Contents

1 Supplementary Text

1.1 Model comparison

We utilized a nested model comparison to determine how well four different computational models explained the observed behavior. These models allowed for uncertainty and novelty to contribute to decision making beyond what could be implemented with a simpler reinforcement learning framework and also allowed incorporating patients' individual preferences. The compared models had two main mechanisms for how novelty, uncertainty, and q-value interacted to create stimulus utility: uncertainty bonuses and novelty initiation biases (see Materials and Methods for detailed model descriptions).

To create the fully nested model analysis, our first model was a simple reinforcement learning model with a learning rate α and an inverse temperature β for the softmax function which determines decision probabilities given the two stimulus utilities. The second model (uncertainty bonus model) we tested worked by adding an uncertainty bonus to stimulus utilities, according to each patient's uncertainty preference. This model also included the learning rate and the inverse temperature parameters. In this model, stimulus utility is equal to a linear combination of the stimulus q-value and the uncertainty bonus. The third model had a learning rate, an inverse temperature, and a novelty bias incorporated into the initial q-value of each stimulus at the beginning of a block. This model attempts to capture patients' tendency to assign biased values to novel stimuli with the purpose of directing exploration (Wittmann et al., 2008). The fourth model we tested (uncertainty bonus + novelty initiation bias) expands on the previous models by incorporating all of its elements, including uncertainty bonus and novelty initiation biases together.

We performed model fitting and comparisons for the four models across the patient population using hierarchical Bayesian inference (Piray et al., 2019) (see Supplementary Fig. 1 for values of fit parameters). The estimated model frequencies for Models 1-4 were (0.006,0.558,0.001,0.433), indicating that Models 2 (uncertainty bonus only) and 4 (uncertainty bonus + novelty initiation bias) shared most of the responsibility in explaining patient behavior in this task.

These behavioral modeling results show that a combination of uncertainty-based and novelty-based mechanisms is appropriate to model the behavior of our subjects. Therefore, for all subsequent neural data analysis we used the variables derived from the model which better explained the data for each session. In total, Model 2 was used for 13/22 sessions, while Model 4 was used for 9/22 sessions.

1.2 Additional behavioral analysis

We summarized model parameter fits to gain more insight into the behavior of the patients as a group (Supplementary Fig. 1 A,B), using the parameters from each session's best explanatory model. For the uncertainty bonus model (Model 2), the mean softmax inverse temperature was 4.02 ± 0.62 and the mean learning rate was 0.35 ± 0.03. The uncertainty intercept (uI) parameter represents how much value was assigned to uncertainty. On average, this parameter was negative (-0.22 ± 0.04, $p < 0.001$, t-test), indicating a slight uncertainty avoidance for sessions best explained by Model 2. For the uncertainty bonus + novelty initiation bias model (Model 4), the mean softmax inverse temperature was 5.1 ± 0.7 and the mean learning rate was 0.34 ± 0.05. The novelty initiation bias indicates a initial tendency to overvalue novel stimuli (0.49 ± 0.02, $p < 0.001$, t-test), which can support early exploratory behavior. Additionally, in this model the uncertainty intercept was not significantly different from zero across sessions (0.0008 ± 0.06, $p = 0.982$, t-test), indicating a variety of uncertainty preferences in patients whose behavior was also influenced by novelty.

To ensure that the models correctly reproduce the effects of expected value, uncertainty and novelty observed in our subjects, we simulated behavior to perform model fits and performed a posterior predictive check analysis. We exposed Models 2 and 4 to the same sequence of trials each patient experienced, with each patient's model fits (for their respective best explanatory model). We then generated a decision for each trial, using the decision probabilities inferred from the model. To directly compare real patient behavior with behavior simulated from model estimates, we fit simulated choices using a model-agnostic logistic regression to map effects of the various variables on choices, and then compared these fits with those obtained with true behavior in the task. To account for variability in probabilistic decisions, we repeated this procedure 100 times and generated a distribution of regression coefficient estimates, which we compared to the actual effects observed in the subjects. To quantify the difference in how these models recovered effects of expected value, uncertainty and novelty on behavior, we computed the 95% confidence intervals for the logistic regression estimates of value feature effects on actual patient decisions (as displayed in Fig. 1, see Materials and Methods). Then, we obtained the proportion of overlap between these confidence intervals and the recovered effect estimates obtained from 100 iterations of simulated sessions in the posterior predictive check analysis (Supplementary Fig. 1 D,E). Using the uncertainty bonus model, we observed an overlap of $[93\%, 23\%, 73\%]$ for the estimates of expected, uncertainty and novelty, respectively. With the uncertainty bonus + novelty initiation bias model, we obtained overlaps of $[75\%, 94\%, 89\%]$. Therefore, while Model 2 was able to more parsimoniously describe behavior in a larger number of sessions, parameter estimates from Model 4 resulted in simulated behavior which more closely resembled qualitative uncertainty preferences from

real behavior. Additionally, both models generated parameter estimates which successfully captured qualitative EV and novelty preference patterns from real behavior. It is possible that the smaller overlap in uncertainty estimates for the uncertainty bonus model is due to the patient cohort for this model concentrating uncertainty averse individuals, leading to negatively biased behavioral estimates, while the estimates from simulated behavior take into account parameters which are hierarchically fit, taking into account the entire patient population.

To quantify whether simulated decisions themselves are well explained by the model, we constructed a R^2 metric which summarizes how high the modeled probabilities were for each simulated decision. We display the results for each of the 100 simulated iterations and each of the 22 sessions in Supplementary Fig. 1C. Intuitively, if a model succeeds at predicting decisions, then the probability assigned by the model to each simulated decision should be higher than for a model which predicts decisions at random. Note that decision probabilities $pChoice$ are obtained from a softmax comparison between the two stimulus utilities. Therefore, we utilize the following definitions (Cockburn et al., 2022): the total log-likelihood of the model, applied to a simulated decision time series containing T trials, is

$$LLE = \sum_{t=1}^{T} \log(pChoice(t)) \tag{1}$$

For a null model which always predicts that the two decisions are equally likely ($pChoice = 0.5$), its log-likelihood is

$$LLE_{null} = \sum_{t=1}^{T} \log(0.5) = T \log(0.5) \tag{2}$$

Finally, we define the R^2 as follows. Note that a model which always predicts decisions perfectly should have $R^2 = 1$ and a model which predicts decisions at random should average $R^2 = 0$, while a model which predicts every decision wrongly would have R^2 diverge to negative infinity.

$$R^2 = 1 - \frac{LLE}{LLE_{null}} \tag{3}$$

1.3 Encoding rejected values

Models of decision making often also depend on maintaining information about the value of options not chosen to evaluate the outcome received. We therefore also examined whether rejected q-values, uncertainty bonuses and novelties were represented (Supplementary Fig. 3 A-C). Unlike selected q-values, preSMA did represent rejected q-values (trial onset 13.9%, $p = 0.002$:, pre-decision: 14.7%, $p = 0.002$, permutation test). Additionally, vmPFC represented rejected uncertainty bonuses (trial onset 14.5%, $p = 0.002$:, pre-decision: 11.0%, $p = 0.004$, permutation test). Neuron count p-values were Bonferroni corrected.

1.4 Encoding unseen values

One hypothesis is that the value of the option that was not presented in a trial is also tracked by neurons. To test this hypothesis, we focused on sessions which contained 3 possible stimuli per block (long version, see Supplementary Table 1), and tracked the integrated utility value of stimuli which were not offered in each trial (unseen option GLM, Supplementary Table 2). We found that vmPFC contained a significant neuron count for the utility unseen stimuli in a trial (trial onset: 14.9%, $p = 0.002$; pre-decision: 11.9%, $p = 0.002$, permutation test).

1.5 Exploratory signaling in vmPFC uncertainty neurons

One potential reason for representing the uncertainty of the selected option is to enable exploratory decision making, which would entail deliberately choosing an item with lower q-value. We therefore divided trials into putative explore and non-explore categories. Trials in which the patient chose the option which had the lower q-value but the higher uncertainty bonus were classified as putative explore trials, while all others were classified as non-explore trials (Supplementary Fig. 3G).

Then, in the sub-populations of vmPFC and preSMA neurons in the pre-decision period that were sensitive to selected uncertainty, we performed a Poisson GLM analysis using the explore trial flag as a regressor (Explore Flag GLM, Supplementary Table 2), correcting for selected uncertainty as a regressor of no interest. We subsequently tested whether neurons whose activity was significantly modulated by the explore flag significantly overlapped with the sub-populations of vmPFC and preSMA neurons that encoded selected uncertainty (Supplementary Fig. 3H). We found a significant overlap in preSMA ($p < 0.001$, Jaccard index test), but not in vmPFC ($p = 0.059$, Jaccard index test). Therefore, a significant proportion of preSMA selected uncertainty neurons signal whether a trial is exploratory or not prior to button press, suggesting a possible neural substrate through which exploratory decisions can be monitored.

2 Supplementary Tables

Positive neurons				
		vmPFC	dACC	preSMA
action q-value	trial onset	5/8 (62.5 %) p=0.727	6/9 (66.7 %) p=0.508	7/18 (38.9 %) p=0.481
	pre-decision	4/7 (57.1 %) p=1	3/6 (50 %) p=1	6/17 (35.3 %) p=0.332
		vmPFC	dACC	preSMA
action unc. bonus	trial onset	2/17 (11.8 %) p=0.002	3/7 (42.9 %) p=1	8/15 (53.3 %) p=1
	pre-decision	5/17 (29.4 %) p=0.143	0/8 (0 %) p=0.007	8/11 (72.7 %) p=0.227
		vmPFC	dACC	preSMA
action novelty	trial onset	7/12 (58.3 %) p=0.774	3/7 (42.9 %) p=1	3/7 (42.9 %) p=1
	pre-decision	3/7 (42.9 %) p=1	4/7 (57.1 %) p=1	5/7 (71.4 %) p=0.453
		vmPFC	dACC	preSMA
sel. q-value	trial onset	11/18 (61.1 %) p=0.481	6/13 (46.2 %) p=1	7/15 (46.7 %) p=1
	pre-decision	10/17 (58.8 %) p=0.629	4/9 (44.4 %) p=1	2/12 (16.7 %) p=0.038
		vmPFC	dACC	preSMA
sel. unc. bonus	trial onset	4/17 (23.5 %) p=0.049	2/9 (22.2 %) p=0.180	8/22 (36.4 %) p=0.286
	pre-decision	5/22 (22.7 %) p=0.016	1/11 (9.09 %) p=0.011	7/16 (43.8 %) p=0.804
		vmPFC	dACC	preSMA
sel. novelty	trial onset	9/16 (56.3 %) p=0.804	3/7 (42.9 %) p=1	6/12 (50 %) p=1
	pre-decision	9/18 (50 %) p=1	3/6 (50 %) p=1	4/9 (44.4 %) p=1
		vmPFC	dACC	preSMA
action utility	trial onset	3/6 (50 %) p=1	3/7 (42.9 %) p=1	12/28 (42.9 %) p=0.572
	pre-decision	1/7 (14.3 %) p=0.125	3/7 (42.9 %) p=1	5/19 (26.3 %) p=0.0636
		vmPFC	dACC	preSMA
sel. utility	trial onset	7/14 (50 %) p=1	6/14 (42.9 %) p=0.791	8/22 (36.4 %) p=0.286
	pre-decision	9/21 (42.9 %) p=0.664	3/10 (30 %) p=0.344	4/19 (21.1 %) p=0.019

Supplementary Table 1: Proportion of neurons significantly coding each variable of interest with a positive correlation (i.e. higher firing rates for higher values of each variable). For the action variables, which had left and right components (i.e. action q-values, uncertainty bonus, novelty, and utility), we only included neurons which coded both the left and right components positively or negatively, excluding neurons which had different signs for the t-score corresponding to left and right components. We obtained p-values for a 2-tailed binomial test for whether either positive or negative coding was larger than expected by chance, assuming a chance level of 0.5.

Name	Model	Periods
Action-value model	$\log(E(Y\|x)) = b_0 + b_1 Q_L + b_2 Q_R + b_3 B_L + b_4 B_R + b_5 N_L + b_6 N_R$	trial onset, pre-decision
Selection-based model	$\log(E(Y\|x)) = b_0 + b_1 Q_{sel} + b_2 Q_{rej} + b_3 B_{sel} + b_4 B_{rej} + b_5 N_{sel} + b_6 N_{rej}$	trial onset, pre-decision
Selection-based utility model	$\log(E(Y\|x)) = b_0 + b_1 U_{sel} + b_2 U_{rej}$	trial onset, pre-decision
Decision and utility model	$\log(E(Y\|x)) = b_0 + b_1 U_L + b_2 U_R + b_3 Decision$	trial onset, pre-decision
Unseen option model	$\log(E(Y\|x)) = b_0 + b_2 U_{unseen} + b_2 U_{sel} + b_3 U_{rej}$	trial onset, pre-decision
Explore flag model	$\log(E(Y\|x)) = b_0 + b_1 Explore + b_2 B_{sel}$	pre-decision
Outcome model	$\log(E(Y\|x)) = b_0 + b_1 O + b_2 Q_{sel} + b_3 \|RPE\|$	outcome

Supplementary Table 2: Models for Poisson GLM single neuron encoding analysis. Action-value model: includes q-value, uncertainty, and novelty, splitting by values of stimuli presented on the left or the right of the screen. In all models, x is a shorthand for the independent variable vector and Y is the spike count in the window of interest. Selection-based model: Same, but splitting by values of stimuli which were chosen or rejected in each trial. Selection-based utility model: includes the utility of the selected and rejected stimuli. Decision and utility model: includes the utility of stimuli presented on the left and the right of the screen, and the decision made on each trial (binary indicator for left/right button press). Unseen option model (fit only in sessions with 3 stimuli, not utilized in sessions with 5 stimuli): includes the q-value of the option not presented in each trial, controlling for the q-values of the selected and rejected options. Explore flag model: includes a binary flag for whether patients made an exploratory decision in each trial or not, correcting by the uncertainty bonus of the selected option. Exploratory trials were those in which the selected option had a higher uncertainty bonus and lower q-value than the rejected option. Outcome model: includes the binary outcome of each trial (win/no-win), selected q-values, and absolute reward prediction errors.

Patient ID	Sex	Age	Left dACC	Right dACC	Left preSMA	Right preSMA	Left vmPFC	Right vmPFC
P71CS	M	40	-1.00,27.97,27.02	7.68,28.87,23.68	-1.89,13.51,38.03	5.09,11.30,39.61	-3.04,32.45,-0.64	5.06,34.29,6.30
P70CS	F	30	-0.84,23.06,26.04	4.25,22.06,22.63	-9.45,11.24,46.67	12.40,11.55,46.60	-0.20,26.30,-6.79	10.50,29.65,-3.63
P69CS	F	41	-6.13,27.44,22.69	-1.66,27.37,22.25	-2.02,23.89,39.02	2.92,32.34,40.50	-3.73,38.30,-11.69	2.26,29.24,-9.80
P67CS	F	38	-4.82,27.63,-11.69	-0.27,28.18,-9.13	-2.63,10.61,37.48	7.38,11.36,36.23	-4.70,27.32,-14.06	-1.21,28.12,-7.68
P65CS	F	55	-4.66, 24.50, 24.48	5.66, 30.60, 25.28	-0.82, 10.87, 52.78	1.53, 18.10, 48.52	-3.69, 45.06, -12.11	6.46, 30.94, -11.70
P64CS	F	63	-7.45, 35.50, 22.73	5.54, 32.54, 30.78	-6.19, 21.34, 47.67	7.46, 22.79, 44.56	-8.18, 30.42, -13.39	6.21, 31.14, -12.75
P63CS	F	48	-5.57, 27.44, 26.20	11.66, 34.65, 33.43	-8.53, 10.02, 41.57	9.46, 28.48, 48.52	-3.52, 45.04, -14.44	9.54, 37.31, -13.49
P62CS	F	25	NaN	9.02, 15.72, 29.33	NaN	5.69, 16.28, 49.18	NaN	11.02, 34.13, -10.05
P61CS	F	52	0.00, 19.15, 22.74	6.82, 20.21, 27.63	-4.55, 14.50, 38.70	14.92, 12.05, 36.59	-3.40,41.84,-15.66	6.99,31.03,-12.39
P60CS	M	67	-8.66, 30.23, 27.13	5.62, 31.09, 20.05	-8.70, 19.60, 44.90	5.55, 26.23, 42.00	-1.49,31.63,-16.97	2.48,33.28,-20.24
P56CS	M	48	-7.01, 23.11, 30.17	16.27, 24.79, 25.90	-6.92, 6.43, 46.70	29.08, 13.01, 45.88	-6.96,34.62,-13.05	27.02, 31.69, -5.48
P55CS	F	43	-4.14, 28.97, 24.60	3.20, 28.05, 21.05	-3.88, 11.08, 46.18	7.58, 13.24, 42.29	-8.64, 36.74, -6.98	6.32, 35.04, -11.70
P54CS	F	59	-3.85, 29.02, 23.70	7.93, 33.96, 26.74	-5.98, 20.87, 49.08	10.72, 26.92, 48.06	-3.53, 36.85, -16.95	3.94, 31.75, -14.51
P53CS	M	60	0.95, 32.02, 27.76	9.47, 36.81, 26.07	-4.30, 12.32, 51.06	10.21, 18.90, 49.06	-0.97, 35.20, -17.12	6.83, 31.74, -14.96
P51CS	M	17	-5.48, 31.63, 27.98	6.65, 30.73, 25.95	-8.81, 8.15, 44.07	11.91, 9.98, 47.01	-1.03, 36.31, -10.91	4.19,30.23,-15.23
P49CS	F	24	-8.72, 34.10, 25.92	6.21, 31.04, 25.99	-5.54, 25.16, 31.55	6.86, 26.19, 38.56	-7.02, 36.94, -13.11	10.87, 35.06, -5.09
P48CS	F	32	-3.34, 28.54, 24.20	5.87, 40.15, 17.96	-7.58, 24.03, 37.93	11.91, 23.85, 40.98	-8.06, 36.04, -6.05	13.66, 37.28, -17.09
P47CS	M	32	-0.67, 25.84, 20.14	1.43, 25.49, 17.20	-7.86, 12.20, 42.04	8.11, 11.45, 45.05	-1.98, 40.24, -15.36	1.02,38.60,-13.11
P43CS	F	42	-2.85, 23.88, 24.58	5.29, 27.07, 24.76	-2.83, 9.40, 43.93	5.33, 8.85, 46.26	-4.26, 38.19, -12.07	5.11,38.48,-16.47
P41CS	F	21	0.18, 26.56, 17.92	12.67, 27.14, 19.01	-0.04, 12.88, 58.69	6.99, 15.53, 57.30	-3.17, 37.28, -7.56	8.67, 38.04, -15.47

Supplementary Table 3: Patient sex, age, and MNI coordinates for microelectrode positioning in dACC, preSMA and vmPFC.

Task version	Patients who performed it
Longer (300 trials)	P60,P61,P62,P63,P64,P65,P67,P69,P70,P71
Shorter (206 trials)	P41,P43,P48,P49,P51,P54,P55,P56

Supplementary Table 4: Patients who performed the longer (300 trials) or shorter (206 trials) version of the task. For all behavioral and neural analyses, datasets from both task versions were pooled.

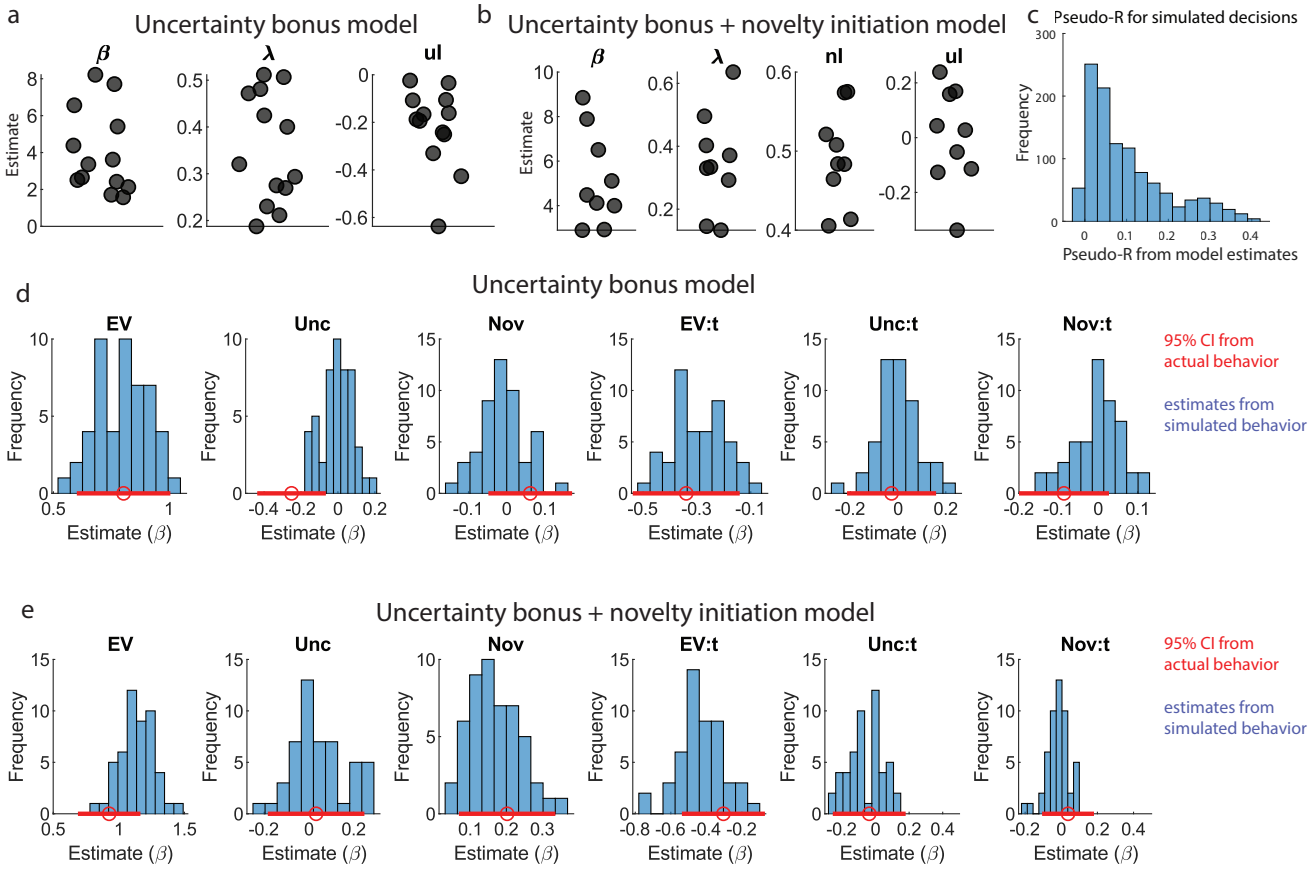
Variable	**Abbreviation**
Q-value	Q
Uncertainty bonus	B
Novelty	N
Utility	U
Outcome	O

Supplementary Table 5: Abbreviations for Poisson GLM regressors: q-value, uncertainty bonus, novelty, utility, and outcome.

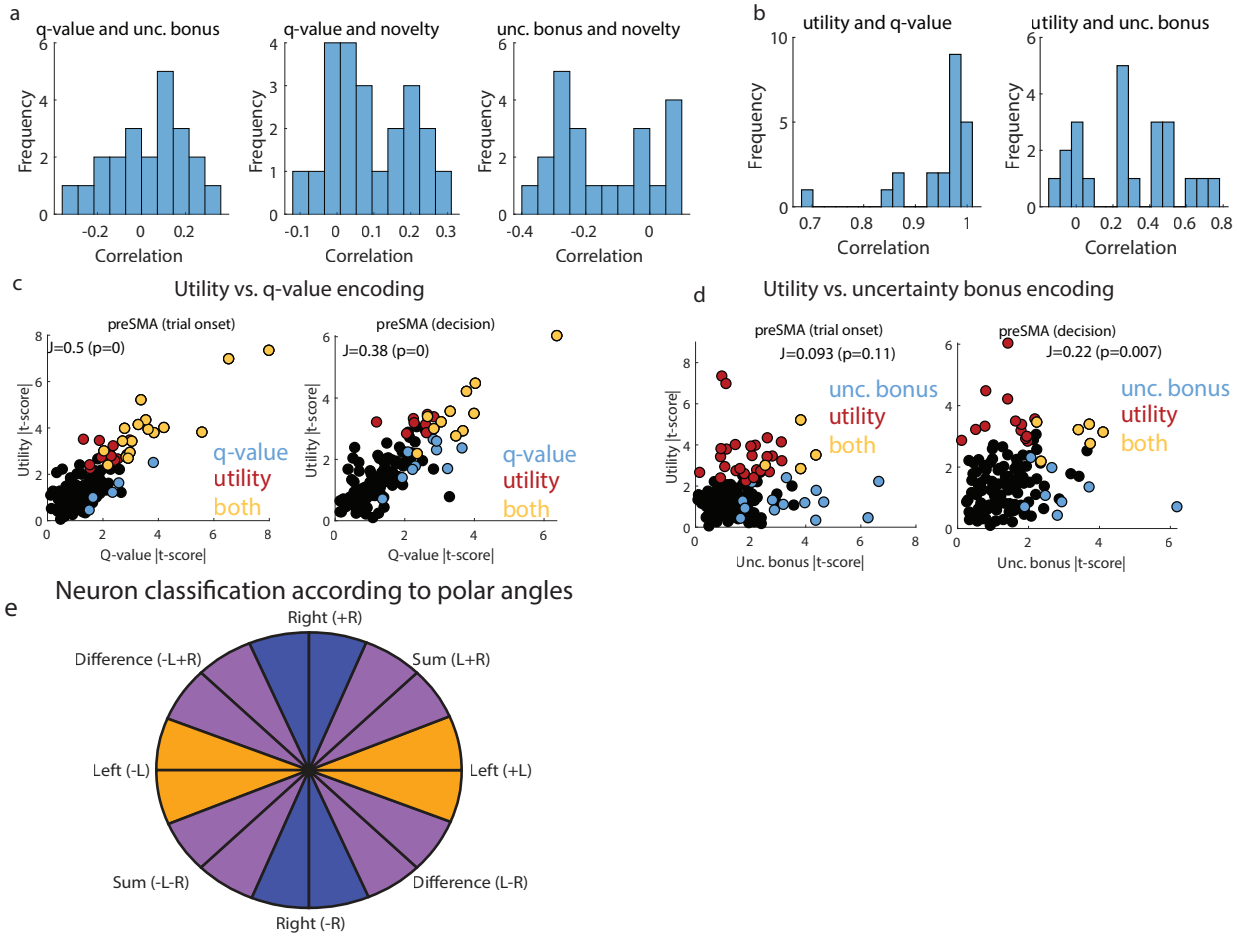
		vmPFC	dACC	preSMA
action q-value	trial onset	12/172 (6.98 %) p=0.180	11/102 (10.8 %) p=0.012	22/136 (16.2 %) p=0.002
	pre-decision	14/172 (8.14 %) p=0.072	7/102 (6.86 %) p=0.258	21/136 (15.4 %) p=0.002
		vmPFC	dACC	preSMA
action unc. bonus	trial onset	17/172 (9.88 %) p=0.002	10/102 (9.8 %) p=0.032	18/136 (13.2 %) p=0.002
	pre-decision	18/172 (10.5 %) p=0.004	8/102 (7.84 %) p=0.154	14/136 (10.3 %) p=0.012
		vmPFC	dACC	preSMA
action novelty	trial onset	11/172 (6.4 %) p=0.258	7/102 (6.86 %) p=0.264	7/136 (5.15 %) p=0.522
	pre-decision	10/172 (5.81 %) p=0.396	7/102 (6.86 %) p=0.248	10/136 (7.35 %) p=0.16
		vmPFC	dACC	preSMA
sel. q-value	trial onset	18/172 (10.5 %) p=0.002	13/102 (12.7 %) p=0.002	14/136 (10.3 %) p=0.014
	pre-decision	16/172 (9.3 %) p=0.018	9/102 (8.82 %) p=0.068	12/136 (8.82 %) p=0.050
		vmPFC	dACC	preSMA
sel. unc. bonus	trial onset	17/172 (9.88 %) p=0.002	9/102 (8.82 %) p=0.070	20/136 (14.7 %) p=0.002
	pre-decision	21/172 (12.7 %) p=0.002	11/102 (10.8 %) p=0.022	15/136 (11 %) p=0.002
		vmPFC	dACC	preSMA
sel. novelty	trial onset	15/172 (8.72 %) p=0.022	7/102 (6.86 %) p=0.246	12/136 (8.82 %) p=0.040
	pre-decision	18/172 (10.5 %) p=0.002	6/102 (5.88 %) p=0.414	8/136 (5.88 %) p=0.398
		vmPFC	dACC	preSMA
action utility	trial onset	8/172 (4.65 %) p=0.664	10/102 (9.8 %) p=0.048	29/136 (21.3 %) p=0.002
	pre-decision	8/172 (4.65 %) p=0.658	7/102 (6.86 %) p=0.270	19/136 (14 %) p=0.002
		vmPFC	dACC	preSMA
decision	trial onset	14/172 (8.14 %) p=0.05	3/102 (2.94 %) p=0.892	11/136 (8.09 %) p=0.084
	pre-decision	11/172 (5.81 %) p=0.362	4/102 (3.92 %) p=0.786	19/136 (14 %) p=0.002
		vmPFC	dACC	preSMA
sel. utility	trial onset	14/172 (8.14 %) p=0.056	14/102 (13.7 %) p=0.002	22/136 (16.2 %) p=0.002
	pre-decision	20/172 (11.6 %) p=0.002	9/102 (8.82 %) p=0.080	17/136 (12.5 %) p=0.002

Supplementary Table 6: Poisson GLM encoding analysis results for each variable of interest, indicating total number of significant neurons relative to total neuron count in each brain area. Uncorrected p-values from permutation tests are reported.

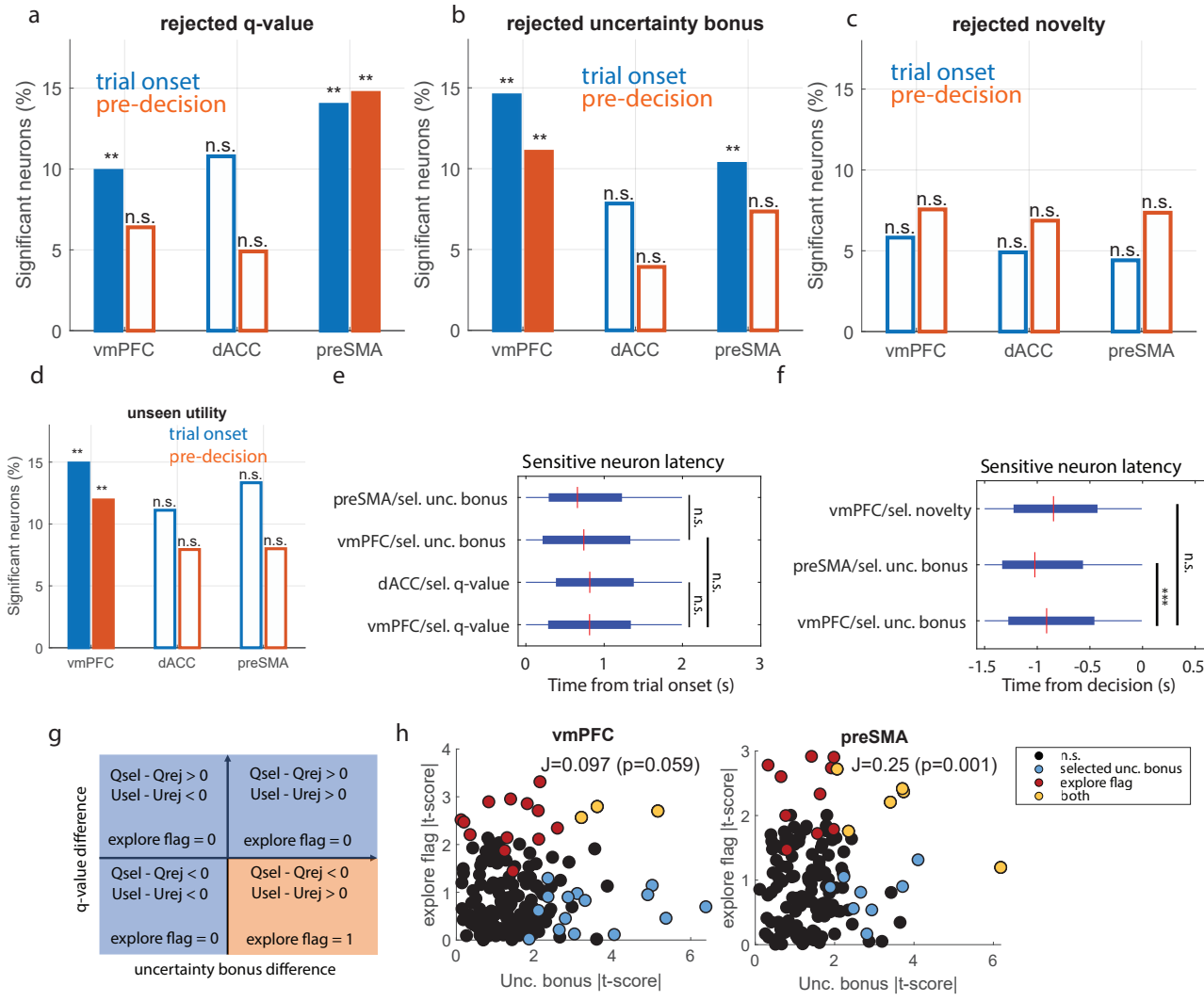
3 Supplementary Figures



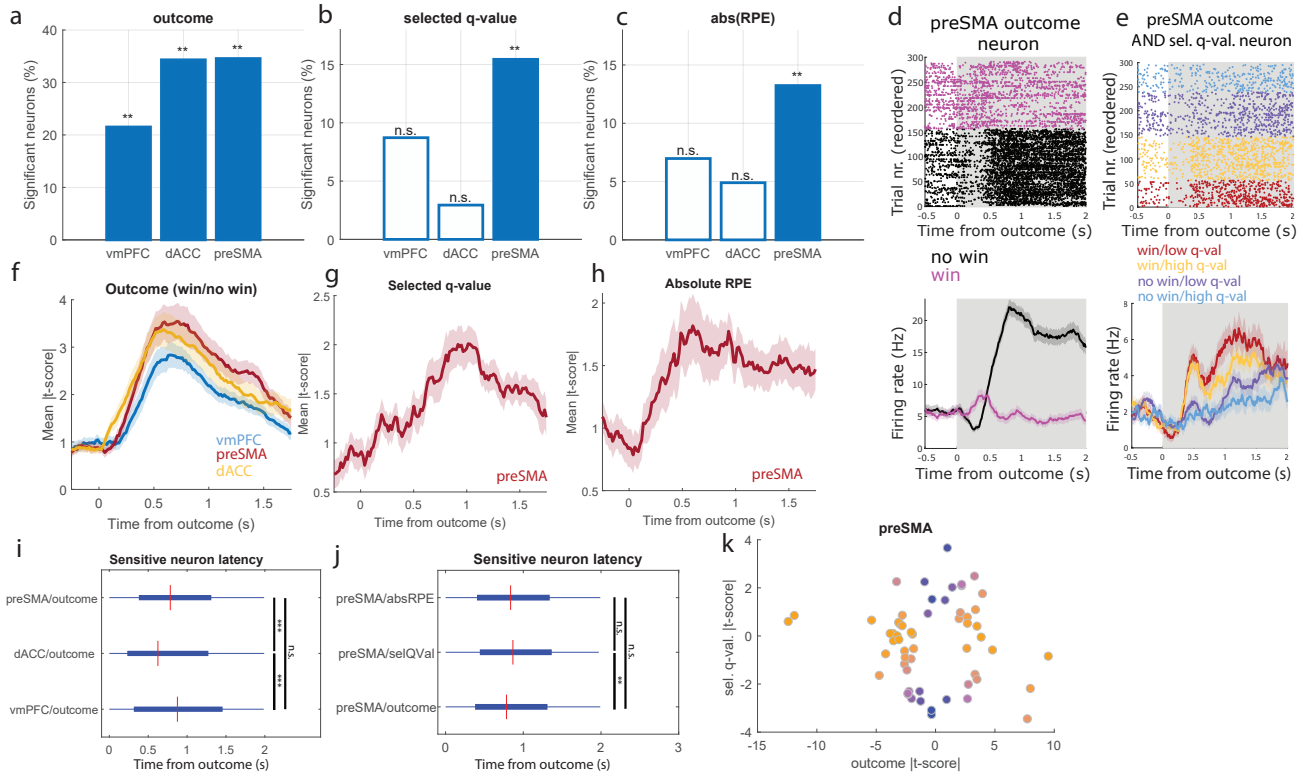
Supplementary Figure 1: Model fits, simulations, and posterior predictive check for exploration models with uncertainty and novelty. (a) Individual uncertainty bonus model parameter fits. Each dot represents a parameter fit for each patient (Left to right: softmax inverse temperature β; learning rate λ; novelty intercept; uncertainty intercept). (b) Same, for model with uncertainty bonus and novelty initiation bonus. (c) Pseudo-R metric for simulated behavior fits. For each simulated session, a pseudo-R score is obtained by comparing its simulated decisions and the decision probabilities assigned to them by the computational model fit to behavior from that session. (d) Posterior predictive check for a model which included a softmax beta, a learning rate, and an uncertainty bonus intercept as free parameters. We fit this model to patient behavior and re-exposed a simulated agent with the obtained model parameters to the same set of trials which patients experienced 50 times, to generate decisions according to the estimated decision probabilities. We then fit a logistic regression for the effect of each variable (left to right: expected values, uncertainty, novelty, and their respective interactions with trial number) on decision in the artificial agents (blue histogram) and compared it to the actual estimate given true decisions concatenated across patients (red line; dot indicates mean and bars indicate 95% confidence interval). (c) Same, for the model which has all the previously mentioned parameters plus a novelty initiation bonus.



Supplementary Figure 2: Summarizing action-value encoding and comparing encoding of utility versus value components. (a) Left: Histogram of correlation between action q-value and action uncertainty bonus trial vectors, across recording sessions. Center: same, for action q-value and action novelty. Right: same, for action uncertainty bonus and action novelty. (b) Histogram of correlation between utility and q-value (left), or utility and uncertainty bonus (right) trial vectors, across recording sessions. (c) Given the sizeable correlations between utility and its components, we mapped out the extent to which utility preSMA neurons also correlated with q-value preSMA neurons, in the trial onset period (left), and the pre-decision period (right). We plot q-value t-scores from the action-value GLM, as well as the utility t-scores from the utility and decision GLM (black: non-sensitive neurons; blue: left q-value neurons; red: left utility neurons; yellow: neurons concurrently sensitive for both). We tested whether this overlap is significant and report the Jaccard index (J), as well as p-values from the Jaccard test of overlap. (d) Same, for utility versus uncertainty bonuses. (e) Neural classification scheme for all action-value encoding analyses. Neurons were classified according to the position of their t-scores along a two-dimensional polar plane: as right coding (blue), left coding (yellow) or mixed coding (purple), which could be either as the sum or difference of values.



Supplementary Figure 3: Rejected and unseen value coding, selected value timing and exploration. (a) Proportion of neurons sensitive to rejected q-value in vmPFC, dACC, and preSMA, in the trial onset (blue) and pre-decision (orange) periods. P-values as follows: vmPFC, trial onset: $p = 0.004$, preSMA, both periods: $p = 0.002$. Stars indicate neuron count significance in a one-sided permutation test (** = p<0.01). (b) Same, for rejected uncertainty bonus. vmPFC, trial onset: $p = 0.002$; vmPFC, pre-decision: $p = 0.004$; preSMA, trial onset: $p = 0.006$. (c) Same, for rejected novelty. (d) Same, for the integrated utility of the option that was not presented in a trial (only for blocks with 3 available options). vmPFC, both periods: $p = 0.002$. (e) Box plots of trial onset aligned latency time across trials for all selected q-value or selected uncertainty bonus neurons in vmPFC, dACC, and preSMA. The red mark indicates the median, and the box extends between the 25th and 75th percentiles of latency times. Bar whiskers extend to the most extreme data points not labeled as outliers, defined as values that are more than 1.5 times the interquartile length away from the edges of the box. Stars indicate significance in a two-sided rank-sum test between latencies for each regressor (* = p<0.05). (f) Same, for pre-decision aligned analyses. vmPFC vs. preSMA selected unc. bonus: $p < 0.001$. (g) Chart indicating how trials were defined as explore or non-explore trials. Trials in which the selected option had lower q-value and higher uncertainty bonus were defined as explore trials (orange) and all other trials were defined as non-explore trials (blue). (h) Scatter plot of selected uncertainty t-scores versus explore flag t-scores from the Poisson GLM analysis. We display neurons sensitive to selected uncertainty bonus (blue), to the explore flag (red), and to both regressors (yellow). We also indicate the Jaccard overlap index and a p-value for the Jaccard test, indicating a significant overlap between uncertainty and exploration coding in individual preSMA neurons. Left: vmPFC neurons; Right: preSMA neurons.



Supplementary Figure 4: Post-feedback encoding. (a) Percentage of outcome neurons in vmPFC, dACC, and preSMA (** = p=0.002 for all regions, one-sided permutation test). (b) Same, for selected q-value. (preSMA: $p = 0.002$) (c) Same, for abs(RPE). (preSMA: $p = 0.002$) (d) Outcome neuron in preSMA. Top: Raster plots. For plotting, we sorted trials by outcome (magenta: win; black: no-win). Bottom: PSTH (bin size = 0.2 s, step size = 0.0625 s). Data are presented as mean values +/- SEM. (e) Same, for an outcome and selected q-value preSMA neuron. Trials were split into outcome/q-value groups: win/low (red); win/high (yellow); no-win/low (purple); no-win/high (blue). (f) Mean absolute t-score in outcome neurons in vmPFC (blue, n=37 neurons), dACC (red, n=35 neurons), and preSMA (yellow, n=47 neurons). (g) Same, for selected q-value neurons. (h) Same, for absolute RPE neurons. (i) Latency times box plot for outcome neurons in vmPFC, dACC, or preSMA. (* = p<0.05; ** = p<0.01; *** = p<0.001, two-sided rank-sum test). vmPFC vs. dACC: $p < 0.001$; dACC vs. preSMA: $p < 0.001$. The red mark indicates the median, and the box extends between the 25th and 75th percentiles of latency times. Bar whiskers extend to the most extreme data points not labeled as outliers, defined as values that are more than 1.5 times the interquartile length away from the edges of the box. (j) Same, comparing preSMA neurons which respond to outcome, selected q-value or absolute RPE. Outcome vs. selected q-value: $p = 0.009$. (k) Scatter plot of t-score for neurons which respond to outcome (yellow), selected q-value (blue) or both (purple).