Individual Differences in Policy Precision: Links to Suicidal Ideation and Default Mode Network Dynamics

Dayoung Yoon[1], Jaejoong Kim[2], Do Hyun Kim[3], Dong Woo Shin[4], Su Hyun Bong[1], Jaewon Kim[1], Hae-Jeong Park[5, 6, 7] Hong Jin Jeon[8] Bumseok Jeong[1, 9]

1 Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
2 Department of Psychiatry and Behavioral Sciences, University of Minnesota, MN 55454, USA
3 Department of Psychiatry, Dankook University College of Medicine, Cheonan, Korea
4 Department of Neurology, Ewha Womans University Mokdong Hospital, Seoul, Korea
5 Graduate School of Medical Science, Brain Korea 21 Project, Department of Nuclear Medicine, Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea
6 Center for Systems and Translational Brain Science, Institute of Human Complexity and Systems Science, Yonsei University, Seoul, Republic of Korea
7 Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea
8 Department of Psychiatry, Depression Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-Ro Gangnam-gu. Seoul, South Korea 06351
9 KAIST Institute for Health Science and Technology and Graduate School of Data Science, KAIST, Daejeon, Republic of Korea

Abstract

Behavior modeling of decision-making processes has deepened our understanding of behavioral impairments in various psychiatric states. However, as research increasingly emphasizes building models to best explain observed decision-making behavior, it may overlook whether these decisions stem from biologically plausible brain functions. To address this, we developed a probabilistic two-armed bandit task model based on the active inference framework. This model was compared to established reinforcement learning (RL) models and demonstrated superior explanatory power in capturing individual choice variability. A key parameter in our model, policy precision, corresponding to the temperature parameter in the RL model, is also optimized based on previous outcomes. This process effectively explains the balance between model-free and model-based strategies. Incorporating the rate of change in policy precision enhanced the model's ability to explain multi-level brain signals and their inter-subject correlations. Notably, this signal positively correlates with the activation of the default mode network while negatively correlating with the dorsal attentional and frontoparietal networks. Lastly, we validated the relevance of behavioral parameters in a population including individuals with major depressive disorder.

Introduction

Computational psychiatry has taken on the challenge of bridging the gap between mental phenomena and their assumed origins in brain function [26, 35]. Central to this effort has been the development of mathematical models. Owing to the intricate cognitive processes involved in decision-making, such as evaluating evidence for potential outcomes and choosing an option accordingly, decision-making has been focused on as a means to explain psychiatric symptoms [7, 32]. In particular, reinforcement learning has provided a major framework accounting for decision-making, leading to the development of various models augmented with additional parameters to more accurately explain the agent's decision [1, 2, 59, 60].

There are various decision-making tasks, and one of the simplest forms is the probabilistic two-armed bandit task. Owing to its simplicity, it has not received much recent attention, as it is considered insufficient for studying complex cognitive processes, such as exploration [9, 50]. However, even in this simple task, individuals display diverse behavioral patterns, which are likely rooted in variations in neural states. Although behavioral models are intended to link these neural and behavioral patterns, most models account for this behavioral diversity merely by randomness in action probabilities and exclude neural state diversity from consideration. To address this limitation, we developed a new model inspired by the principles of active inference framework (AIF), a unified theory of action and perception that attempts to elucidate a series of decision-making processes by minimizing a singular objective function, free energy [38]. Free energy is a measure of the discrepancy between an agent's internal model (its beliefs about the world) and the actual sensory data it receives. In AIF, it serves as a proxy for surprise (unexpectedness), which is something agents cannot compute directly.

The reinforcement learning (RL) model assumes that people develop expected values (EV) for each choice option that represent the reward (or punishment) they expect to receive following each choice. In this process of learning, a Rescolar-wagner model is generally implemented, which uses a fixed learning rate parameter for updating expected value [44].

$$EV_{i,t+1} = EV_{i,t} + lr \cdot (r(t) - EV_{i,t}) \tag{1}$$

Choices are made according to the probability calculated from EVs, which are updated and learned through observations, using the softmax action selection rule. The degree to which the option with the highest EV is chosen is determined by γ referred to as the 'temperature parameter.'

$$P(a_{i,t}) = \frac{e(\gamma \cdot EV_{i,t})}{\Sigma_{i=1}^{K} e(\gamma \cdot EV_{i,t})} \tag{2}$$

However, this method has limitation; the probability of selecting each choice is decided only based on their value. In other words, this model assumes that differences in selection behavior among individuals arise from increased randomness in selection probability caused by discounted values, rather than differences in individual characteristics. Regarding the temperature parameter, when it is estimated from an individual's data, the more trials in which a person changes their choice, the lower the estimated temperature parameter will be. Such individuals are classified as having a reduced reliance on value-based selection, which is often studied in the context of impulsivity in decision-making processes [19]. Therefore, this method may indeed be well-suited to explain the observed behavioral data.

However, the model lacks indicators that reflect individual characteristics on a trial-by-trial basis. It cannot answer questions like, "Why did the person make that

specific choice at that specific moment?". Especially when variability in choices among individuals increases, the model's only explanatory factor for such variability is the increased randomness in selection probability caused by discounted values. It does not capture individual characteristics. In other words, the parameter whose function should explain choice variability in specific trials as a reflection of individual differences is absent.

Even though the temperature parameter can modulate the level of decisiveness, it has a limitation: it explains only the general tendency of whether decisions follow learned value or not, rather than considering the choice at each trial individually. For example, under two conditions that share same EVs, the probability of action can be differentiated by individual responses to the outcome of previous trial. Another choice strategy called the win-stay-lose-shift (WSLS) strategy is totally dependent on the previous outcome [24, 37, 52]. In many previous studies, researchers found that both methods are expected to be utilized in decision-making, and many studies have attempted to harmonize both strategies [2, 60]. In the WSLS-RL model, the probability of selecting each choice is calculated by mixing the probability calculated from both strategies, weighting each probability with a parameter w.

$$P(a_{i,t}) = P(a_{i,t})_{WSLS} \cdot w + P(a_{i,t})_{RL} \cdot (1 - w) \tag{3}$$

In this model, the parameter w and probability of lose-shift (P_{WSLS}) correspond to the parameters that can explain the choice variability in specific trials. However, this method relies heavily on additional parameters and cannot explain the mechanism of the exchange between both strategies according to the context of the state.

Meanwhile, two aspects of AIF allow for a flexible model to address this limitation. First, rather than using EV directly, the expected free energy (EFE) is applied as the input to calculate choice probability. EFE represents the distance between the probability distribution of this EV and each agent's preference distribution for EV (Equation 5). This approach not only accounts for utility differences [1], but also captures variations in EV precision, which can influence choice. For instance, when the mean of EV for two options is identical, calculating EFE with higher precision in one case and lower precision in another leads to a smaller EFE difference between the two options in the latter case, making the choice probability more similar to chance (Figure1A). Second, the policy precision parameter (γ in Equation 2), which corresponds to the temperature parameter in a softmax function modulates the effect of EFE on action probability. This parameter is updated as the differences between EFE and free energy (Equation 7, Figure 1B). Here, EFE reflects model-based (MB) strategy in RL, as it is calculated based on accumulated evidence and preferences for future outcomes, whereas free energy corresponds to the prediction error, representing a model-free (MF) strategy. The update of policy precision allows the probability of action selection to be more influenced by either EFE or free energy. Thus, without additional weighting parameters, the model can modulate the alteration between both strategies as a function of state uncertainty. This aligns with previous research findings indicating that uncertainty influences the arbitration process between MF and MB, such that high levels of uncertainty lead participants to rely more on the MF strategy [29]. A detailed explanation of the model is provided in the Methods section.

We suggest a new behavioral model using AIF in the probabilistic two-armed bandit task and demonstrate its advantages in explaining choice selection, especially when inter-subject variations in responses are large. In addition, the validity of the policy precision signal (rate of change in policy precision, Figure 1D) is confirmed through two experiments: one involving the analysis of neural signal data and another examining whether behavior parameters provide better explanatory power in describing traits of depressive patients.

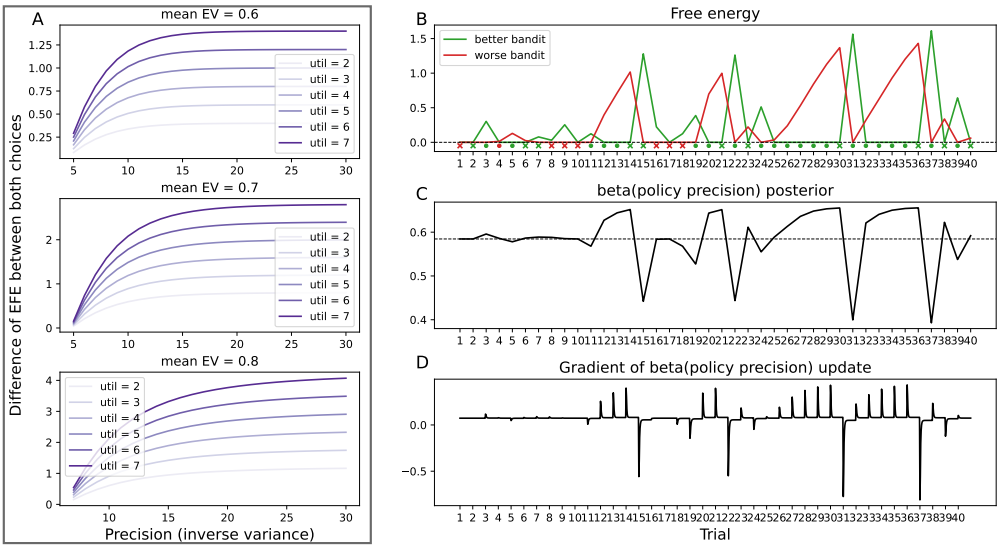

Figure 1. Simulated expected free energy and example estimated results of AIF model. (A) Differences in expected free energy values between both actions increase as the precision of the state increases, regardless of the expected value of the state. (B) to (D) show the estimated results for one participant. (B) Free energy values for both actions in each trial. These values are generally higher during the later phase of the trials because the variance of the state distribution is lower. In this phase, if the outcomes are different from the prediction, the distribution's distance to the predicted state is large. Conversely, free energy values before approximately the tenth trial are lower because the precision of prediction is low so the prediction error is also low. Each bandit's free energy tends to increase after it is chosen and a losing outcome is observed. (C) As the policy precision is also updated according to the free energy, updated posterior values tend to follow the worse bandit's free energy while opposing the free energy of the better bandit. (D) While policy precision is updated, the rate of change is hypothesized to represent neural discharge.

In Experiment 1, using neural signal data obtained via task-functional magnetic resonance imaging (fMRI), we used two methods to confirm whether the policy precision signal is related to neural activity (Figure 4). First, we conducted a conventional GLM approach, where models are generated by including the stimuli of behavioral task as regressors. The second method addresses the limitations of the conventional GLM by using inter-subject correlation [36]. We examined whether the variance in brain imaging data across individuals could be explained by the variance in behavioral signals and, similarly, whether model performance measured by Akaike Information Criterion (AIC) improved by adding policy precision signal.
In Experiment 2, we conducted the decision-making task involving a population that included patients diagnosed with major depressive disorder (MDD) and healthy controls (HC). In patients with MDD, the risk of suicide is nearly 20 times higher than that in the general population [56]. According to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5), "impairment in decision-making" is an important criterion for diagnosing MDD as "indecisiveness". However, whether it is associated with an increase in suicide in patients with MDD remains unclear. Previous studies have analyzed patients with MDD and serious suicidal behavior and found that their ability to learn and choose an action that could result in a rewarding outcome is impaired [3, 8, 17]. We aimed to verify the feasibility of finding similar clinical differences in the new model.

Results

Model construction and parameter recovery

We constructed 12 models, comprising three learning models combined with four choice models. One of learning models utilized scalar state value explained in Equation 1 (*rw* in Figure 2), whereas the others utilized beta (*betads* in Figure 2) and normal distributions (*normds* in Figure 2) for state value. The later corresponds to the Kalman-filter model [28]. The four choice models included two previous models of softmax rule (*sft* in Figure 2) and the WSLS-RL (*wsls* in Figure 2) model. The other two models used EFE, referred to as the *efeX* and *efegc* models. And *efeX* model excluded the process of policy precision update, making it similar to the softmax model, whereas the *efegc* model resembled the WSLS-RL model. The parameter compositions of each model are shown in Figure 2 with descriptions of each parameter and model construction detailed in the Methods and Materials section. Parameter estimation was done using the variational Bayes method [21], and parameter recovery was examined with simulation data that is generated from the distributions of each estimated parameter. Except for the beta distribution model, other models show poor parameter recovery performance in at least one action model combination (Figure 2). Consequently, we decided to focus on beta distribution models in subsequent analyses.

0.1 Identifying the best models for observed data and inter-subject action variability through model comparison

Log-likelihood values for each model were calculated as the sum of the log probabilities of the actions selected by each participant on each trial. These action probabilities were obtained from parameter estimation. The protected exceedance probability (pxp) values of each model were calculated, and the model with the highest value was considered optimal. The model with softmax rule was best for both loss and reward task (Figure 3G). However, this method had critical limitations in the two-armed bandit task: one choice is better than the other, and the task's objective is to learn which is better. As a

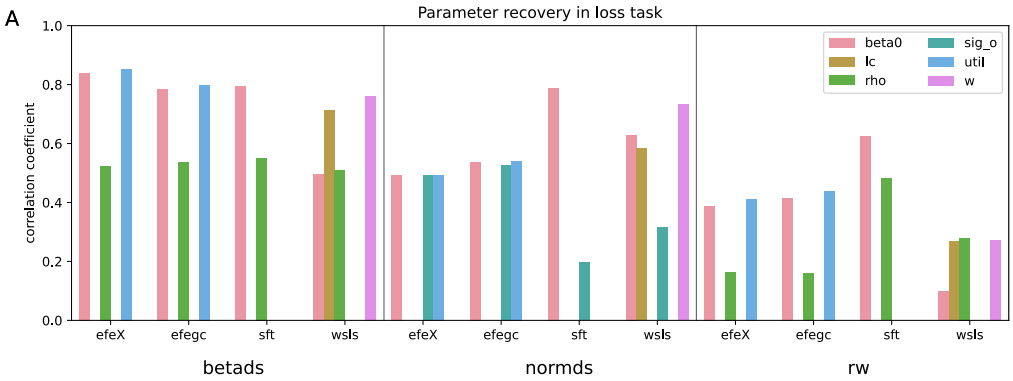

Figure 2. Parameter recovery results of 12 (three learning x four choice models) models in Experiment 1. (A) Three learning models are divided by vertical lines. The correlation coefficients of each parameter that is included in the models are plotted. Beta distribution model (betads) among three learning models shows consistent performances across four choice models in both tasks.
beta0; prior of policy precision in EFE models and temperature parameter in softmax and WSLS-RL models, lc; probability of lose-shift, rho; decay parameter; sig_o; standard deviation of observation, util; reward sensitivity in reward task and loss aversiveness in loss task, w: weight of lose-shift probability

result, the better bandit is chosen far more frequently than the other. In such cases, models that estimate action probabilities overly favoring the better bandit, are more likely to be evaluated as optimal, even if they lack a balanced perspective. For instance, when the action probability for the the better bandit (p) is estimated as a large value close to 1, the selection probability for the other bandit decreases to (1 - p).
Since the worse bandit accounts for only a small proportion of the total trials (as it is rarely selected), the sum of the log probabilities for the selected bandit can still yield a large final log-likelihood value. This poses a problem because it creates a bias in model evaluation favoring models that heavily prioritize the better bandit, even if they fail to represent the data comprehensively.
This can be seen in Figures 3. The majority of participants identified the better bandit by the tenth trial and predominantly selected it in subsequent trials, resulting in low variability in selection across participants. The estimated action probabilities of selecting the better bandit are higher (approaching 1.0) in the softmax model compared to those in other models (around 0.9). The high action probabilities for the better bandit are estimated at the expense of those for the worse bandit. Since the better bandit is predominantly chosen, the cumulative log-likelihood across all trials becomes highly dependent on the probability of the better bandit, resulting in the softmax model achieving the highest cumulative log-likelihood across all trials. In this way, it cannot be concluded that the best model optimally explains the cognitive process in the infrequent cases where choices deviate from the expected pattern, which varies across individuals. These cases correspond to the trials after receiving a lose outcome, which is rare because the probability of a loss when selecting the better bandit is only 0.3. We referred to these as high free energy (HF) trials (marked as orange dots in Figures 3A-F) because the free energy of choosing the better bandit increases on these trials due to the difference between outcome and prediction. Specific trials following consecutive HF trials exhibited an increase in selection variability (marked as red dots in Figures 3A-F). On the first trial, the selection variability was highest because there was no difference in value between bandits, suggesting that this early variance is driven by random selection. In contrast, the increase in variance observed after the tenth trial is attributed to a

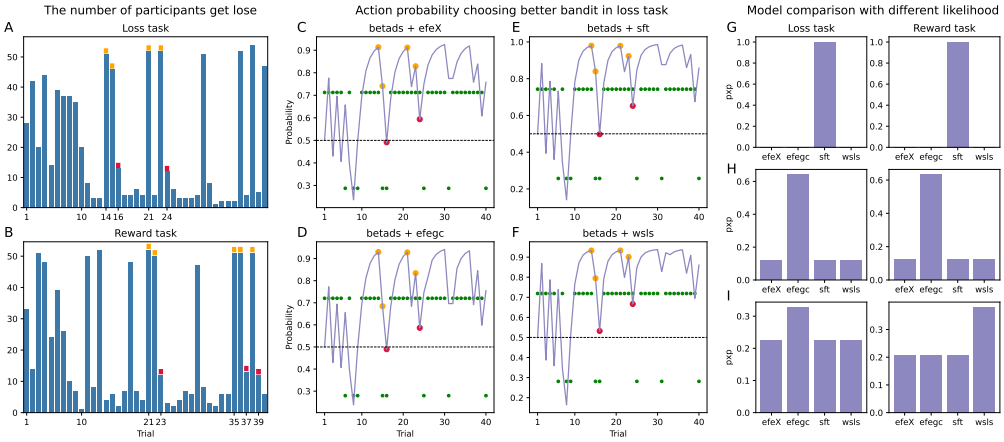

Figure 3. Action probability estimated by each action model and model comparison. (A) and (B): The histograms show the number of participants who lose in each trial out of 54 participants. (C) to (F): Action probability choosing the better bandit in each trial. In trials where green dots are above the line, the participant chose the better bandit. Orange dots over the line are marked on the trials of consecutive high free energy (HF) trials, and red dots indicate their subsequent trials, where the variance of choices across participants increases. (G) to (I): Model comparison results with different sum of log-likelihoods: (G) from all trials, (H) only trials following consecutive HF trials, and (I) for trials when individuals chose the worse bandit after the tenth trial.

distinct mechanism related to the direct influence of the previous trial's result, which is encoded as the lose-shift probability (*lc* in Figure 2) in the WSLS-RL model. The softmax model accounts for this by the increased selection randomness due to a temporary reduction in the value difference between two options, similar to what occurred on the first trial. To evaluate whether other models, particularly those incorporating the impact of prior outcomes, better explain the data from trials with high selection variability across participants, we performed a model comparison using log-likelihoods summed only for trials following consecutive HF trials (Figures 3H). In this case, the *efegc* model outperformed all other models.
Additionally, we conducted a model comparison using log-likelihoods summed over trials where the worse bandit was chosen after the tenth trial. This approach aimed to identify which models best capture the mechanisms underlying occasional changes in choice behavior. As anticipated, models that accounted for the impact of prior losing outcomes, such as the *efegc* and WSLS-RL models, demonstrated superior performance compared to the softmax model in these specific contexts (Figure 3I).

Neuroimaging analysis

We found that *efegc* and WSLS-RL models better explain individually differentiated actions compared to the softmax model. This improvement stems from the additional parameters, *lc* and *w*, in the WSLS-RL model and the inclusion of an additional process in *efegc* that optimizes policy precision by modulating the dependency between MF and MB strategies. Notably, *efegc* does not introduce any additional parameters but includes an additional signal represented by the rate of change in policy precision. To assess whether this policy precision signal is suitable for explaining brain function, we conducted an analysis comparing behavioral signals and neural signals across three levels: node, edge, and network. This approach captures regional brain activity and changes in connectivity between regions and alterations in large-scale networks critical

to brain function (Figure 4). We estimated four key states using the graph Laplacian mixture model (GLMM) as described by [45]. A detailed description of the neural signal extraction process is provided in the Materials and Methods.
Using the mean activation patterns, we grouped the regions into a priori resting-state networks (RSNs) based on [54] (Supplementary Figure 3A-D). State 1 was positively correlated with activation of the visual and somatomotor (SOM) networks, dorsal attention network (DAN), and ventral attention network and negatively correlated with the default mode network (DMN) and frontoparietal network (FPN). State 2 was positively correlated with the DMN and negatively correlated with State 1's network. State 3 was positively correlated with the DAN and FPN and negatively correlated with the DMN and SOM network. State 4 shows generally low activations of all networks except for the limbic (LIM) network and an anterior-to-posterior gradient. We found that adding policy precision signals to the GLM resulted in lower AIC values, which means better performance (Figure 4 I - N). Using beta coefficients from the best model, we performed one-sample t-test to identify which regional activations or interactions are related to the policy precision signals (Figure 5). Across both tasks, similar regions included in DMN were positively correlated with policy precision signal (Figure 5 B), whereas regions included in the FPN and DAN showed negative correlations. At the interaction among functional networks level, both tasks showed similar patterns of positive significance in integration between the SOM network and the FPN and DMN (Figure 5 C). Lastly, at the state level, State 2 showed a positive correlation with policy precision signals ($p = 2.25 \times 10^{-6}$) whereas State 3 exhibited a negative correlation ($p = 0.0039$) which corresponds to the result of node level.

Model validation in population including the patients with depressive

Finally, we investigated whether the parameters estimated from the new model could better discriminate clinical characteristics. To this end, we conducted the behavioral task with a population that included patients diagnosed with MDD. The task composition was slightly modified by setting the probability of winning on the better bandit to 0.75 and mixing loss and reward tasks within the same session. The same models were used for parameter estimation, with an additional learning model introduced to account for the alternating loss and reward tasks. Since the trials alternated randomly, task trial effects were incorporated into the decay parameter. As in previous analyses, most models—except for the beta distribution learning model—exhibited poor parameter recovery. Thus, subsequent analyses were conducted using beta distribution models (Supplementary Figure 3). Among these, the model considering decay as a function of trials with softmax showed the best performance in model comparison. However, as noted earlier, the highest likelihood of data under a model does not always guarantee its reliability in reflecting brain function. When we performed a model comparison with the log-likelihood summed only for trials following consecutive HF trials, the model considering decay as trials with *efegc* showed the best performance. Additionally, results with the log-likelihood summed over trials where the worse bandit was chosen after the tenth trial showed the *efegc* and WSLS-RL models performed best for loss and reward tasks, respectively (Supplementary Figure 3J, K). Parameter values of each model were then compared between the MDD group and the HC group using analysis of covariance (ANCOVA), controlling for age, sex, education level and task set. To investigate whether behavioral parameters can reveal differences between the HC and MDD patients that are not captured by the Hamiltonian depression rating scale (HAM-D) scores [61], which measures the severity of MDD, the HAM-D score was also included as an independent variable in an ANCOVA. Post-hoc

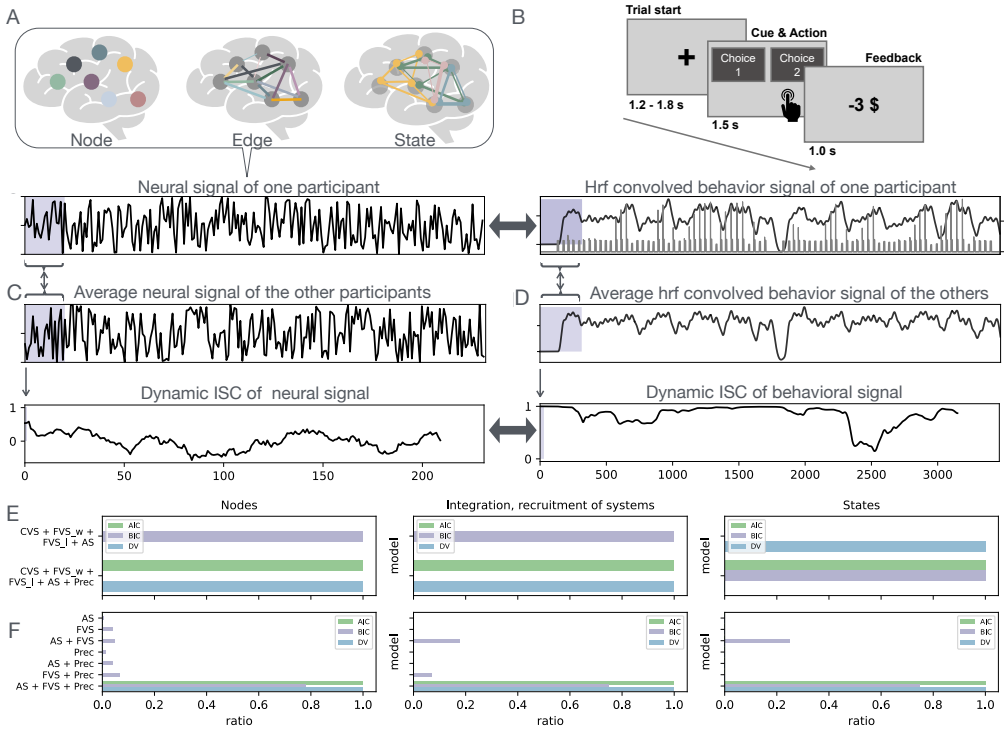

Figure 4. Schematic illustration of analysis for explaining neural signals by behavioral signals. (A) From functional magnetic resonance imaging (fMRI) data measured during decision-making tasks, neural signals are extracted at three different levels. (B) The task begins with the gaze fixation on a central cross for a random duration ranging from 1.2 to 1.8 s; subsequently, the meaningful experimental stimulus is presented. The "Cue" means the presentation of two option figures, and it lasts for 1.5 s during which the participant had to choose one. After 1.5 s, the results of the choice are presented as the monetary reward or loss. We defined this presentation result as "Feedback", and it lasts for 1.0 s. The images of each choice are shown as meaningless black-and-white pictures, and different sets of images are shown during each part of the task. Selection should be done in 1.5 s during which the Cue images are being presented on the screen. Policy precision signal is generated by the process of calculation as in Figure 1D, then aligned to the time-point of Feedback presentation because the gradient update is assumed to occur by free energy that are induced by Feedback. (C) and (D) By calculating the inter-subject correlation (ISC) of time-series over the sliding window, neural and behavioral dynamic inter-subject correlation (dISC) were obtained, respectively. (E) Performance comparison of conventional GLM. (F) Performance of dISC GLM. The x-axis means the ratio of the number of regions where each model was measured as best (smallest values of each criterion), and the y-axis shows the independent variables included in each GLM model.
CVS:cue visual stimulus; FVS:feedback visual stimulus; FVS_w:feedback visual stimulus of win; FVS_l:feedback visual stimulus of lose; AS:action selection; Prec:policy precision signal

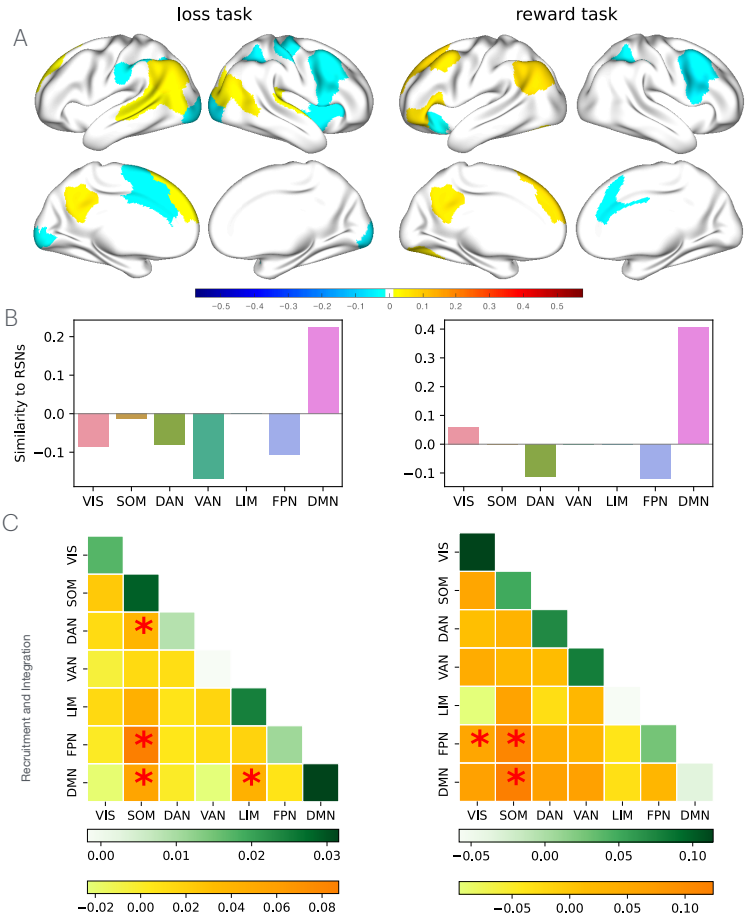

Figure 5. Regions and connectivity that are related to the policy precision signal. (A) Results of a one-sample t-test using beta coefficient values result from GLM to investigate the regions that are related to the policy precision signal. Colored regions have significant correlations with policy precision signal ($p < 0.05$) (B) Cosine similarity between beta coefficient values of significantly related regions and the seven networks of intrinsic functional connectivity from [54]. Regions included within the default mode network (DMN) show positive correlations, while regions included in the dorsal attention network (DAN) and frontoparietal network (FPN) show negative correlations. (C) Results of a one-sample t-test using beta coefficient values resulted from the GLM to investigate the system interactions that are related to the policy precision signal. Integrations between somato-motor (SOM) and the FPN and DMN have positive correlations (Asterisks denotes statistical significance ($p < 0.05$).

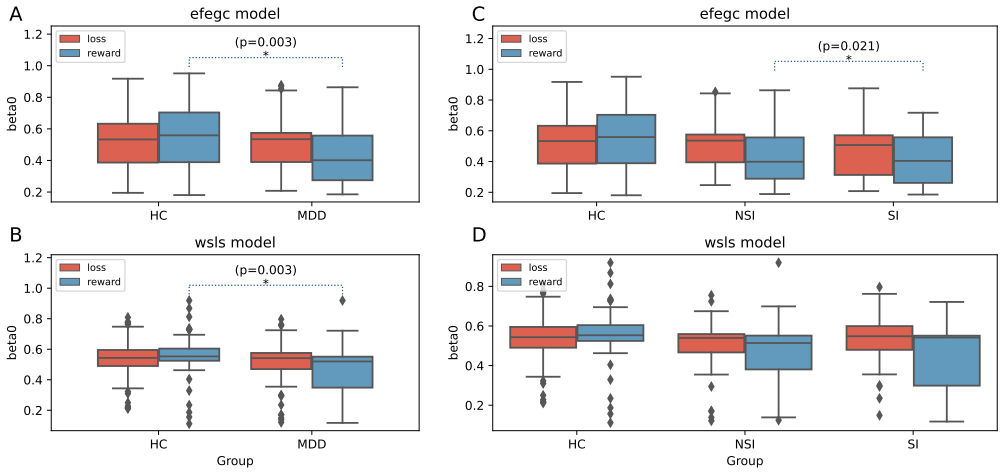

Figure 6. Differences in parameter values estimated from four behavior models across MDD and control groups. (A), (B) ANCOVA results comparing *beta0* estimated from *efegc* and WSLS-RL models between control (HC) and major depressive disorder (MDD) group. (C), (D) ANCOVA results comparing *beta0* estimated from *efegc* and WSLS-RL models between HC and MDD without suicidal ideation (NSI) and with suicidal ideation (SI) group. Asterisks (*) denotes statistical significance in ANCOVA.

analyses including pairwise comparisons between groups were then performed. In this comparison, parameter *beta0* from all action models showed significant difference only in the reward task, with lower values in the MDD group (Figure 6A, B, and Supplementary Figure 3A, B). The *beta0* represents the prior policy precision in EFE models and the temperature parameter in other models. Further analysis divided the MDD group into subgroups: MDD without suicidal ideation (NSI), MDD with suicidal ideation (SI). Only EFE models showed significant differences across these groups in reward task parameters (Figure 6C, D and Supplementary Figure 3C, D). There was no significant difference in HAM-D scores between the NSI and SI groups ($p = 0.13$). Additionally, across all analyses, the HAM-D score was not significant in explaining between-group differences in behavioral parameters (p ¿ 0.4). The decay parameter *rho*, which is included in all models, showed a significant difference between the HC and MDD groups in the reward task (smaller in MDD than in HC), except in the softmax model, but no significant differences among HC, NSI, and SI groups. Owing to the high positive correlation between the *beta0* and *util* parameters (Supplementary Figure 3A), the results for the *util* parameter were similar to those for *beta0* in the EFE models. Lastly, other parameters in the WSLS-RL model, *lc* and *w*, showed significant differences between the HC and MDD groups in the reward task ($p = 0.018$, 0.038, respectively), but no significant differences among HC, NSI, and SI groups.

Discussion

To explain psychiatric symptoms or real human cognitive actions, it is essential to utilize a biologically plausible model rather than rely solely on models that demonstrate the best performance. This approach ensures that the model is not only accurate but also reflective of the complexities of neural processes. Our model included distributional representation during the learning process, calculated expected energy during action selection, and optimized policy precision by observation of each trial. Although this may seem overly complex, the inclusion of biologically plausible parameters and signals

significantly enhances the explanatory power in neural signal data and validates the advantage in clinical data.

The main advantage of this model was its ability to account for the context-dependent exchange between MF and MB strategies using a single free energy function, eliminating the need for an additional weighting parameter. By optimizing policy precision based on the results of previous trials, the model improved both the performance and the explanatory power of free parameter in distinguishing clinical characteristics (Figure 6). In addition, the rate of change in policy precision served as a meaningful signal in explaining neural signals (Figure 4E) and improved explanatory power when incorporating inter-subject correlations (ISC) across participants, in relation to neural signal variability (Figure 4F).

In neural signal analysis, regions positively correlated with policy precision are dominantly affiliated with the DMN, the occurrence of State 2, dominated by the DMN, and also showed a positive correlation. Although the DMN is generally regarded as a resting network activated during rest, mind-wandering, and self-referential tasks [13, 14, 42], recent studies suggest that its role extends to focusing on intrinsic information rather than just resting [4, 40, 43]. DMN has been suggested to detect the associative relevance of internal and external stimuli and enables the encoding of values [47]. From this viewpoint, an alternative proposal argues that understanding the self-referential function of the DMN entails linking it to the choice of the best actions within the environmental context, necessitating recursive exploration [31]. Following this suggestion, there is a concept outlining how the DMN functions at elevated levels of inferential hierarchy, which diminishes the free energy associated with lower levels [6]. In addition to the DMN, the hippocampus is activated in State 2 (Supplementary Figure 3J). Recent studies examining the function of the hippocampus focusing on the internal model [34, 57] support the notion that the role of DMN is linked to the internal models.

Unlike loss decision, the *beta0* in reward decision significantly decreases in patients with MDD, and this relation is easily captured in all models (Figure 6A, B, and Supplementary Figure 3A, B). These results are consistent with previous studies reporting impairment of reward value utilization in patients with MDD [25, 48]. These results are obtained independent of HAM-D scores, suggesting that the behavioral parameters might reflect additional differences that cannot be captured in HAM-D scale. Additionally, in analysis including MDD subgroups, significant differences between the NSI and SI groups were observed only for the *beta0* parameter in the EFE models. In contrast, no significant differences were found in other models or for the *rho* parameter, which plays a role in the learning process. These results suggest that the SI in the MDD group is not attributable to differences in the process of learning the value of a situation, but rather to differences in the process of determining action based on the learned values (MB). And the lower *beta0* value in the SI group may implies that it may lead to relatively impulsive (MF) behavior. However, since this subgroup analyses included a small number of people for NSI and SI group (n = 33 and 23 respectively), future studies with larger samples are needed to confirm our finding.

It is necessary to create a behavior model that can reflect the various mental states of people, to better explain mental states through the model. In addition, there are limitations to solely understanding individual characteristics through parameters obtained from behavioral modeling. Therefore, high-dimensional brain signal data that reflects individual diversity must be integrated. This makes behavioral modeling even more critical. Rather than focusing solely on better explaining behavioral outcomes, we need to prioritize the development of biologically plausible models, identify their neural correlates through brain signals, and understand mental symptoms in terms of the differences or abnormalities that arise from these connections. We believe that this will explain the differences in various brain states through models, and will serve as a

foundation for personalized neuromodulatory treatment through accurate correlation between behavior and individual brain activity.

Materials and Methods

Participants

0.1.1 Experiment 1

A total of 54 volunteers participated in Experiment 1 (15 females and 39 males; mean [standard deviation] age: 22.4 [3.2] years; all participants were right-handed and had no history of psychiatric or neurological disorders). All participants provided written informed consent for all the procedures and data usage before the start of the study. The experimental procedures were approved by the Ethics Committee of Korea Advanced Institute of Science and Technology.

0.1.2 Experiment 2

A total of 58 patients with MDD (25 SI and 33 NSI) and 59 HCs without any current or past history of psychiatric illness, aged 18 to 34 years, were recruited through the outpatient clinic of the Samsung Medical Center in Seoul, South Korea between July 2018 and October 2020 MDD was diagnosed by two psychiatrists (H. Kim and H.J. Jeon) based on clinical interviews. Those whose diagnosis was categorized as MDD per DSM-5 were included in the MDD group. The exclusion criteria were: (1) MDD with psychotic features; (2) comorbidity of any other psychiatric illnesses including bipolar disorder, schizophrenia, delusional disorder, delirium, neurocognitive disorder, intellectual disability, and other mental disorders due to another medical condition; (3) history of substance-related disorders except for tobacco-related disorders within 12 months; (4) primary neurologic illness or history of brain damage; and (5) a history of major physical illness. HCs were recruited from the community using an advertisement of the Clinical Trial Center of Samsung Medical Center. The control group enrolled those who had no history of major depressive episodes and scored 7 or less on the HAM-D scale, using the same exclusion criteria that were applied to participants in the MDD group. The study design was approved by the Institutional Review Board of Samsung Medical Center (IRB No. 2018-04-137). All participants provided written informed consent before participating in this study in accordance with the Declaration of Helsinki. SI was assessed with the suicidality module of the Mini International Neuropsychiatric Interview [30]. Among the patients with MDD, the SI group included participants who answered "yes" to the question "Did you think about suicide during the past one month?" (SI group, n=25), whereas the no SI group included those who answered 'no' to the same question (NSI group, n=33). Depression severity was measured using the 17-item Hamilton depression rating scale (HAM-D) [61]. At the one-month follow-up visit, we differentiated the MDSI group into TSI (n = 8) and SSI (n = 15) group to determine whether patients' SI disappeared or was sustained.

0.2 fMRI data acquisition

In Experiment 1, all participants underwent MRI using 3.0-T MRI (Siemens Verio Syngo Scanner). Both T1-weighted (T1w; echo time [TE] = 2.02 ms, repetition time [TR] = 2400 ms, field of view (FOV) = 224 × 224 mm, and 0.7 mm iso-voxel) and T2-weighted (T2w; TE = 330 ms, TR = 2200 ms, FOV = 224 × 224 mm, and 0.7 mm iso-voxel) structural images were scanned. A total of 240 echo-planar imaging scans of BOLD responses (multiband factor = 4, TE = 32 ms, TR = 1.5 s, flip angle = 50°,

FOV = 225 × 221 mm, matrix size = 110 × 107, 2.0 mm iso-voxel with no gap, 64 slices, phase encoding = anterior to posterior) were performed using a 32-channel head coil. Field map images (TR = 731 ms, TE: 4.92 ms and 7.38 ms, flip angle = 90°) were also acquired to correct distorted images. In every trial, the participants performed the experimental task with both hands to press an MRI-compatible buttons to make their decision. In the pretraining session, prior to the first scan, the participants were acquainted with the apparatus and tasks performed during scanning. Sequences executed during this pretraining session were not encountered later in the experiment.
Preprocessing and processing of the neuroimaging data were performed using AFNI (v23.1.00) [10, 11], FSL (v6.0.5.2), FreeSurfer (v7.1.1), Workbench (v1.5.0), and Advanced Normalization Tools (ANTs). Following the Human Connectome Project, minimal processing pipeline for structural images, structural T1w and T2w images [23], were processed involving automatic segmentation of surfaces and reconstruction using FreeSurfer 7.1.1 (https://surfer.nmr.mgh.harvard.edu/) [12, 20]. The generated volumes and surface files were converted to the CIFTI file format using ciftify (https://edickie.github.io/ciftify/) [16] including MSMSulc surface realignment [46].
Raw time-series of task fMRI were performed with slice-timing correction, realignment, and skull stripping using AFNI commands (3dTshift, 3dvolreg, and 3dAutomask) [53]. Realigned data were unwarped using a fieldmap in FSL. A bandpass filter (highpass = 0.01) was applied to the unwarped functional data and motion artifact time-series before regression. Using algorithms in ANTs, we registered the unwarped functional data into a 2-mm MNI152 T1w template space via skull-stripped T1w images preprocessed with AFNI. The eroded white matter (WM) and ventricular cerebrospinal fluid masks segmented using FreeSurfer were moved to the functional image space using the transformation matrix generated via ANTs in the previous step. By applying principal component analysis (PCA: 3dpca in AFNI) to the eroded ventricular mask in bandpass-filtered, scaled functional data, and ventricular PCA time-series were acquired. Using the fast ANATICOR algorithm in AFNI [27] on an eroded WM mask, a voxel-wise local WM regressor was generated.
3dTproject in AFNI, all regressors generated were projected into a regression matrix. Finally, the bandpass-filtered, scaled functional data were denoised, including pre-whitening, to remove the autocorrelation property using 3dREMLfit (residual maximum likelihood) in AFNI. Using transformation matrices and warping images generated by ANT algorithms, the denoised functional data were directly moved to a 2-mm MNI152 T1w template space and projected to the CIFTI format using the ciftify package [16].

0.3 Neural signal extraction

After preprocessing, we acquired the averaged ROI time-series extracted from the Schaefer Atlas [49], including 100 regions covering the cortex, and the Melbourne Subcortex Atlas [55] covering the subcortex. Additionally, to include other subcortical regions related to the release of important neurotransmitters, including dopamine and norepinephrine, we added some regions to the CIT168 atlas [39] and Harvard Ascending Arousal Network Atlas [18] to obtain signals from a total of 147 ROIs.
To investigate the dynamic changes in different levels of neural interactions, we obtained neural signals from both network and edge level. At the edge levels, we used phase coherence to calculate the connectivity between the two regions. This method is adept at capturing temporally delayed relationships and exhibits high sensitivity in detecting ultraslow oscillatory processes that underlie the formation of RSNs, a phenomenon corroborated by recent experimental and computational research [15, 41]. In our study, we used a recently developed a data-driven approach named Leading Eigenvector Dynamics Analysis (LEiDA). LEiDA can reduce the dimensionality by considering only

the relative phase of BOLD signals and capturing instantaneous phase-locking (PL) patterns [5, 58]. However, in this study, we did not employ k-means clustering to identify representative states based on the leading eigenvector. Instead, we used the outer product of the leading eigenvectors of the PL patterns to reconstruct the connectivity patterns at each time point. K-means clustering was not used because it does not consider the graph-based connectivity of brain structures that generate signals when distinguishing clusters.

Therefore, to extract the representation of the overall brain state corresponding to network-level interactions, we employed the Graph Laplacian mixture model (GLMM). This method identifies meta-networks representing both intra- and inter-network interactions and learns the graph structure of the states by estimating the Laplacian matrices capturing dynamics of brain states which are consistent with the timing of various task paradigms in the HCP dataset [33, 45]. Four key states are estimated and depicted in Supplementary Figure 3.

0.4 Behavior task and modeling

0.4.1 Experiment 1

Participants faced two choices: one with a 70% likelihood of winning and the other with a 30% probability. The number of trials was 40 per part; thus, a total of 80 trials were performed by each participant. They were asked to infer which choice was better to win as trial progressed to maximize the accumulated reward. Participants were expected to learn about the reward value of each option by trial-and-error and utilize computed reward values to make their decision. Therefore, the probability of one choice being better represents the hidden state value inferred by the agents during the task. This state could be modeled using either a distribution of beta or normal distribution. The inferred value over one choice could be updated in every trial, even if that choice was not selected, because winning or losing by selecting one choice simultaneously means losing or winning by selecting the other choice. The expected value can be represented by either scalar probability or distribution form. Using a scalar value corresponds to the Rescolar-wagner model. Meanwhile, using a distribution form, we can utilize the variance of state when calculating behavioral measures in the model, such as learning rate or free energy. Each trial corresponds to Bernoulli trial with outcomes of win or lose, and state value $q(\theta)$ can be represented by a beta distribution, where θ is the probability of one choice being better bandit. The beta distribution is parameterized as:

$$\theta \sim Beta(\alpha, \beta),$$

where α and β are equal to one plus the number of win or lose outcomes, respectively, for a given choice. In this case, the learning process as the task progresses can be represented simply by adding one according to the outcome of each trial, as the likelihood of outcome at each trial is binomial distribution, which is conjugate to the beta distribution. When selecting choice a results in a win or selecting choice b results in a lose:

$$q(\theta)_{t+1} = Beta(\rho \cdot \alpha_t + 1, \rho \cdot \beta_t)$$

When selecting choice b results in a win or selecting choice a results in a lose:

$$q(\theta)_{t+1} = Beta(\rho \cdot \alpha_t, \rho \cdot \beta_t + 1)$$

where ρ is the decay parameter ($0 < \rho < 1$). And $q(\theta)$ can be represented by normal distribution as $\theta \sim Norm(\mu, \sigma^2)$ with a prior mean of 0.5 and typically large standard deviation of 10. In this case, corresponding to the Kalman-filter model [28], likelihood of outcome at each trial can be represented as normal distribution, and the standard

deviation of the observation (σ_o in Equation 4) either can be fixed or can be defined as free parameter. That modulates how much prior belief will be updated:

$$\mu_{t+1} = \mu_t + \frac{\sigma_t^2 \cdot \sigma_o^2}{\sigma_t^2 + \sigma_o^2} \cdot (r(t) - \mu_t) \tag{4}$$

As an action model, we used a conventional softmax function (Equation 2), WSLS-RL (Equation 3), and two models based on active inference scheme utilizing EFE. Formally, we express the EFE of a choice a on trial t as:

$$G_t(a) = D_{KL}[q(o_t|a_t = a)||p(o_t)] + E_{q(\theta)}[H[o_t|\theta, a_t = a]] \tag{5}$$

preference distribution of outcomes, $p(o_t)$, is encoded using *util* parameter for each participant as

$$p(o_t)[win, lose] = [\frac{\exp^{util}}{\exp^{util}+1}, \frac{1}{\exp^{util}+1}]. \tag{6}$$

The probability of selecting choice a is then calculated as:

$$P_t(a) = \sigma(-\gamma \cdot G_t(a)) \tag{7}$$

In addition, the precision of beliefs about actions γ, itself can be inferred instead of a fixed value [22]. In the *efeX* model, the following adjustment excludes the *beta0* parameter (β_0) which is used as γ in Equation 7, whereas in the *efegc* model, the initial value of γ is calculated from the fixed prior value of β_0 [51]. Thus, the probability that the agent will choose each policy π_0 will be

$$\pi_0 \leftarrow \sigma(-\gamma G), \tag{8}$$

As precision weights the EFE to determine the probability of choosing each policy, the results of trial t are given as either a win or a lose, and the variational free energy (VFE) of each policy F over that observation is calculated. Subsequently, the probability of choosing each policy after a subsequent observation is:

$$\pi \leftarrow \sigma(-F - \gamma G), \tag{9}$$

and the error of the agent's EFE (G_{error}) is given as:

$$G_{error} \leftarrow (\pi - \pi_0) \cdot (-G). \tag{10}$$

The above equation reflects the level of (dis)agreement between the EFE (G and π_0) and the VFE of the observation (F and π) [51]. Thus, the β whose inverse means γ is updated based on G_{error}.

$$\begin{aligned} G_{error} &\leftarrow (\pi - \pi_0) \cdot (-G). \\ \gamma_{\text{update}} &\leftarrow 1/\gamma - 1/\beta_0 + G_{error} \\ 1/\gamma &\leftarrow 1/\gamma - \gamma_{\text{update}}/\psi \end{aligned} \tag{11}$$

Therefore, the policy precision reflects the reliability of the EFE based on new observations that represent either winning or losing outcomes. If the result of a chosen action betrays expectations, the precision decreases, causing the probability of action selection to rely less on the 490 expected free energy. Simultaneously, this increases the effect of free energy that is calculated from the previous outcome. This process explains the interplay between model-based and model-free strategies.

0.4.2 Experiment 2

The overall paradigm of the decision-making task is similar to that of Experiment 1. However, participants performed 80 trials at once, and each trial's decision was either loss or reward specific. In reward decision-making task trials, choice options are presented with blue boundary, while red boundary are used in loss trials (Supplementary Figure 3). One option was associated with a high probability of reward (75% of reward) while another option had a high probability of non-reward. Reward and loss schedules were semi-randomized and counterbalanced with two sets of schedules (74 participants for set 1 and 43 participants for set 2). The same behavior models were used as in Experiment 1, with the addition of one learning model that considered decay along the trial sequence. For example, the second and third loss tasks occurred on the third and ninth trials of the 80-trial sequence. Considering decaying as the number of trials between those two trials, it's as much as six trials instead of one. Therefore,

$$q(\theta)_5 = Beta(\rho^6 \cdot \alpha_4 + 1, \rho^6 \cdot \beta_t),$$

when selecting choice a results in a win.

0.5 GLM analysis

To identify the shared dynamics of neural engagement across participants who performed the same experimental task, we calculated the dISC for each level of the time-series. Following established methods [**?**], a sliding window with a length of three repetition times (TRs), corresponding to approximately to the length of one trial (3.7-4.3s) and a 1 TR overlap. This was calculated by correlating the neural signal of one participant for each region within each window with the mean signal within the same window for the same region from the remaining participants (leave-one-out method) (Figure 4C). The ISC method increases the signal-to-noise ratio (SNR) to detect stimulus-induced inter-regional correlations. The improvement in the SNR arises from filtering out intrinsic neural dynamics (for example, responses arising from intrinsic cognitive processes unrelated to ongoing stimulus processing) and non-neural artifacts (for example, respiratory rate and motion) that can influence network correlation patterns within the brain but are not correlated across participants. The time-series of neural signal from node i of participant j is $x_j^i(t)$, and this signal consists of a linear combination of the shared component c(t), idiosyncratic responses to the stimulus for participant j $\mathrm{id}_j^i(t)$, and spontaneous activity unrelated to the stimulus $\varepsilon_j^i(t)$. Therefore the time-series of participants a and b can be expressed as Equation [12], with shared c(t).

$$\begin{aligned} x_a^i(t) &= \alpha_a^i \cdot c(t)^i + \delta_a^i \cdot \mathrm{id}_a^i(t) + \varepsilon_a^i(t) \\ x_b^i(t) &= \alpha_b^i \cdot c(t)^i + \delta_b^i \cdot \mathrm{id}_b^i(t) + \varepsilon_b^i(t) \end{aligned} \tag{12}$$

$$ISC_{a,b}^i \approx \sqrt{\alpha_a^i \cdot \alpha_b^i}$$

In the previous ISC study, the correlation between participants a and b was approximated to the $\sqrt{\alpha_a^i \cdot \alpha_b^i}$ because id(t) and $\varepsilon(t)$ are not systematically correlated across participants, whereas c(t) is perfectly correlated [36]. However, in the current study, the idiosyncratic response conforms to the policy precision signal, in that, id(t) is not perfectly uncorrelated across participants. This is shown in the bottom plot of Figure 4D showing the ISC of the policy precision signals varied from 0 to 1, with a considerably high value. In the original logic of ISC, this idiosyncratic response should be averaged to a small value close to zero due to its inconsistent characteristic, such

that only c(t) can be isolated; however, correlations across idiosyncratic responses of the participants exist. Therefore, the neural ISC values corresponded to the values including not only the effect of shared component but also the idiosyncratic response component in the extent of ISC of the policy precision signal for participant j. Thus, the neural ISC value of each node i can be approximated as:

$$ISC^i \approx \beta_0 + \beta_1 \cdot \text{bISC} \tag{13}$$

where we can assume that the intercept term β_0 corresponds to the effect of α^i in Equation [12] and the β_1 corresponds to the effect of ISC of behavioral signal (bISC) such as policy precision. If the bISC is 0, the result is equivalent to the third line of Equation [12]. This can be applied to the time-series data of each participant (j) using dynamic ISC (dISC) of neural and dISC of behavioral signals (dbISC in Equation [14]) for performing GLM analysis.

$$dISC^i_j(t) = \beta_{0j} + \beta_{1j} \cdot \text{dbISC}_j(t) \tag{14}$$

Since action timings (AS) and feedback visual stimulus (FVS) are also different across participants, we performed a GLM analysis that included dISC of these signals, both with and without dISC of policy precision signal. We then compared the model performance using AIC, BIS, and deviance scores.
To validate the relevance of policy precision signals in explaining neural signals, we performed conventional GLM to identify which regions or interactions are related to this signal. Using beta coefficient values estimated from the GLM, we conducted a one-sample t-test to examine if those values are significantly positive or negative.

1 acknowledgments

This study was supported by the Medical Scientist Training Program from the Ministry of Science & ICT of Korea, and the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science & ICT (NRF-2022M3E5E8081200 and NRF-2016M3C7A1914448).

2 Data Availability

The data underlying this article will be shared upon reasonable request to the corresponding author. The code will be available at https://github.com/ydy-dyd-y/AIF_dFC

3 Supporting Information

Supplementary Figures S1–S6, and Tables S1 are presented in Supplementary file.

4 Competing Interests

The authors declare no competing interests.

5 Technical Terms

Bayesian brain Theory that considers the brain as a statistical organ of inference predicting current and future events based on experience.

prior belief The probability distribution or density of the causes of data that encode beliefs about those causes prior to observing the data.

posterior belief The conditional probability results from updating the prior belief with information.

Precision In statistics, the inverse variance or precision matrix quantifies certainty about variables and is also called the concentration matrix.

Free energy An information theory measure that provides an upper bound on the surprise when sampling data from a generative model.

Policy A sequence of actions, guiding its decision-making process to achieve specific objectives.

computational model An explanation of a complex system's operation using an algorithmic or mechanistic approach.

dynamic functional connectivity Changes of functional connectivity and statistical dependencies between different brain areas over a short time.

two-armed bandit task A classic reinforcement learning task that models decision-making by choosing the better option between two choices.

prediction error The difference between the observed outcomes and predicted value.

Brain state The evolving pattern of self-organized neural activity specific to a condition.

Phase coherence Cosine difference of phases between each pair of brain areas.

References

1. W. Y. Ahn, J. R. Busemeyer, E. J. Wagenmakers, and J. C. Stout. Comparison of decision learning models using the generalization criterion method. In *Cognitive Science*, volume 32, 2008.

2. W. Y. Ahn, G. Vasilev, S. H. Lee, J. R. Busemeyer, J. K. Kruschke, A. Bechara, and J. Vassileva. Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users. *Frontiers in Psychology*, 5(AUG), 2014.

3. V. M. Brown, J. Wilson, M. N. Hallquist, K. Szanto, and A. Y. Dombrovski. Ventromedial prefrontal value signals and functional connectivity during decision-making in suicidal behavior and impulsivity. *Neuropsychopharmacology*, 45(6), 2020.

4. R. L. Buckner and D. C. Carroll. Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 2007.

5. J. Cabral, D. Vidaurre, P. Marques, R. Magalhães, P. Silva Moreira, J. Miguel Soares, G. Deco, N. Sousa, and M. L. Kringelbach. Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest. *Scientific Reports*, 7(1), 12 2017.

6. R. L. Carhart-Harris and K. J. Friston. The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas, 2010.

7. C. Chen, T. Takahashi, S. Nakagawa, T. Inoue, and I. Kusumi. Reinforcement learning in depression: A review of computational research, 2015.

8. L. Clark, A. Y. Dombrovski, G. J. Siegle, M. A. Butters, C. L. Shollenberger, B. J. Sahakian, and K. Szanto. Impairment in Risk-Sensitive Decision-Making in Older Suicide Attempters With Depression. *Psychology and Aging*, 26(2), 2011.

9. J. D. Cohen, S. M. McClure, and A. J. Yu. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. In *Philosophical Transactions of the Royal Society B: Biological Sciences*, volume 362, 2007.

10. R. W. Cox. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3), 1996.

11. R. W. Cox and J. S. Hyde. Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, 10(4-5), 1997.

12. A. M. Dale, B. Fischl, and M. I. Sereno. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9(2), 1999.

13. J. S. Damoiseaux, S. A. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America*, 103(37), 2006.

14. M. Dastjerdi, B. L. Foster, S. Nasrullah, A. M. Rauschecker, R. F. Dougherty, J. D. Townsend, C. Chang, M. D. Greicius, V. Menon, D. P. Kennedy, and J. Parvizi. Differential electrophysiological response during rest, self-referential, and non-self-referential tasks in human posteromedial cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2011.

15. G. Deco and M. Kringelbach. Correction to Metastability and Coherence: Extending the Communication through Coherence Hypothesis Using a Whole-Brain Computational Perspective [Trends in Neurosciences, 39, 3, (2016), 125-135], doi: 10.1016/j.tins.2016.01.001, 2016.

16. E. W. Dickie, A. Anticevic, D. E. Smith, T. S. Coalson, M. Manogaran, N. Calarco, J. D. Viviano, M. F. Glasser, D. C. Van Essen, and A. N. Voineskos. Ciftify: A framework for surface-based analysis of legacy MR acquisitions. *NeuroImage*, 197, 2019.

17. A. Y. Dombrovski, M. N. Hallquist, V. M. Brown, J. Wilson, and K. Szanto. Value-Based Choice, Contingency Learning, and Suicidal Behavior in Mid- and Late-Life Depression. *Biological Psychiatry*, 85(6), 2019.

18. B. L. Edlow, E. Takahashi, O. Wu, T. Benner, G. Dai, L. Bu, P. E. Grant, D. M. Greer, S. M. Greenberg, H. C. Kinney, and R. D. Folkerth. Neuroanatomic connectivity of the human ascending arousal system critical to consciousness and its disorders. *Journal of Neuropathology and Experimental Neurology*, 71(6), 2012.

19. C. Findling, V. Skvortsova, R. Dromnelle, S. Palminteri, and V. Wyart. Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature Neuroscience*, 22(12), 2019.

20. B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 2002.

21. K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny. Variational free energy and the Laplace approximation. *NeuroImage*, 34(1), 2007.

22. K. Friston, P. Schwartenbeck, T. FitzGerald, M. Moutoussis, T. Behrens, and R. J. Dolan. The anatomy of choice: Dopamine and decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 2014.

23. M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, and M. Jenkinson. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–124, 10 2013.

24. R. J. Herrnstein, H. Rachlin, and D. I. Laibson. The matching law: Papers in psychology and economics. *New York: Cambridge, MA. Kacelnik, A*, 53, 1997.

25. Q. J. Huys, M. Browning, M. P. Paulus, and M. J. Frank. Advances in the computational understanding of mental illness, 1 2021.

26. Q. J. Huys, T. V. Maia, and M. J. Frank. Computational psychiatry as a bridge from neuroscience to clinical applications, 2 2016.

27. H. J. Jo, Z. S. Saad, W. K. Simmons, L. A. Milbury, and R. W. Cox. Mapping sources of correlation in resting state FMRI, with artifact detection and removal. *NeuroImage*, 52(2), 2010.

28. R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering, Transactions of the ASME*, 82(1), 1960.

29. D. Kim, G. Y. Park, J. P. ODoherty, and S. W. Lee. Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning. *Nature Communications*, 10(1), 2019.

30. K. Kim, S. W. Kim, W. Myung, C. E. Han, M. Fava, D. Mischoulon, G. I. Papakostas, S. W. Seo, H. Cho, J. K. Seong, and H. J. Jeon. Reduced orbitofrontal-thalamic functional connectivity related to suicidal ideation in patients with major depressive disorder. *Scientific Reports*, 7(1), 2017.

31. K. Lehmann, D. Bolis, K. J. Friston, L. Schilbach, M. J. Ramstead, and P. Kanske. An Active-Inference Approach to Second-Person Neuroscience. *Perspectives on Psychological Science*, 2023.

32. T. V. Maia and M. J. Frank. From reinforcement learning models to psychiatric and neurological disorders. In *Nature Neuroscience*, volume 14, 2011.

33. H. P. Maretic and P. Frossard. Graph Laplacian Mixture Model. *IEEE Transactions on Signal and Information Processing over Networks*, 6:261–270, 2020.

34. K. J. Miller, M. M. Botvinick, and C. D. Brody. Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*, 20(9):1269–1276, 9 2017.

35. P. R. Montague, R. J. Dolan, K. J. Friston, and P. Dayan. Computational psychiatry, 2012.

36. S. A. Nastase, V. Gazzola, U. Hasson, and C. Keysers. Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, 14(6), 2019.

37. A. R. Otto, E. G. Taylor, and A. B. Markman. There are at least two kinds of probability matching: Evidence from a secondary task. *Cognition*, 118(2), 2011.

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

39. W. M. Pauli, A. N. Nili, and J. Michael Tyszka. Data Descriptor: A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Scientific Data*, 5, 2018.

40. J. M. Pearson, S. R. Heilbronner, D. L. Barack, B. Y. Hayden, and M. L. Platt. Posterior cingulate cortex: Adapting behavior to a changing world, 2011.

41. A. Ponce-Alvarez, G. Deco, P. Hagmann, G. L. Romani, D. Mantini, and M. Corbetta. Resting-State Temporal Synchronization Networks Emerge from Connectivity Topology and Heterogeneity. *PLoS Computational Biology*, 11(2), 2015.

42. J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, and S. E. Petersen. Functional Network Organization of the Human Brain. *Neuron*, 72(4), 2011.

43. M. E. Raichle and M. E. Raichle. Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, 2(10), 2001.

44. R. A. Rescorla and A. R. Wagner. A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. *Clasical conditioning II: current research and theory*, 1972.

45. I. Ricchi, A. Tarun, H. P. Maretic, P. Frossard, and D. Van De Ville. Dynamics of functional network organization through graph mixture learning. *NeuroImage*, 252, 2022.

46. E. C. Robinson, S. Jbabdi, M. F. Glasser, J. Andersson, G. C. Burgess, M. P. Harms, S. M. Smith, D. C. Van Essen, and M. Jenkinson. MSM: A new flexible framework for multimodal surface matching. *NeuroImage*, 100, 2014.

47. M. Roy, D. Shohamy, and T. D. Wager. Ventromedial prefrontal-subcortical systems and the generation of affective meaning, 2012.

48. S. Rupprechter, A. Stankevicius, Q. J. Huys, J. D. Steele, and P. Seriès. Major Depression Impairs the Use of Reward Values for Decision-Making. *Scientific Reports*, 8(1), 2018.

49. A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. T. Yeo. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, 28(9):3095–3114, 9 2018.

50. E. Schulz and S. J. Gershman. The algorithmic architecture of exploration in the human brain, 2019.

51. R. Smith, K. J. Friston, and C. J. Whyte. A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, 107, 2022.

52. M. Steyvers, M. D. Lee, and E. J. Wagenmakers. A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, 53(3), 2009.

53. P. A. Taylor, G. Chen, D. R. Glen, J. K. Rajendra, R. C. Reynolds, and R. W. Cox. FMRI processing with AFNI: Some comments and corrections on "Exploring the Impact of Analysis Software on Task fMRI Results". *bioRxiv*, 2018.

54. B. T. Thomas Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fisch, H. Liu, and R. L. Buckner. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165, 9 2011.

55. Y. Tian, D. S. Margulies, M. Breakspear, and A. Zalesky. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nature Neuroscience*, 23(11), 2020.

56. G. Turecki, D. A. Brent, D. Gunnell, R. C. O'Connor, M. A. Oquendo, J. Pirkis, and B. H. Stanley. Suicide and suicide risk, 2019.

57. V. S. Turner, R. O. O'Sullivan, and M. A. Kheirbek. Linking external stimuli with internal drives: A role for the ventral hippocampus, 10 2022.

58. J. Vohryzek, G. Deco, B. Cessac, M. L. Kringelbach, and J. Cabral. Ghost Attractors in Spontaneous Brain Activity: Recurrent Excursions Into Functionally-Relevant BOLD Phase-Locking States. *Frontiers in Systems Neuroscience*, 14, 4 2020.

59. D. A. Worthy, M. J. Hawthorne, and A. R. Otto. Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic Bulletin and Review*, 20(2), 2013.

60. D. A. Worthy and W. Todd Maddox. A comparison model of reinforcement-learning and win-stay-lose-shift decision-making processes: A tribute to W.K. Estes. *Journal of Mathematical Psychology*, 59(1), 2014.

61. J. S. Yi, S. O. Bae, Y. M. Ahn, and D.-b. Park. Validity and Reliability of the Korean Version of the Hamilton Depression Rating Scale (K-HDRS). *Journal of the Korean Neuropsychiatric Association*, 44(4), 2005.