrinsic motivation? a typology of computational approaches," *Frontiers in neurorobotics*, vol. 1, p. 108, 2007.

[27] J. Schmidhuber, "Formal theory of creativity, fun, and intrinsic motivation (1990–2010)," *IEEE transactions on autonomous mental development*, vol. 2, no. 3, pp. 230–247, 2010.

[28] A. Barto, M. Mirolli, and G. Baldassarre, "Novelty or surprise?" *Frontiers in psychology*, vol. 4, p. 61898, 2013.

[29] T. H. FitzGerald, P. Schwartenbeck, M. Moutoussis, R. J. Dolan, and K. Friston, "Active inference, evidence accumulation, and the urn task," *Neural computation*, vol. 27, no. 2, pp. 306–328, 2015.

[30] T. Lu, D. Pál, and M. Pál, "Contextual multi-armed bandits," in *Proceedings of the Thirteenth international conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 485–492.

[31] R. Martínez-Cancino, A. Delorme, D. Truong, F. Artoni, K. Kreutz-Delgado, S. Sivagnanam, K. Yoshimoto, A. Majumdar, and S. Makeig, "The open eeglab portal interface: High-performance computing with eeglab," *NeuroImage*, vol. 224, p. 116778, 2021.

[32] L. Esch, C. Dinh, E. Larson, D. Engemann, M. Jas, S. Khan, A. Gramfort, and M. S. Hämäläinen, "Mne: software for acquiring, processing, and visualizing meg/eeg data," *Magnetoencephalography: From Signals to Dynamic Cortical Networks*, pp. 355–371, 2019.

[33] R. D. Pascual-Marqui, "Discrete, 3d distributed, linear imaging methods of electric neuronal activity. part 1: exact, zero error localization," *arXiv preprint arXiv:0710.3341*, 2007.

[34] S. Khan, J. A. Hashmi, F. Mamashli, K. Michmizos, M. G. Kitzbichler, H. Bharadwaj, Y. Bekhti, S. Ganesan, K.-L. A. Garel, S. Whitfield-Gabrieli *et al.*, "Maturation trajectories of cortical resting-state networks depend on the mediating frequency band," *NeuroImage*, vol. 174, pp. 57–68, 2018.

[35] S. I. Vrieze, "Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic)." *Psychological methods*, vol. 17, no. 2, p. 228, 2012.

[36] P. I. Frazier, "A tutorial on bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

[37] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[38] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo, "Active inference and epistemic value," *Cognitive neuroscience*, vol. 6, no. 4, pp. 187–214, 2015.

[39] J. Harper, S. M. Malone, and W. G. Iacono, "Theta-and delta-band eeg network dynamics during a novelty oddball task," *Psychophysiology*, vol. 54, no. 11, pp. 1590–1605, 2017.

[40] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[41] S. J. Gershman, D. M. Blei, K. A. Norman, and P. B. Sederberg, "Decomposing spatiotemporal brain patterns into topographic latent sources," *NeuroImage*, vol. 98, pp. 91–102, 2014.

[42] S. J. Gershman, "Deconstructing the human algorithms for exploration," *Cognition*, vol. 173, pp. 34–42, 2018.

[43] J. Gläscher, N. Daw, P. Dayan, and J. P. O'Doherty, "States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning," *Neuron*, vol. 66, no. 4, pp. 585–595, 2010.

[44] N. D. Daw, Y. Niv, and P. Dayan, "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control," *Nature neuroscience*, vol. 8, no. 12, pp. 1704–1711, 2005.

[45] G. E. Crooks, "Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems," *Journal of Statistical Physics*, vol. 90, pp. 1481–1487, 1998.

[46] K. Lehmann, D. Bolis, M. J. Ramstead, K. Friston, and P. Kanske, "An active inference approach to second-person neuroscience," 2022.

[47] M. Laskin, K. Lee, A. Stooke, L. Pinto, P. Abbeel, and A. Srinivas, "Reinforcement learning with augmented data," *Advances in neural information processing systems*, vol. 33, pp. 19 884–19 895, 2020.

[48] J. X. Wang, Z. Kurth-Nelson, D. Tirumala, H. Soyer, J. Z. Leibo, R. Munos, C. Blundell, D. Kumaran, and M. Botvinick, "Learning to reinforcement learn," *arXiv preprint arXiv:1611.05763*, 2016.

[49] V. Raja, D. Valluri, E. Baggs, A. Chemero, and M. L. Anderson, "The markov blanket trick: On the scope of the free energy principle and active inference," *Physics of Life Reviews*, vol. 39, pp. 49–72, 2021.

[50] A. Wald, "An essentially complete class of admissible decision functions," *The Annals of Mathematical Statistics*, pp. 549–555, 1947.

[51] L. D. Brown, "A complete class theorem for statistical problems with finite sample spaces," *The Annals of Statistics*, pp. 1289–1300, 1981.

[52] L. Wang, J. Zheng, S. Huang, and H. Sun, "P300 and decision making under risk and ambiguity," *Computational intelligence and neuroscience*, vol. 2015, pp. 1–1, 2015.

[53] Y. Lin, L. Duan, P. Xu, X. Li, R. Gu, and Y. Luo, "Electrophysiological indexes of option characteristic processing," *Psychophysiology*, vol. 56, no. 10, p. e13403, 2019.

[54] A. R. Bland and A. Schaefer, "Electrophysiological correlates of decision making under varying levels of uncertainty," *Brain research*, vol. 1417, pp. 55–66, 2011.

[55] C. Botelho, C. Fernandes, C. Campos, C. Seixas, R. Pasion, H. Garcez, F. Ferreira-Santos, F. Barbosa, J. Maques-Teixeira, and T. O. Paiva, "Uncertainty deconstructed: conceptual analysis and state-of-the-art review of the erp correlates of risk and ambiguity in decision-making," *Cognitive, Affective, & Behavioral Neuroscience*, pp. 1–21, 2023.

[56] S. Sun, S. Zhen, Z. Fu, D.-A. Wu, S. Shimojo, R. Adolphs, R. Yu, and S. Wang, "Decision ambiguity is mediated by a late positive potential originating from cingulate cortex," *NeuroImage*, vol. 157, pp. 400–414, 2017.

[57] Y. Zheng, W. Yi, J. Cheng, and Q. Li, "Common and distinct electrophysiological correlates of feedback processing during risky and ambiguous decision making," *Neuropsychologia*, vol. 146, p. 107526, 2020.

[58] Z. Guo, J. Chen, S. Liu, Y. Li, B. Sun, and Z. Gao, "Brain areas activated by uncertain reward-based decision-making in healthy volunteers," *Neural regeneration research*, vol. 8, no. 35, pp. 3344–3352, 2013.

[59] K. Smith, J. Dickhaut, K. McCabe, and J. V. Pardo, "Neuronal substrates for choice under ambiguity, risk, gains, and losses," *Management science*, vol. 48, no. 6, pp. 711–718, 2002.

[60] Y. Li, G. Vanni-Mercier, J. Isnard, F. Mauguiere, and J.-C. Dreher, "The neural dynamics of reward value and risk coding in the human orbitofrontal cortex," *Brain*, vol. 139, no. 4, pp. 1295–1309, 2016.

[61] E. T. Rolls, Z. Wan, W. Cheng, and J. Feng, "Risk-taking in humans and the medial orbitofrontal cortex reward system," *NeuroImage*, vol. 249, p. 118893, 2022.

[62] A. L. Krain, A. M. Wilson, R. Arbuckle, F. X. Castellanos, and M. P. Milham, "Distinct neural mechanisms of risk and ambiguity: a meta-analysis of decision-making," *Neuroimage*, vol. 32, no. 1, pp. 477–484, 2006.

[63] M. P. Paulus, N. Hozack, B. Zauscher, J. E. McDowell, L. Frank, G. G. Brown, and D. L. Braff, "Prefrontal, parietal, and temporal cortex networks underlie decision-making in the presence of uncertainty," *Neuroimage*, vol. 13, no. 1, pp. 91–100, 2001.

[64] S. A. Huettel, A. W. Song, and G. McCarthy, "Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices," *Journal of Neuroscience*, vol. 25, no. 13, pp. 3304–3311, 2005.

[65] M. Ragni, I. Franzmeier, S. Maier, and M. Knauff, "Uncertain relational reasoning in the parietal cortex," *Brain and Cognition*, vol. 104, pp. 72–81, 2016.

[66] T. B. Williams, C. J. Burke, S. Nebe, K. Preuschoff, E. Fehr, and P. N. Tobler, "Testing models at the neural level reveals how the brain computes subjective value," *Proceedings of the National Academy of Sciences*, vol. 118, no. 43, p. e2106237118, 2021.