

The Neural Correlates of Ambiguity and Risk in Human Decision-Making under an Active Inference Framework

Reviewed Preprint

v2 • August 23, 2024

Revised by authors

Reviewed Preprint

v1 • February 22, 2024

Shuo Zhang, Yan Tian, Quanying Liu , Haiyan Wu 

Department of Biomedical Engineering, Southern University of Science and Technology Shenzhen, China • Centre for Cognitive and Brain Sciences and Department of Psychology, University of Macau, Macau, China

 https://en.wikipedia.org/wiki/Open_access
 Copyright information

Abstract

Active inference integrates perception, decision-making, and learning into a united theoretical frame-work, 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 (ambiguity and risk), 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 ambiguity and risk, 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 ambiguity, 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 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 decision under ambiguity and risk.

eLife assessment

The study addresses a central question in systems neuroscience (validation of active inference models of exploration) using a combination of behavior, neuroimaging, and modelling. The data provided are **useful** but **incomplete** due to issues with multiple comparisons and lack of model validation.

<https://doi.org/10.7554/eLife.92892.2.sa4>

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 ambiguity - the kind of uncertainty that can be reduced through sampling, and risk - 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 risk from ambiguity in uncertainty or actions from states (perception). In our work, we develop a contextual multi-armed bandit task to enable participants to actively reduce ambiguity, avoid risk, 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 ambiguity and risk. Ambiguity is represented by the uncertainty about model parameters associated with choosing a particular action, while risk is signified by the variance of the environment's hidden states. The value of reducing ambiguity, the value of avoiding risk, and extrinsic value together constitute expected free energy (see Section 2.1).

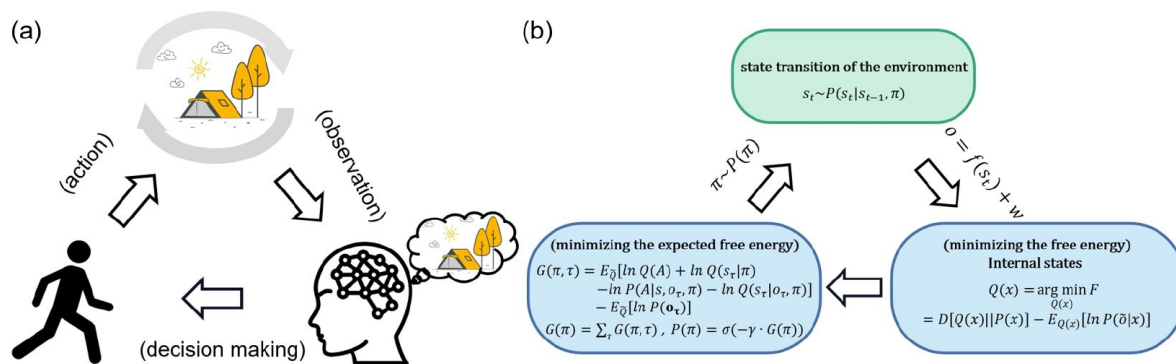


Figure 1

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.

Our study aim to utilize the active inference framework to investigate how the brain represents the decision-making process and how the brain disassociates the representations of ambiguity and risk (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 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](#)).

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 optimise 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 $q(s)$ 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 ambiguity, minimizing risk, 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]:

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

where o is observations or sensory inputs ($\tilde{o}$ is the history of observations), s is the hidden states of the environment ($\tilde{s}$ 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, $\text{Cat}()$ is the categorical distribution, $\text{Dir}()$ is the Dirichlet distribution, and $\Gamma()$ is the Gamma distribution.

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

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

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

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 $R(\tilde{o}, \tilde{s}, \tilde{u})$	The generative process generates observations and next states (state transitions of the environment) based on current states and actions
P	a generative model $P(\tilde{o}, \tilde{s}, \pi, \eta)$	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.

Table 1

Ingredients for computational modeling of active inference

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):

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

Here, $x = \tilde{s}, \pi, A, B, \beta$, 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):

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

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, $E_Q[\ln P(\tilde{o}|s)]$, 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.

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

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

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

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

In Eq.(7), it is important to note that we anticipate observations that have not yet occurred. Consequently, we designate $\tilde{Q} = Q(o_\tau, s_\tau, A|\pi)$. 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.

$$G(\pi, \tau) = \underbrace{E_{\tilde{Q}}[\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)] + \underbrace{\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)}_{\text{Value of avoiding risk}}]}_{\text{Epistemic value}} - \underbrace{E_{\tilde{Q}}[\ln P(o_\tau)]}_{\text{Extrinsic value}} \quad (8)$$

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 (reducing ambiguity) and the hidden states (avoiding risk), which are to be illuminated by future observations.

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

$$G(\pi, \tau) = AL \cdot E_{\tilde{Q}}[\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)] + AI \cdot E_{\tilde{Q}}[\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)] - EX \cdot E_{\tilde{Q}}[\ln P(o_\tau)] \quad (9)$$

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 [34]. 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):

$$a_0 = \begin{bmatrix} 1 & 1 & \text{prior} & \text{prior} & 0 & 0 & 0 & 0 \\ 0 & 0 & \text{prior} & \text{prior} & 0 & 0 & 0 & 0 \\ 0 & 0 & \text{prior} & \text{prior} & 0 & 0 & 0 & 0 \\ 0 & 0 & \text{prior} & \text{prior} & 0 & 0 & 0 & 0 \\ 0 & 0 & \text{prior} & \text{prior} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (10)$$

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

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 twoarmed 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 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 ambiguity and avoiding risk.

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 “Risk” option offered greater informational value and reduced ambiguity. 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 “Risk” option in “Context 1” for a higher expected reward, and the “Safe” option in “Context 2” where the informational value of the “Risk” option was outweighed by the difference in expected rewards between the “Safe” option and the “Risk” 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).

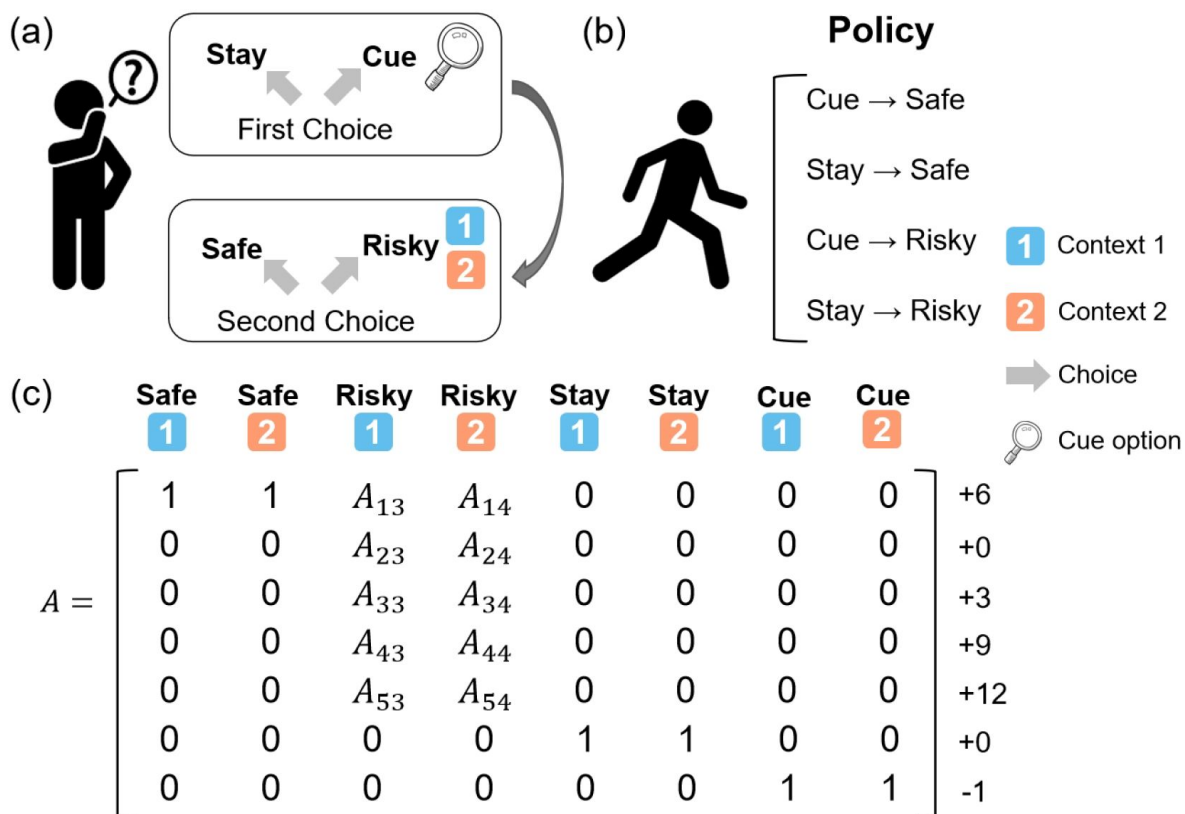


Figure 2

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).

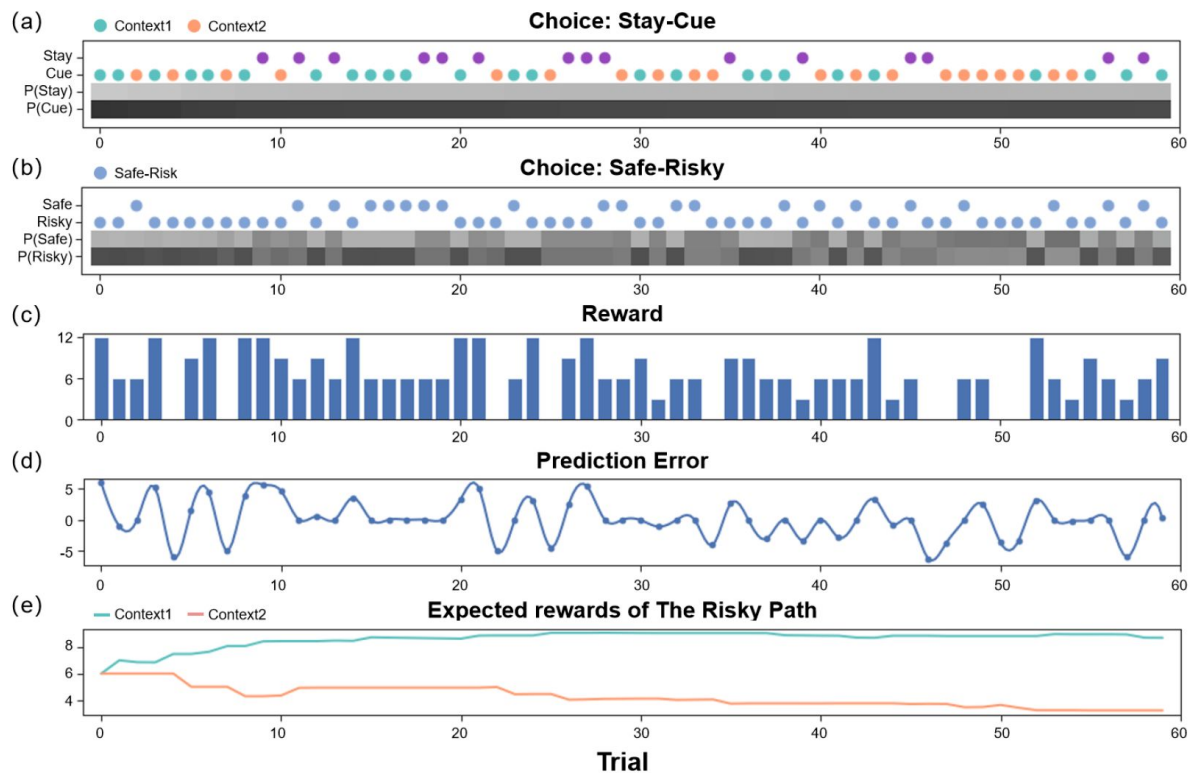


Figure 3

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.

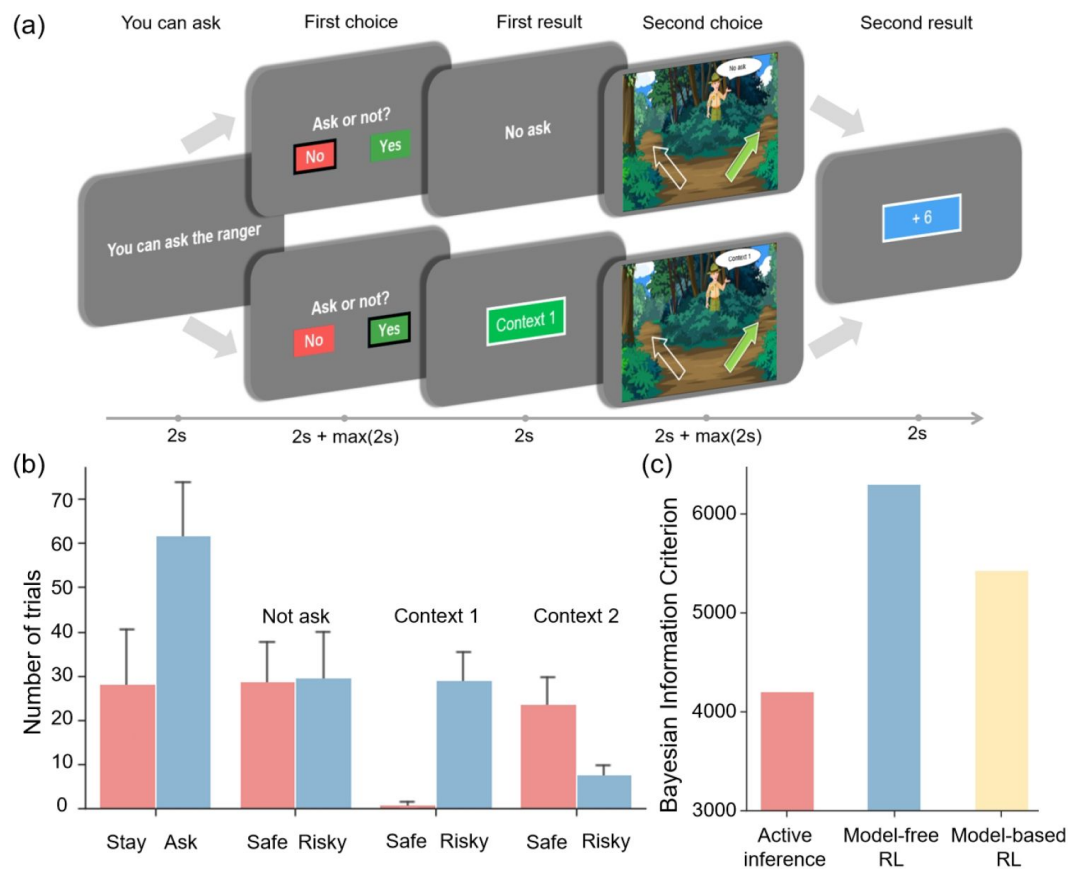


Figure 4

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.

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:

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 ambiguity more, and increasing AI meant participants would like to learn the hidden state of the environment and avoid risk 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").

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.

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 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 5 (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 5 (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 5 (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 correlations over trials between constituent terms in expected free energy (the value of avoiding risk, the value of reducing ambiguity, extrinsic value, and expected free energy itself) and neural responses in source space. Additionally, we also investigated the neural correlate of (the degree of) risk, (the degree of) ambiguity, and prediction

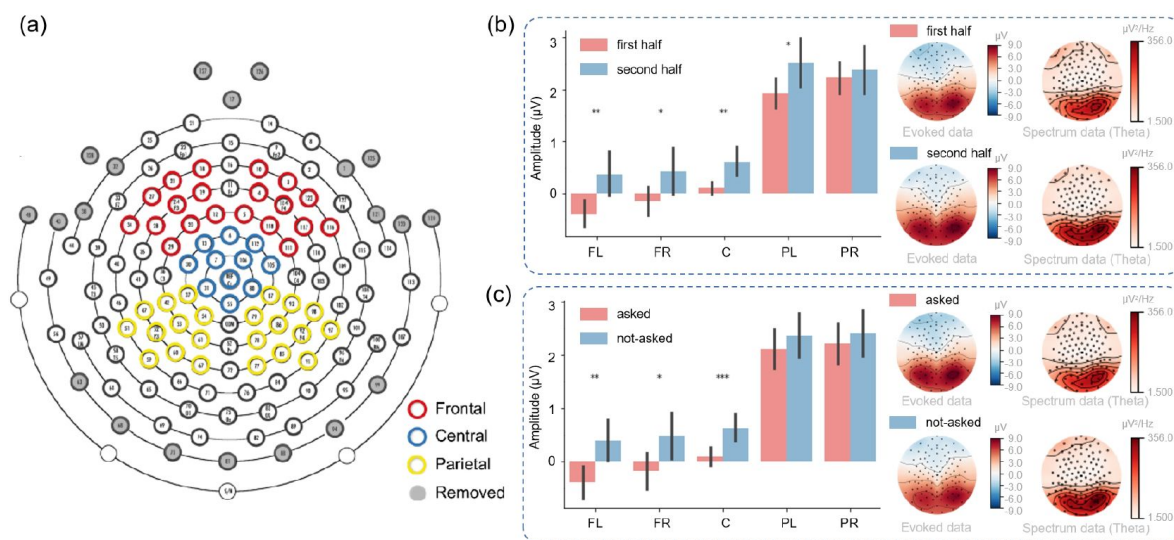


Figure 5

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.

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 ambiguity, 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. Note that we did not attempt to correct for multiple comparisons; largely, because the correlations observed were sustained over considerable time periods, which would be almost impossible under the null hypothesis of no correlations.

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 avoiding risk” ($AI \cdot E_{\tilde{Q}}[\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)]$) and “extrinsic value” ($EX \cdot E_{\tilde{Q}}[\ln P(o_\tau)]$).

We found a robust correlation ($p < 0.01$) between the “expected free energy” regressor and the frontal pole (**Figure 6 (a)** [↗](#)). In addition, the superior temporal gyrus also displayed strong correlations with expected free energy. With respect to the “value of avoiding risk” regressor, we identified a strong correlation ($p < 0.05$) with the temporal lobe, including the inferior temporal gyrus, middle temporal gyrus, and superior temporal gyrus (**Figure 6 (b)** [↗](#)). In addition, the frontal pole and rostral middle frontal gyrus also displayed strong correlations with the value of avoiding risk. For the “extrinsic value” regressor, we observed a strong correlation ($p < 0.05$) with the frontal pole (see Figure S1(a)). In addition, the lateral orbitofrontal cortex, rostral middle frontal gyrus, superior frontal gyrus, middle 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 “avoiding risk” regressor, as it corresponded to resolving uncertainties about hidden states. We assumed that the brain learning hidden states (avoiding risk) corresponded to the value of avoiding risk. Thus, the “avoiding risk” regressor could be $AI \cdot (\ln Q(s_t|\pi) - \ln Q(s_t|o_t, \pi))$

For “avoiding risk”, we observed a robust correlation ($p < 0.05$) within the medial orbitofrontal cortex (**Figure 7 (a)** [↗](#)). In addition, the lateral orbitofrontal cortex, middle temporal gyrus, and superior temporal gyrus also displayed strong correlations with avoiding risk.

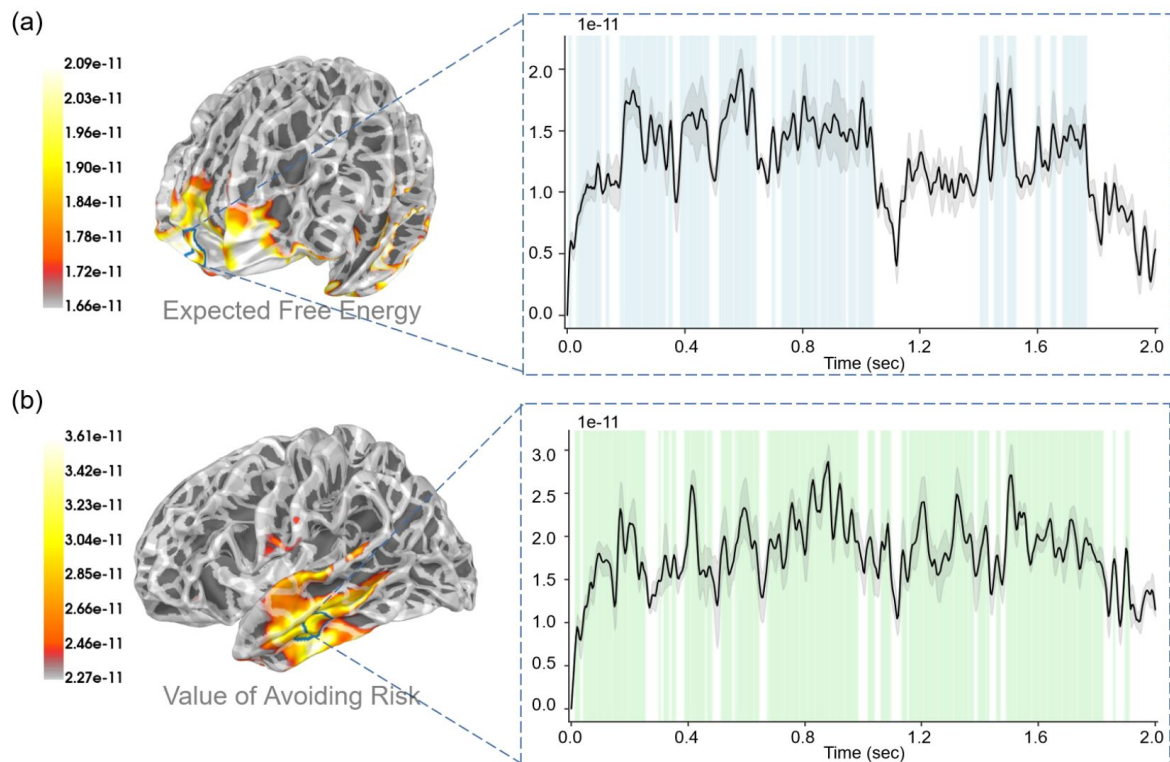


Figure 6

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 expected free energy, the black line indicates the average intensity of this region, and the gray-shaded region indicates the range of variation. The blue-shaded region indicates $p < 0.01$. (B) The regression intensity (β) of the value of avoiding risk. The right panel indicates the regression intensity between the middle temporal gyrus (5, left half) and the value of avoiding risk, the black line indicates the average intensity of this area, and the gray-shaded region indicates the range of variation. The green-shaded region indicates $p < 0.05$.

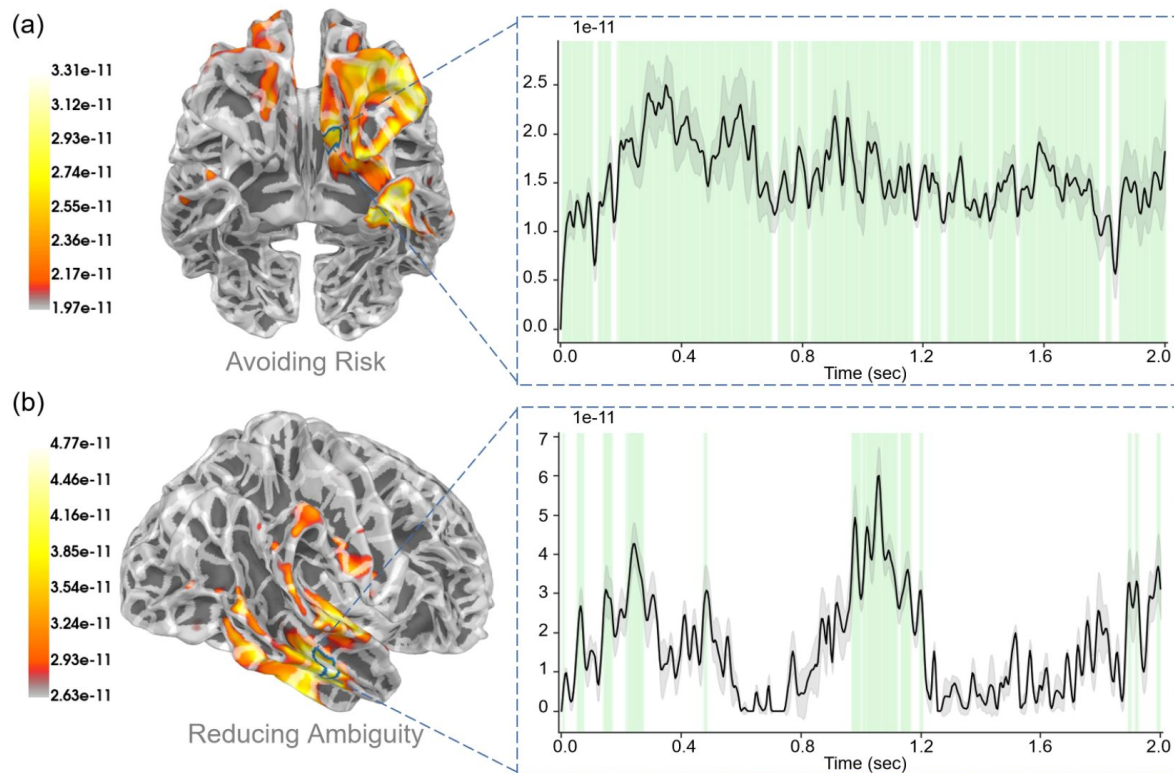


Figure 7

The source estimation results of avoiding risk and reducing ambiguity in the two result stages. (A) The regression intensity (β) of avoiding risk in the "First result" stage. The right panel indicates the regression intensity between the medial orbitofrontal cortex (5, left half) and avoiding risk, the black line indicates the average intensity of this region, and the gray-shaded region indicates the range of variation. The green-shaded region indicates $p < 0.05$. (B) The regression intensity (β) of reducing ambiguity in the "Second result" stage. The right panel indicates the regression intensity between the middle temporal gyrus (5, right half) and reducing ambiguity, the black line indicates the average intensity of this region, and the gray-shaded region indicates the range of variation. The green-shaded region indicates $p < 0.05$.

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 ambiguity” ($AL \cdot E_{\tilde{Q}}[\ln Q(A) - \ln P(A|s_{\tau}, o_{\tau}, \pi)]$) “extrinsic value” ($EX \cdot E_{\tilde{Q}}[\ln P(o_{\tau})]$), and “ambiguity” ($E_{\tilde{Q}}[\ln Q(A) - \ln P(A|s_{\tau}, o_{\tau}, \pi)]$).

For “expected free energy” (Figure 8(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 ambiguity”, we found that the middle temporal gyrus showed strong correlations ($p < 0.05$). In addition, the inferior temporal gyrus, insula, rostral middle frontal gyrus, and superior temporal gyrus also displayed strong correlations with the value of reducing ambiguity. For “extrinsic value”, strong correlations ($p < 0.001$) were evident in the rostral middle frontal gyrus (see Figure S1(b)). In addition, the caudal middle frontal gyrus, pars opercularis, pars triangularis, and precentral gyrus also displayed strong correlations with extrinsic values. In the “Second choice” stage, participants made choices under different degrees of ambiguity. For “ambiguity”, we found strong correlations ($p < 0.05$) in the frontal pole (see Figure S3). In addition, the rostral middle frontal gyrus and superior frontal gyrus also displayed strong correlations with ambiguity. 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” ($r_t - E_{\tilde{Q}}[\ln P(o_{\tau})]$) and “reducing ambiguity” ($\ln Q(A) - \ln P(A|s_t, o_t, \pi)$). Here, we also assumed that learning model parameters (reducing ambiguity) corresponded to the value of reducing ambiguity.

For “extrinsic value”, we observed strong correlations ($p < 0.05$) in the lateral occipital gyrus, paracentral lobule, postcentral gyrus, and superior parietal lobule (see Figure S2(a)). For “prediction error”, we observed strong correlations ($p < 0.05$) in the bank of the superior temporal sulcus, inferior temporal gyrus, and lateral occipital gyrus. For “reducing ambiguity”, we observed strong correlations ($p < 0.05$) in the middle temporal gyrus, parahippocampal gyrus, postcentral gyrus, and precentral gyrus (see Figure S2(b)).

4 Discussion

In this study, we utilized active inference to explore the neural correlates involved in the human decision-making process under ambiguity and risk. 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.

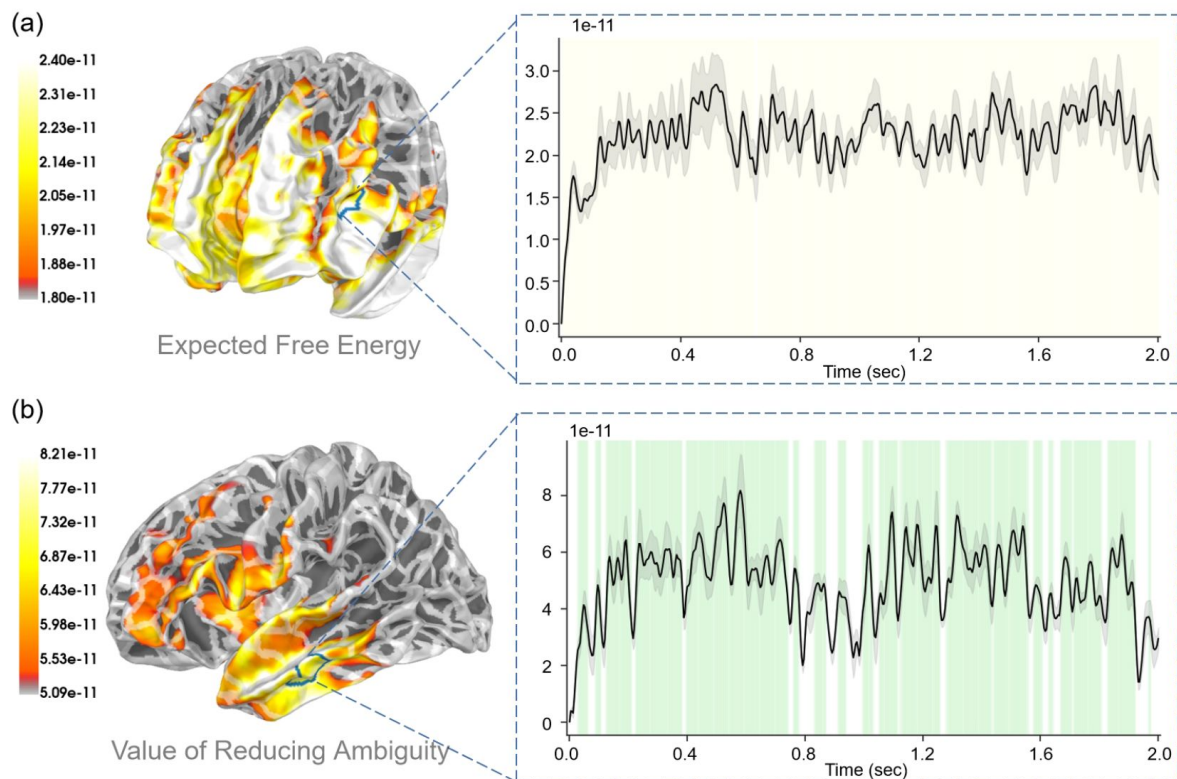


Figure 8

The source estimation results of expected free energy and the value of reducing ambiguity 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-region indicates $p < 0.001$. (b) The regression intensity (β) of the value of reducing ambiguity. The right panel indicates the regression intensity between the middle temporal gyrus (5, left half) and the value of reducing ambiguity, the black line indicates the average intensity of this region, and the gray-shaded region indicates the range of variation. The green-shaded region indicates $p < 0.05$.

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 [40]. 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 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 [41, 42]. 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 [43] and neuroscientific underpinnings [44], 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 [45] 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 [46] 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 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 [47]. 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 [48, 49], 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 [50–53]. 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 5).

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 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-risk actions. This elevated risk was reflected in increased positive brain activities.

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 greater neural response for the second half of the trials than the first half of the trials (Figure 5(b)). It may indicate a generally greater neural response for the lower ambiguity trials, which may contrast with previous studies showing greater neural response for higher ambiguity trials in previous studies [53, 54]. 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. [55] 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 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 avoiding risk. 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 6). Our results also showed the orbitofrontal cortex, middle temporal gyrus, and superior temporal gyrus were correlated with the value of avoiding risk. Previous study [56] demonstrated that the frontal pole was strongly activated in the “risk” condition and “ambiguous” condition during decision-making. Another

study also demonstrated that the frontal pole played an important role in the interaction between beliefs (risk and ambiguity) and payoffs (gains and losses). About the orbitofrontal cortex, a previous study [57] found that both the medial and lateral orbitofrontal cortex encoded risk and reward probability while the lateral orbitofrontal cortex played a dominant role in coding experienced value. Another study [58] indicated that the medial orbitofrontal cortex was related to risk-taking and risk-taking was driven by specific orbitofrontal cortex reward systems.

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 regions within the temporal lobe played a crucial role in both valuing the uncertainty about 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 evaluating the ambiguity regarding lexical semantics [59, 60]. Studies on macaques also identified the role of the inferior temporal lobe in representing blurred visual objects [61]. 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 [62], 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 [63], thereby indicating the critical role of the middle temporal gyrus in future-oriented decision-making tasks, particularly those involving future thinking [64–66].

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 risk avoidance 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 middle temporal gyrus, rostral middle frontal gyrus, and inferior temporal gyrus, in evaluating the value of reducing ambiguity. Compared with the “First choice stage” where participants needed to evaluate the value of avoiding risk, We found that there was a high overlap in the brain regions involved in the value of avoiding risk and the brain regions involved in the value of reducing ambiguity. These results suggest that some brain regions may evaluate both the value of reducing ambiguity and the value of avoiding risk [67]. In addition, our results showed that the frontal pole was correlated with the degree of ambiguity. This result was consistent with a previous study [56] that the frontal pole was strongly activated in the “ambiguous” condition.

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 middle temporal gyrus and parahippocampal cortex in learning the state transition function and reducing ambiguity (**Figure 7 (b)**). This result indicated that the middle temporal gyrus played an important role in dealing with uncertainty in both choosing stages and the result (learning) stages. Participants made their decisions in different contexts and there was a previous study [68] emphasizing the role of parahippocampal-prefrontal communication in context-modulated behavior.

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 [69]. 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 avoiding risk, the value of

reducing ambiguity, and the degree of ambiguity, but not the degree of risk. Future work should design a task with different degrees of risk to investigate how the brain encodes risk. 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.

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 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), while the temporal lobe regions were mainly engaged in evaluating the value of avoiding risk. Furthermore, the middle temporal gyrus and rostral middle frontal gyrus were predominantly involved in evaluating the value of reducing ambiguity and the middle temporal gyrus also encoded the degree of ambiguity. The orbitofrontal cortex primarily participated in learning the hidden states of the environment (avoiding risk), while the frontal pole was more engaged in learning the model parameters of the environment (reducing ambiguity). In essence, our findings suggest that active inference is capable of investigating human behaviors in decision-making under uncertainty. 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] Friston K. (2010) **The free-energy principle: a unified brain theory?** *Nature reviews neuroscience* **11**:127–138
- [2] Friston K. J., Stephan K. E. (2007) **Free-energy and the brain** *Synthese* **159**:417–458
- [3] Friston K., FitzGerald T., Rigoli F., Schwartenbeck P., Pezzulo G., et al. (2016) **Active inference and learning** *Neuro-science & Biobehavioral Reviews* **68**:862–879
- [4] Friston K., FitzGerald T., Rigoli F., Schwartenbeck P., Pezzulo G. (2017) **Active inference: a process theory** *Neural computation* **29**:1–49
- [5] Kirchhoff M., Parr T., Palacios E., Friston K., Kiverstein J. (2018) **The markov blankets of life: autonomy, active inference and the free energy principle** *Journal of The royal society interface* **15**
- [6] Parr T., Pezzulo G., Friston K. J. (2022) **Active inference: the free energy principle in mind, brain, and behavior**
- [7] Friston K. J., Daunizeau J., Kiebel S. J. (2009) **Reinforcement learning or active inference?** *PloS one* **4**
- [8] Schwartenbeck P., Passecker J., Hauser T. U., FitzGerald T. H., Kronbichler M., Friston K. J. (2019) **Computational mechanisms of curiosity and goal-directed exploration** *Elife* **8**
- [9] O'Reilly C. A., Tushman M. L. (2011) **Organizational ambidexterity in action: How managers explore and exploit** *California management review* **53**:5–22
- [10] Wilson R. C., Geana A., White J. M., Ludvig E. A., Cohen J. D. (2014) **Humans use directed and random exploration to solve the explore-exploit dilemma** *Journal of Experimental Psychology: General* **143**
- [11] Gershman S. J. (2019) **Uncertainty and exploration** *Decision* **6**
- [12] Daw N. D., O'doherty J. P., Dayan P., Seymour B., Dolan R. J. (2006) **Cortical substrates for exploratory decisions in humans** *Nature* **441**:876–879
- [13] Badre D., Doll B. B., Long N. M., Frank M. J. (2012) **Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration** *Neuron* **73**:595–607
- [14] Cavanagh J. F., Figueroa C. M., Cohen M. X., Frank M. J. (2012) **Frontal theta reflects uncertainty and unexpect-edness during exploration and exploitation** *Cerebral cortex* **22**:2575–2586
- [15] Payzan-LeNestour E., Dunne S., Bossaerts P., O'Doherty J. P. (2013) **The neural representation of unexpected uncertainty during value-based decision making** *Neuron* **79**:191–201
- [16] Guha S., Munagala K., Shi P. (2010) **Approximation algorithms for restless bandit problems** *Journal of the ACM (JACM)* **58**:1–50

- [17] Tomov M. S., Truong V. Q., Hundia R. A., Gershman S. J. (2020) **Dissociable neural correlates of uncertainty underlie different exploration strategies** *Nature Communications* **11**
- [18] Fan H., Gershman S. J., Phelps E. A. (2022) **Trait somatic anxiety is associated with reduced directed exploration and underestimation of uncertainty** *Nature Human Behaviour* :1–12
- [19] Schulz E., Konstantinidis E., Speekenbrink M. (2015) **Exploration-exploitation in a contextual multi-armed bandit task** *International conference on cognitive modeling* :118–123
- [20] Schulz E., Konstantinidis E., Speekenbrink M. (2015) **Learning and decisions in contextual multi-armed bandit tasks** *CogSci*
- [21] Molinaro G., Collins A. G. (2023) **Intrinsic rewards explain context-sensitive valuation in reinforcement learning** *PLoS Biology* **21**
- [22] Levy I., Snell J., Nelson A. J., Rustichini A., Glimcher P. W. (2010) **Neural representation of subjective value under risk and ambiguity** *Journal of neurophysiology* **103**:1036–1047
- [23] Galdo M., Bahg G., Turner B. M. (2020) **Variational bayesian methods for cognitive science** *Psychological methods* **25**
- [24] Friston K. (2013) **Active inference and free energy** *Behavioral and brain sciences* **36**
- [25] Buckley C. L., Kim C. S., McGregor S., Seth A. K. (2017) **The free energy principle for action and perception: A mathematical review** *Journal of Mathematical Psychology* **81**:55–79
- [26] Oudeyer P.-Y., Kaplan F. (2007) **What is intrinsic motivation? a typology of computational approaches** *Frontiers in neurorobotics* **1**
- [27] Schmidhuber J. (2010) **Formal theory of creativity, fun, and intrinsic motivation (1990–2010)** *IEEE transactions on autonomous mental development* **2**:230–247
- [28] Barto A., Mirolli M., Baldassarre G. (2013) **Novelty or surprise?** *Frontiers in psychology* **4**
- [29] FitzGerald T. H., Schwartenbeck P., Moutoussis M., Dolan R. J., Friston K. (2015) **Active inference, evidence accumulation, and the urn task** *Neural computation* **27**:306–328
- [30] Lu T., Pál D., Pál M. (2010) **Contextual multi-armed bandits** *Proceedings of the Thirteenth international conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings* :485–492
- [31] Martínez-Cancino R., Delorme A., Truong D., Artoni F., Kreutz-Delgado K., Sivagnanam S., Yoshimoto K., Majumdar A., Makeig S. (2021) **The open eeglab portal interface: High-performance computing with eeglab** *NeuroImage* **224**
- [32] Esch L., Dinh C., Larson E., Engemann D., Jas M., Khan S., Gramfort A., Hämäläinen M. S. (2019) **Mne: software for acquiring, processing, and visualizing meg/eeg data** *Magnetoencephalography: From Signals to Dynamic Cortical Networks* :355–371
- [33] Pascual-Marqui R. D. (2007) **Discrete, 3d distributed, linear imaging methods of electric neuronal activity. part 1: exact, zero error localization** *arXiv*
- [34] Khan S. *et al.* (2018) **Maturation trajectories of cortical resting-state networks depend on the mediating frequency band** *NeuroImage* **174**:57–68

- [35] Vrieze S. I. (2012) **Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic)** *Psychological methods* **17**
- [36] Frazier P. I. (2018) **A tutorial on bayesian optimization** *arXiv*
- [37] Sutton R. S., Barto A. G. (2018) **Reinforcement learning: An introduction**
- [38] Friston K., Rigoli F., Ognibene D., Mathys C., Fitzgerald T., Pezzulo G. (2015) **Active inference and epistemic value** *Cognitive neuroscience* **6**:187–214
- [39] Harper J., Malone S. M., Iacono W. G. (2017) **Theta-and delta-band eeg network dynamics during a novelty oddball task** *Psychophysiology* **54**:1590–1605
- [40] Gershman S. J. (2018) **Deconstructing the human algorithms for exploration** *Cognition* **173**:34–42
- [41] Gläscher J., Daw N., Dayan P., O’Doherty J. P. (2010) **States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning** *Neuron* **66**:585–595
- [42] Daw N. D., Niv Y., Dayan P. (2005) **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control** *Nature neuroscience* **8**:1704–1711
- [43] Crooks G. E. (1998) **Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems** *Journal of Statistical Physics* **90**:1481–1487
- [44] Lehmann K., Bolis D., Ramstead M. J., Friston K., Kanske P. (2022) **An active inference approach to second-person neuroscience** *PsyArXiv*
- [45] Laskin M., Lee K., Stooke A., Pinto L., Abbeel P., Srinivas A. (2020) **Reinforcement learning with augmented data** *Advances in neural information processing systems* **33**:19–19
- [46] Wang J. X., Kurth-Nelson Z., Tirumala D., Soyer H., Leibo J. Z., Munos R., Blundell C., Kumaran D., Botvinick M. (2016) **Learning to reinforcement learn** *arXiv*
- [47] Raja V., Valluri D., Baggs E., Chemero A., Anderson M. L. (2021) **The markov blanket trick: On the scope of the free energy principle and active inference** *Physics of Life Reviews* **39**:49–72
- [48] Wald A. (1947) **An essentially complete class of admissible decision functions** *The Annals of Mathematical Statistics* :549–555
- [49] Brown L. D. (1981) **A complete class theorem for statistical problems with finite sample spaces** *The Annals of Statistics* :1289–1300
- [50] Wang L., Zheng J., Huang S., Sun H. (2015) **P300 and decision making under risk and ambiguity** *Computational intelligence and neuroscience* **2015**:1–1
- [51] Lin Y., Duan L., Xu P., Li X., Gu R., Luo Y. (2019) **Electrophysiological indexes of option characteristic processing** *Psychophysiology* **56**
- [52] Bland A. R., Schaefer A. (2011) **Electrophysiological correlates of decision making under varying levels of uncertainty** *Brain research* **1417**:55–66

- [53] Botelho C., Fernandes C., Campos C., Seixas C., Pasion R., Garcez H., Ferreira-Santos F., Barbosa F., Maques-Teixeira J., Paiva T. O. (2023) **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* :1–21
- [54] Sun S., Zhen S., Fu Z., Wu D.-A., Shimojo S., Adolphs R., Yu R., Wang S. (2017) **Decision ambiguity is mediated by a late positive potential originating from cingulate cortex** *NeuroImage* **157**:400–414
- [55] Zheng Y., Yi W., Cheng J., Li Q. (2020) **Common and distinct electrophysiological correlates of feedback processing during risky and ambiguous decision making** *Neuropsychologia* **146**
- [56] Guo Z., Chen J., Liu S., Li Y., Sun B., Gao Z. (2013) **Brain areas activated by uncertain reward-based decisionmaking in healthy volunteers** *Neural regeneration research* **8**:3344–3352
- [57] Li Y., Vanni-Mercier G., Isnard J., Mauguiere F., Dreher J.-C. (2016) **The neural dynamics of reward value and risk coding in the human orbitofrontal cortex** *Brain* **139**:1295–1309
- [58] Rolls E. T., Wan Z., Cheng W., Feng J. (2022) **Risk-taking in humans and the medial orbitofrontal cortex reward system** *NeuroImage* **249**
- [59] Hoenig K., Scheef L. (2005) **Mediotemporal contributions to semantic processing: fmri evidence from ambiguity processing during semantic context verification** *Hippocampus* **15**:597–609
- [60] Vitello S., Warren J. E., Devlin J. T., Rodd J. M. (2014) **Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences** *Frontiers in human neuroscience* **8**
- [61] Emadi N., Esteky H. (2013) **Neural representation of ambiguous visual objects in the inferior temporal cortex** *PloS one* **8**
- [62] McIntosh A. R., Rajah M. N., Lobaugh N. J. (2003) **Functional connectivity of the medial temporal lobe relates to learning and awareness** *Journal of Neuroscience* **23**:6520–6528
- [63] Palombo D. J., Keane M. M., Verfaellie M. (2015) **The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking** *Hippocampus* **25**:345–353
- [64] Okuda J., Fujii T., Ohtake H., Tsukiura T., Tanji K., Suzuki K., Kawashima R., Fukuda H., Itoh M., Yamadori A. (2003) **Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes** *Neuroimage* **19**:1369–1380
- [65] Palombo D., Keane M., Verfaellie M. (2016) **Using future thinking to reduce temporal discounting: Under what circumstances are the medial temporal lobes critical?** *Neuropsychologia* **89**:437–444
- [66] Schacter D. L., Addis D. R. (2009) **On the nature of medial temporal lobe contributions to the constructive simulation of future events** *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**:1245–1253

- [67] Krain A. L., Wilson A. M., Arbuckle R., Castellanos F. X., Milham M. P. (2006) **Distinct neural mechanisms of risk and ambiguity: a meta-analysis of decision-making** *Neuroimage* **32**:477–484
- [68] Peng X., Burwell R. D. (2021) **Beyond the hippocampus: the role of parahippocampal-prefrontal communication in context-modulated behavior** *Neurobiology of learning and memory* **185**
- [69] Williams T. B., Burke C. J., Nebe S., Preuschoff K., Fehr E., Tobler P. N. (2021) **Testing models at the neural level reveals how the brain computes subjective value** *Proceedings of the National Academy of Sciences* **118**

Editors

Reviewing Editor

Maria Chait

University College London, London, United Kingdom

Senior Editor

Michael Frank

Brown University, Providence, United States of America

Reviewer #1 (Public Review):

Summary:

This paper presents a compelling and comprehensive study of decision-making under uncertainty. It addresses a fundamental distinction between belief-based (cognitive neuroscience) formulations of choice behavior with reward-based (behavioral psychology) accounts. Specifically, it asks whether active inference provides a better account of planning and decision making, relative to reinforcement learning. To do this, the authors use a simple but elegant paradigm that includes choices about whether to seek both information and rewards. They then assess the evidence for active inference and reinforcement learning models of choice behavior, respectively. After demonstrating that active inference provides a better explanation of behavioral responses, the neuronal correlates of epistemic and instrumental value (under an optimized active inference model) are characterized using EEG. Significant neuronal correlates of both kinds of value were found in sensor and source space. The source space correlates are then discussed sensibly, in relation to the existing literature on the functional anatomy of perceptual and instrumental decision-making under uncertainty.

<https://doi.org/10.7554/eLife.92892.2.sa3>

Reviewer #2 (Public Review):

Summary:

Zhang and colleagues use a combination of behavioral, neural, and computational analyses to test an active inference model of exploration in a novel reinforcement learning task.

Strengths:

The paper addresses an important question (validation of active inference models of exploration). The combination of behavior, neuroimaging, and modeling is potentially powerful for answering this question.

I appreciate the addition of details about model fitting, comparison, and recovery, as well as the change in some of the methods.

Weaknesses:

The authors do not cite what is probably the most relevant contextual bandit study, by Collins & Frank (2018, PNAS), which uses EEG.

The authors cite Collins & Molinaro as a form of contextual bandit, but that's not the case (what they call "context" is just the choice set). They should look at the earlier work from Collins, starting with Collins & Frank (2012, EPN).

Placing statistical information in a GitHub repository is not appropriate. This needs to be in the main text of the paper. I don't understand why the authors refer to space limitations; there are none for eLife, as far as I'm aware.

In answer to my question about multiple comparisons, the authors have added the following: "Note that we did not attempt to correct for multiple comparisons; largely, because the correlations observed were sustained over considerable time periods, which would be almost impossible under the null hypothesis of no correlations." I'm sorry, but this does not make sense. Either the authors are doing multiple comparisons, in which case multiple comparison correction is relevant, or they are doing a single test on the extended timeseries, in which case they need to report that. There exist tools for this kind of analysis (e.g., Gershman et al., 2014, NeuroImage). I'm not suggesting that the authors should necessarily do this, only that their statistical approach should be coherent. As a reference point, the authors might look at the aforementioned Collins & Frank (2018) study.

I asked the authors to show more descriptive comparison between the model and the data. Their response was that this is not possible, which I find odd given that they are able to use the model to define a probability distribution on choices. All I'm asking about here is to show predictive checks which build confidence in the model fit. The additional simulations do not address this. The authors refer to figures 3 and 4, but these do not show any direct comparison between human data and the model beyond model comparison metrics.

<https://doi.org/10.7554/eLife.92892.2.sa2>

Reviewer #3 (Public Review):

Summary:

This paper aims to investigate how the human brain represents different forms of value and uncertainty that participate in active inference within a free-energy framework, in a two-stage decision task involving contextual information sampling, and choices between safe and risky rewards, which promotes shifting between exploration and exploitation. They examine neural correlates by recording EEG and comparing activity in the first vs second half of trials and between trials in which subjects did and did not sample contextual information, and perform a regression with free-energy-related regressors against data "mapped to source space."

Strengths:

This two-stage paradigm is cleverly designed to incorporate several important processes of learning, exploration/exploitation and information sampling that pertain to active inference. Although scalp/brain regions showing sensitivity to the active-inference related quantities do not necessarily suggest what role they play, they are illuminating and useful as candidate regions for further investigation. The aims are ambitious, and the methodologies impressive. The paper lays out an extensive introduction to the free energy principle and active inference to make the findings accessible to a broad readership.

Weaknesses:

In its revised form the paper is complete in providing the important details. Though not a serious weakness, it is important to note that the high lower-cutoff of 1 Hz in the bandpass filter, included to reduce the impact of EEG noise, would remove from the EEG any sustained, iteratively updated representation that evolves with learning across trials, or choice-related processes that unfold slowly over the course of the 2-second task windows.

<https://doi.org/10.7554/eLife.92892.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:

Reviewer #1 (Public Review):

Summary:

This paper presents a compelling and comprehensive study of decision-making under uncertainty. It addresses a fundamental distinction between belief-based (cognitive neuroscience) formulations of choice behaviour with reward-based (behavioural psychology) accounts. Specifically, it asks whether active inference provides a better account of planning and decision-making, relative to reinforcement learning. To do this, the authors use a simple but elegant paradigm that includes choices about whether to seek both information and rewards. They then assess the evidence for active inference and reinforcement learning models of choice behaviour, respectively. After demonstrating that active inference provides a better explanation of behavioural responses, the neuronal correlates of epistemic and instrumental value (under an optimised active inference model) are characterised using EEG. Significant neuronal correlates of both kinds of value were found in sensor and source space. The source space correlates are then discussed sensibly, in relation to the existing literature on the functional anatomy of perceptual and instrumental decision-making under uncertainty.

Strengths:

The strengths of this work rest upon the theoretical underpinnings and careful deconstruction of the various determinants of choice behaviour using active inference. A particular strength here is that the experimental paradigm is designed carefully to elicit both information-seeking and reward-seeking behaviour; where the information-seeking is itself separated into resolving uncertainty about the context (i.e., latent states) and the contingencies (i.e., latent parameters), under which choices are made. In other words, the paradigm - and its subsequent modelling - addresses both inference and learning as necessary belief and knowledge-updating processes that underwrite decisions.

The authors were then able to model belief updating using active inference and then look for the neuronal correlates of the implicit planning or policy selection. This speaks to a

further strength of this study; it provides some construct validity for the modelling of belief updating and decision-making; in terms of the functional anatomy as revealed by EEG. Empirically, the source space analysis of the neuronal correlates licences some discussion of functional specialisation and integration at various stages in the choices and decision-making.

In short, the strengths of this work rest upon a (first) principles account of decision-making under uncertainty in terms of belief updating that allows them to model or fit choice behaviour in terms of Bayesian belief updating - and then use relatively state-of-the-art source reconstruction to examine the neuronal correlates of the implicit cognitive processing.

Response: We are deeply grateful for your careful review of our work and for the thoughtful feedback you have provided. Your dedication to ensuring the quality and clarity of the work is truly admirable. Your comments have been invaluable in guiding us towards improving the paper, and We appreciate your time and effort in not just offering suggestions but also providing specific revisions that I can implement. Your insights have helped us identify areas where I can strengthen the arguments and clarify the methodology.

Comment 1:

The main weaknesses of this report lies in the communication of the ideas and procedures. Although the language is generally excellent, there are some grammatical lapses that make the text difficult to read. More importantly, the authors are not consistent in their use of some terms; for example, uncertainty and information gain are sometimes conflated in a way that might confuse readers. Furthermore, the descriptions of the modelling and data analysis are incomplete. These shortcomings could be addressed in the following way.

First, it would be useful to unpack the various interpretations of information and goal-seeking offered in the (active inference) framework examined in this study. For example, it will be good to include the following paragraph:

"In contrast to behaviourist approaches to planning and decision-making, active inference formulates the requisite cognitive processing in terms of belief updating in which choices are made based upon their expected free energy. Expected free energy can be regarded as a universal objective function, specifying the relative likelihood of alternative choices. In brief, expected free energy can be regarded as the surprise expected following some action, where the expected surprise comes in two flavours. First, the expected surprise is uncertainty, which means that policies with a low expected free energy resolve uncertainty and promote information seeking. However, one can also minimise expected surprise by avoiding surprising, aversive outcomes. This leads to goal-seeking behaviour, where the goals can be regarded as prior preferences or rewarding outcomes.

Technically, expected free energy can be expressed in terms of risk plus ambiguity - or rearranged to be expressed in terms of expected information gain plus expected value, where value corresponds to (log) prior preferences. We will refer to both decompositions in what follows; noting that both decompositions accommodate information and goal-seeking imperatives. That is, resolving ambiguity and maximising information gain have epistemic value, while minimising risk or maximising expected value have pragmatic or instrumental value. These two kinds of values are sometimes referred to in terms of intrinsic and extrinsic value, respectively [1-4]."

Response 1: We deeply thank you for your comments and corresponding suggestions about our interpretations of active inference. In response to your identified weaknesses and

suggestions, we have added corresponding paragraphs in the Methods section (The free energy principle and active inference, line 95-106):

“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 (Oudeyer et al., 2007; Schmidhuber et al., 2010). 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 ambiguity, minimizing risk, and maximizing information gain has epistemic value while maximizing expected value have pragmatic or instrumental value. These two types of values can be referred to in terms of intrinsic and extrinsic value, respectively (Barto et al., 2013; Schwartenbeck et al., 2019).”

Oudeyer, P. Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in neurorobotics*, 1, 108.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE transactions on autonomous mental development*, 2(3), 230-247.

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise?. *Frontiers in psychology*, 4, 61898.

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, 8, e41703.

Comment 2:

The description of the modelling of choice behaviour needs to be unpacked and motivated more carefully. Perhaps along the following lines:

"To assess the evidence for active inference over reinforcement learning, we fit active inference and reinforcement learning models to the choice behaviour of each subject. Effectively, this involved optimising the free parameters of active inference and reinforcement learning models to maximise the likelihood of empirical choices. The resulting (marginal) likelihood was then used as the evidence for each model. The free parameters for the active inference model scaled the contribution of the three terms that constitute the expected free energy (in Equation 6). These coefficients can be regarded as precisions that characterise each subjects' prior beliefs about contingencies and rewards. For example, increasing the precision or the epistemic value associated with model parameters means the subject would update her beliefs about reward contingencies more quickly than a subject who has precise prior beliefs about reward distributions. Similarly, subjects with a high precision over prior preferences or extrinsic value can be read as having more precise beliefs that she will be rewarded. The free parameters for the reinforcement learning model included..."

Response 2: We deeply thank you for your comments and corresponding suggestions about our description of the behavioral modelling. In response to your identified weaknesses and suggestions, we have added corresponding content in the Results section (Behavioral results, line 279-293):

“To assess the evidence for active inference over reinforcement learning, we fit 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) (Vrieze 2012) as the evidence for each model. The free parameters for the active inference model (AL, AI, EX, prior, and α) scaled the contribution of the three terms that constitute the expected free energy in Eq.9. These coefficients can be regarded as precisions that characterize each participant's prior beliefs about contingencies and rewards. For example, increasing α means participants would update their beliefs about reward contingencies more quickly, increasing AL means participants would like to reduce ambiguity more, and increasing AI means participants would like to learn the hidden state of the environment and avoid risk more. The free parameters for the model-free reinforcement learning model are the learning rate α and the temperature parameter γ and the free parameters for the model-based are the learning rate α , the temperature parameter γ and prior (the details for the model-free reinforcement learning model can be seen in Eq.S1-11 and the details for the model-based reinforcement learning model can be seen Eq.S12-23 in the Supplementary Method). The parameter fitting for these three models was conducted using the BayesianOptimization' package in Python (Frazier 2018), first randomly sampling 1000 times and then iterating for an additional 1000 times.”

$$G(\pi, \tau) = AL \cdot E_Q[\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)] + AI \cdot E_Q[\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)] - EX \cdot E_Q[\ln P(o_\tau)] \quad (9)$$

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods*, 17(2), 228.

Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Comment 3:

In terms of the time-dependent correlations with expected free energy - and its constituent terms - I think the report would benefit from overviews of these analyses with something like the following:

"In the final analysis of the neuronal correlates of belief updating - as quantified by the epistemic and intrinsic values of expected free energy - we present a series of analyses in source space. These analyses tested for correlations between constituent terms in expected free energy and neuronal responses in source space. These correlations were over trials (and subjects). Because we were dealing with two-second timeseries, we were able to identify the periods of time during decision-making when the correlates were expressed.

In these analyses, we focused on the induced power of neuronal activity at each point in time, at each brain source. To illustrate the functional specialisation of these neuronal correlates, we present whole-brain maps of correlation coefficients and pick out the most significant correlation for reporting fluctuations in selected correlations over two-second periods. These analyses are presented in a descriptive fashion to highlight the nature and variety of the neuronal correlates, which we unpack in relation to the existing EEG literature in the discussion. Note that we did not attempt to correct for multiple comparisons; largely, because the correlations observed were sustained over considerable time periods, which would be almost impossible under the null hypothesis of no correlations."

Response 3: We deeply thank you for your comments and corresponding suggestions about our description of the regression analysis in the source space. In response to your suggestions, we have added corresponding content in the Results section (EEG results at source level, line 331-347):

“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 correlations over trials between constituent terms in expected free energy (the value of avoiding risk, the value of reducing ambiguity, extrinsic value, and expected free energy itself) and neural responses in source space. Additionally, we also investigated the neural correlate of (the degree of) risk, (the degree of) ambiguity, 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 ~ 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 ambiguity, 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. Note that we did not attempt to correct for multiple comparisons; largely, because the correlations observed were sustained over considerable time periods, which would be almost impossible under the null hypothesis of no correlations.”

Comment 4:

There was a slight misdirection in the discussion of priors in the active inference framework. The notion that active inference requires a pre-specification of priors is a common misconception. Furthermore, it misses the point that the utility of Bayesian modelling is to identify the priors that each subject brings to the table. This could be easily addressed with something like the following in the discussion:

"It is a common misconception that Bayesian approaches to choice behaviour (including active inference) are limited by a particular choice of priors. As illustrated in our fitting of choice behaviour above, priors are a strength of Bayesian approaches in the following sense: under the complete class theorem [5, 6], any pair of choice behaviours 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 choice behaviour in terms of some priors. This means that one can, in principle, characterise any given behaviour in terms of the priors that explain that behaviour. In our example, these were effectively priors over the precision of various preferences or beliefs about contingencies that underwrite expected free energy."

Response 4: We deeply thank you for your comments and corresponding suggestions about the prior of Bayesian methods. In response to your suggestions, we have added corresponding content in the Discussion section (The strength of the active inference framework in decision-making, line 447-453):

“However, it may be the opposite. As illustrated in our fitting results, priors can be a strength of Bayesian approaches. Under the complete class theorem (Wald 1947; Brown 1981), 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.”

Wald, A. (1947). An essentially complete class of admissible decision functions. *The Annals of Mathematical Statistics*, 549-555.

Brown, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *The Annals of Statistics*, 1289-1300.

Reviewer #2 (Public Review):

Summary:

Zhang and colleagues use a combination of behavioral, neural, and computational analyses to test an active inference model of exploration in a novel reinforcement learning task.

Strengths:

The paper addresses an important question (validation of active inference models of exploration). The combination of behavior, neuroimaging, and modeling is potentially powerful for answering this question.

Response: We want to express our sincere gratitude for your thorough review of our work and for the valuable comments you have provided. Your attention to detail and dedication to improving the quality of the work are truly commendable. Your feedback has been invaluable in guiding us towards revisions that will strengthen the work. We have made targeted modifications based on most of the comments. However, due to factors such as time and energy constraints, we have not added corresponding analyses for several comments.

Comment 1:

The paper does not discuss relevant work on contextual bandits by Schulz, Collins, and others. It also does not mention the neuroimaging study of Tomov et al. (2020) using a risky/safe bandit task.

Response 1:

We deeply thank you for your suggestions about the relevant work. We now discuss and cite these representative papers in the Introduction section (line 42-55):

“The decision-making process frequently involves grappling with varying forms of uncertainty, such as ambiguity - the kind of uncertainty that can be reduced through sampling, and risk - 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 (Daw et al., 2006; Badre et al., 2012; Cavanagh et al., 2012).

These studies utilized different forms of multi-armed bandit tasks, e.g the restless multi-armed bandit tasks (Daw et al., 2006; Guha et al., 2010), risky/safe bandit tasks (Tomov et al., 2020; Fan et al., 2022; Payzan et al., 2013), contextual multi-armed bandit tasks (Schulz et al., 2015; Schulz et al., 2015; Molinaro et al., 2023). However, these tasks either separate risk from ambiguity in uncertainty, or separate action from state (perception). In our work, we develop a contextual multi-armed bandit task to enable participants to actively reduce ambiguity,

avoid risk, 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 (Levy et al., 2010) 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 ambiguity and risk. Ambiguity is represented by the uncertainty about model parameters associated with choosing a particular action, while risk is signified by the variance of the environment's hidden states. The value of reducing ambiguity, the value of avoiding risk, and extrinsic value together constitute expected free energy (see Section 2.1)."

Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095), 876-879.

Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron*, 73(3), 595-607.

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral cortex*, 22(11), 2575-2586.

Guha, S., Munagala, K., & Shi, P. (2010). Approximation algorithms for restless bandit problems. *Journal of the ACM (JACM)*, 58(1), 1-50.

Tomov, M. S., Truong, V. Q., Hundia, R. A., & Gershman, S. J. (2020). Dissociable neural correlates of uncertainty underlie different exploration strategies. *Nature communications*, 11(1), 2371.

Fan, H., Gershman, S. J., & Phelps, E. A. (2023). Trait somatic anxiety is associated with reduced directed exploration and underestimation of uncertainty. *Nature Human Behaviour*, 7(1), 102-113.

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, J. P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron*, 79(1), 191-201.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2015, April). Exploration-exploitation in a contextual multi-armed bandit task. In *International conference on cognitive modeling* (pp. 118-123).

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2015, November). Learning and decisions in contextual multi-armed bandit tasks. In *CogSci*.

Molinaro, G., & Collins, A. G. (2023). Intrinsic rewards explain context-sensitive valuation in reinforcement learning. *PLoS Biology*, 21(7), e3002201.

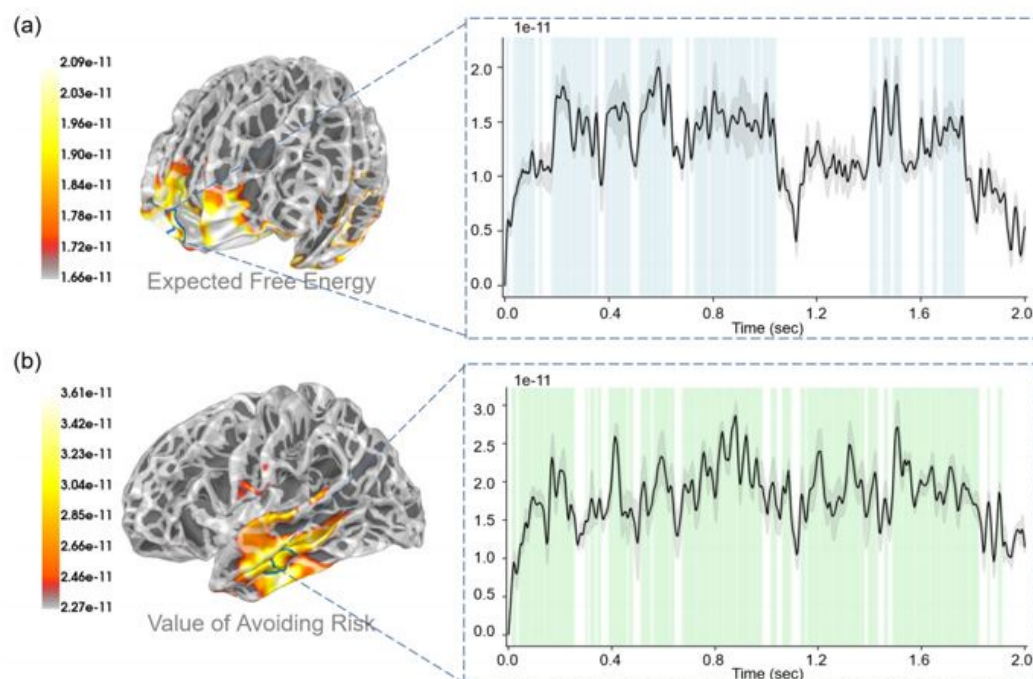
Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *Journal of neurophysiology*, 103(2), 1036-1047.

Comment 2:

The statistical reporting is inadequate. In most cases, only p-values are reported, not the relevant statistics, degrees of freedom, etc. It was also not clear if any corrections for multiple comparisons were applied. Many of the EEG results are described as "strong" or "robust" with significance levels of $p < 0.05$; I am skeptical in the absence of more details, particularly given the fact that the corresponding plots do not seem particularly strong to me.

Response 2: We deeply thank you for your comments about our statistical reporting. We have optimized the fitting model and rerun all the statistical analyses. As can be seen (Figure 6, 7, 8, S3, S4, S5), the new regression results are significantly improved compared to the previous ones. Due to the limitation of space, we place the other relevant statistical results, including t-values, std err, etc., on our GitHub (<https://github.com/andlab-um/FreeEnergyEEG>). Currently, we have not conducted multiple comparison corrections based on Reviewer 1's comments (Comments 3) "Note that we did not attempt to correct for multiple comparisons; largely, because the correlations observed were sustained over considerable time periods, which would be almost impossible under the null hypothesis of no correlations".

Author response image 1.



Comment 3:

The authors compare their active inference model to a "model-free RL" model. This model is not described anywhere, as far as I can tell. Thus, I have no idea how it was fit, how many parameters it has, etc. The active inference model fitting is also not described anywhere. Moreover, you cannot compare models based on log-likelihood, unless you are talking about held-out data. You need to penalize for model complexity. Finally, even if active inference outperforms a model-free RL model (doubtful given the error bars in Fig. 4c), I don't see how this is strong evidence for active inference per se. I would want to see a much more extensive model comparison, including model-based RL algorithms which are not based on active inference, as well as model recovery analyses confirming that the models can actually be distinguished on the basis of the experimental data.

Response 3: We deeply thank you for your comments about the model comparison details. We previously omitted some information about the comparison model, as classical reinforcement learning is not the focus of our work, so we put the specific details in the supplementary materials. Now we have placed relevant information in the main text (see the part we have highlighted in yellow). We have now added the relevant information regarding the model comparison in the Results section (Behavioral results, line 279-293):

“To assess the evidence for active inference over reinforcement learning, we fit 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) as the evidence for each model. The free parameters for the active inference model (AL, AI, EX, prior, and α) scaled the contribution of the three terms that constitute the expected free energy in Eq.9. These coefficients can be regarded as precisions that characterize each participant's prior beliefs about contingencies and rewards. For example, increasing α means participants would update their beliefs about reward contingencies more quickly, increasing AL means participants would like to reduce ambiguity more, and increasing AI means participants would like to learn the hidden state of the environment and avoid risk more. The free parameters for the model-free reinforcement learning model are the learning rate α and the temperature parameter γ and the free parameters for the model-based are 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 in Python, first randomly sampling 1000 times and then iterating for an additional 1000 times.”

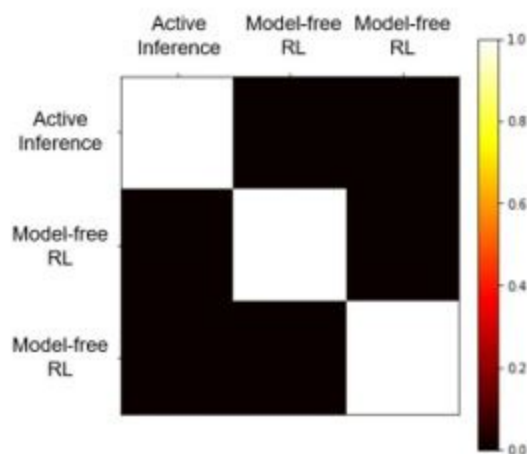
We have now incorporated model-based reinforcement learning into our comparison models and placed the descriptions of both model-free and model-based reinforcement learning algorithms in the supplementary materials. We have also changed the criterion for model comparison to Bayesian Information Criterion. As indicated by the results, the performance of the active inference model significantly outperforms both comparison models.

Sorry, we didn't do model recovery before, but now we have placed the relevant results in the supplementary materials. From the result figures, we can see that each model fits its own generated simulated data well:

“To demonstrate how reliable our models are (the active inference model, model-free reinforcement learning model, and model-based reinforcement learning model), we run some simulation experiments for model recovery. We use these three models, with their own fitting parameters, to generate some simulated data. Then we will fit all three sets of data using these three models.

The model recovery results are shown in Fig.S6. This is the confusion matrix of models: the percentage of all subjects simulated based on a certain model that is fitted best by a certain model. The goodness-of-fit was compared using the Bayesian Information Criterion. We can see that the result of model recovery is very good, and the simulated data generated by a model can be best explained by this model.”

Author response image 2.



Comment 4:

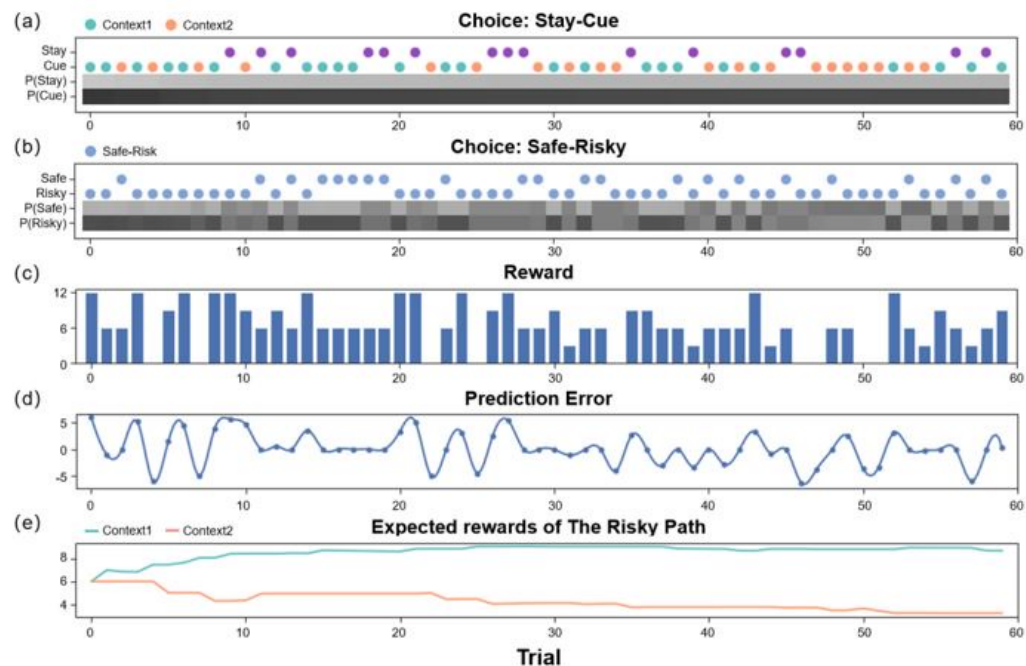
Another aspect of the behavioral modeling that's missing is a direct descriptive comparison between model and human behavior, beyond just plotting log-likelihoods (which are a very impoverished measure of what's going on).

Response 4: We deeply thank you for your comments about the comparison between the model and human behavior. Due to the slight differences between our simulation experiments and real behavioral experiments (the "you can ask" stage), we cannot directly compare the model and participants' behaviors. However, we can observe that in the main text's simulation experiment (Figure 3), the active inference agent's behavior is highly consistent with humans (Figure 4), exhibiting an effective exploration strategy and a desire to reduce uncertainty. Moreover, we have included two additional simulation experiments in the supplementary materials, which demonstrate that active inference may potentially fit a wide range of participants' behavioral strategies.

$$G(\pi, \tau) = AL \cdot E_{\hat{Q}}[\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)] + AI \cdot E_{\hat{Q}}[\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)] - EX \cdot E_{\hat{Q}}[\ln P(o_\tau)] \quad (9)$$

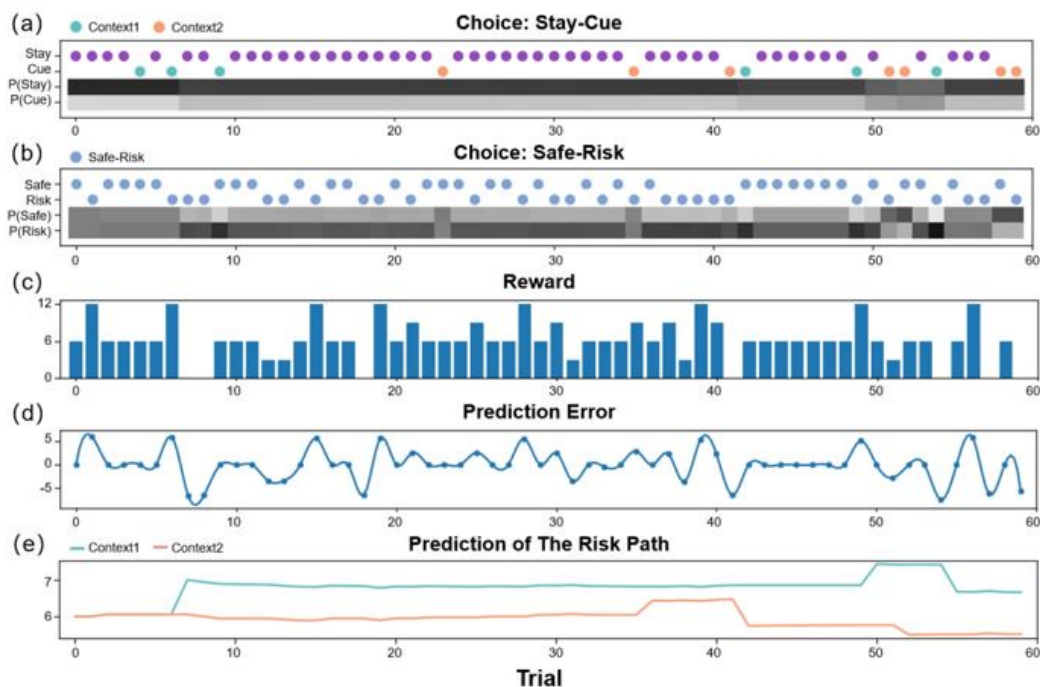
Author response image 3.

(An active inference agent with AL=AI=EX=0. It can accomplish tasks efficiently like a human being, reducing the uncertainty of the environment and maximizing the reward.)



Author response image 4.

(An active inference agent with $AL=AI=0$, $EX=10$. It will only pursue immediate rewards (not choosing the "Cue" option due to additional costs), but it can also gradually optimize its strategy due to random effects.)



Author response image 5.

(An active inference agent with $EX=0$, $AI=AL=10$. It will only pursue environmental information to reduce the uncertainty of the environment. Even in "Context 2" where immediate rewards are scarce, it will continue to explore.)

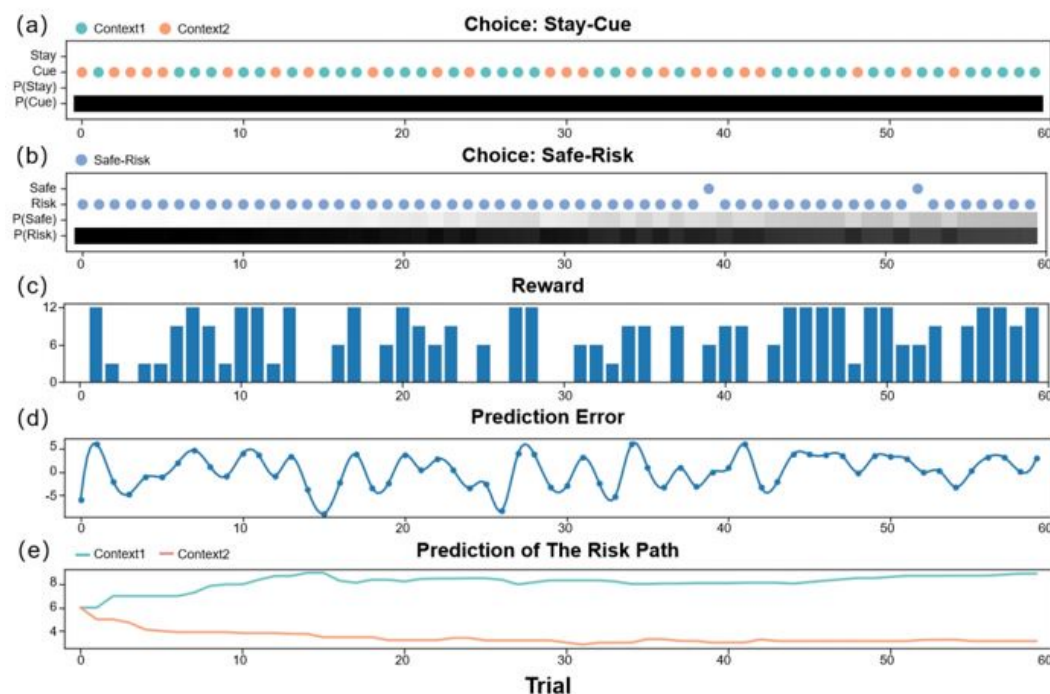
Figure (a) shows the decision-making of active inference agents in the Stay-Cue choice. Blue corresponds to agents choosing the "Cue" option and acquiring "Context 1"; orange corresponds to agents choosing the "Cue" option and acquiring "Context 2"; purple corresponds to agents choosing the "Stay" option and not knowing the information about the hidden state of the environment. The shaded areas below correspond to the probability of the agents making the respective choices.

Figure (b) shows the decision-making of active inference agents in the Stay-Cue choice. The shaded areas below correspond to the probability of the agents making the respective choices.

Figure (c) shows the rewards obtained by active inference agents.

Figure (d) shows the reward prediction errors of active inference agents.

Figure (e) shows the reward predictions of active inference agents for the "Risky" path in "Context 1" and "Context 2".



Comment 5:

The EEG results are intriguing, but it wasn't clear that these provide strong evidence specifically for the active inference model. No alternative models of the EEG data are evaluated.

Overall, the central claim in the Discussion ("we demonstrated that the active inference model framework effectively describes real-world decision-making") remains unvalidated

| *in my opinion.*

Response 5: We deeply thank you for your comments. We applied the active inference model to analyze EEG results because it best fit the participants' behavioral data among our models, including the new added results. Further, our EEG results serve only to verify that the active inference model can be used to analyze the neural mechanisms of decision-making in uncertain environments (if possible, we could certainly design a more excellent reinforcement learning model with a similar exploration strategy). We aim to emphasize the consistency between active inference and human decision-making in uncertain environments, as we have discussed in the article. Active inference emphasizes both perception and action, which is also what we wish to highlight: during the decision-making process, participants not only passively receive information, but also actively adopt different strategies to reduce uncertainty and maximize rewards.

Reviewer #3 (Public Review):

Summary:

This paper aims to investigate how the human brain represents different forms of value and uncertainty that participate in active inference within a free-energy framework, in a two-stage decision task involving contextual information sampling, and choices between safe and risky rewards, which promotes a shift from exploration to exploitation. They examine neural correlates by recording EEG and comparing activity in the first vs second half of trials and between trials in which subjects did and did not sample contextual information, and perform a regression with free-energy-related regressors against data "mapped to source space." Their results show effects in various regions, which they take to indicate that the brain does perform this task through the theorised active inference scheme.

Strengths:

This is an interesting two-stage paradigm that incorporates several interesting processes of learning, exploration/exploitation, and information sampling. Although scalp/brain regions showing sensitivity to the active-inference-related quantities do not necessarily suggest what role they play, it can be illuminating and useful to search for such effects as candidates for further investigation. The aims are ambitious, and methodologically it is impressive to include extensive free-energy theory, behavioural modelling, and EEG source-level analysis in one paper.

Response: We would like to express our heartfelt thanks to you for carefully reviewing our work and offering insightful feedback. Your attention to detail and commitment to enhancing the overall quality of our work are deeply admirable. Your input has been extremely helpful in guiding us through the necessary revisions to enhance the work. We have implemented focused changes based on a majority of your comments. Nevertheless, owing to limitations such as time and resources, we have not included corresponding analyses for a few comments.

Comment 1:

Though I could surmise the above general aims, I could not follow the important details of what quantities were being distinguished and sought in the EEG and why. Some of this is down to theoretical complexity - the dizzying array of constructs and terms with complex interrelationships, which may simply be part and parcel of free-energy-based theories of active inference - but much of it is down to missing or ambiguous details.

Response 1: We deeply thank you for your comments about our work's readability. We have significantly revised the descriptions of active inference, models, research questions, etc. Focusing on active inference and the free energy principle, we have added relevant basic descriptions and unified the terminology. We have added information related to model comparison in the main text and supplementary materials. We presented our regression results in clearer language. Our research focused on the brain's representation of decision-making in uncertain environments, including expected free energy, the value of reducing ambiguity, the value of avoiding risk, extrinsic value, ambiguity, and risk.

Comment 2:

In general, an insufficient effort has been made to make the paper accessible to readers not steeped in the free energy principle and active inference. There are critical inconsistencies in key terminology; for example, the introduction states that aim 1 is to distinguish the EEG correlates of three different types of uncertainty: ambiguity, risk, and unexpected uncertainty. But the abstract instead highlights distinctions in EEG correlates between "uncertainty... and... risk" and between "expected free energy .. and ... uncertainty." There are also inconsistencies in mathematical labelling (e.g. in one place 'p(s|o)' and 'q(s)' swap their meanings from one sentence to the very next).

Response 2: We deeply thank you for your comments about the problem of inconsistent terminology. First, we have unified the symbols and letters (P, Q, s, o, etc.) that appeared in the article and described their respective meanings more clearly. We have also revised the relevant expressions of "uncertainty" throughout the text. In our work, uncertainty refers to ambiguity and risk. Ambiguity can be reduced through continuous sampling and is referred to as uncertainty about model parameters in our work. Risk, on the other hand, is the inherent variance of the environment and cannot be reduced through sampling, which is referred to as uncertainty about hidden states in our work. In the analysis of the results, we focused on how the brain encodes the value of reducing ambiguity (Figure 8), the value of avoiding risk (Figure 6), and (the degree of) ambiguity (Figure S5) during action selection. We also analyzed how the brain encodes reducing ambiguity and avoiding risk during belief update (Figure 7).

$$G(\pi, \tau) = \underbrace{E_{\tilde{Q}} \left[\frac{\ln Q(A) - \ln P(A|s_{\tau}, o_{\tau}, \pi)}{\text{Value of reducing ambiguity}} + \frac{\ln Q(s_{\tau}|\pi) - \ln Q(s_{\tau}|o_{\tau}, \pi)}{\text{Value of avoiding risk}} \right]}_{\text{Epistemic value}} - \underbrace{E_{\tilde{Q}} [\ln P(o_{\tau})]}_{\text{Extrinsic value}} \quad (8)$$

$$G(\pi, \tau) = AL \cdot E_{\tilde{Q}} [\ln Q(A) - \ln P(A|s_{\tau}, o_{\tau}, \pi)] + AI \cdot E_{\tilde{Q}} [\ln Q(s_{\tau}|\pi) - \ln Q(s_{\tau}|o_{\tau}, \pi)] - EX \cdot E_{\tilde{Q}} [\ln P(o_{\tau})] \quad (9)$$

Comment 3:

Some basic but important task information is missing, and makes a huge difference to how decision quantities can be decoded from EEG. For example:

- How do the subjects press the left/right buttons - with different hands or different fingers on the same hand?

Response 3: We deeply thank you for your comments about the missing task information. We have added the relevant content in the Methods section (Contextual two-armed bandit task and Data collection, line 251-253):

“Each stage was 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. The participants are required to use any two fingers of one hand to press the buttons (left arrow and right arrow on the keyboard).”

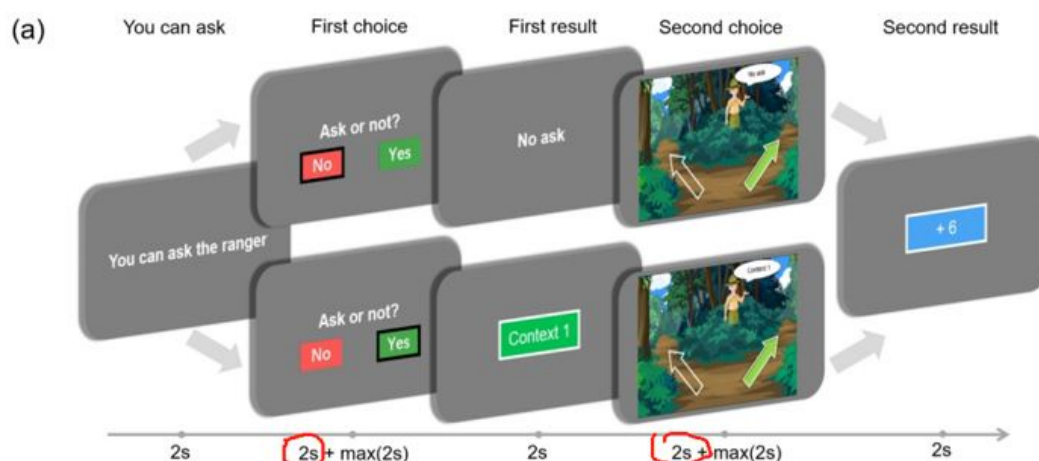
Comment 4:

- Was the presentation of the Stay/cue and safe/risky options on the left/right sides counterbalanced? If not, decisions can be formed well in advance especially once a policy is in place.

Response 4: The presentation of the Stay/cue and safe/risky options on the left/right sides was not counterbalanced. It is true that participants may have made decisions ahead of time. However, to better study the state of participants during decision-making, our choice stages consist of two parts. In the first two seconds, we ask participants to consider which option they would choose, and after these two seconds, participants are allowed to make their choice (by pressing the button).

We also updated the figure of the experiment procedure as below (We circled the time that the participants spent on making decisions).

Author response image 6.



Comment 5:

- What were the actual reward distributions ("magnitude X with probability p , magnitude y with probability $1-p$ ") in the risky option?

Response 5: We deeply thank you for your comments about the missing task information. We have placed the relevant content in the Methods section (Contextual two-armed bandit task and Data collection, line 188-191):

“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%).”

Comment 6:

The EEG analysis is not sufficiently detailed and motivated.

For example,

- why the high lower-filter cutoff of 1 Hz, and shouldn't it be acknowledged that this removes from the EEG any sustained, iteratively updated representation that evolves with learning across trials?

Response 6: We deeply thank you for your comments about our EEG analysis. The 1Hz high-pass filter may indeed filter out some useful information. We chose a 1Hz high-pass filter to filter out most of the noise and prevent the noise from affecting our results analysis. Additionally, there are also many decision-related works that have applied 1Hz high-pass filtering in EEG data preprocessing (Yau et al., 2021; Cortes et al., 2021; Wischniewski et al., 2022; Schutte et al., 2017; Mennella et al., 2020; Giustiniani et al., 2020).

Yau, Y., Hinault, T., Taylor, M., Cisek, P., Fellows, L. K., & Dagher, A. (2021). Evidence and urgency related EEG signals during dynamic decision-making in humans. *Journal of Neuroscience*, 41(26), 5711-5722.

Cortes, P. M., García-Hernández, J. P., Iribe-Burgos, F. A., Hernández-González, M., Sotelo-Tapia, C., & Guevara, M. A. (2021). Temporal division of the decision-making process: An EEG study. *Brain Research*, 1769, 147592.

Wischniewski, M., & Compen, B. (2022). Effects of theta transcranial alternating current stimulation (tACS) on exploration and exploitation during uncertain decision-making. *Behavioural Brain Research*, 426, 113840.

Schutte, I., Kenemans, J. L., & Schutter, D. J. (2017). Resting-state theta/beta EEG ratio is associated with reward-and punishment-related reversal learning. *Cognitive, Affective, & Behavioral Neuroscience*, 17, 754-763.

Mennella, R., Vilarem, E., & Grèzes, J. (2020). Rapid approach-avoidance responses to emotional displays reflect value-based decisions: Neural evidence from an EEG study. *NeuroImage*, 222, 117253.

Giustiniani, J., Nicolier, M., Teti Mayer, J., Chabin, T., Masse, C., Galmès, N., ... & Gabriel, D. (2020). Behavioral and neural arguments of motivational influence on decision making during uncertainty. *Frontiers in Neuroscience*, 14, 583.

Comment 7:

- Since the EEG analysis was done using an array of free-energy-related variables in a regression, was multicollinearity checked between these variables?

Response 7: We deeply thank you for your comments about our regression. Indeed, we didn't specify our regression formula in the main text. We conducted regression on one variable each time, so there was no need for a multicollinearity check. We have now added the relevant content in the Results section ("EEG results at source level" section, line 337-340):

"The linear regression was run by the "mne.stats.linear regression" function in the MNE package (Activity ~ 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 ambiguity, etc.)."

Comment 8:

- In the initial comparison of the first/second half, why just 5 clusters of electrodes, and why these particular clusters?

Response 8: We deeply thank you for your comments about our sensor-level analysis. These five clusters are relatively common scalp EEG regions to analyze (left frontal, right frontal, central, left parietal, and right parietal), and we referred previous work analyzed these five clusters of electrodes (Laufs et al., 2006; Ray et al., 1985; Cole et al., 1985). In addition, our work pays more attention to the analysis in source space, exploring the corresponding functions of specific brain regions based on active inference models.

Laufs, H., Holt, J. L., Elfont, R., Krams, M., Paul, J. S., Krakow, K., & Kleinschmidt, A. (2006). Where the BOLD signal goes when alpha EEG leaves. *Neuroimage*, 31(4), 1408-1418.

Ray, W. J., & Cole, H. W. (1985). EEG activity during cognitive processing: influence of attentional factors. *International Journal of Psychophysiology*, 3(1), 43-48.

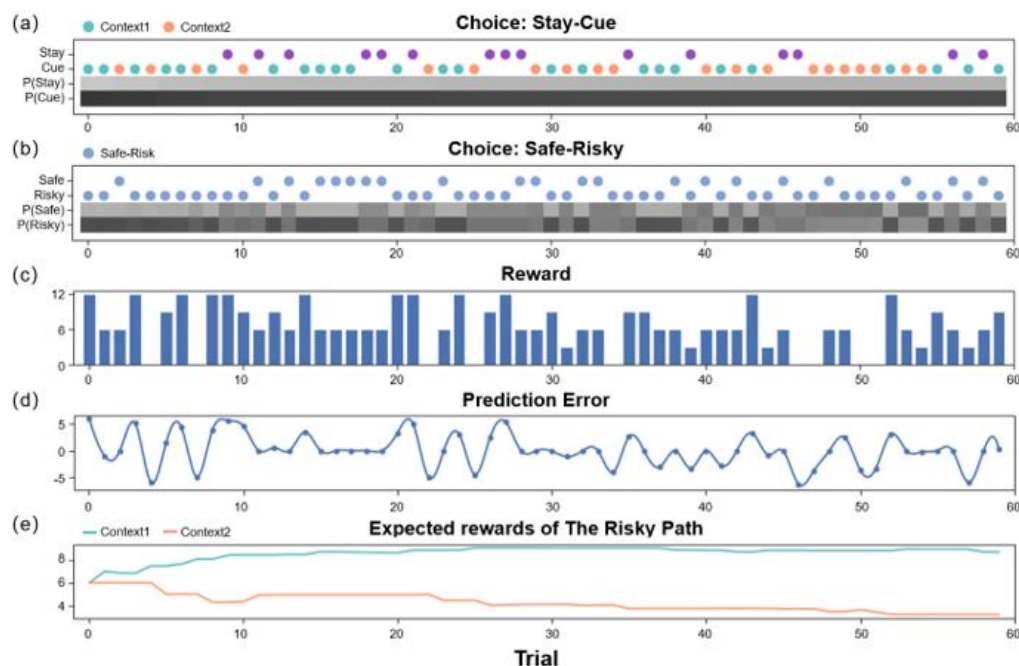
Cole, H. W., & Ray, W. J. (1985). EEG correlates of emotional tasks related to attentional demands. *International Journal of Psychophysiology*, 3(1), 33-41.

Comment 9:

How many different variables are systematically different in the first vs second half, and how do you rule out less interesting time-on-task effects such as engagement or alertness? In what time windows are these amplitudes being measured?

Response 9 (and the Response for Weaknesses 11): There were no systematic differences between the first half and the second half of the trials, with the only difference being the participants' experience. In the second half, participants had a better understanding of the reward distribution of the task (less ambiguity). The simulation results can well describe these.

Author response image 7.



As shown in Figure (a), agents can only learn about the hidden state of the environment ("Context 1" (green) or "Context 2" (orange)) by choosing the "Cue" option. If agents choose the

"Stay" option, they will not be able to know the hidden state of the environment (purple). The risk of agents is only related to wh

ether they choose the "Cue" option, not the number of rounds. Figure (b) shows the Safe-Risky choices of agents, and Figure (e) is the reward prediction of agents for the "Risky" path in "Context 1" and "Context 2". We can see that agents update the expected reward and reduce ambiguity by sampling the "Risky" path. The ambiguity of agents is not related to the "Cue" option, but to the number of times they sample the "Risky" path (rounds).

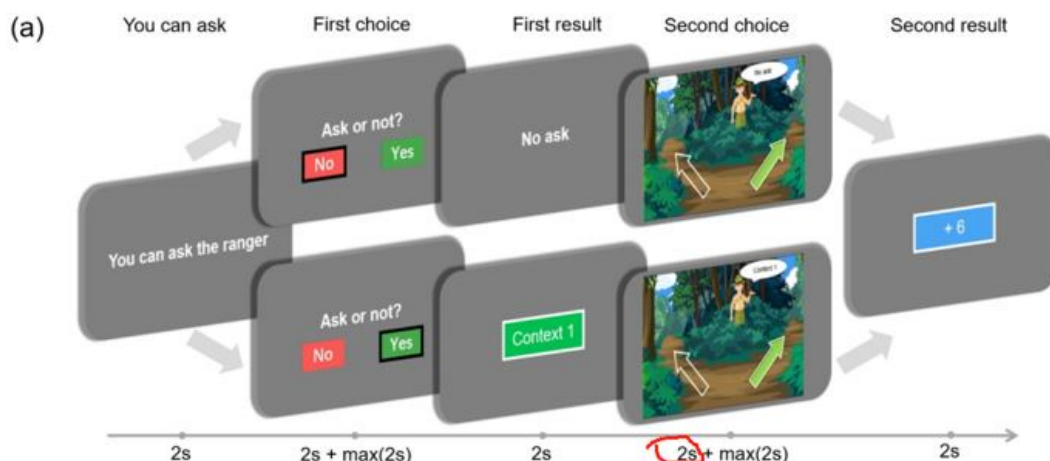
In our choosing stages, participants were required to think about their choices for the first two seconds (during which they could not press buttons). Then, they were asked to make their choices (press buttons) within the next two seconds. This setup effectively kept participants' attention focused on the task. And the two second during the "Second choice" stage when participants decide which option to choose (they cannot press buttons) are measured for the analysis of the sensor-level results.

Comment 10:

In the comparison of asked and not-asked trials, what trial stage and time window is being measured?

Response 10: We have added relevant descriptions in the main text. The two second during the "Second choice" stage when participants decide which option to choose (they cannot press buttons) are measured for the analysis of the sensor-level results.

Author response image 8.



Comment 11:

Again, how many different variables, of the many estimated per trial in the active inference model, are different in the asked and not-asked trials, and how can you know which of these differences is the one reflected in the EEG effects?

Response 11: The difference between asked trials and not-asked trials lies only in whether participants know the specific context of the risky path (the level of risk for the participants). A simple comparison indeed cannot tell us which of these differences is reflected in the EEG

effects. Therefore, we subsequently conducted model-based regression analysis in the source space.

Comment 12:

The authors choose to interpret that on not-asked trials the subjects are more uncertain because the cue doesn't give them the context, but you could equally argue that they don't ask because they are more certain of the possible hidden states.

Response 12: Our task design involves randomly varying the context of the risky path. Only by choosing to inquire can participants learn about the context. Participants can only become increasingly certain about the reward distribution of different contexts of the risky path, but cannot determine which specific context it is. Here are the instructions for the task that we will tell the participants (line 226-231).

"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."

Comment 13:

- The EEG regressors are not fully explained. For example, an "active learning" regressor is listed as one of the 4 at the beginning of section 3.3, but it is the first mention of this term in the paper and the term does not arise once in the methods.

Response 13: We have accordingly revised the relevant content in the main text (as in Eq.8). Our regressors now include expected free energy, the value of reducing ambiguity, the value of avoiding risk, extrinsic value, prediction error, (the degree of) ambiguity, reducing ambiguity, and avoiding risk.

$$G(\pi, \tau) = \underbrace{E_{\tilde{Q}} \left[\frac{\ln Q(A) - \ln P(A|s_{\tau}, o_{\tau}, \pi)}{\text{Value of reducing ambiguity}} + \frac{\ln Q(s_{\tau}|\pi) - \ln Q(s_{\tau}|o_{\tau}, \pi)}{\text{Value of avoiding risk}} \right]}_{\text{Epistemic value}} - \underbrace{E_{\tilde{Q}} [\ln P(o_{\tau})]}_{\text{Extrinsic value}} \quad (8)$$

Comment 14:

- In general, it is not clear how one can know that the EEG results reflect that the brain is purposefully encoding these very parameters while implementing this very mechanism, and not other, possibly simpler, factors that correlate with them since there is no engagement with such potential confounds or alternative models. For example, a model-free reinforcement learning model is fit to behaviour for comparison. Why not the EEG?

Response 14: We deeply thank you for your comments. Due to factors such as time and effort, and because the active inference model best fits the behavioral data of the participants, we did not use other models to analyze the EEG data. At both the sensor and source level, we observed the EEG signal and brain regions that can encode different levels of uncertainties

(risk and ambiguity). The brain's uncertainty driven exploration mechanism cannot be explained solely by a simple model-free reinforcement learning approach.

Recommendations for the authors:

Response: We have made point-to-point revisions according to the reviewer's recommendations, and as these revisions are relatively minor, we have only responded to the longer recommendations here.

Reviewer #1 (Recommendations For The Authors)

I enjoyed reading this sophisticated study of decision-making. I thought your implementation of active inference and the subsequent fitting to choice behaviour - and study of the neuronal (EEG) correlates - was impressive. As noted in my comments on strengths and weaknesses, some parts of your manuscript with difficult to read because of slight collapses in grammar and an inconsistent use of terms when referring to the mathematical quantities. In addition to the paragraphs I have suggested, I would recommend the following minor revisions to your text. In addition, you will have to fill in some of the details that were missing from the current version of the manuscript. For example:

Recommendation 1:

Which RL model did you use to fit the behavioural data? What were its free parameters?

Response 1: We have now added information related to the comparison models in the behavioral results and supplementary materials. We applied both simple model-free reinforcement learning and model-based reinforcement learning. The free parameters for the model-free reinforcement learning model are the learning rate α and the temperature parameter γ , while the free parameters for the model-based approach are the learning rate α , the temperature parameter γ , and the prior.

Recommendation 2:

When you talk about neuronal activity in the final analyses (of time-dependent correlations) what was used to measure the neuronal activity? Was this global power over frequencies? Was it at a particular frequency band? Was it the maximum amplitude within some small window et cetera? In other words, you need to provide the details of your analysis that would enable somebody to reproduce your study at a certain level of detail.

Response 2: In the final analyses, we used the activity amplitude at each point in the source space for our analysis. Previously, we had planned to make our data and models available on GitHub to facilitate easier replication of our work.

Reviewer #3 (Recommendations For The Authors)

Recommendation 1:

It might help to explain the complex concepts up front, to use the concrete example of the task itself - presumably, it was designed so that the crucial elements of the active inference framework come to the fore. One could use hypothetical choice patterns in this task to exemplify different factors such as expected free energy and unexpected uncertainty at work. It would also be illuminating to explain why behaviour on this task is fit better by the active inference model than a model-free reinforcement learning model.

Response 1: Thank you for your suggestions. We have given clearer explanations to the three terms in the active inference formula: the value of reducing ambiguity, the value of avoiding

risk, and the extrinsic value (Eq.8), which makes it easier for readers to understand active inference.

$$G(\pi, \tau) = \underbrace{E_{\tilde{Q}} \left[\frac{\ln Q(A) - \ln P(A|s_\tau, o_\tau, \pi)}{\text{Value of reducing ambiguity}} + \frac{\ln Q(s_\tau|\pi) - \ln Q(s_\tau|o_\tau, \pi)}{\text{Value of avoiding risk}} \right]}_{\text{Epistemic value}} - \underbrace{E_{\tilde{Q}} [\ln P(o_\tau)]}_{\text{Extrinsic value}} \quad (8)$$

In addition, we can simply view active inference as a computational model similar to model-based reinforcement learning, where the expected free energy represents a subjective value, without needing to understand its underlying computational principles or neurobiological background. In our discussion, we have argued why the active inference model fits the participants' behavior better than our reinforcement learning model, as the active inference model has an inherent exploration mechanism that is consistent with humans, who instinctively want to reduce environmental uncertainty (line 435-442).

“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 (Laskin et al., 2020) 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 (Wang et al., 2016) 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 to different task environments (Friston et al., 2017). It can configure different generative models to address distinct tasks, and compute varied forms of free energy and expected free energy.”

Laskin, M., Lee, K., Stooke, A., Pinto, L., Abbeel, P., & Srinivas, A. (2020). Reinforcement learning with augmented data. *Advances in neural information processing systems*, 33, 19884-19895.

Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., ... & Botvinick, M. (2016). Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*.

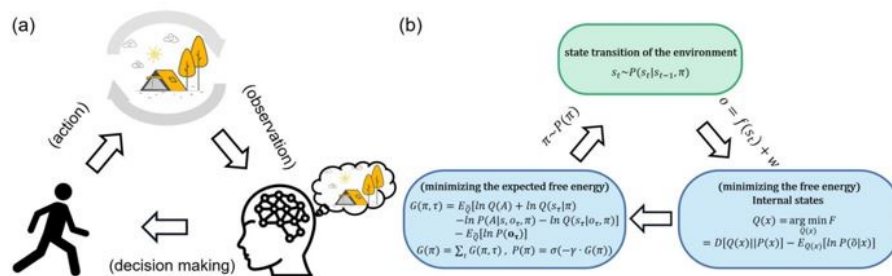
Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. *Neural computation*, 29(1), 1-49.

Recommendation 2:

Figure 1A provides a key example of the lack of effort to help the reader understand. It suggests the possibility of a concrete example but falls short of providing one. From the caption and text, applied to the figure, I gather that by choosing either to run or to raise one's arms, one can control whether it is daytime or nighttime. This is clearly wrong but it is what I am led to think by the paper.

Response 2: Thank you for your suggestion, which we had not considered before. In this figure, we aim to illustrate that "the agent receives observations and optimizes his cognitive model by minimizing variational free energy → the agent makes the optimal action by minimizing expected free energy → the action changes the environment → the environment generates new observations for the agent." We have now modified the image to be simpler to prevent any possible confusion for readers. Correspondingly, we removed the figure of a person raising their hand and the shadowed house in Figure a.

Author response image 9.



Recommendation 3:

I recommend an overhaul in the labelling and methodological explanations for consistency and full reporting. For example, line 73 says sensory input is 's' and the cognitive model is 'q(s),' and the cause of the sensory input is 'p(s | o)' but on the very next line, the cognitive model is 'p(s | o)' and the causes of sensory input are 'q(s).' How this sensory input s relates to 'observations' or 'o' is unclear, and meanwhile, capital S is the set of environmental states. P seems to refer to the generative distribution, but it also means probability.

Response 3: Thank you for your advice. Now we have revised the corresponding labeling and methodological explanations in our work to make them consistent. However, we are not sure how to make a good modification to P here. In many works, P can refer to a certain probability distribution or some specific probabilities.

Recommendation 4:

Even the conception of a "policy" is unclear (Figure 2B). They list 4 possible policies, which are simply the 4 possible sequences of steps, stay-safe, cue-risky, etc, but with no contingencies in them. Surely a complete policy that lists 'cue' as the first step would entail a specification of how they would choose the safe or risky option BASED on the information in that cue

Response 4: Thank you for your suggestion. In active inference, a policy actually corresponds to a sequence of actions. The policy of "first choosing 'Cue' and then making the next decision based on specific information" differs from the meaning of policy in active inference.

Recommendation 5:

I assume that the heavy high pass filtering of the EEG (1 Hz) is to avoid having to baseline-correct the epochs (of which there is no mention), but the authors should directly acknowledge that this eradicates any component of decision formation that may evolve in any way gradually within or across the stages of the trial. To take an extreme example, as Figure 3E shows, the expected rewards for the risky path evolve slowly over the course of 60 trials. The filter would eliminate this.

Response 5: Thank you for your suggestion. The heavy high pass filtering of the EEG (1 Hz) is to minimize the noise in the EEG data as much as possible.

Recommendation 6:

There is no mention of the regression itself in the Methods section - the section is incomplete.

Response 6: Thank you for your suggestion. We have now added the relevant content in the Results section (EEG results at source level, line 337-340):

“The linear regression was run by the "mne.stats.linear regression" function in the MNE package (Activity ~ 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).”

Recommendation 7:

On Lines 260-270 the same results are given twice.

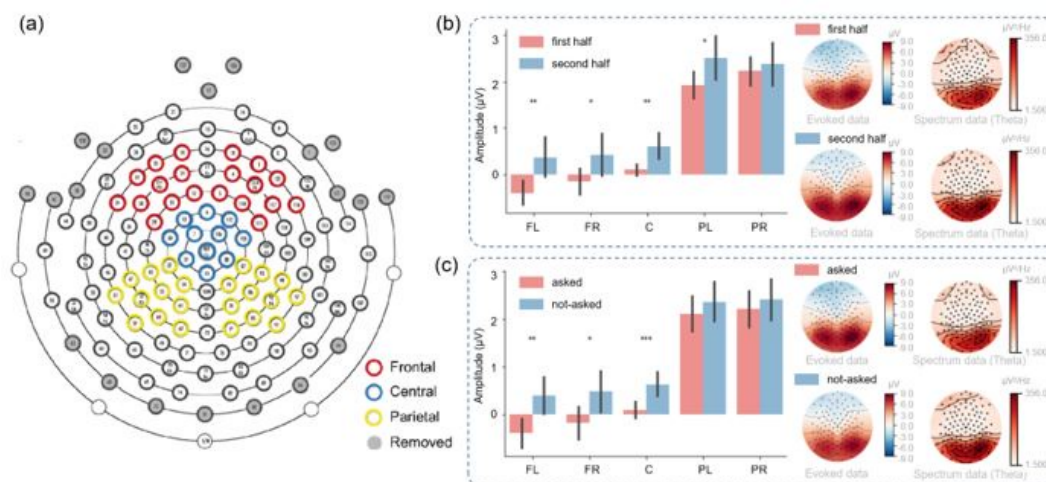
Response 7: Thank you for your suggestion. We have now deleted redundant content.

Recommendation 8:

Frequency bands are displayed in Figure 5 but there is no mention of those in the Methods. In Figure 5b Theta in the 2nd half is compared to Delta in the 1st half- is this an error?

Response 8: Thank you for your suggestion. It indeed was an error (they should all be Theta) and now we have corrected it.

Author response image 10.



<https://doi.org/10.7554/eLife.92892.2.sa0>