THE NEURAL CORRELATES OF AMBIGUITY AND RISK IN HUMAN DECISION-MAKING UNDER AN ACTIVE INFERENCE FRAMEWORK

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

September 18, 2023

ABSTRACT

Active inference integrates perception, decision-making, and learning into a united theoretical framework, providing an efficient way to trade off exploration and utilization by minimizing (expected) free energy. In this study, we asked how the brain represents values, uncertainty, and resolves the uncertainty under the active inference framework in the exploration-exploitation trade-off. 25 participants performed a contextual two-step two-armed bandit task, with electroencephalogram (EEG) recordings. By comparing the fitting results from the active inference and reinforcement learning model, we show that active inference can better capture the exploration instinct of humans, which helps resolve the uncertainty of the environment. The EEG sensor-level results show that the activity in the frontal, central, and parietal regions is associated with uncertainty, while activity in the frontal and central brain regions is associated with risk. The EEG source-level results indicate that the expected free energy is encoded in the lateral occipital cortex and the uncertainty in the middle temporal pole. Our study dissociates the expected free energy and the uncertainty in active inference theory and their neural correlates, suggesting the reliability 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 decision under different kinds of uncertainty.

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 uncertainty used to describe a system, which can be understood as the difference between the real system state and the estimated system state [2]. In addition, expected free energy can also 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 variational 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 belief 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*).

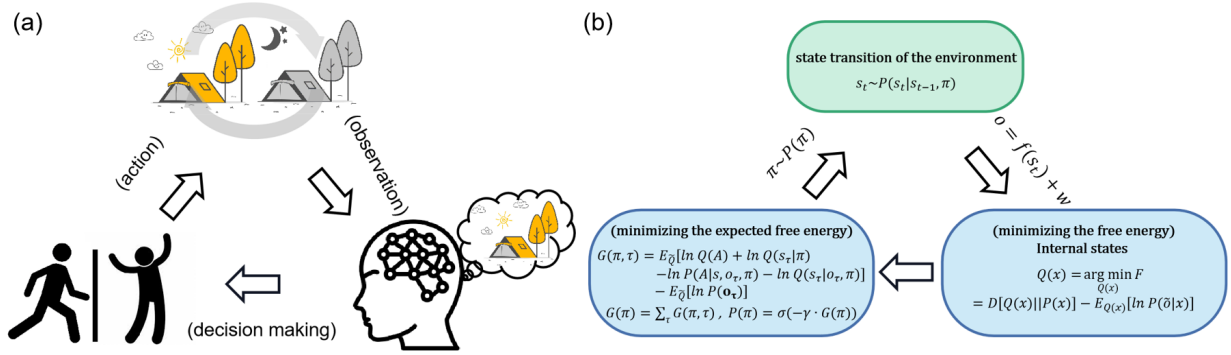

Figure 1: Active inference. (a) Qualitatively, agents receive observations from the environment and use these observations to optimize the internal cognitive model of the environment. Then agents actively sample the environment states by action (choose actions that would make them in more favorable states). The environment changes its state according to agents' actions and again agents receive new observations from the environment. (b) Quantificationally, agents optimize the internal cognitive model by minimizing the variational free energy. Then agents select policies (actions) by minimizing the expected free energy to minimize the surprise in the future.

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 active inference framework to describe complex cognitive processes, which provide theoretical insights into behavioral dynamics [4–7]. For instance, it is theoretically deduced in the active inference framework on 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 the action that offers extra information about the current environment, while *exploitation* is to take action to maximize the potential reward given the current belief. The exploration-exploitation trade-off captures an inherent tension and uncertainty, 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 levels.

The decision-making process frequently involves grappling with varying forms of uncertainty, such as ambiguity - the element of uncertainty that can be mitigated through sampling, risk - the inherent uncertainty presented by a stable environment, and unexpected uncertainty - the uncertainty pertaining to environmental changes. Studies have investigated these different forms of uncertainty in decision-making, focusing especially on their neural correlates [12–15]. However, it remains an open question whether the brain represents these different types of uncertainty distinctly [16](**Aim 1**). In addition to the representation of uncertainties, the brain may also encode the value of resolving these uncertainties [16](**Aim 2**). The active inference framework presents a theoretical approach to resolving these research gaps. Within this framework, ambiguity is represented by the information gain about model parameters associated with choosing a particular action, while risk is signified by the variance of hidden environmental states. In active inference, the representations of uncertainty naturally translate into representations of the value of reducing ambiguity or avoiding risk(see *Section 2.1*), which means that those representations may show common associated neural signatures [1].

Our study, therefore, aimed to determine how the human brain represents different uncertainties (**Q1**), and how humans encode the policies or value of resolving these uncertainties (**Q2**). To achieve these aims, we utilized the active inference framework to examine the exploration-exploitation trade-off, with behavioral data and electroencephalogram (EEG) neural recordings (see *Methods*). We designed a contextual two-armed bandit task (see *Section 2.2* and Figure 4 (a)), wherein participants were instructed to maximize cumulative rewards. They were offered various policies to either avoid risk, reduce ambiguity, or maximize immediate rewards (see *Methods*). With aims to address the two questions, 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 ambiguity and avoiding risk during action selection, and 4) updates the information about the environment during belief update (source-level EEG evidence)(see *Section 3.3*) .

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 environment states 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 true cognitive model encoded in the brain (minimizing the free energy)

2 Methods

2.1 The free energy principle and active inference

The free energy principle can sample different states of the environment by choosing different actions to obtain the sensory input we like that this process is termed as *active inference*. Under the active inference framework, the free energy can be viewed as the objective function of the system, i.e. the free energy to minimize. By minimizing free energy, we can optimize decisions and reduce uncertainty. In active inference, variational inference is used to estimate model parameters (minimize variational free energy), guide the agent's actions (minimize expected free energy), and compute an objective function by maximizing the expected log-likelihood (see Figure 1 (b)). This process can be viewed as an optimization problem that seeks to find the best model parameters and action strategy to maximize the log-likelihood. By minimizing the objective function, optimal model parameters can be estimated and better decisions can be made [17]. This principle bridges the sensory input, cognitive processes, and action output, enabling us to quantitatively describe the neural processes of learning about the environment. For example, the brain receives sensory input s from the environment, and the cognitive model encoded by the human brain $q(s)$ makes an inference on the cause of sensory input $p(s|o)$. In the free energy principle, minimizing free energy refers to minimizing the difference (*e.g.*, KL divergence) between the cognitive model ($p(s|o)$) encoded by our brain and the causes of the sensory input ($q(s)$). Thus, free energy is an information-theoretic quantity that constrains the evidence for the data model. Free energy can be minimized by the following two means [18]:

- 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 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 the states which are preferred. Minimizing free energy through action is one of the advantages of the free energy principle over Bayesian inference, which can only passively optimize cognition.

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]:

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

where o is the sensory inputs, s is the hidden states of the environment, π is the policy of the agent, A is the likelihood matrix mapping from hidden states to outcomes, B is the transition probability for hidden states under the action of a policy in time t, d is the prior expectation of each state at the beginning of each trial, γ is the inverse temperature pf beliefs about policies, β is the prior expectation of temperature of beliefs about policies, a is the concentration parameters of likelihood, $Cat()$ is the categorical distribution, $Dir()$ is the Dirichlet distribution, and $\Gamma()$ is the Gamma distribution.

The posterior probability of the corresponding hidden state 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 construe their environmental circumstances. 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 or action sequence. 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 probabilities incorporated within this model are parameterized using a Dirichlet distribution [19]. 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 outcomes into the generative model. Therefore, the generative model plays a pivotal role in stipulating the probabilities and uncertainties related to the potential states and outcomes.

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):

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

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):

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

The initial term, denoted as $D_{KL}(q(s)||p(s))$, is conventionally referred to as "complexity". This term, reflecting the divergence between $q(s)$ and $p(s)$, quantifies the volume of information intended to be encoded within $q(s)$ that is not inherent in $p(s)$. The subsequent term, $E_q[logp(o|s)]$, designated as "accuracy", represents the conditional probability of receiving an observation given each state.

The minimization of variational free energy facilitates a progressive alignment between the approximate posterior distribution of hidden states, as encoded within the brain's representation of the environment, and the actual posterior

distribution of the environment, conditioned on observed data. However, it is noteworthy that our policy beliefs are predominantly future-oriented. Our aspiration entails designing policies that possess the potential to effectively guide us toward achieving the future state that we desire. Thus, it follows that these policies should aim to minimize future free energy, or in other words, the expected free energy. 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:

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

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

In Eq. (7), it is important to note that we anticipate observations that have not yet transpired. Consequently, we designate $\widetilde{Q} = Q(o_\tau, s_\tau, A|\pi)$. Within the context of expected free energy, if we establish a relationship between $lnP(o_\tau)$ and preference, it enables us to express expected free energy in terms of cognitive 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.

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

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

Belief updates play a dual role by facilitating both inference and learning processes. The inference is here understood as the optimization of expectations concerning 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 mapping function A and the concentration parameter a:

$$\begin{aligned} A &= Cat(a) \\ a &= a + \alpha \sum_\tau o_\tau \otimes s_\tau \end{aligned} \tag{9}$$

where α is the learning rate.

2.2 Contextual two-armed bandit task

In this study, we developed a "contextual two-armed bandit task", which is based on the conventional multi-armed bandit task [20]. In this task, participants are instructed to explore two paths that offer rewards with the aim of maximizing them. One path provides constant rewards in each trial, labeled the "safe path," while the other, referred to as the "risky path," probabilistically offers varying amounts of rewards. The risky path has two distinct contexts, "Context 1" (high-reward context) and "Context 2" (low-reward context), each corresponding to different reward distributions.

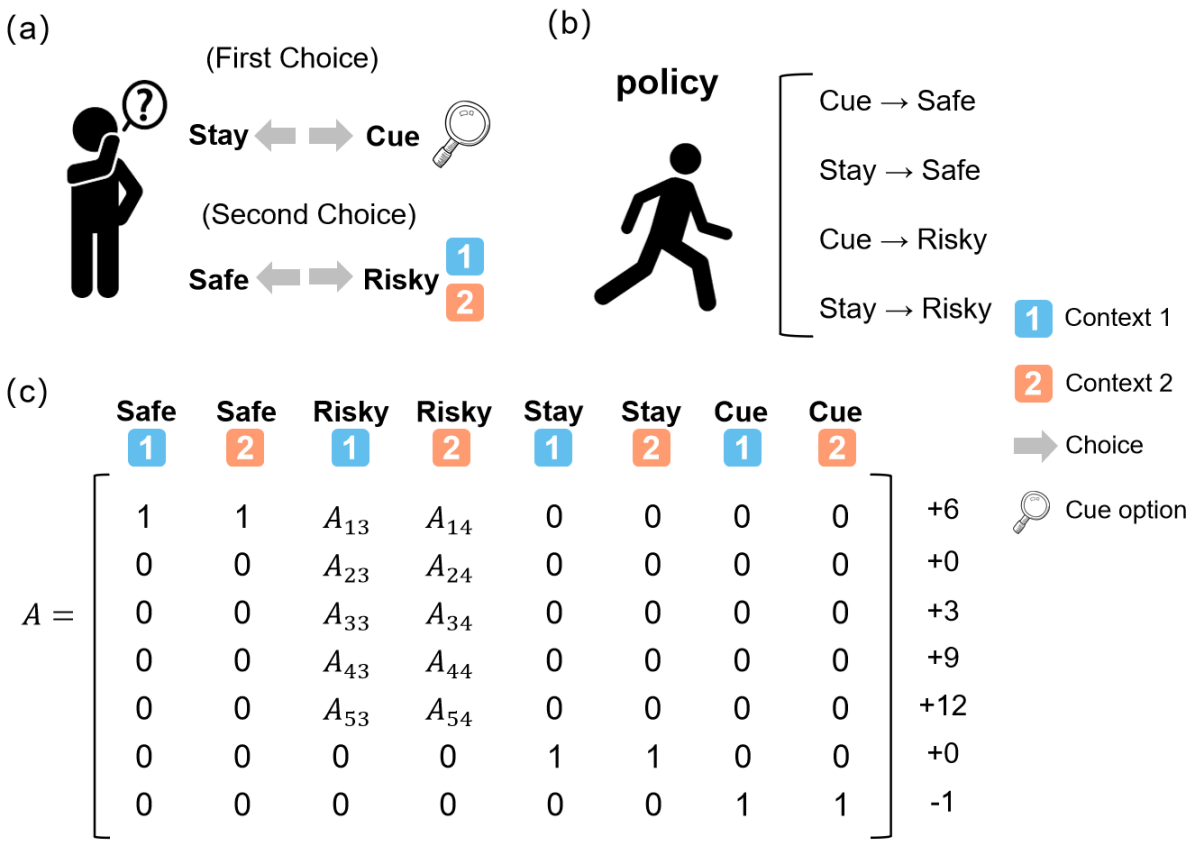

Figure 2: Generative model of the contextual two-armed bandit task. (a) There are 2 stages in this task. The first choice is: “Stay" and “Cue". The “Stay" option gives you nothing while the “Cue" option gives you a -1 reward and the context information about the “Risky" option in the current trial. The second choice is: “Safe" and “Risky". The “Safe" option 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 A-matrix maps from 8 hidden states (columns) to 7 observable outcomes (rows).

The context associated with the risky path undergoes random alternations in each trial, and participants can determine the specific context of the current trial’s risky path by accessing a cue, although this comes with a cost. Additionally, participants must discern and comprehend the reward distributions of both contexts. For a comprehensive overview of the specific parameter settings, please refer to Figure 2.

In the task, active inference agents with different parameter configurations can exhibit different decision-making policies, as demonstrated in a simulation experiment (see Figure 3). By adjusting parameters such as prior, learning rate, and precision, agents can operate under different policies. Agents with a low learning rate (and a relatively high proportion of epistemic value) will 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 is obtained, the agent will evaluate the actual values of various policies and select the optimal policy for exploitation. In the experimental setup, the optimal policy requires 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 (and a smaller proportion of epistemic value) may become trapped in a local optimum and consistently opt for the safe path, especially if the initial high-reward scenarios encountered yield minimal rewards.

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

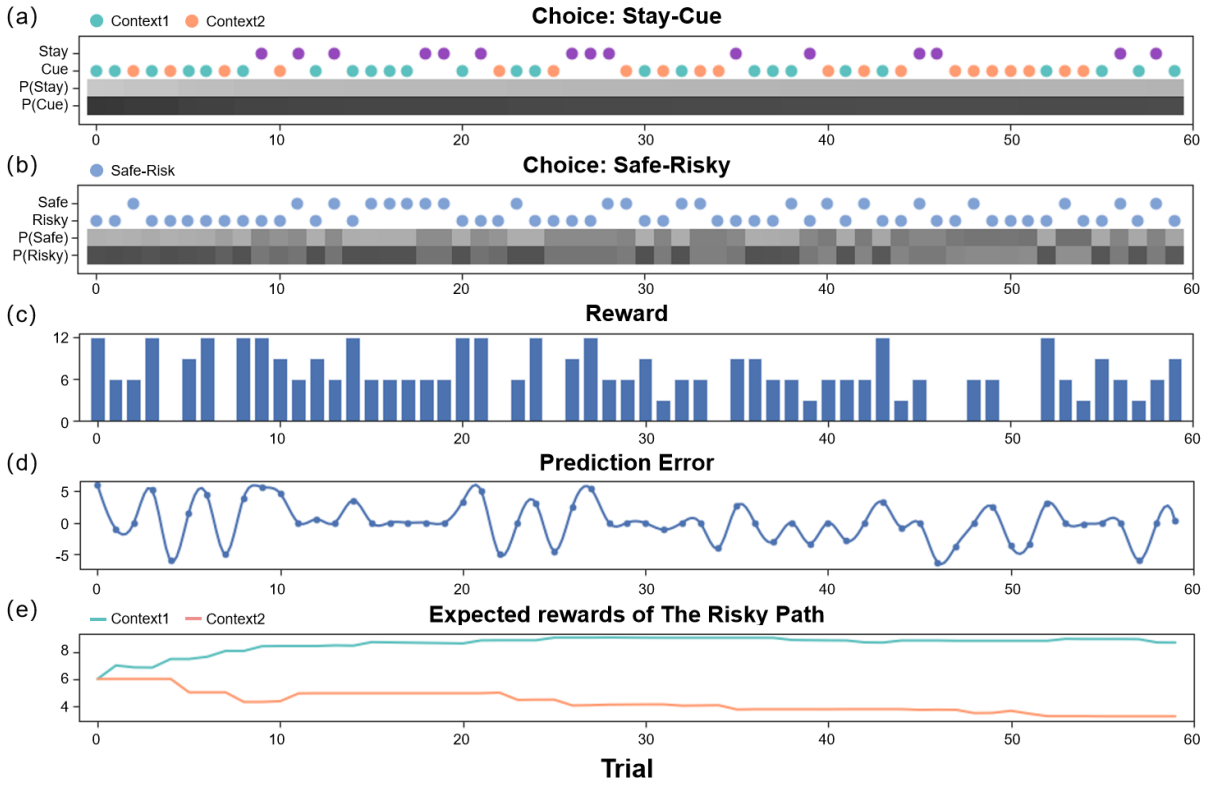

Figure 3: The simulation experiment results. This figure demonstrates how an agent selects actions and updates beliefs over 60 trials in the active inference framework. The first two panels (a-b) display the agent's policy and depict how the policy probabilities are updated (choosing between the stay 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 shadow represents the agent's confidence, with darker shadows 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.

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).

2.3.2 Data collection

In our experiment, to diversify the data, we incorporated an additional "you can ask" stage prior to the stay-cue option. This measure was implemented to ensure that each participant encountered trials under conditions of heightened uncertainty. Participants were presented with the following experimental scenario and 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 per exploration, but it includes 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 informed of the introduction which included basic information about the experiment and press the spacebar to proceed(e.g., the total number of apples collected being linked to the monetary reward they would receive). For each trial, the experimental procedure is illustrated in Figure 4 (a), and comprises five stages:

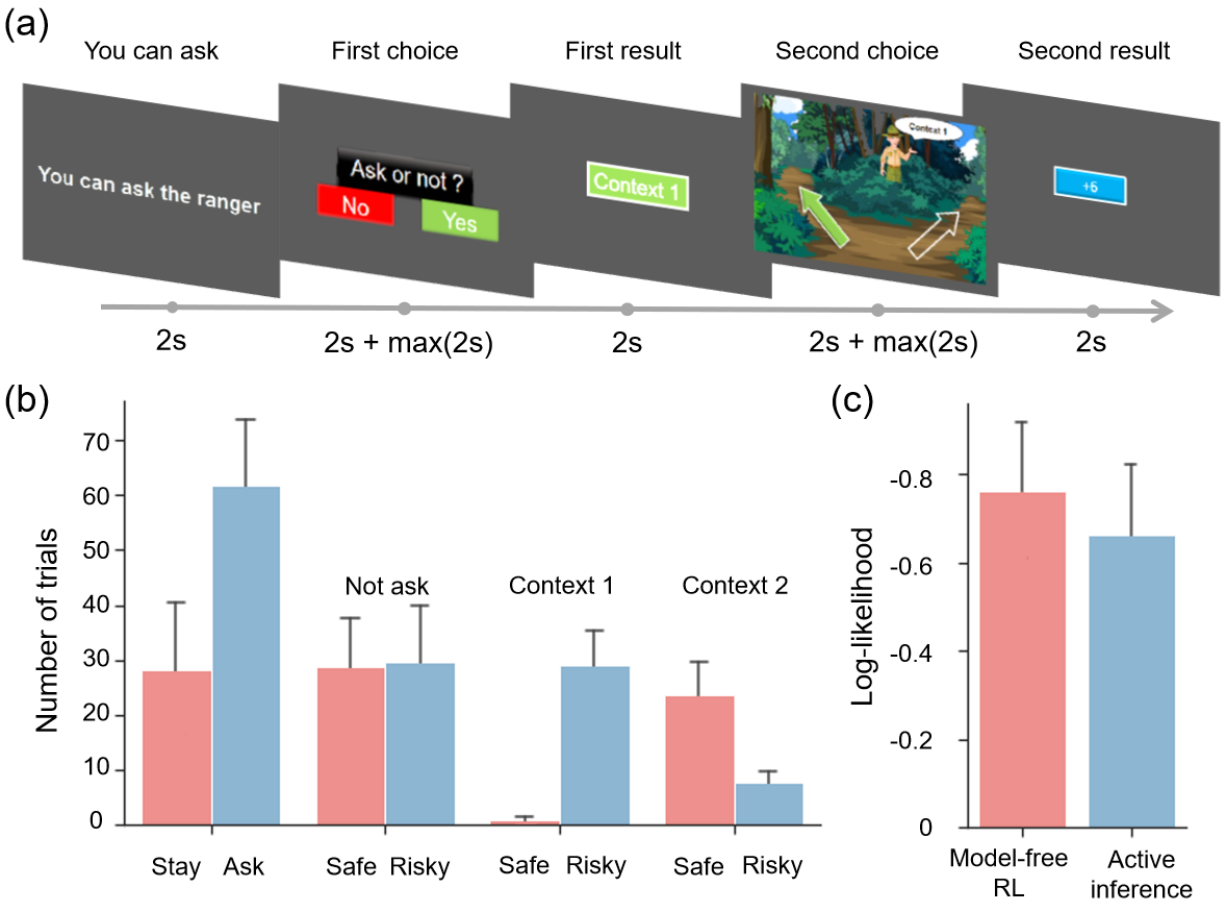

Figure 4: The experiment task and behavioral result. Panel (a) outlines the five stages of the experiment, which include the “You can ask" stage to determine if participants can request information from the ranger, the “First choice" stage to decide whether to ask the ranger for information, the “First 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. Panel (b) displays the number of times each option was selected. Finally, panel (c) compares model-free RL and active inference models.

1. “You can ask" stage: Participants are given the option to ask the ranger for information, which lasts for 2 seconds.

2. “First choice" stage: Participants must decide whether to press the right or left button to ask the ranger for information, at the cost of an apple. This stage also lasts 2 seconds and 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. 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 respective path. This stage again lasts for 2 seconds and 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. And this stage corresponds to the belief update in active inference.

Each stage is separated by a jitter ranging from 0.6 to 1.0 seconds. The entire experiment consists of a single block with a total of 120 trials.

2.3.3 EEG processing

The processing of EEG signals was conducted using the EEGLAB toolbox [21] in the Matlab and the MNE package [22]. The preprocessing of EEG data involved multiple steps, including data selection, downsampling, high- and low-

pass filtering, and independent component analysis (ICA) decomposition. Data segments encompassing a 2-second interval before and after the experiment were chosen. 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. To accomplish this, we employed the head model and source space available in the “fsaverage" of the MNE package. To localize the sources, we utilized eLORETA [23] and mapped the EEG data to the source space.

We segmented the data into five time intervals that corresponded to the five stages of the experiment. The first stage, known as the “You can ask" stage, involved identifying participants’ willingness to access cues. 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, focused on obtaining the results of cue access. The fourth stage, known as the “Second choice" stage, involved choosing between visiting the safe or risky path. Finally, the fifth stage, named the “Second result" stage, encompassed receiving rewards. Each interval lasted two, and this categorization allowed us to investigate brain activity responses to two distinct choices at different stages of the task. Specifically, we examined the processes of prediction (action selection) and outcome (belief update) within the framework of active inference.

3 Results

3.1 Behavioral results

The active inference framework was employed to fit the participants’ behavioral policies. To account for differing preferences regarding resolving uncertainty and rewards, we incorporated fitting coefficients into the three terms constituting the expected free energy (active learning, active inference, and extrinsic value). Subsequently, these fitting parameters were integrated into the active inference model, enabling the extraction of expected free energy, prediction error, and other variables for each trial. We then could perform linear regression to analyze the associated EEG signals for each brain region.

The model comparison results demonstrated that active inference provided a performance to fit participants’ behavioral data compared to the basic model-free reinforcement learning (Figure 4 (c)). Notably, the active inference captured the participants’ exploratory inclinations better compared to model-free RL [24,25]. This was evident in our experimental observations (Figure 4 (b)) where participants significantly favored consulting the ranger over opting to stay. Consulting the ranger, which provided environmental information, emerged as a more beneficial policy within the context of this task.

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

3.2 EEG results at sensor level

As depicted in Figure 5 (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 amidst 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 differential responses under such uncertainty.

Usually, in the first half of the experimental trials, participants would display 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 5 (b)). We postulate that when participants have constructed the statistical model of the environment during the second half of the trials, brains could effectively utilize the statistical model to make superior decisions and exhibit more positive activities.

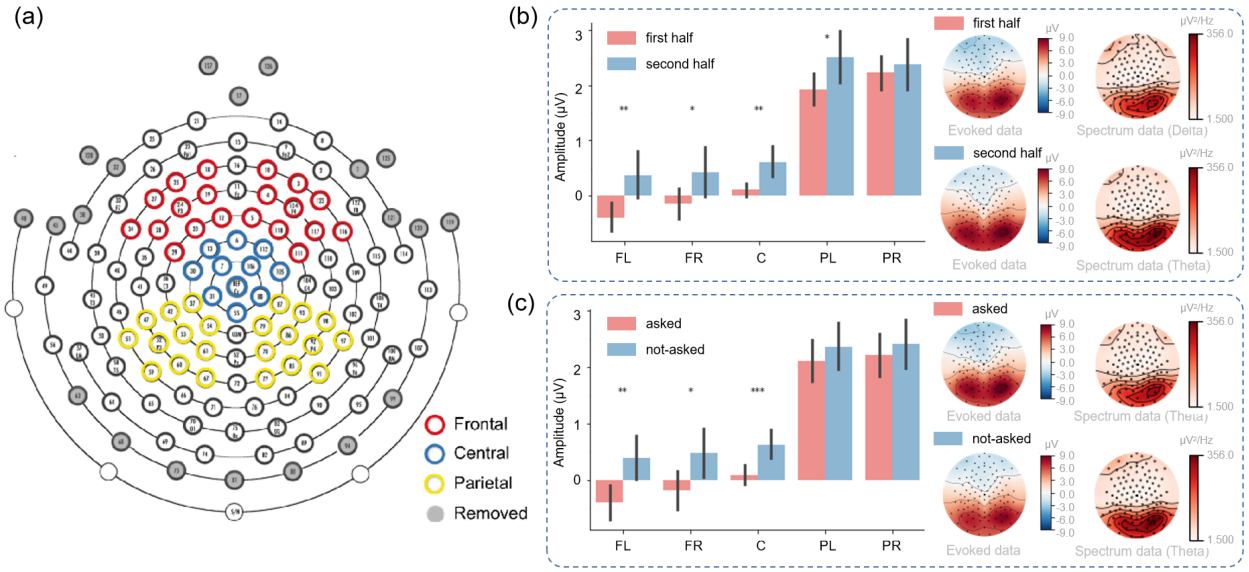

Figure 5: EEG results in at the sensor level. (a) The electrode distribution. (b) The signal amplitude of different brain regions in the first and second half of the experiment in the "Second choice" stage. The right panel shows the visualization of the evoked data and spectrum data. (c) The signal amplitude 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 right panel shows the visualization of the evoked data and spectrum data

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 5 (c)), participants displayed 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 may suggest that when participants knew the hidden states, they could effectively integrate the information with the environmental statistical structure to make superior decisions and exhibit more positive brain activities. The right panel of Figure 5 (c) reveals a higher signal in the delta band during not-asked trials, suggesting a correlation between theta band signal and uncertainty about the hidden states [26]. 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 the range in the "First choice" stage. In our settings, participants would display greater uncertainty about the hidden states of the environment for the *not-asked* trials, compared to the *asked* trials. We identified statistically significant differences between the two conditions in the signal amplitude over the left frontal region ($p < 0.01$), the right frontal region ($p < 0.05$), and the central region ($p < 0.001$)(Figure 5 (c)), suggesting a role for these areas in encoding the hidden states of the environment. The right panel of Figure 5 (c) also reveals a higher signal in the delta band during *not-asked* trials, suggesting a correlation between theta band signal and uncertainty about the hidden states [26].

3.3 EEG results at source level

In order to uncover the functional roles of various brain regions in the decision, we employed a generalized linear model (GLM) to fit the EEG source signal. The GLM included several regressors to capture different aspects of the decision-making process, namely expected free energy, active learning, active inference, extrinsic value (reward), and reward prediction error. Incorporating these regressors, enables us to know how each of these factors influenced EEG activity and contributed to the decision-making process.

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 approaching the ranger to gather information regarding the present situation of the risky path, the latter choice coming at a cost.

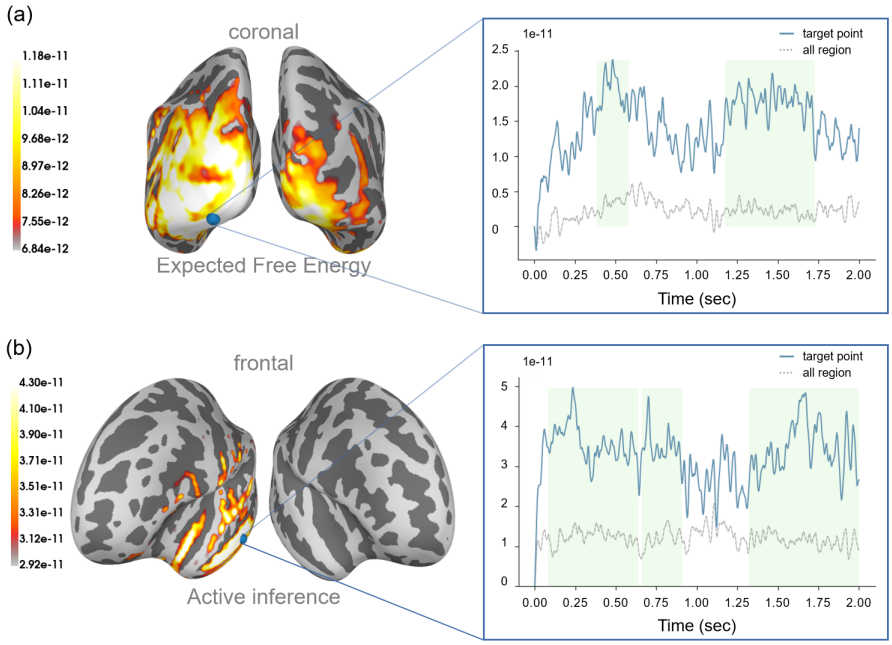

Figure 6: The source estimation results of the "First choice" stage of expected free energy and active inference. (A) The regression coefficients (β) of the expected free energy regressor. The blue point indicates the most correlated brain region (the lateral occipital cortex, left hemisphere, MNI: [-9.9, -96.8, 9.8]). The right panel shows the neural activity of the blue point and the shadow indicates that the neural activity of the blue point in these time intervals (0.380s to 0.581s and 1.172s to 1.724s) is significantly correlated with expected free energy ($p < 0.05$ and the time intervals are longer than 0.2s). (B) The regression coefficients (β) of the active inference regressor. The blue point indicates the most correlated brain region (the middle temporal gyrus, left hemisphere, MNI: [-63.5, -23.5, -13.8]). The right panel shows the neural activity of the blue point and the green shadow indicates that the neural activities of the blue point in these time intervals (0.08s to 0.636s, 0.657s to 0.906s, and 1.32s to 2.00s) are significantly correlated with active inference ($p < 0.05$ and the time intervals are longer than 0.2s).

We found a robust correlation ($p < 0.05$) between the "expected free energy" regressor and activities of the lateral occipital cortex (Figure 6 (a)). In addition, the rostral middle frontal cortex and the middle temporal gyrus, also displayed correlations with expected free energy. With respect to the "extrinsic value" regressor, we identified a strong correlation ($p < 0.05$) with the activities over the inferior temporal gyrus. Furthermore, the superior temporal gyrus, insula, and precentral area. For the "active inference" regressor (Figure 6 (b)), we observed a strong negative correlation ($p < 0.05$) with activities of the middle temporal gyrus, the inferior temporal gyrus, superior temporal gyrus, and precentral area. Interestingly, we observed that during the "First choice" stage, expected free energy and extrinsic value regressors were strongly correlated both at the beginning and the end. However, expected free energy correlations appeared earlier than those of extrinsic value at the beginning, suggesting that the brain initially encodes reward values before integrating these values with information values (active inference and active learning) for decision-making. In the case of the "active learning" regressor, we found strong correlations in the middle temporal gyrus, lateral occipital cortex, inferior temporal gyrus, and inferior parietal lobule.

3.3.2 "First result" stage – belief update

During the "First result" stage, participants were presented with the outcome of their initial 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 correlates with the "active inference" regressor (Figure 7 (a) (b)), as it corresponds to resolving uncertainties about hidden states. We observed a robust correlation ($p < 0.05$) within certain regions of the middle temporal gyrus and superior frontal cortex. Additionally, other brain regions such as the temporal pole, lateral orbitofrontal cortex, and rostral middle frontal cortex also displayed time-dependent correlations. Towards the conclusion of the "First result" stage, we noted subtle correlations with active inference in parietal and occipital regions. This observation suggests that following the assimilation of environmental state information, the brain starts

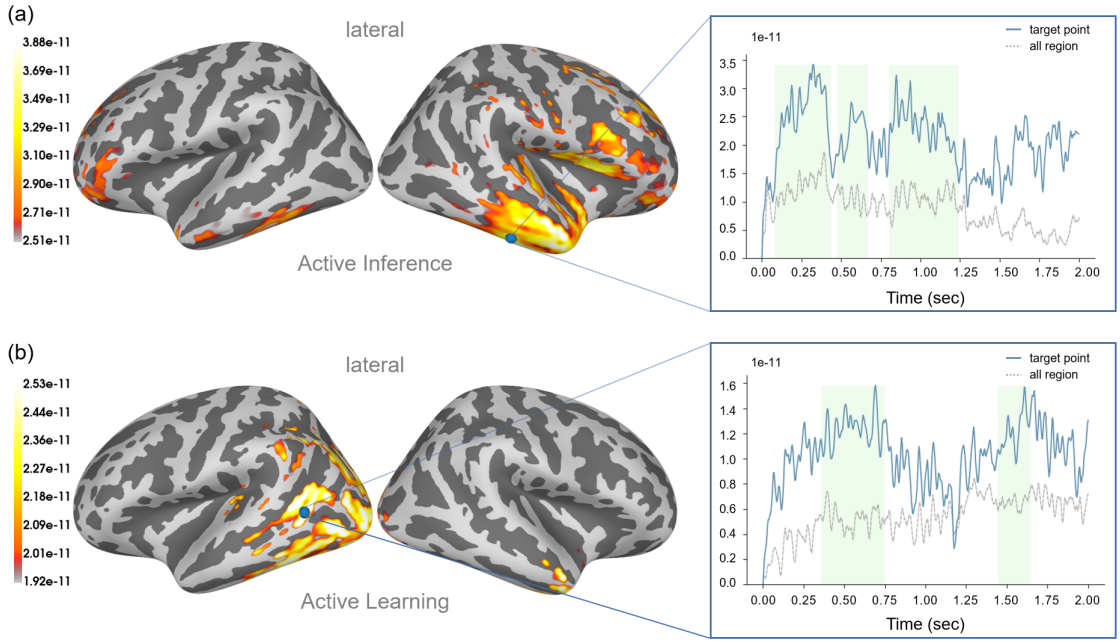

Figure 7: The source estimation results of the two result stages of active inference and active learning. (A) The regression coefficients (β) of the active inference regressor in the "First result" stage. The blue point indicates the most correlated brain region (the middle temporal gyrus, right hemisphere, MNI: [52.6, -32.3, -19.7]). The right panel shows the neural activity of the blue point and the shadow indicates that the neural activity of the blue point in these time intervals (0.076s to 0.436s, 0.47s to 0.67s, and 0.80s to 1.24s) is significantly correlated with expected free energy ($p < 0.05$ and the time intervals are longer than 0.2s). (B) The regression coefficients (β) of the active learning regressor in the "Second result" stage. The blue point indicates the most correlated brain region (the intersection of the middle temporal gyrus and the bankssts, left hemisphere, MNI: [-52.5, -56.6, 5.5]). The right panel shows the neural activity of the blue point and the shadow indicates that the neural activity of the blue point in these time intervals (0.355s to 0.745 and 1.441 to 1.651) is significantly correlated with active inference ($p < 0.05$ and the time intervals are longer than 0.2s).

utilizing a reward model to assist in decision-making.

3.3.3 "Second choice" stage – action selection

During the "Second choice" stage, participants choose between the risky and safe paths based on the available information, with the aim of maximizing rewards. This requires a balance between exploration and exploitation, similar to the "First choice" stage. First, for the "expected free energy" regressor (Figure 8(a) (b)), we identified strong correlations ($p < 0.01$) in particular regions of the lateral occipital cortex. Additionally, correlations were observed in various periods within other brain regions such as the superior parietal gyrus, inferior parietal gyrus, and rostral middle frontal cortex. Generally, the correlations between regressors and brain signals were more pronounced in the "Second choice" stage compared to the "First choice" stage. Regarding the "extrinsic value" regressor, we found that certain regions of the middle temporal gyrus showed strong correlations during the time interval from 0.104 to 0.168. Furthermore, other brain regions such as the inferior temporal gyrus, and insula, exhibited some degree of correlation at different time periods. For the "active learning" regressor (Figure 8(c) (d)), strong correlations ($p < 0.05$) were evident in the lateral occipital cortex, the parietal lobule, and temporal pole.

3.3.4 "Second result" stage – belief update

During the "Second result" stage, participants obtain 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. For the "extrinsic value" regressor, we observed strong correlations ($p < 0.05$) in specific regions of the rostral middle frontal gyrus and the lateral orbitofrontal cortex. Additionally, other brain regions such as the pars orbitalis, inferior temporal gyrus, lateral occipital sulcus, caudal middle frontal gyrus, and frontal pole demonstrated varying degrees of

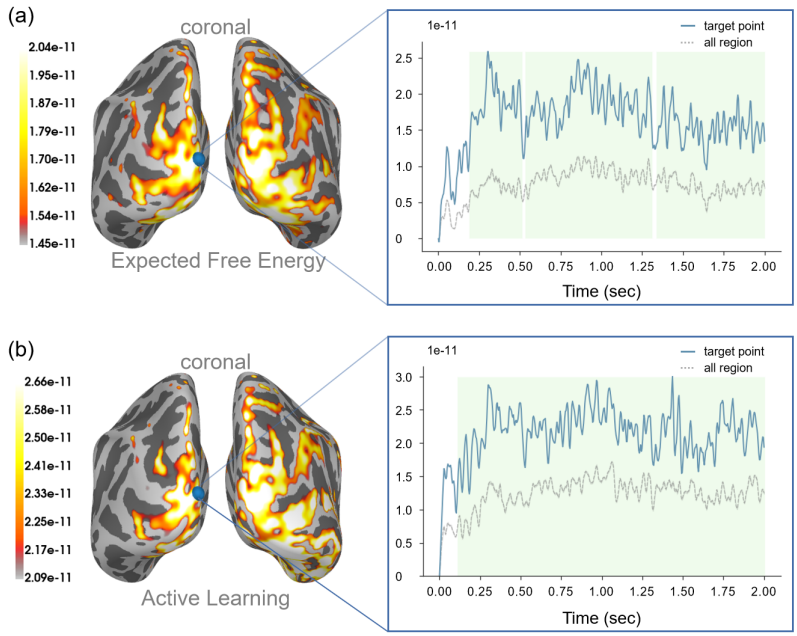

Figure 8: The source estimation results of the "Second choice" stage of expected free energy and active learning. (a) The regression coefficients (β) of the expected free energy regressor. The blue point indicates the most correlated brain region (the lateral occipital cortex, left hemisphere, MNI: [-5.5, -92.8, 15.4]). The right panel shows the neural activity of the blue point and the shadow indicates that the neural activity of the blue point in these time intervals (0.188s to 0.516s, 0.532s to 1.312s, and 1.336s to 2.00s) is significantly correlated with expected free energy ($p < 0.05$ and the time intervals are longer than 0.2s). (b) The regression coefficients (β) of the active learning regressor. The blue point indicates the most correlated brain region (the lateral occipital cortex, left hemisphere, MNI: [-9.9, -96.8, 9.8]). The right panel shows the neural activity of the blue point and the shadow indicates that the neural activity of the blue point in the time intervals (0.108s to 2.00s) is significantly correlated with expected free energy ($p < 0.05$).

correlation with "extrinsic value" across different time periods. With regards to the "active learning" regressor (Figure 7 (c) (d))), strong correlations ($p < 0.05$) were identified at the middle temporal gyrus, and the superior parietal lobule.

4 Discussion

In this study, we utilized active inference to explore the different cognitive types and associated neural components involved in human exploration strategies during the decision-making process. By employing a contextual two-bandit task, we demonstrated that the active inference model framework effectively describes real-world decision-making. Our findings indicate that active inference not only provides explanations and distinctions for uncertainties during decision-making, 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 exploration strategies vary significantly depending on the situation at hand. Such strategies can be viewed as a blend of *directed exploration*, where options with higher levels of uncertainty or ambiguity are favored, and *random exploration*, where actions are chosen with absolute randomness [27]. In the framework of active inference, the randomness in exploration is derived from the precision parameter employed during policy selection. As the precision parameter increases, the randomness in agents' actions also escalates. On the other hand, the directed exploration aspect stems from the computation of expected free energy. Policies that lead to the exploration of more ambiguous options, hence yielding higher information gain, are assigned increased expected free energy by the model [3, 4, 11].

Our model-fitting results of decision behavior indicate that people show high variance in the exploration strategies4 (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 [28, 29]. 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 representation of the environment to guide the agent's actions. Our simulation results showed these model-based behaviors that the agent constructs an environment model and uses the model to maximize rewards3. To integrate model-free learning, a habitual term was added in [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

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 [30] and neuroscientific underpinnings [31], allowing for a lucid understanding of an agent's interaction within its environment. This model effectively harmonizes cognitive and practical values by emphasizing both within the context of free energy, thereby facilitating the modeling of each participant's cognitive and practical value.

Active inference offers a superior exploration mechanism compared with basic model-free reinforcement learning (Figure 4 (c)). Since traditional reinforcement learning models predicate their policies solely on the state, this setting leads to difficulty in extracting temporal information [32] 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 [33] 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 ambiguity. A further advantage of active inference lies in its adaptability of 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 [34]. 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].

4.3 Representing uncertainties at the sensor level

In the previous work, 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 [35–38]. In our study, the sensor level results reveal more positive activities in multiple brain regions during the second half trials compared to the first half, and similarly, during not-asked trials as opposed to asked trials5.

In our setting, after the first half of the trials, participants had learned some information about the environmental statistical structure, thus experiencing less ambiguity in the latter half of the trials. This increased understanding enabled them to better utilize the statistical structure for decision-making as compared to 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-risk actions. This elevated risk was reflected in increased positive brain activities.

The aspects of ambiguity and risk, 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 more positive amplitude for the second half of the trials than the first half of the trials5 (b). It may indicate a generally more positive amplitude of EEG for the lower ambiguity trials, which may contract with previous studies showing more positive amplitude for higher ambiguity trials in previous studies [38, 39]. For example, a late positive potential (LPP) was identified in their work, which differentiated levels of ambiguity, with the amplitude of the LPP serving as an index for perceptual ambiguity levels. However, the ambiguity in their task was defined as the perceptual difficulty of distinguishing, while our definition of ambiguity corresponds to the information gained from certain policies. Furthermore, Zheng et al. [40] used a wheel-of-fortune task to examine the ERP and oscillatory correlations of neural feedback processing under conditions of risk and ambiguity. 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 bear an additional cost in exchange for information regarding the environment’s state (epistemic value). Here, the primary source of epistemic value stemmed from resolving the uncertainty of the hidden staterisk. The occipital cortex appears to play a critical role in this process by combining extrinsic value with epistemic value (expected free energy) to guide decision-making (Figure 6). Previous study [41] has demonstrated significant activations in the lateral occipital complex during perceptual decision-making, indicating that the human brain may use perceptual persistence to facilitate reward-related decisions.

As for the “First result" stage, participants learned about the environment’s hidden states. Our results indicated that the regions within the temporal lobe played a crucial role in both valuing the uncertainty of hidden states and learning information about these hidden states (Figure 7 (a)). Other studies have similarly demonstrated the importance of the temporal pole and the inferior temporal areas in processing the ambiguity regarding lexical semantics [42, 43]. Studies on macaques also identified the role of the inferior temporal lobe in representing blurred visual objects [44]. Throughout the “First result" stage, participants are processing the state information relevant to the current trial. The middle temporal gyrus is postulated to play a key role in processing this state information and employing it to construct an environmental model. This aligns with previous findings [45], which suggest that the middle temporal gyrus collaborates with other brain regions to facilitate conscious learning. Moreover, studies have also identified deficits in episodic future thinking in patients with damage to the middle temporal gyrus (MTG) [46], thereby indicating the critical role of MTG in future-oriented decision-making tasks, particularly those involving future thinking [47–49].

In the “Second choice stage", participants chose between a safe path and a risky path, contingent on perceived value. When knowing the environment’s hidden states, participants tended to resolve the uncertainty of model parameters by opting for the risky path. Conversely, without knowledge of the hidden states, participants gravitated towards risk avoidance by selecting the safe path. Our results highlighted the significance of the occipital lobe regions, in conjunction with the temporal and parietal lobes, in both valuing the uncertainty of model parameters and learning about these parameters (Figure 8). These results are consistent with another study that demonstrates activation in the superior parietal, right precentral gyrus, postcentral gyrus, and superior frontal regions during decision-making involving ambiguity and risk [50]. Since the superior parietal region was involved in the integration of visual motion information and extraction of episodic memory [51–53], it may play a crucial role in our decision task as well, where participants need to extract statistical relationship from different visual input at different times.

For the “Second result stage", participants got rewards according to their actions, constructing the model-free action value function and the model-based state transition function. Our results highlighted the role of the frontal cortex in learning the action value function and the role of the middle temporal gyrus in learning the state transition function (Figure 7 (b)). Notably, the correlations’ significance between “active inference" and the middle temporal gyrus reached its peak later than the correlations’ significance between “extrinsic value" and the orbitofrontal cortex. This temporal disparity may suggest that the brain processes reward information earlier than environmental information. However, this is contrary to previous findings that the temporospatial factors derived from PCA on the effect model-based prediction error peaked earlier than those on the effect model-free prediction error [54]. Future work should look deeper into where and when the human brain processes different information in decision tasks.

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 serve as a better representation of the brain’s actual value employed to guide actions [55]. Our results pointed to a strong correlation between the expected free energy and activations in the occipital lobe. Such a result may be explained by that the lateral occipital sulcus plays a key role in the persistence of perceptual information and represents delayed rewards [41, 56].

5 Conclusion

In the current study, we introduced the active inference framework as a means 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 during the initial trials 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 lateral occipital gyrus was primarily involved in action selection (expected free energy), while the temporal lobe regions were mainly engaged in valuing

the information related to the hidden states of the environment. Furthermore, the middle temporal gyrus and lateral occipital gyrus were prominently involved in valuing the information related to the environmental model parameters. The temporal pole regions primarily participated in learning the hidden states of the environment (active inference), while the middle temporal gyrus was more engaged in learning the model parameters of the environment (active learning). In essence, our findings suggest that active inference is capable of investigating human behaviors in decision-making under uncertainty, reducing ambiguity, avoiding risk, and maximizing rewards. 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), Shenzhen-Hong Kong-Macao Science and Technology Innovation Project (SGDX2020110309280100), 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] 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.

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

[26] 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.

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

[28] 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.

[29] 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.

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

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

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

[38] 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.

[39] 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.

[40] 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.

[41] M. G. Philiastides and P. Sajda, "Eeg-informed fmri reveals spatiotemporal characteristics of perceptual decision making," *Journal of Neuroscience*, vol. 27, no. 48, pp. 13 082–13 091, 2007.

[42] K. Hoenig and L. Scheef, "Mediotemporal contributions to semantic processing: fmri evidence from ambiguity processing during semantic context verification," *Hippocampus*, vol. 15, no. 5, pp. 597–609, 2005.

[43] S. Vitello, J. E. Warren, J. T. Devlin, and J. M. Rodd, "Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences," *Frontiers in human neuroscience*, vol. 8, p. 530, 2014.

[44] N. Emadi and H. Esteky, "Neural representation of ambiguous visual objects in the inferior temporal cortex," *PloS one*, vol. 8, no. 10, p. e76856, 2013.

[45] A. R. McIntosh, M. N. Rajah, and N. J. Lobaugh, "Functional connectivity of the medial temporal lobe relates to learning and awareness," *Journal of Neuroscience*, vol. 23, no. 16, pp. 6520–6528, 2003.

[46] D. J. Palombo, M. M. Keane, and M. Verfaellie, "The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking," *Hippocampus*, vol. 25, no. 3, pp. 345–353, 2015.

[47] J. Okuda, T. Fujii, H. Ohtake, T. Tsukiura, K. Tanji, K. Suzuki, R. Kawashima, H. Fukuda, M. Itoh, and A. Yamadori, "Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes," *Neuroimage*, vol. 19, no. 4, pp. 1369–1380, 2003.

[48] D. Palombo, M. Keane, and M. Verfaellie, "Using future thinking to reduce temporal discounting: Under what circumstances are the medial temporal lobes critical?" *Neuropsychologia*, vol. 89, pp. 437–444, 2016.

[49] D. L. Schacter and D. R. Addis, "On the nature of medial temporal lobe contributions to the constructive simulation of future events," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1521, pp. 1245–1253, 2009.

[50] 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.

[51] M. Iacoboni and E. Zaidel, "Interhemispheric visuo-motor integration in humans: the role of the superior parietal cortex," *Neuropsychologia*, vol. 42, no. 4, pp. 419–425, 2004.

[52] A. D. Wagner, B. J. Shannon, I. Kahn, and R. L. Buckner, "Parietal lobe contributions to episodic memory retrieval," *Trends in cognitive sciences*, vol. 9, no. 9, pp. 445–453, 2005.

[53] D. M. Wolpert, S. J. Goodbody, and M. Husain, "Maintaining internal representations: the role of the human superior parietal lobe," *Nature neuroscience*, vol. 1, no. 6, pp. 529–533, 1998.

[54] T. D. Sambrook, B. Hardwick, A. J. Wills, and J. Goslin, "Model-free and model-based reward prediction errors in eeg," *NeuroImage*, vol. 178, pp. 162–171, 2018.

[55] 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.

[56] M. M. Owens, J. C. Gray, M. T. Amlung, A. Oshri, L. H. Sweet, and J. MacKillop, "Neuroanatomical foundations of delayed reward discounting decision making," *NeuroImage*, vol. 161, pp. 261–270, 2017.