Supplementary Materials for Closed-loop theta stimulation in orbitofrontal cortex prevents reward-based learning

Knudsen EB & Wallis, JD

Neuron

Supplementary Figure 1. Calculating the learning cycle.

Supplementary Figure 2. Recording locations.

Supplementary Figure 3. Example LFP traces.

Supplementary Figure 4. Additional OFC LFP analyses.

Supplementary Figure 5. Schematic of the closed-loop stimulation method.

Supplementary Figure 6. Effects of stimulation on LFP.

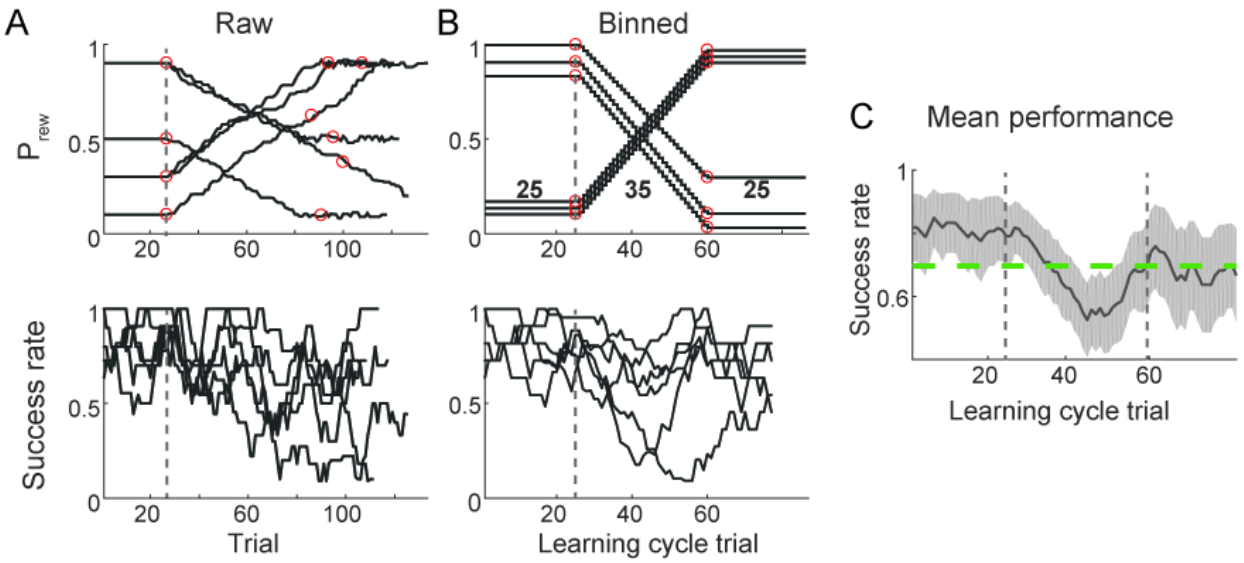
Supplementary Figure 7. Effects of open-loop stimulation on behavior.

Supplementary Figure 8. Reinforcement learning (RL) modeling.

Supplementary Figure 9. Single neuron examples of value encoding.

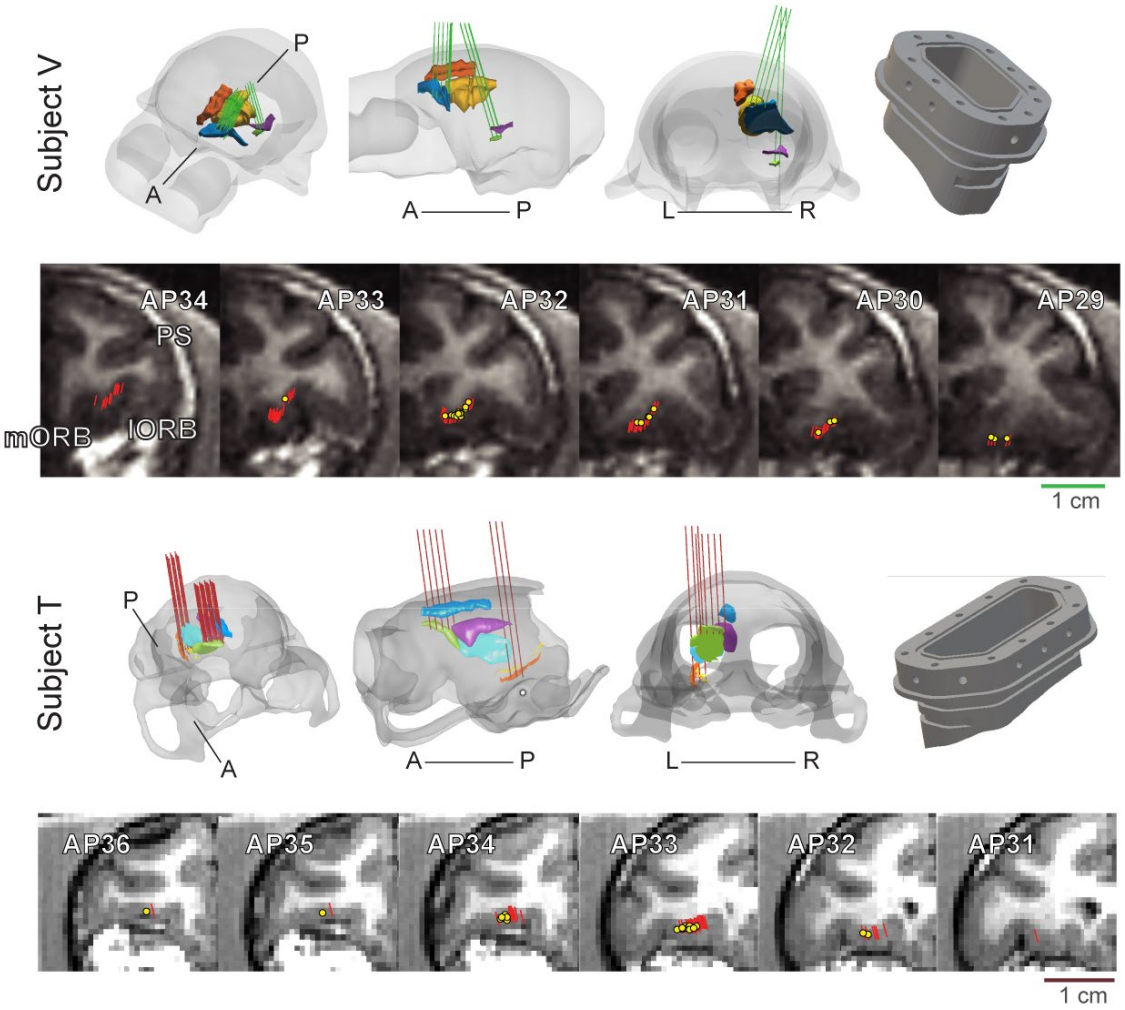
Supplementary Figure 10. Effects of stimulation on single neuron firing rates.

Supplementary Figure 11. HPC-OFC interactions, subject T.



Supplementary Figure 1. Related to Figure 1. Calculating the learning cycle.

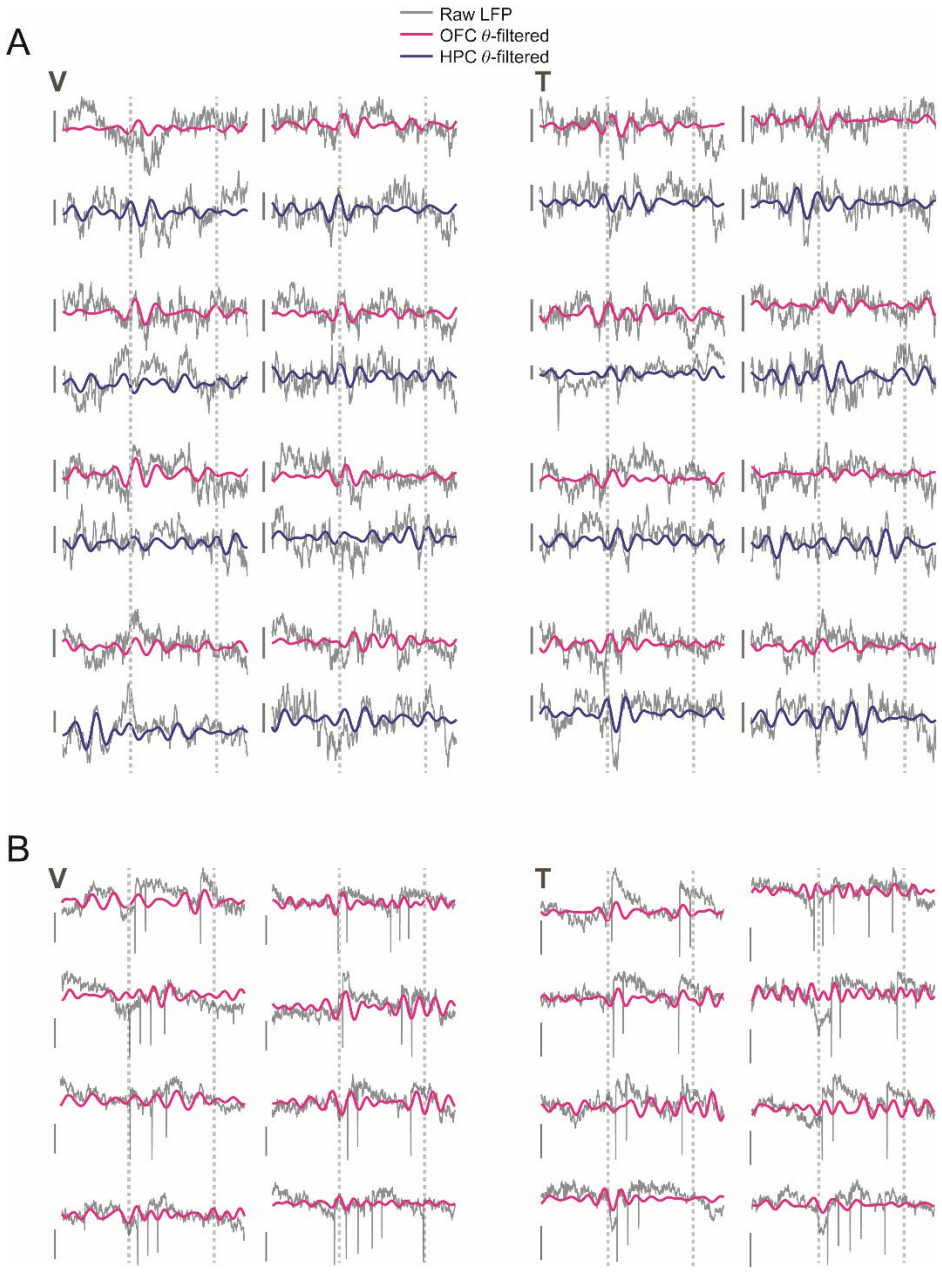
Method to calculate the standardized learning cycle for a single session from subject V. A) One picture value (top) and subject's success rates (bottom) across six successive learning cycles from this session. The duration of each learning cycle differs by around 20 trials. The beginning (dashed vertical lines) and end of each drift period are indicated with red circles. B) Drift values were binned into 35 bins bracketed by 25 pre- and post-drift trials. Each learning cycle is now the same length. C) Mean (± s.e.m.) success rate for the same session across the six standardized learning cycles. Vertical dashed gray lines indicate onset and offset of the learning cycle. Horizontal dashed green line indicates criterion performance.



Supplementary Figure 2. Related to STAR Methods, Neurophysiological recording. Recording locations.

Subjects were scanned in a 3T MRI scanner to generate 3D models of the skull and determine trajectories to reach target brain areas. For each subject, we plotted these renders in isometric (left), sagittal (middle), and coronal (right) views. Both subjects were implanted with unilateral (V: left hemisphere, T: right hemisphere) polyether ether ketone (PEEK) recording chambers for acute neurophysiology. Colored lines on 3D models denote electrode trajectories used to target OFC and HPC. Red lines on MRI scans indicate final placement of multisite recording probes.

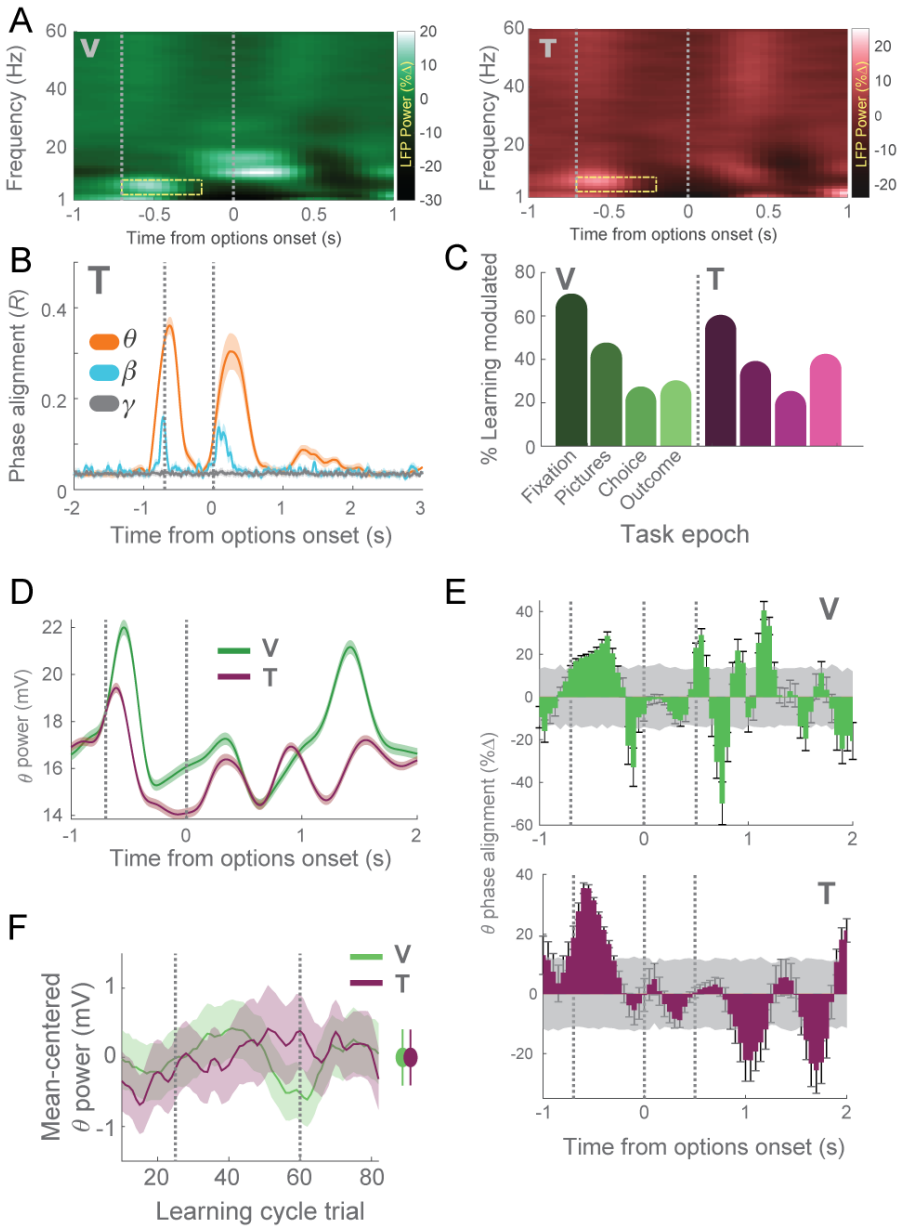
Yellow dots indicate locations of stimulation electrodes. A = anterior, P = posterior, L = left, R = right.



Supplementary Figure 3. Related to Figure 2. Example LFP traces.

A) Example raw LFP traces from OFC and HPC. Black traces show notch-filtered LFP, colored traces show theta-filtered oscillations. Vertical gray dashed lines correspond to the onsets of

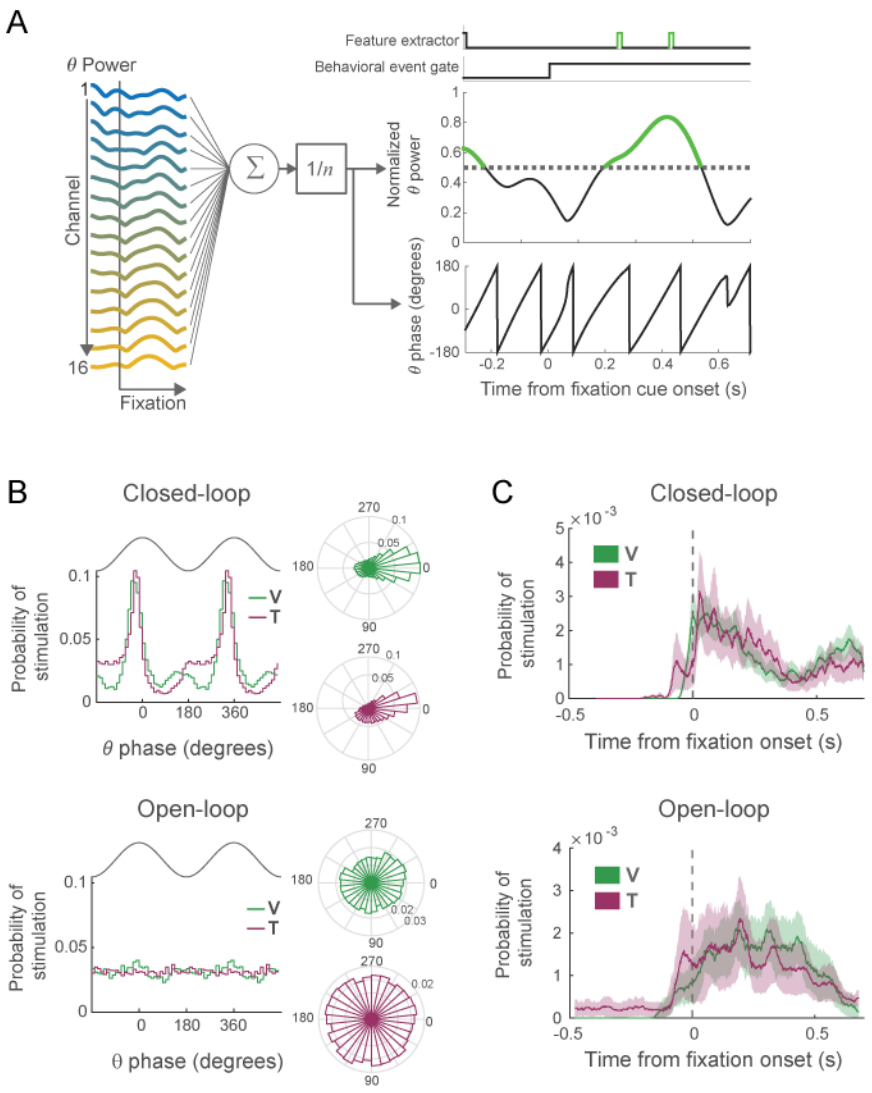
fixation and options, respectively. Scale bar denotes 100 mV. B) Example LFP traces recorded on closed-loop stimulation trials. Black traces show notch-filtered LFP pink traces show theta-filtered oscillations. The dashed vertical lines correspond to the onsets of fixation and options, respectively. Scale bar denotes 150 mV. The stimulation pulse often causes a broadband increase in power, but the stimulation artifact is filtered out from the theta band.



Supplementary Figure 4. Related to Figure 2. Additional OFC LFP analyses.

A) Mean percent change in broadband OFC LFP power during the fixation epoch relative to intertrial interval power. Yellow dashed box indicates the 4-8 Hz frequency band. B) Cross-trial phase alignment in theta (4 - 8 Hz), beta (13 - 30 Hz), and gamma (30 - 60 Hz) bands in subject T. Convention follows Figure 2D. C) Proportion of OFC channels where the cross-trial theta phase alignment was modulated by learning, defined as a significant difference in phase alignment between stable and drift trials within the first 400 ms of each epoch (paired t-test evaluated at $p <$

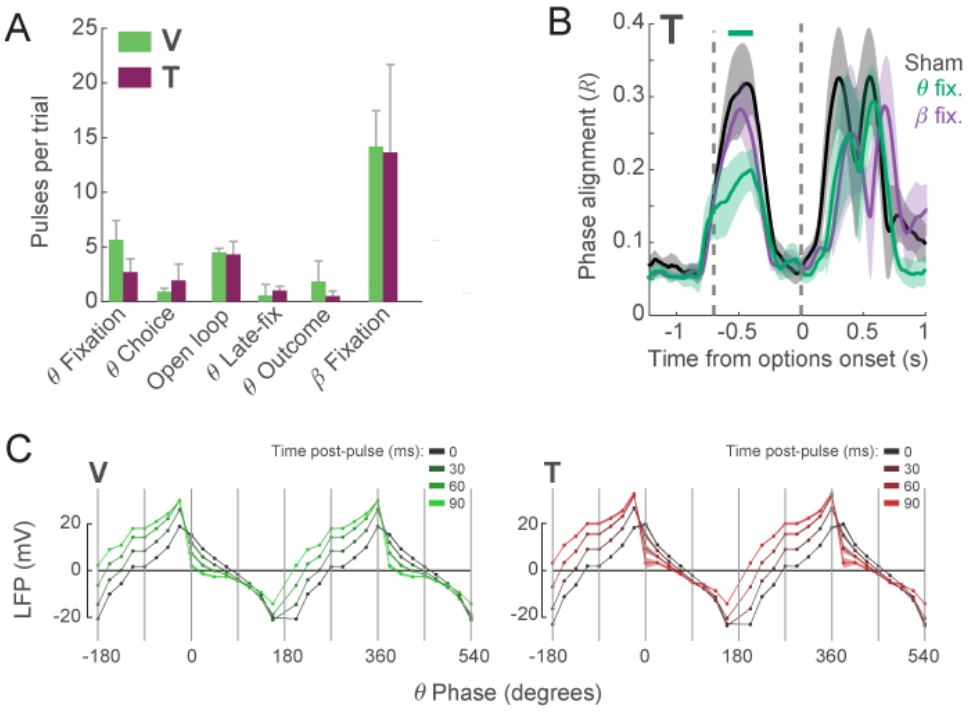
0.01). Many channels showed a significant effect of learning on theta, particularly during fixation. D) Mean (± s.e.m.) theta power over the course of the trial. Vertical dashed lines indicate the fixation and options onset, respectively. E) Percent change in cross-trial phase alignment mid-drift (averaged over the middle third of the drift period) compared to trials with stable contingencies. Gray shaded regions denote three standard deviations of the shuffled percent change over 25 bootstraps at each time point. Vertical dashed lines indicate fixation onset, options onset, and 500 ms after options onset which lies within the hold period of the median choice. The change in the amount of cross-trial phase alignment with learning was strongest and most consistent across subjects during the fixation epoch. F) Change in theta power across the learning cycle. Convention follows Figure 2E. Error bars at right correspond to ± 1 s.e.m. of bootstrapped distributions. There was no effect of learning on theta power.



Supplementary Figure 5. Related to STAR Methods, Electrical microstimulation. Schematic of the closed-loop stimulation method.

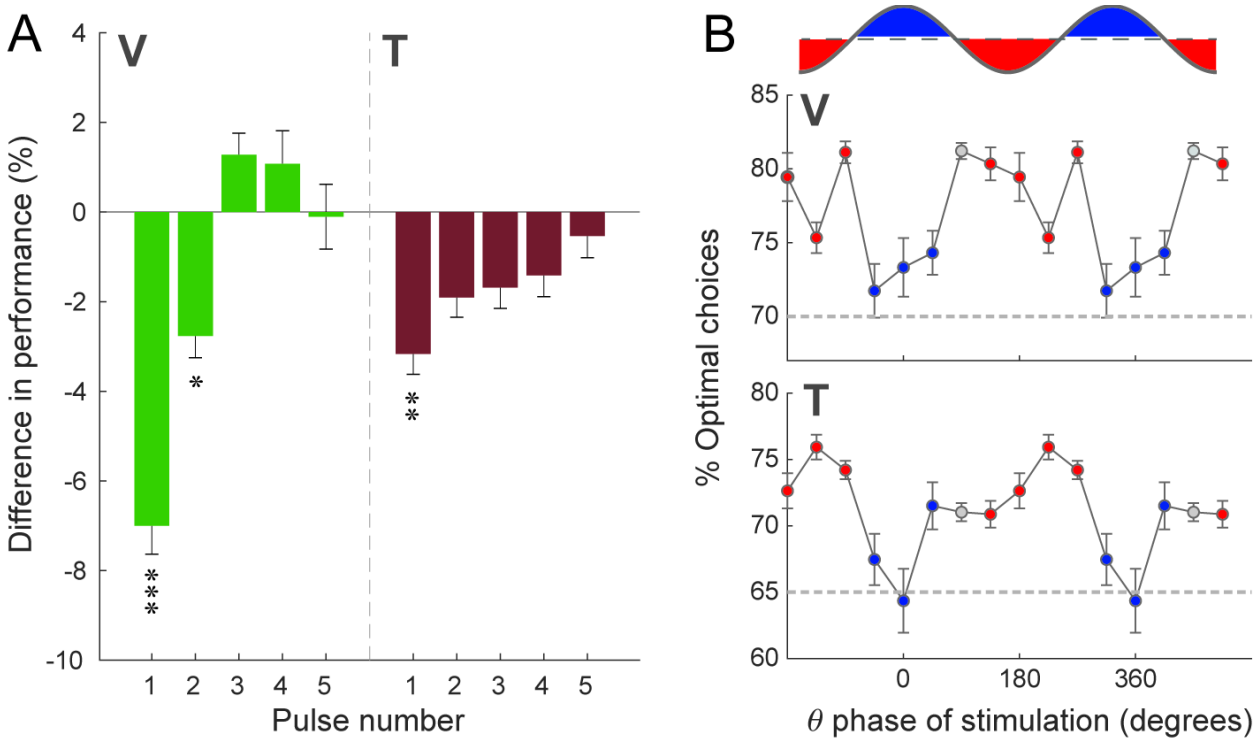
A) We extracted LFP activity from a 16-channel probe in real time and computed instantaneous power and phase using the Hilbert transform. We then thresholded mean power in the frequency band of interest at approximately the midpoint of the dynamic range (gray dashed line). Each cycle of phase that power remained above threshold generated a single pulse triggered on the peak of the Hilbert transform (right, top panel; “Feature extractor”). The Hilbert transform varies from -180° at the trough of the theta oscillation to +180° at the trough of the next wave in the oscillation. Because of the lag in our system (64 ± 3 ms, or approximately half a theta cycle) this

ensured that our stimulation was delivered close to the peak of the theta oscillation. This signal was integrated with the “Behavioral event gate” signal to ensure that stimulation pulses only occurred during the behavioral epoch of interest. B) Histograms and polar plots showing the distribution of stimulation pulses as a function of theta phase in both closed-loop and open-loop conditions. For clarity, two full cycles of theta are illustrated. Stimulation pulses in the closed-loop condition cluster around the peaks of the theta oscillation, whereas they are uniformly distributed in the open-loop condition. C) Distribution of the time of stimulation pulses in the closed-loop and open-loop (bottom) conditions. Because the behavioral event gate was triggered by the presentation of the fixation cue, a small proportion of the stimulation pulses occurred before the animal had acquired fixation (15% subject V, 12% subject T). However, there was no difference in the time to acquire fixation on trials where stimulation pulses occurred prior to acquisition of fixation compared to those where they did not (permutation test, 10000 iterations, time to acquire fixation when first pulse occurred prior to fixation onset vs. after fixation onset; V: 267 ± 25 ms vs. 278 ± 11 ms, $p > 0.1$; T: 367 ± 47 ms vs. 346 ± 18 ms, $p > 0.1$).



Supplementary Figure 6. Related to Figure 4. Effects of stimulation on LFP.

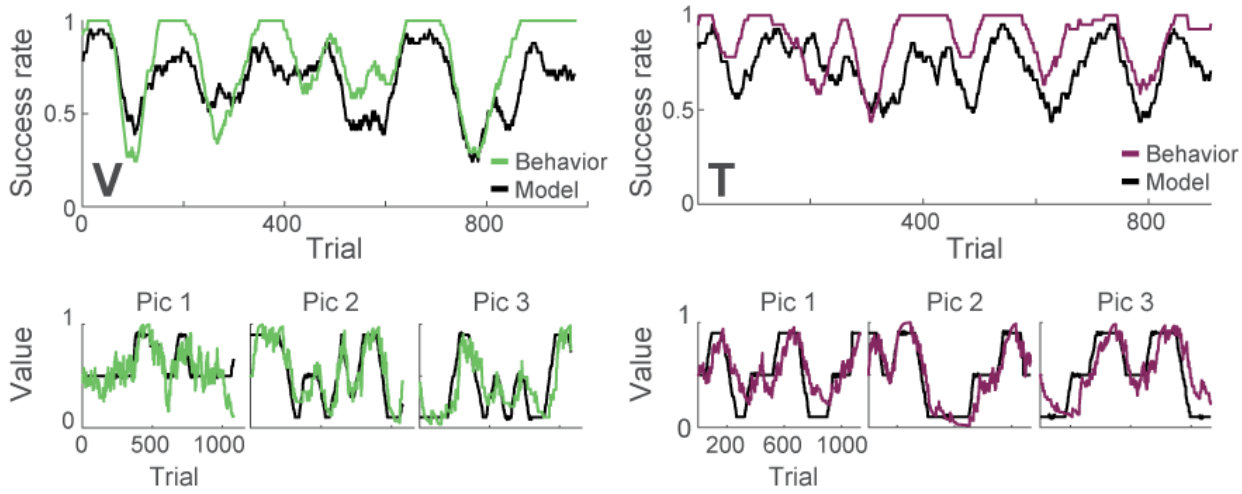
A) Mean number of stimulation pulses delivered per trial for all experimental conditions for both subjects. B) Effects of theta and beta fixation epoch stimulation on cross-trial theta phase alignment for subject T. Convention follows Figure 4A. Theta stimulation disrupted theta phase alignment, whereas beta stimulation had no effect. C) The effect of delivering pulses on the mean (± s.e.m.) LFP amplitude at different phases of the theta oscillation. The different lines illustrate different time intervals following the stimulation pulse. Note that the shaded error interval is too small to display.



Supplementary Figure 7. Related to Figure 4. Effects of open-loop stimulation on behavior.

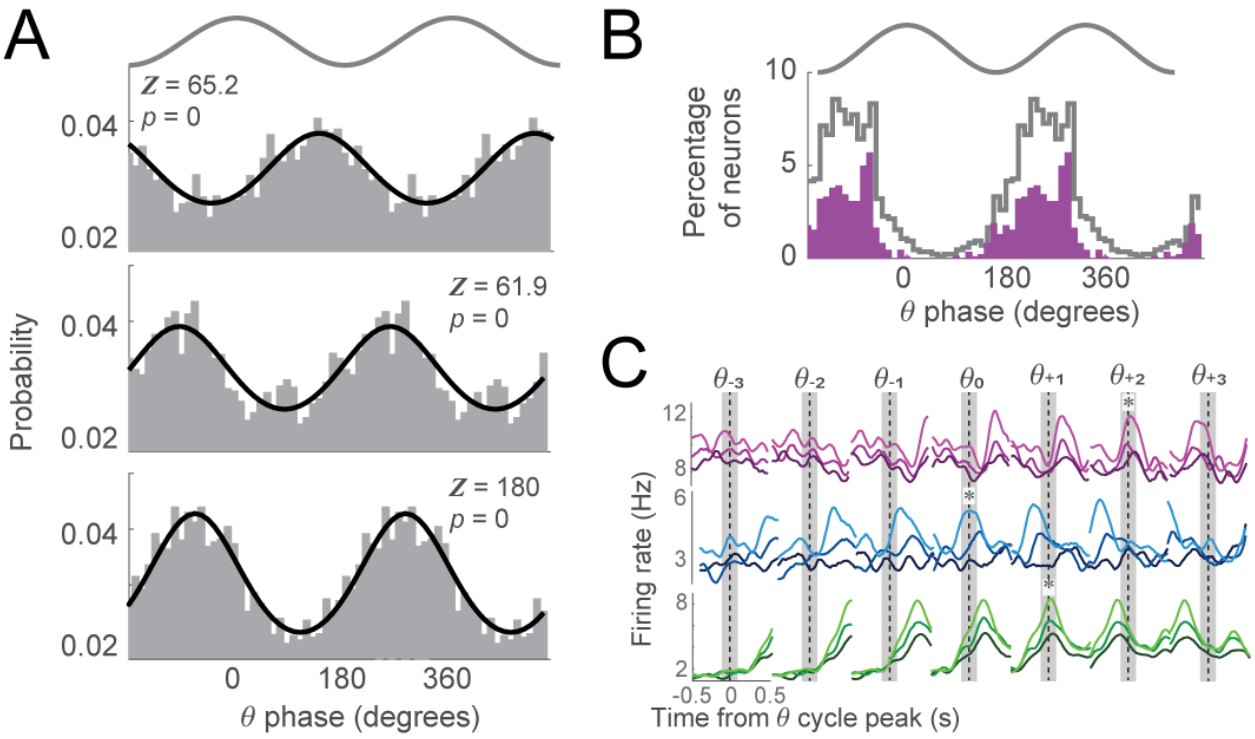
A) The difference in choice performance following stimulation pulses on the positive phase of theta relative to the negative phase. A 2-way ANOVA with factors of Valence and Pulse Number showed a significant interaction in Subject V ($F_{52408,3}$ = 11.6, p = 1.2 x 10^{-7}), which a simple effects analysis revealed was due to a particularly disruptive effect on choice behavior when either the first or second pulse was delivered during the positive phase of theta (first pulse: p < 1 x 10^{-8}, second pulse: p = 0.02, all others p > 0.05). In subject T there was a significant main effect of Valence ($F_{70414,1}$ = 17, p = 3 x 10^{-5}), but the interaction was not significant ($F_{70414,3}$ = 0.6, p = 0.6). B) Mean choice performance as a function of the phase of stimulation of the first stimulation pulse. Data centered on the zero crossing of the oscillation were excluded from the analysis (gray datapoints). In both subjects, stimulation delivered on the positive phase of theta (blue datapoints) significantly disrupted choice behavior compared to stimulation delivered on the negative phase (red datapoints). Because of the strong cross-trial phase alignment of the theta oscillation to

fixation onset, we ensured that these effects were driven by theta phase *per se* and not solely by the timing of the stimulation pulse relative to fixation onset. We compared two logistic models to predict the optimality of behavioral choice: a full model containing the pulse time and the sine and cosine components of theta phase at the time of the pulse, and a reduced model containing only the time parameter. Pulse timing alone significantly predicted choice behavior in both subjects (V: normalized β_{time} = -0.09, p = 3 x 10^{-7}, $d.f.$ = 13822; T: β_{time} = -0.12, p = 2 x 10^{-8}, $d.f.$ = 7982). However, at least one phase component in each subject (sine θ in V, cosine θ in T) significantly predicted behavioral choice in the full model (V: $\beta_{sin\theta}$ = 0.063, p = 0.001; T: $\beta_{cos\theta}$ = 0.058, p = 0.005). In both subjects, the fit of data was better explained in the full model relative to the reduced model (via Wilks' theorem; V: χ^2 = 14, p = 0.001; T: χ^2 = 8.7, p = 0.01).



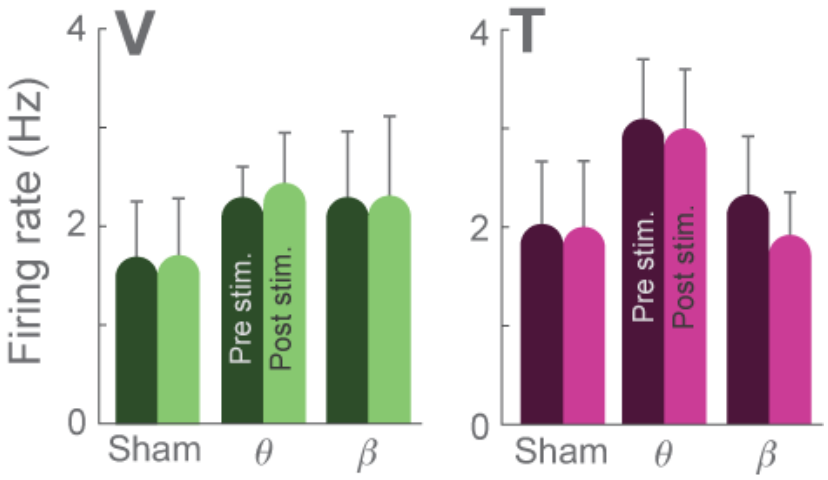
Supplementary Figure 8. Related to Figure 5. Reinforcement learning (RL) modeling.

Two example sessions showing RL fits for subjects V (left) and T (right). The top plot shows each subject's success rate across the session (black trace; 20 trial sliding average) and the best fit of the data described by the model (colored traces). The bottom plots show the objective (black) and model-derived picture values (colored). Across all sessions, the inverse temperature (β) parameter, which measures how sensitive choices are to approximate values, was 3.4 ± 0.3 for subject V and 3.4 ± 0.2 for T. The learning rate (α), which determines how much value is updated following an outcome, was 0.09 ± 0.01 for subject V and 0.08 ± 0.05 for T.



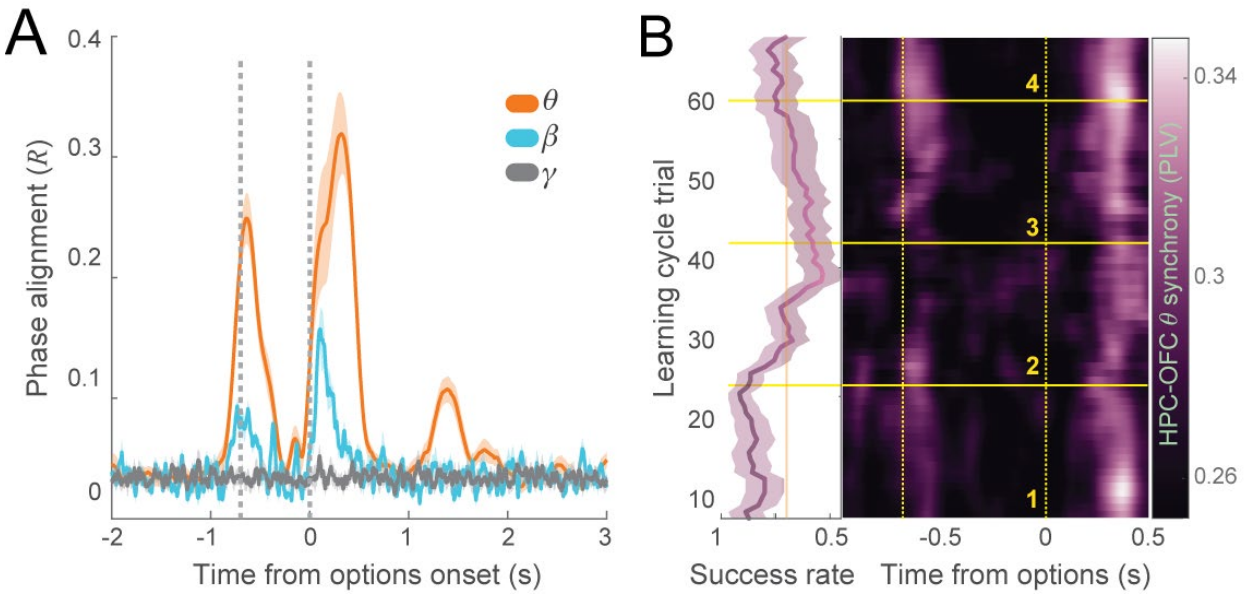
Supplementary Figure 9. Related to Figure 5. Single neuron examples of value encoding.

A) Three example OFC neurons whose firing rates significantly correlated with the phase of theta. Z indicates the results of Rayleigh's Z-test that tests whether a circular distribution is non-uniform. B) Distribution of phase locking of the OFC population. Neurons whose spikes were significantly locked to theta are shown in purple. Most neurons fired preferentially during the rising phase of the theta oscillation. C) Spike density histograms of three value neurons that are phase-locked to theta: dark to light shading denotes low to high value. Each histogram is synchronized to specific theta cycles, where θ_0 is the first theta cycle immediately following fixation onset. The theta cycle with peak value encoding is denoted by an asterisk.



Supplementary Figure 10. Related to Figure 5. Effects of stimulation on single neuron firing rates.

Mean firing rate of neurons in a 100 ms window immediately before a stimulation pulse (dark colors) and a 100 ms window after a stimulation pulse (light colors). There was no effect of stimulation on neuronal firing rates.



Supplementary Figure 11. Related to Figure 6. HPC-OFC interactions, subject T.

A) Mean (± s.e.m.) cross-trial theta phase alignment across all electrodes in HPC from subject T (N=288). Convention follows Figure 2D. B) HPC-OFC theta phase synchrony for all pairwise combinations (N=2528) in subject T. Convention follows Figure 6C. Data is the source of the bars in Figure 6D. Like subject V, the decrease in success rate disrupts HPC-OFC synchrony, but as performance stabilizes, synchrony exceeds pre-learning levels.