# Adaptive Reinforcement Learning is causally supported by Anterior Cingulate Cortex and Striatum

Robert Louis Treuting[1,2], Kianoush Banaie Boroujeni[3,4], Charles Grimes Gerrity[5], Adam Neumann[3], Paul Tiesinga[6], Thilo Womelsdorf[1,2,3,5]

[1]*Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37240*
[2]*Vanderbilt Brain Institute, Nashville, TN 372404*
[3]*Department of Psychology, Vanderbilt University, Nashville, TN 37240*
[4]*Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08540*
[5]*Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN 37240*
[6]*Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen 6525 EN, Netherlands.*

**Abbreviated Title:** Stimulation modifies adaptive reinforcement learning

**Corresponding Authors**: Thilo Womelsdorf (thilo.womelsdorf@vanderbilt.edu)

**Keywords:** Cognitive flexibility, nonhuman primate, task difficulty, motivation, attentional control, working memory, prediction error

**Conflict of Interest Statement**: The authors declare no competing financial interests.

## Short Summary

The anterior cingulate cortex (ACC) and the striatum are core nodes of a network supporting reinforcement learning, but how they augment reinforcement learning with adaptive cognitive strategies has remained unresolved. This study uses electrical microstimulation to show that ACC and the striatum have causal roles adjusting the exploration-exploitation balance and optimize reinforcement learning in multidimensional environments, while not altering attentional filtering or working memory strategies.

**Summary**

Learning the reward structure of complex environments can be achieved using reinforcement learning processes augmented with cognitive strategies. Among these strategies is the adjustment of the exploration-exploitation trade-off to increase exploration when behavior gets stuck and increase exploitation when reward contingencies remain stable. Here we tested how the anterior cingulate cortex (ACC) and the striatum causally support adaptive cognitive strategies augmenting reinforcement learning. We electrically microstimulated the ACC or the striatum in nonhuman primates at the time they chose multidimensional objects to learn about their reward values, while varying target feature uncertainty and the motivational saliency of the chosen objects. We found that stimulation of the ACC and the striatum affected adaptive strategies and reinforcement learning when target feature uncertainty was high, but in opposite ways. ACC-stimulation impaired learning and sustaining correct responses, while striatum-stimulation on average improved learning from rewarding outcomes. Behavioral modeling showed that stimulation affected the same mechanisms but in opposite ways. ACC stimulation impaired the monitoring of outcome uncertainty for adapting exploration-exploitation and reduced the ability to lower prediction errors during learning, while striatum stimulation enhanced the monitoring of outcome uncertainty and prediction-error based updating of object values. Stimulation did not alter the use of working memory or attentional filtering as alternative learning strategy. These opposing behavioral stimulation effects were associated with the ACC having populations of neurons that fired stronger during choices that were more uncertain, had a lower value, and that tracked error history, while striatum neurons more likely encoded higher values and more certain choices. In summary, microstimulation during object choices suggest that ACC and the striatum causally

adapt exploration-exploitation levels to guide exploration towards reward-relevant objects during periods of uncertainty.

## Introduction

Reinforcement learning (RL) reduces the uncertainty about the value of options. This uncertainty reduction can be inefficient in volatile environments and in multidimensional environments with many possible reward-relevant features as is typical for the real world. For these volatile and multidimensional environments, reinforcement learning can be augmented with adaptive strategies and extended to incorporate cognitive strategies [1,2]. Previous studies found that in multidimensional environments, subjects' learning progress is better predicted with adaptive strategies that allowed for more exploration and dynamic learning rates when value uncertainty increases, or that allows working memory to guide decision making when the expected values of options are ambiguous or low [3-7]. These computational insights suggest that adaptive learning strategies might be supported by brain systems tracking the uncertainty of outcomes, such as the anterior cingulate cortex (ACC) [8,9], and that integrate reward information across different learning systems, such as the striatum [10]. Here, we set out to test how the ACC and the striatum causally contribute to adaptive RL by transiently microstimulating these areas while subjects learned to reduce reward uncertainty in a multidimensional learning task.

Studies testing the causal contribution of the ACC to learning have shown that disrupting the ACC impairs the ability to choose among objects in volatile environments when object values vary probabilistically and are uncertain [11-13], but impairments are also apparent in learning and set shifting paradigms that require shifting choices between objects that have otherwise fixed reward

associations [14-17]. For all these paradigms, ACC disruption does not result in persevering with the same non-optimal choices, but rather in a difficulty sustaining attending to and choosing objects once they are rewarded. It has been suggested that one reason for this pattern of deficits after ACC disruption is an impaired integration of reward information over time to estimate the value of currently available objects [18,19] and to guide the next choice to the option with the highest predicted value [12,20]. It has remained unresolved, however, whether the ACC's role in tracking reward history is used to adaptively change RL during periods of uncertainty.

One possibility to adaptively improve RL is to use cognitive strategies such as working memory (WM) of recently rewarded objects [4,6,21]. Both the ACC and the striatum have been implicated in WM processes. Striatal lesions impair not only reward learning but also WM [22], while a role of the ACC in WM can be inferred from a disruption study that found the ACC mediates choosing counterfactual stimuli that are not available as options for a current choice but which would be the best valued alternative option in future trials based on their past values [12]. The maintenance of object values for unavailable options reflects a WM process that computational studies have shown to augment reinforcement learning processes when subjects learn values in environments with multiple multidimensional objects[17,23] or learn complex sensori-motor associations [3,6,21,24,25]. While these studies suggest that WM processes support learning, it is unknown whether the ACC or striatum causally contribute to a learning strategy that combines WM and reinforcement learning.

Another versatile way to improve RL learning in volatile and multidimensional environments is to optimize the exploration-exploitation balance given the currently experienced levels of outcome

100 uncertainty [25,26]. Both the ACC and the striatum may support this mechanism. The ACC has been
101 implicated in guiding exploration during periods of uncertainty in learning experiments [9,25,27] and
102 prolonged ACC disruptions impairs exploratory fixational sampling of objects [17], suggesting that
103 the ACC may be able to dynamically adjust the exploration-exploitation balance during learning.
104 In the striatum, neuronal firing tracks during learning the certainty of value-based choice [28] and
105 choice outcomes [29], while dopaminergic tone modulates the degree of exploration to exploitation
106 [30]. Direct microstimulation of the anterior striatum increases the value of an object that is processed
107 during the stimulation event and has been shown to increase exploitation of that object [31], while
108 object-specific stimulation had no effect in another study [32]. Whether these acute stimulation
109 effects could support adaptive reinforcement learning in multidimensional environments that
110 require object specific credit assignment processes remains unresolved.
111
112 To address these questions and determine possible casual contributions of the ACC and the
113 striatum to learning in multidimensional environments, we devised a task that increased the
114 demands on exploration vs exploitation of objects by varying the number of object features
115 subjects needed to evaluate during learning. In addition, we varied the motivational saliency of
116 choosing objects by varying the gains and losses for correct and erroneous choices. We found that
117 microstimulation affected learning particularly when feature uncertainty was high.
118 Microstimulation in the ACC disrupted learning and in the striatum it improved learning.
119 Computational modeling [23] suggests that stimulation affected not only standard reinforcement
120 learning mechanisms but also the exploration-exploitation tradeoff, inducing more exploration
121 when stimulating the ACC, and inducing more efficient exploitation of values when stimulating
122 the striatum. The pattern of results suggest that the ACC and striatum are causally important for

enabling adaptative reinforcement learning in complex, multidimensional environments.

## Results

We applied in each of two monkeys electrical microstimulation to the fundus of the anterior cingulate cortex (area 24c) and to the striatum (head of the caudate) (48 sessions per subject, 24 sessions per area). Microstimulation was delivered during at least the first twelve trials and continued until subjects learned the new feature-reward rule. A brief 0.3 s stimulation pattern was initiated when monkeys fixated an object for longer than 0.3 s, which indicated choosing that object, and receive either token rewards or losses for that choice (**Figure 1A**). In a given block, stimulation was administered either when the rewarded target object was chosen (*StimReinf+* condition), or when the non-rewarded object was chosen (*StimReinf-* condition) (**Figure 1A**). Monkeys learned which object was rewarded through trial-and-error. In each trial three objects were shown. Subjects explored them with fixational eye movements until they chose an object by sustaining fixation of an object for 0.7 s. Choices were followed by visual and auditory feedback and by the addition of tokens for correct responses or the subtraction of already attained tokens for incorrect responses. Tokens were green coin symbols in a token bar presented at the top of the screen. Accumulating eight tokens resulted in fluid reward. The rewarded target feature remained constant for a block of 30-50 trials before it switched to a new feature that was not shown before (**Figure 1B**). Blocks differed pseudo-randomly in their motivational context (+2 or +3 tokens gained for correct choices; loss of -1 or -3 tokens for errors; token outcomes were deterministically provided for choices with an outcome probability of 1), and their learning difficulty. Difficulty levels corresponded to the size of the feature space, i.e. the number of visual features that distinguished objects, increasing from features of one, two, to three feature dimensions (1D, 2D,

146 and 3D learning conditions) (**Figure 1C**). Monkeys F / I completed on average 33.5 / 36 blocks

147 during ACC stimulation sessions and 32.0 / 35.8 blocks during striatum stimulation sessions.

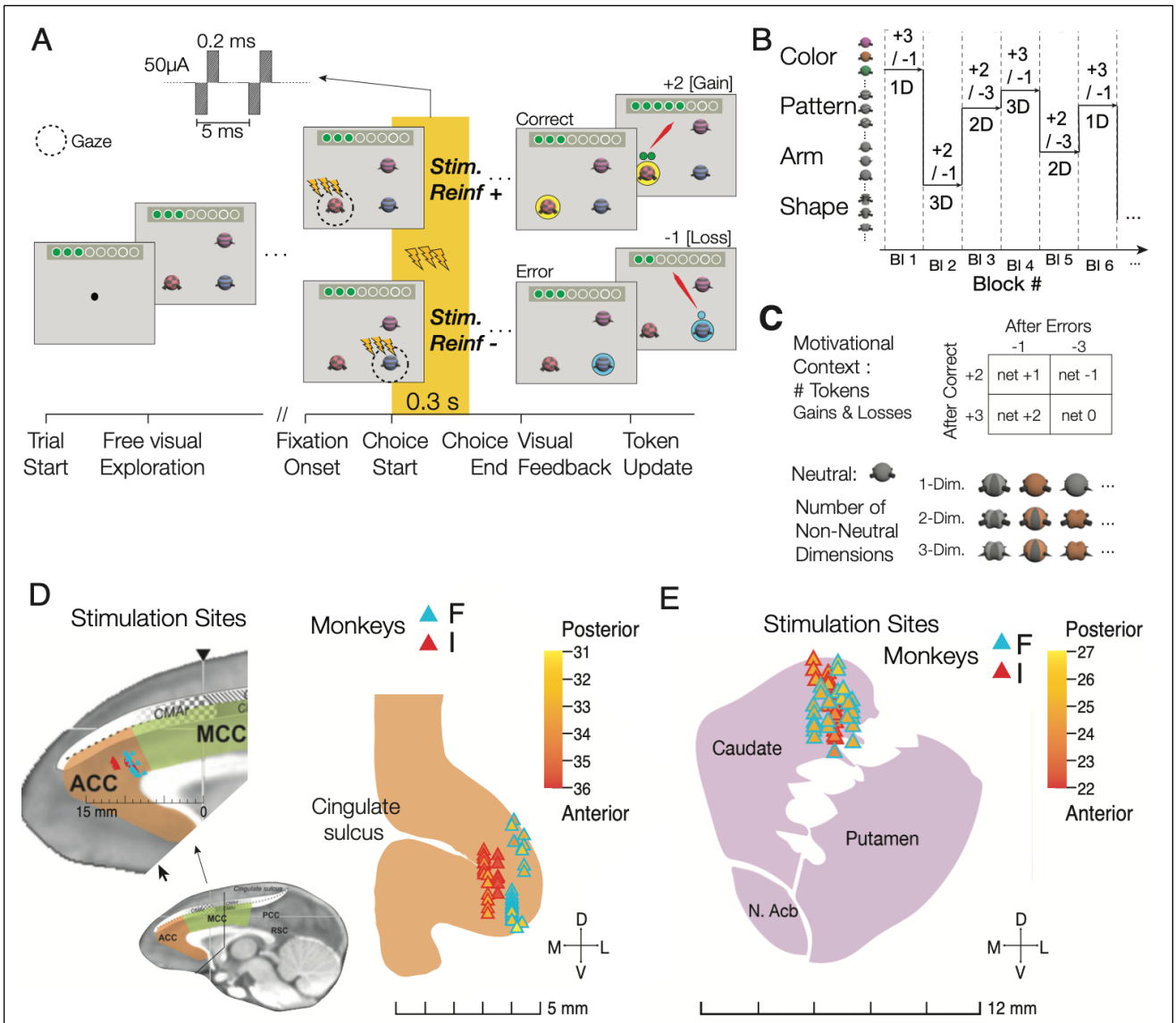


**Figure 1. Task paradigm and stimulation locations**. (**A**) On each trial monkeys freely explored three stimuli until they fixated on one stimulus for longer than 0.3 s which indicated they committed choosing that stimulus. Electrical (or sham) stimulation was delivered 0.3-0.6 s after the onset of the fixation of the chosen stimulus. Stimulation was triggered either when the positive reinforced (rewarded) object was chosen (*StimReinf+)* or when a negative reinforced stimulus (resulting in loosing tokens) was chosen (*StimReinf-).* Stimulation was only delivered during learning new feature-reward associations, i.e. prior to reaching a 70% performance criterion. 0.75 s after the choice-fixation onset a halo behind the chosen object provided visual feedback (yellow/blue for correct/error choices), followed 0.5 s later by auditory feedback and the reveal of tokens gained or lost. (**B**) The task switched the rewarded feature (y-axis) in blocks of 30-50 trials. (**C**) Blocks randomly varied the motivational context from gaining +2 or +3 for choosing the object with the rewarded feature, and the loss of 1 or 3 tokens otherwise. Blocks also varied randomly whether features from one, two, or three feature dimensions of the objects changed from trial to trial (1D, 2D, and 3D conditions). (**D**) Stimulation locations in the ACC for monkeys F / I (blue / red) in the reference frame of Procyk et al, 2016 (left) and shown on a coronal view (right). Coloring indicates the anterior-posterior distance to the interaural line. (**E**) Stimulation locations in the striatum (head of the caudate nucleus) for monkeys F / I (blue / red).

148 **Microstimulation sites within a connected cingulate-striatal circuit**

149 The anatomical locations for microstimulation in the ACC (**Figure 1D**) focused on the cognitive
150 subfield of the ACC that has prominent connectivity to the lateral prefrontal cortex, the anterior
151 striatum, and executive function networks [33-35]. Neural activity at this ACC site codes for values
152 of covertly attended stimuli [36], values of choice options [37,38], and prediction errors of visual features
153 [29,39]. The prediction error responses suggest this ACC site tracks previous trials' outcomes which
154 could inform triggering a behavioral change of choice strategies when outcomes are unfavorable
155 or unexpected [29,39-41]. The anatomical locations for microstimulation in the anterior striatum
156 (**Figure 1E**) focused on the head of the caudate nucleus, which has neurons with activity tracking
157 the expected value of visual objects during learning [28,42] and is anatomically connected to the ACC
158 [43-45], with neuronal spiking activity synchronizing with spiking activity in the ACC during
159 attention, choice, and outcome periods [35,46].

160

161 **Microstimulation affects learning in the multidimensional environment**

162 Without microstimulation (sham condition) monkeys successfully learned feature-reward rules at
163 >70% correct in the 1D/2D/3D conditions of the ACC sessions after an average of 4.0/11.6/20.2
164 trials (monkey F) and 4.0/11.3/15.02 trials (monkey I) and in the striatum sessions after an average
165 of 4.0/11.3/19.7 trials (monkey F) and 3.9/10.8/18.8 trials (monkey I) (**Figure 2**). Microstimulation
166 systematically affected learning of both subjects in both areas. The average performance curves of
167 each of the monkeys showed that in the difficult 3D task condition the learning criterion was
168 reached later when the ACC was microstimulated (**Figure 2A**), and learning criterion was reached
169 earlier when the striatum was microstimulated (**Figure 2B**), while there was no consistent
170 modulation for the easier 1D and 2D conditions. The specific microstimulation effects on learning

171 in the multidimensional (3D) condition were similarly found when using different thresholds for
172 defining the trials-to-criterion and confirmed statistically. ACC microstimulation overall impaired
173 learning dependent on dimensionality (2-way ANOVA: stim. condition x dimensionality
174 interaction, Fstat = 3.06, p = 0.0159). Impaired learning was evident in each subject in the 3D
175 condition in blocks with microstimulation in correct trials (*StimReinf*+) compared to sham, but not
176 with stimulation during erroneous choices (*StimReinf*-) (**Figure 2C**) (3D condition *StimReinf*+ vs
177 sham, monkey F: Fstat = 4.15, p = 0.0448; monkey I: Fstat = 5.94, p = 0.0166; *StimReinf*- vs sham:
178 monkey F: Fstat = 1.89, p = 0.1733; monkey I: Fstat = 0.43, p = 0.5121). Monkeys F and I needed
179 on average 3.05 and 8.19 trials longer to reach criterion performance in the *StimReinf*+ than in the
180 sham condition (trials-to-criterion in the *StimReinf*+ / *StimReinf*- / sham condition for monkey F:
181 23.26 (SE: 2.54) / 19.30 (SE: 2.27) / 20.21 (SE: 1.70); for monkey I: 23.22 (SE: 2.71) / 17.29 (SE:
182 1.95) / 15.03 (SE: 1.40)). While monkey F appeared to have slower learning in the 2D condition
183 with *StimReinf*- stimulation (**Figure 2C**), this did not reach significance (FDR-corrected pairwise
184 t-test for *StimReinf*- vs sham, monkey F: p = 0.1224).
185
186 In the striatum, microstimulation affected behavior also as a function of dimensionality, but with
187 an opposite sign relative to the ACC (**Figure 2C**). Striatal microstimulation overall improved
188 learning dependent on dimensionality (2-way ANOVA: stimulation condition x dimensionality
189 interaction, Fstat = 2.78, p = 0.0256). Learning was improved in the 3D condition in blocks with
190 stimulation in correct trials (*StimReinf*+) compared to sham, but not with stimulation during
191 erroneous choices (*StimReinf*-) in monkey F (**Figure 2C**) (3D condition *StimReinf*+ vs sham,
192 monkey F: Fstat = 3.96, p = 0.0494; *StimReinf*- vs sham: monkey F: Fstat = 3.22, p = 0.0756),
193 while monkey I showed a qualitatively similar result that did not reach statistical significance (*see*

194 below; 3D condition *StimReinf*+ vs sham, monkey I: Fstat = 1.57, p = 0.2137; *StimReinf*- vs sham:
195 monkey I: Fstat = 0.17, p = 0.6787). Monkeys F and I were on average 5.93 and 3.12 trials faster
196 to reach criterion performance in the *StimReinf*+ than in the sham condition (trials-to-criterion in
197 the *StimReinf*+ / *StimReinf*- / sham condition for monkey F: 13.92 (SE: 1.68) / 15.92 (SE: 1.80) /
198 19.85 (SE: 1.33); for monkey I: 15.73 (SE: 1.99) / 16.78 (SE: 2.29) / 18.85 (SE: 1.65).
199
200 We next tested whether microstimulation affected learning differently in blocks with high versus
201 low gains (gain of 3 tokens in G3-L1 and G3-L3 blocks versus 2 tokens in G2-L1 and G2-L3
202 blocks,) or losses (blocks with loss of 1 versus 3 tokens, i.e. G2-L1 and G3-L1 versus G2-L3 and
203 G3-L3). In the ACC, there was no 2-way interaction of microstimulation with either the gain or
204 loss factor, but there was a 3-way interaction with dimensionality and gains, signifying overall
205 slower learning in the ACC in the 3D condition with high versus low gains (**Figure 2D**,**E**) (3-way
206 ANOVA, overall: Fstat = 2.41, p = 0.0476). Post-hoc t-tests showed that this effect reflected slower
207 learning in blocks with larger gains primarily in monkey I, and less but qualitatively similar in
208 monkey F (**Figure 2E**) (trials-to-criterion with gains of 3 token in the *StimReinf*+ / sham condition
209 for monkey I: 25.5 / 14.3, p = 0.0053; for monkey F: 23.6 / 18.1, p = 0.3304). There was no overall
210 3-way interaction of stimulation condition x dimensionality x loss (**Figure 2F**,**G**) (3-way ANOVA,
211 overall: Fstat = 1.25, p = 0.2801). However, finer visual inspections showed that in the ACC
212 learning slowed with *StimReinf*+ in the 3D conditions in blocks with high loss (-3 tokens) (**Figure**
213 **2G**) (FDR-corrected t-test, p = 0.0099). This effect in blocks with higher loss was qualitatively
214 similar in both monkeys (trials-to-criterion in the *StimReinf*+ / sham conditions for monkey F: 25.8
215 / 20.3, p = 0.4003; for monkey I: 27.4 / 14.9, p = 0.0099).
216

217 In the striatum, there was no 2-way or 3-way interaction of microstimulation with either varying
218 gains or losses and dimensionality (**Figure 2D-G**). Visual inspection showed the only
219 improvement was evident in the low (-1 tokens) loss condition in 3D condition
220 with microstimulation on correct trials (*StimReinf+*), where overall (monkeys combined) learning
221 was faster, but this effect was not significant in either animal (FDR-corrected t-test, overall: p =
222 0.0496; trials-to-criterion in the *StimReinf+* / sham conditions for monkey F: 12.6 / 9.1, p = 0.0756;
223 for monkey I: 15.3 / 21.2, p = 0.3977) (**Figure 2F**).
224
225
226 The impairment with ACC microstimulation and the improvement with striatum microstimulation
227 in the 3D condition specifically affected the learning of new feature-reward associations.
228 Microstimulation (*1*) did not change the overall likelihood that the learning criterion was reached
229 within a block (**Figure S1A**), (*2*) did not affect the plateau performance once learning was
230 completed (**Figure S1B**), (*3*) had no overall effect on choice reaction times during learning or
231 when learning was completed (**Figure S1C**,**D**), (*4*) did not have a systematic influence on the
232 number of objects subjects fixated (sampled) prior to choosing an object before or after reaching
233 criterion performance (**Figure S1E**,**F**), and (*5*) did not systematically alter the duration of object
234 fixations prior to making a choice (**Figure S2A,B** and **Supplemental Results**). There were only
235 selective conditions when the task was difficult (3D conditions) and the gains and losses were low
236 in which stimulation of the striatum resulted in shorter choice reaction times or reduced number
237 of object fixations with reduced viewing durations (**Figure S1C-F, Figure, S2B**).
238
239 The stimulation effects on learning might vary over time within and across experimental sessions.

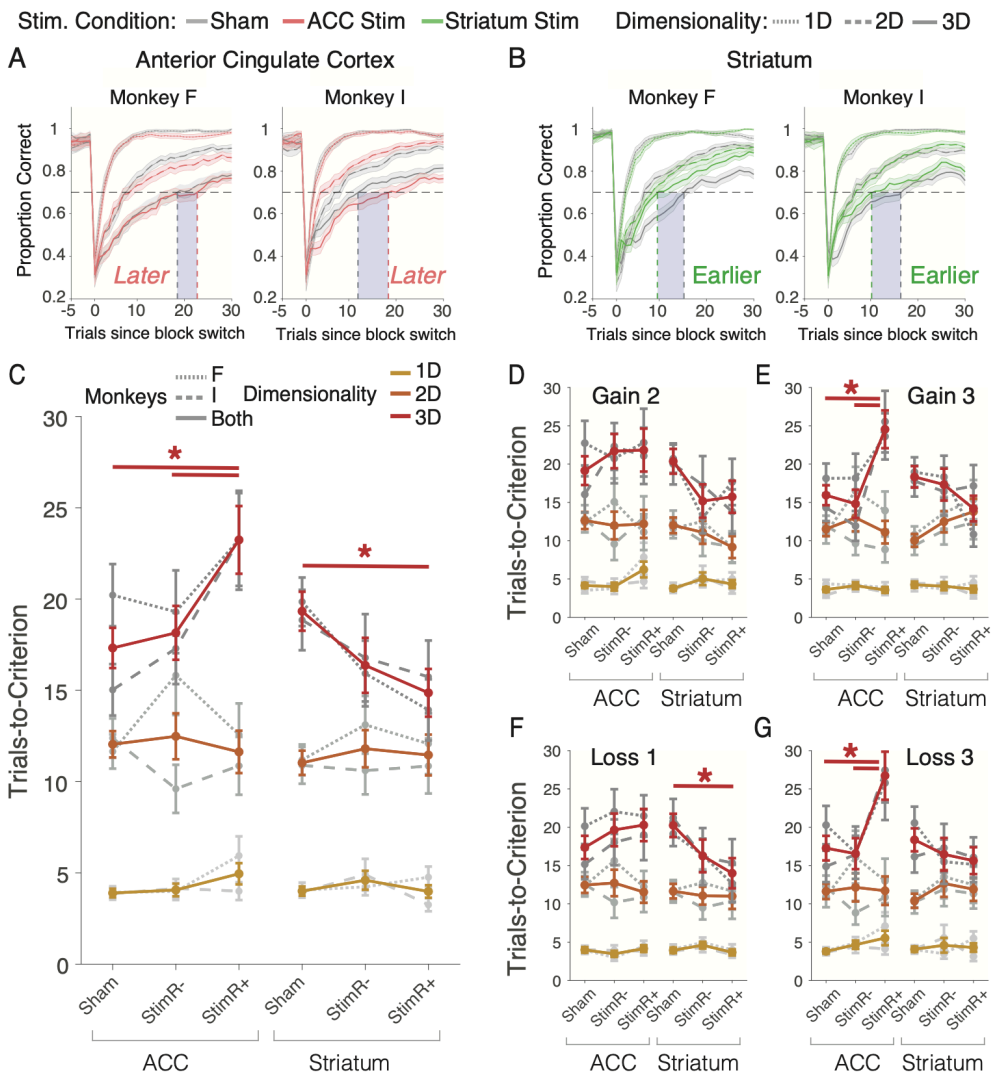


**Figure 2. Stimulation in anterior cingulate cortex impairs and in striatum improves learning in multidimensional (3D) conditions.** (**A**,**B**) Average Learning curves for blocks with stimulation (*StimReinf+* and *StimReinf-* combined) and sham in the ACC (A) and the striatum (B) for each monkey in the 1D, 2D, and 3D condition. Vertical grey lines denote the trial at which performance reached criterion performance (first trial with 70% correct over subsequent 12 trials) in the 3D condition, which was *later* with stimulation vs sham in the ACC and *earlier* with stimulation vs sham in the striatum. (**C**) Average number of trials to reach 70% performance criterion (y-axis) for the 1D/2D/3D dimensionality conditions for the sham, *StimReinf+*, *StimReinf-* stimulation conditions in the ACC (left) and the striatum (right). Colored lines reflect both monkeys, grey lines show individual monkeys. Horizontal bars mark slower learning in *StimReinf+* than sham condition for high dimensionality blocks. (**D-G**) Same format as C for 1D/2D/3D blocks where correct trials led to gains of +2 tokens (*D*), +3 tokens (*E*), and when erroneous choices led to a loss of 1 token (*F*) and 3 tokens (*G*). Error bars are SEM. Asterisks (*) indicate FDR-corrected $p < 0.05$ using t-tests with both monkeys pooled.

We found that stimulation slowed learning in ACC and improved learning in the striatum in the 3D condition across twenty-four sessions and from earlier to later blocks during sessions in monkey I, while monkey F showed more prominent slowing of learning (ACC) in earlier sessions

and in the first two thirds of experimental sessions (**Figure S2C-J**). One source of varying efficacy of the stimulation effects could be how strong the stimulation pulses modulated neural activity at a particular stimulation site. We extracted the post-stimulation multiunit activity peaks in the stimulation and sham conditions for individual sessions and correlated them with the behavioral stimulation effects on learning. However, we found no correlation between the strength of the stimulation effects on neural activity and behavioral learning in the ACC (**Figure S3A-D**), or the striatum (**Figure S3E-H**).

**Subjects similarly contribute effective stimulation sessions to the average stimulation effects**

Another source of variability of stimulation effects is the anatomical site of stimulation. Prior studies have reported heterogeneous effects of stimulation in the ACC and in the striatum on

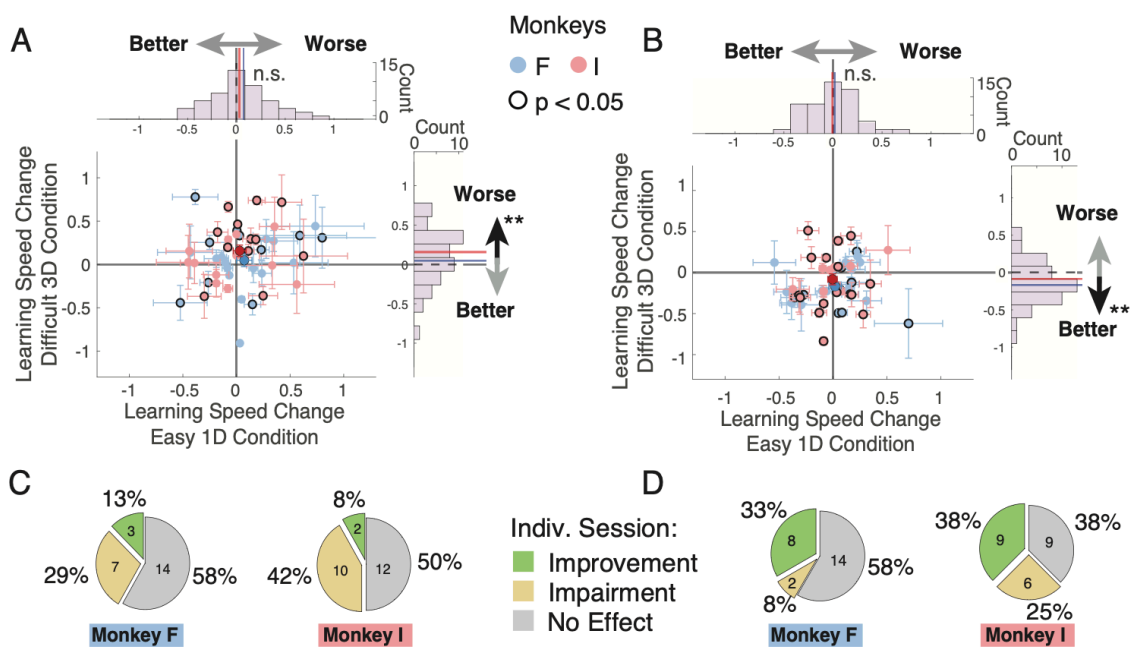


**Figure 3. The majority of stimulation sessions in the ACC impairs learning and in the striatum improves learning.** (**A**) The scatterplot shows for each experimental session in the ACC the difference in learning speed (trial-to-criterion) with stimulation vs sham in the 3D condition (*y-axis*) and the 1D condition (*x-axis*). Black outlined markers denote sessions with individually significant difference of stim vs. sham. Blue /Red data are from subject F/I. Errors bars are STD. Solid point. is mean. Histogram summarized the effects for 1D (upper) and 3D (rightmost) condition. Values <0 shows slower learning with stimulation than sham. (**B**) Pie charts summarizing the proportion of sessions in which stimulation improved (green) or impaired (yellow) learning for each subject. (**C**,**D**) Same format as (*A*,*C*) for experiments in the striatum.

254 approach/avoidance behavior with anatomically nearby sites having sometimes opposite
255 stimulation effects on how stimuli are approached [47,48]. We evaluated the anatomical heterogeneity
256 of stimulation effects by quantifying the effect of microstimulation on learning in the 3D relative
257 to the 1D condition for individual experimental sessions. This analysis revealed some
258 heterogeneity of stimulation effects but also confirmed an overall prevalence of sessions showing
259 impaired learning with ACC microstimulation (**Figure 3A**) and improved learning with striatum
260 microstimulation (**Figure 3B**). We calculated the number of effective sessions as those sessions
261 for which the behavioral stimulation effect was statistically significant and found that each subject
262 contributed a similar proportion of effective sessions to the main effects. In the ACC,
263 microstimulation impaired learning in 7 (29%) and 10 (42%) sessions in monkey F and I,
264 respectively, while learning was improved in fewer sessions (3 and 2 in monkey F and I,
265 respectively; improved vs impaired sessions in ACC for both monkeys combined: $p = 0.0036$; for
266 monkeys F / I: $p = 0.1551$ / $p = 0.0077$, Chi-square tests) (**Figure 3C**). In the striatum, effective
267 sessions were also consistent with an overall learning improvement. Most effective sessions in the
268 striatum showed improved learning (8 and 9 in monkey F and I, 33% and 38%, respectively) while
269 fewer sessions (n=2, monkey F) or a similar number of sessions (n=9, monkey I) showed impaired
270 learning (**Figure 3D**) (Chi-square test, improved vs impaired sessions in monkeys combined: $p =$
271 $0.0363$, for monkey F: $p = 0.0330$, for monkey I: $p = 0.3502$). We confirmed that the impairment
272 in the effective stimulation sessions in the ACC was selectively evident in the 3D condition
273 (**Figure S4A-E**). Similarly, sessions with improved learning in the striatum showed the
274 improvement specifically in the 3D task condition (**Figure S4F-J**). In separate analyses we tested
275 whether sessions with a significant improvement or impairment of learning were clustered
276 anatomically in the ACC or the striatum, but did not find anatomical clustering of these functional

effects.

**Stimulation affects positive credit assignment**

The results so far show that robust stimulation effects on learning particularly with microstimulation in rewarded trials (*StimReinf*+ condition) (**Figure 2C-G**, **S4A-J**), which suggest that stimulation modulates the positive credit assignment process that associates visual feature with the positive reward expectations, and less with the assignment of negative credit after error trials, i.e. recognizing which features were incorrectly chosen and should be avoided. We tested this suggestion by focusing on the 3D condition on trials in a block prior to reaching the learning criterion and quantifying how subjects used outcomes for the trial-by-trial adjustment of performance following correct trials (i.e. after gaining tokens) compared to error trials (i.e. after losing tokens). The analysis was done separately for correct and error trials in the *StimReinf*+ and the *StimReinf*- condition. In the ACC, microstimulation in the *StimReinf*+ condition overall impaired performance around trials with gains, while there was no consistent effect after losses (main effect of stimulation (β2) for gains: t-stat = -2.7409, p = 0.0061; for losses: t-stat = -1.594, p = 0.1109) (**Figure 4A**,**B**). Similarly, for the *StimReinf*- condition an impairment was visible as a trend around gains, but there was no effect around losses (**Figure 4C,D**) (main effect of stimulation (β2) for gains: t-stat = -1.91, p = 0.0561, t-stat = -0.8764; for losses: p = 0.3807). These trial-by-trial effects around gains and losses of *StimReinf*+ and *StimReinf*- in the ACC were evident in each subject (**Figure S5**). Restricting the analysis to the 17 effective sessions confirmed that stimulation during rewarded (*StimReinf+)* and non-rewarded (*StimReinf*-) trials impaired performance prominently around gains, but not reliably after losses (**Figure S4K-N**).

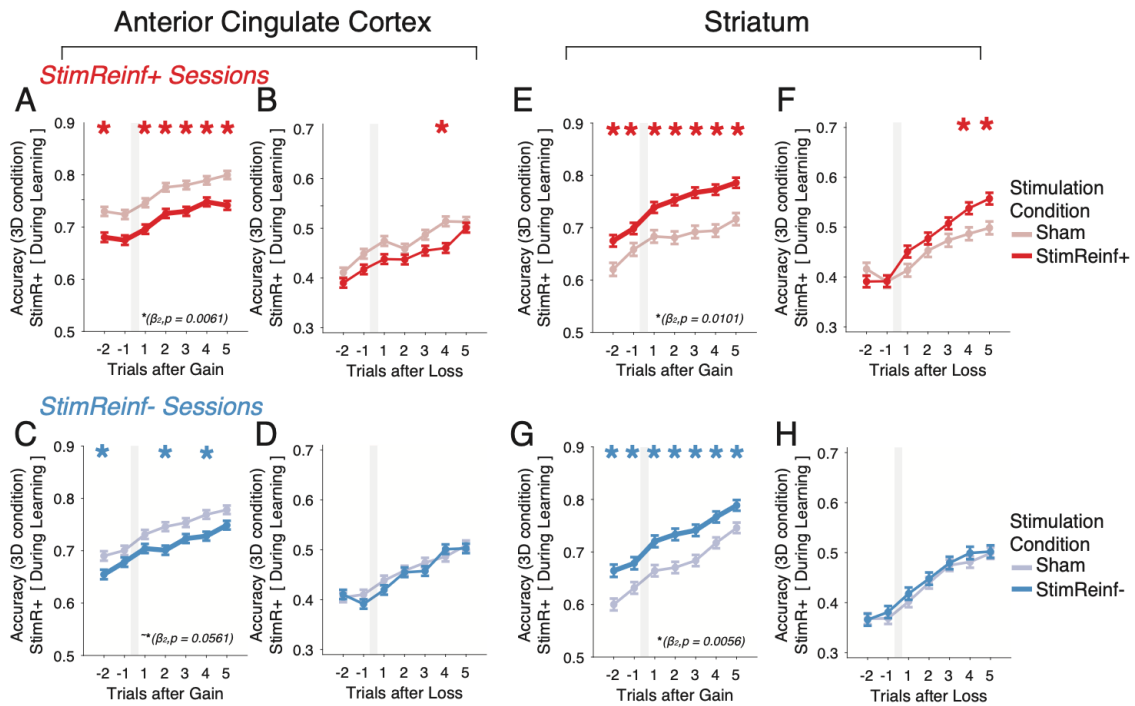


**Figure 4. Stimulation modulates behavior following gains in the 3D condition.** (**A**) Average accuracy (*y-axis*) for trials before and after a correct trial that gained subjects tokens for blocks with stimulation in the ACC on correct trials (*StimReinf*+) and sham in the 3D condition for trials until the mean trials criterion in the ACC / *StimReinf*+ condition. *'s denote FDR-corrected p<0.05 (permutation test) significant differences of stim vs sham. The p-value at bottom refers to a significant coefficient using the Wald test after stimulation and sham conditions were fitted to a logistic regression model, produced from a generalized linear model with a binomial distribution and logit link function (see Methods). (**B**) Same as (*A*) for accuracy before and after incorrect trials that lost tokens. (**C**,**D**). Same format as (*A*,*B*) for experimental sessions with *StimReinf*- stimulation and sham. (**E-H**). Same format as (*A-D*) for experimental sessions in the striatum with *StimReinf*+ stimulation (*E*,*F*), and with *StimReinf*- stimulation (*G*,*H*).

A similar result pattern was evident in the striatum but with the opposite sign relative to the ACC. Microstimulation in the *StimReinf*+ and in the *StimReinf*- condition improved trial-by-trial adjustment following gains (**Figure 4E**,**G**), but around losses, there was only an effect when restricting the analysis to the 17 sessions with a stimulation induced learning improvement (for *StimReinf*+: main effect of stimulation (β2) for gains: t-stat = 2.5716, p = 0.0101; for losses: t-stat = 0.7338, p = 0.4631; for *StimReinf*-: main effect of stimulation (β2) for gains: t-stat = 2.7683, p = 0.0056; for losses: t-stat = 0.5640, p = 0.5727) (**Figure 4F,H; Figure S4O-R**). The result pattern was similar in both subjects (**Figure S5**). Taken together, these findings suggest that microstimulation in the ACC interferes, and in the striatum facilitates most prominently with

309 positive credit assignment, and that similar effects for negative credit assignment become evident
310 only in the striatum when considering sessions with the strongest effect sizes.
311
312 The trial-by-trial effects of microstimulation on performance during learning of feature-reward
313 associations raises the question how fast the brief 0.3 s electrical stimulation pulses affected
314 behavior. Statistically reliable behavioral effects of microstimulation in the 3D condition emerged
315 after the sixth stimulation trial in the ACC in the *StimReinf+* condition (**Figure S6A**,**B**), while in
316 the striatum performance improvements became evident already after the second (*StimReinf+*) and
317 first (*StimReinf-*) stimulation trial (**Figure S6C**,**D**).
318
319 **Microstimulation in ACC disrupts sustaining correct responses**
320 Lesions studies suggest that an intact ACC is important for sustaining correct responses [11,14], and
321 that the striatum supports preventing persevering non-rewarded choices [49]. To test whether
322 microstimulation affected these behavioral patterns we quantified how stimulation changed the
323 likelihood of correct choices in trials after committing one error (after an 'E'), after a single correct
324 trial ('EC'), after two correct trials ('ECC'), etc. This $EC_n$ analysis showed that in the ACC
325 microstimulation reduced performance after sequences of three correct choices ('ECCC') in the
326 *StimReinf+* condition, consistent with a disruption of sustaining rewarded choices (FDR-correct
327 p-values < 0.05, permutation test; **Figure S6E**). There was no $EC_n$ effects with microstimulation
328 in the striatum (**Figure S6G**,**H**). There were no differences of stimulation compared to sham
329 conditions with regard to sequences of erroneous choices ($CE_n$ analysis).
330
331 **Microstimulation modulates reinforcement learning and adaptive exploration**

332 Microstimulation impaired versus improved learning after stimulation of the ACC versus the
333 striatum. At the cognitive-behavioral level this dissociation could either indicate that the ACC and
334 the striatum implement different learning mechanisms, or that they modulate the same learning
335 mechanisms but in an opposite way. To evaluate these alternatives we fit the trial-by-trial
336 performance of the subjects with models that account for choices of subjects with either one, two,
337 or three computationally distinct learning mechanisms that previous studies have shown to be
338 important for learning in environments with multidimensional objects [2,4,23,50]. For this analysis we
339 focused on the 3D condition and the effective sessions where stimulation impaired learning in the
340 ACC (**Fig. 3B**) and improved learning in the striatum (**Fig. 3D**).
341
342 We evaluated as first mechanism a Rescorla Wagner RL model augmented with a selective
343 forgetting (SF) parameter. This '*SF model*' used a parameter ($\omega^{SelectiveDecay}$) to decay the values
344 of features from non-chosen objects akin to an attentional attenuation of non-chosen stimuli, which
345 is a mechanism that speeds up learning among multidimensional options [23,29,50] (**Figure 5A**). As a
346 second mechanism we evaluated a RL model with adaptive exploration (AE). This '*AE model*'
347 tracked prediction errors over trials in a separate error-trace variable, which it weighted with an
348 error-trace weight ($\alpha^{\mathrm{ErrorTrace}}$) that increased exploration over exploitation when erroneous choices
349 accumulated, and reduced exploration when errors were rare [23,25,51]. The mechanism reflects meta-
350 learning insight about the level of experienced outcome uncertainty (in the form of prediction
351 errors), increasing exploratory choices when uncertainty increases. As third mechanism we
352 evaluated working memory (WM), which supports learning by uploading an object when its choice
353 led to reward, decaying its representation with parameter $\omega^{WM}$, and biases choosing objects that
354 have the strongest WM representation (see methods). WM can support fast learning as it rapidly

updates objects after an outcome without needing to compute prediction errors for slowly updating value expectations [3,4,6]. We evaluated each of the three model mechanisms and their combinations

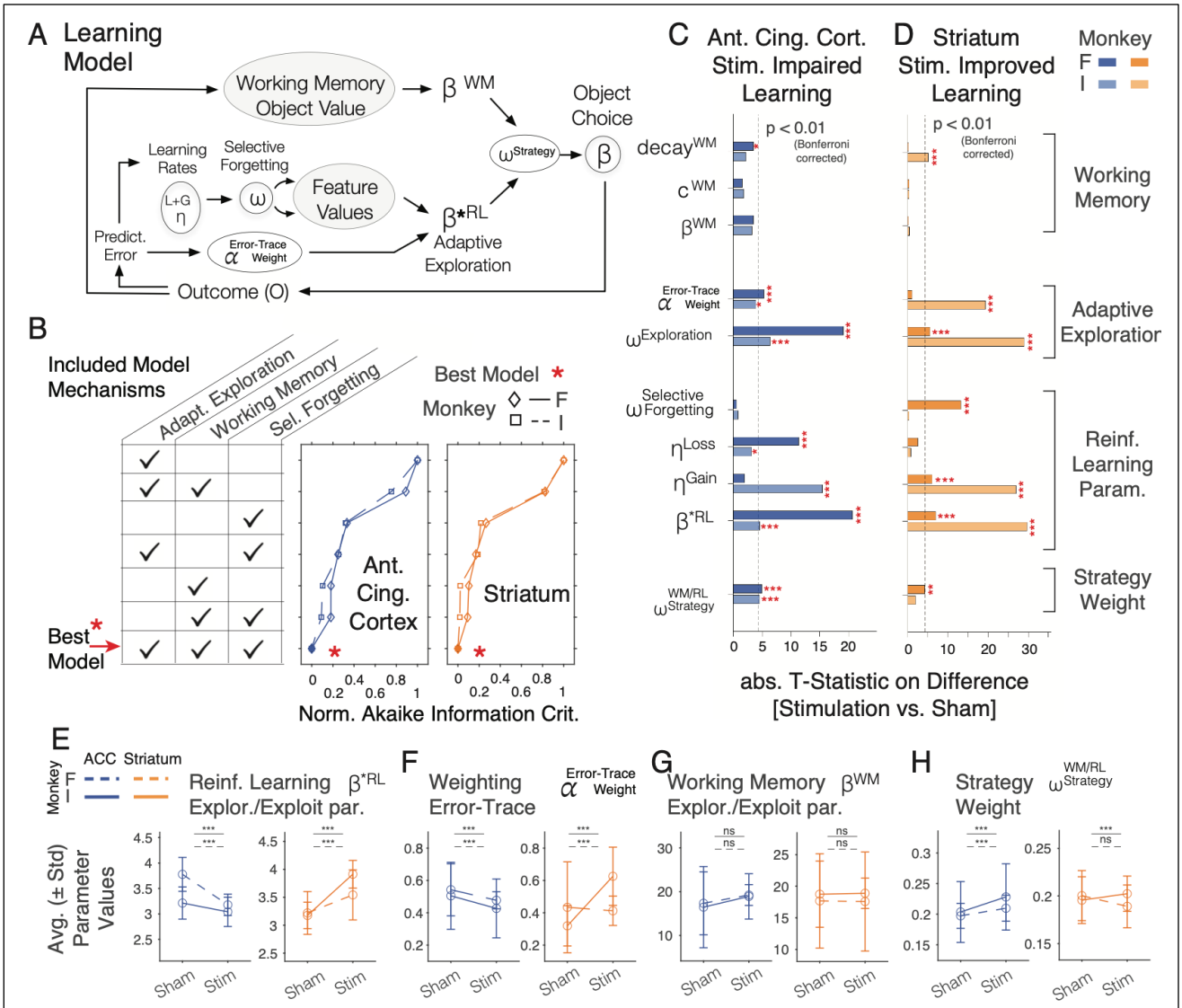


**Figure 5. A triple-strategy model accounts for learning at high (3D) feature uncertainty.** (**A**) Overview of the triple-strategy model combining adaptive exploration / selective forgetting / working memory (AE-SF-WM). The model learns to make choices of objects by a combination of (i) working memory (WM) and by using expected feature values of an RL module, which uses (ii) an adaptive mechanism to adjust exploration during learning and (iii) a selective forgetting of values from non-chosen features. (**B**) The AE-SF-WM model best fit choices in the effective high (3D) feature-uncertainty condition across sessions and stimulation / sham conditions in the ACC (blue) and striatum (orange) for both monkeys (solid and dashed line). The lowest normalized Akaike Information Criterion corresponds to the best model. (**C,D**) The effect size (absolute value of t-statistics) and Bonferroni corrected p-values for the difference of parameter values of the model fit separately to the sham and stimulation conditions in the 3D condition in ACC (*C*) and in the striatum (*D*). The parameters are defined in Methods. (**E-H**) Change in fitting parameters between stimulation and sham condition for ACC (blue) and striatum (orange) for both monkeys (F dot-dashed; I solid line) for (E) the adaptive exploration-exploitation value $\beta^{*RL}$, (F) the error trace weight $\alpha^{ErrorTrace}$, (G) the working memory softmax constant $\beta^{WM}$ and (H) the strategy weight $\omega^{\frac{WM}{RL}Strategy}$. Stars in panel (C-H) denote standard significance levels.

357 (**Figure 5B**). For models combining the RL models (i.e. the SF or AE model) with the WM module
358 we used a weighting variable ($w^{\frac{WM}{RL}Strategy}$), that determined for each trial how strong the RL
359 component contributed to choices relative to the objects represented in WM (with values lower
360 than 0.5 reflecting a larger influence of RL) (*see* Methods).
361
362 We found that all three model mechanisms contributed to accounting for choices made by the
363 subjects during task performance. This result is based on evaluating all combinations of model
364 mechanisms (**Figure 5B**), which showed that the behavior across the stimulation and sham blocks
365 of each monkey and in each area (ACC and striatum) is best accounted for by the full *AE-SF-WM*
366 *model* combining all three learning mechanisms as evaluated with the Akaike Information
367 Criterion (AIC) that corrects model performance for the number of free parameters used (**Figure**
368 **5B**; AIC's for models rank-ordered from worst to best: (7) only adaptive exploration AE, (6)
369 AE+WM, (5) only selective forgetting SF, (4) AE+SF, (3) only WM, (2) WM+SF, and for the
370 best ranked model (1) SF+AE+WM for monkey 1 in ACC: 3138.9 / 3119.9 / 3024.5 / 3010.2 /
371 2998.7 / 2998.6 / 2967.5; for monkey 2 in ACC: 4819.6 / 4738.5 / 4596.3 / 4574.9 / 4525.6 / 4521.3
372 / 4491.6; for monkey 1 in striatum: 3492.5 / 3478.8 / 3496.5 / 3580.9 / 3489.9 / 3494.5 / 3744.0;
373 for monkey 2 in striatum: 3331.7 / 3329.1 / 3332.8 / 3426.6 / 3329.5 / 3333.7 / 3573.7).
374
375 We next fitted the full *AE-SF-WM model* separately to the performance of stimulation and sham
376 blocks to evaluate which model parameter values were affected by microstimulation.
377 Microstimulation in ACC and the striatum affected similar model parameters, impairing (in ACC)
378 and improving (in striatum) learning through changes of adaptive exploration parameters ($\alpha$
379 $^{\text{ErrorTrace}}$, $\omega^{\text{Exploration}}$), and RL parameters (learning rates for prediction errors), but less consistent

380 or no changes in WM, and no changes of selective forgetting (**Figure 5C**,**D**). In particular, for the
381 ACC, microstimulation changed parameters in each subject that implement adaptive exploration,
382 altered RL learning rates, and increased WM/RL strategy weighting (**Figure 5C**, t-tests,
383 Bonferroni corrected significance for monkeys F / I in ACC: adaptive exploration weight
384 $\omega^{Exploration}$: 6.4 / 19.1, p<0.001 / p<0.001; RL learning rate for correct outcomes: t-stat: -15.5 / -
385 1.9, p<0.001 / ns; RL learning rate for incorrect outcomes: t-stat: 3.2 / 11.3, p=0.0199 / p<0.001;
386 AE error-trace weight $\alpha^{ErrorTrace}$: t-stat: -3.9 / -5.3, p=0.002 / p<0.001; exploration-exploitation
387 level $\beta^{*RL}$: t-stat: 4.6 / 20.7, p<0.001 / p<0.001; $w^{\frac{WM}{RL}Strategy}$: t-stat: -4.5 / -5.0, p=0.001 /
388 p<0.001). The selective forgetting parameter value was unaffected by stimulation (**Figure 5C**;
389 $\omega^{SelectiveForgetting}$ for monkey F / I: t-stat: 0.9 / -0.5, ns / ns). The WM parameters for temporal
390 decay, capacity and the exploration-exploitation level were not modulated or only moderately
391 affected by stimulation in one of the subjects (**Figure 5C**; for monkey F / I : WM decay: t-stat: 2.2
392 / -3.5, ns / p=0.007; exploration-exploitation level $\beta^{WM}$ t-stat: -3.2 / -3.5; p=0.002 / p=0.008; WM
393 capacity: t-stat: 1.8 / 1.6; ns / ns).
394
395 A similar set of model mechanisms was modulated by stimulation in the striatum. For each subject,
396 striatal stimulation affected the adaptive exploration parameter values and RL learning rates for
397 correct outcomes (**Figure 5D**; for monkeys F / I: adaptive exploration weight $\omega^{Exploration}$: t-stat:
398 19.3 / -1.01, p<0.001 / ns; RL learning correct outcomes: t-stat: 27.0 / 6.0, p<0.001 / p<0.001; RL
399 learning rate after error outcomes: t-stat: 0.76 / -2.53, ns / ns; exploration-exploitation level $\beta^{*RL}$:
400 t-stat: -29.7 / -6.95, p<0.001 / p<0.001). In contrast, WM parameters, selective forgetting and the
401 strategy weight were either unaffected or inconsistently modulated across subjects (for monkeys
402 F / I: $\beta^{WM}$ t-stat: -0.35 / 0.13, ns / ns; WM capacity: t-stat: 0.3 / -0.27, ns / ns; WM decay: t-stat:

403 5.05 / 0.14, p<0.001 / ns; $\omega^{SelectiveForgetting}$: t-stat: -0.33 / -13.18, ns / p<0.001; $w^{\frac{WM}{RL}Strategy}$: t-
404 stat: -1.92 / 4.2, ns / p<0.001) (**Figure 5D**).
405
406 Both, the ACC and the striatum modulated adaptive exploration and the updating of feature values
407 by scaling prediction errors with learning rates, but they did so in opposite ways. Stimulation in
408 ACC increased exploration (lower $\beta^{*RL}$) and reduced weighting of recent errors to adapt
409 exploration ($\alpha^{ErrorTrace}$) (**Figure 5E**,**F**), while in the striatum it enhanced exploitation (higher
410 $\beta^{*RL}$), i.e. the use of expected values to make a choice, which was linked with a stronger weighting
411 of recent prediction errors ($\alpha^{ErrorTrace}$) in subject F (**Figure 5E**,**F**), and to stronger weighting of
412 exploitation over exploration in subject I ($\omega^{Exploration}$) (**Figure 5D**). These changes occurred
413 while the model parameter that determined how strong the WM representations informed choices
414 ($\beta^{WM}$) remained unchanged in both subjects and brain areas (**Figure 5G**). While there were
415 moderate increases of the strategy weight, $w^{\frac{WM}{RL}Strategy}$ with microstimulation, the weight values
416 of ~0.2 reflected an overall low contribution of WM over RL representations to inform choices in
417 both, stimulation and sham conditions (values of $w^{\frac{WM}{RL}Strategy}$ ~ 0.5 would reflect equal weighting
418 of WM and RL, whereas values<0.5 reflect less weight of the WM module), showing that
419 stimulation did not change the dominant RL learning strategy (**Figure 5H**).
420
421 Taken together, microstimulation affected AE and RL model mechanisms in the ACC and the
422 striatum (visualized in **Figure 6A**,**B**). To discern the consequences of changed model mechanisms
423 on behavior we calculated the trial-by-trial changes of behavior (**Figure 6C**,**D**), and how the model
424 variables of the *AE-SF-WM model* related to them. In experimental blocks with microstimulation
425 (compared to sham) choice probabilities were reduced in ACC, but enhanced in the striatum

426 (**Figure 6E,F**), prediction errors remained elevated in ACC, but were more effectively reduced in

427 the striatum (**Figure 6G,H**), and subjects showed consistently more random exploration across

428 trials with stimulation in ACC, but more effective exploitation with stimulation in the striatum

429 (**Figure 6I,J**). A different pattern was found for the choice strategy, which temporarily increased

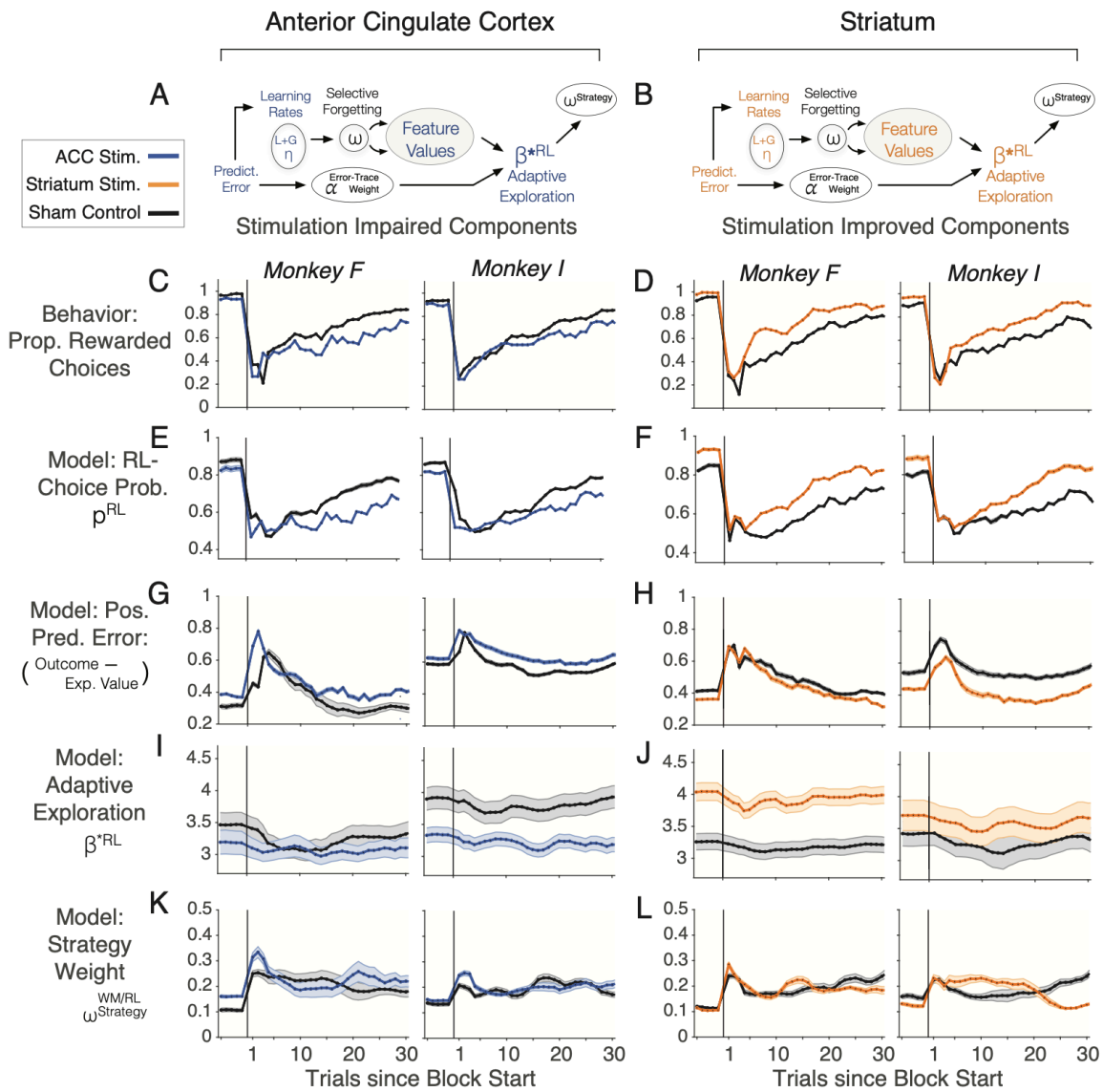


**Figure 6. Stimulation has opposite effects in ACC and striatum on reinforcement learning and adaptive exploration.** (**A**,**B**) Visual summary of the model mechanisms with coloring reflecting modulation with microstimulation in the ACC (*A*) and the striatum (*B*). See Figure 5A for the full model. (**C**,**D**) Trial-by-trial changes in behavioral accuracy for each monkey in ACC (*C*) and the striatum (*D*) when monkeys learned choices in the 3D condition in sham blocks (black lines) and stimulation blocks (colored lines). Analysis used the seventeen ACC sessions and seventeen striatum sessions in which stimulation affected behavior. (**E-L**) Same format as (*A*,*B*) showing estimated values from the AE-SF-WM model for the choice probability of the RL component (*E*,*F*), the positive prediction errors (*G*,*H*), the adaptive exploration-exploitation value $\beta^{*RL}$ (*I*,*J*), and the strategy weight $\omega^{\frac{WM}{RL}Strategy}$ (*K*,*L*). Error shading is SD across fifty bootstrap model fits.

430 for a few trials at the beginning of a block with stimulation in the ACC but stayed at a low (<0.5)
431 level and thereafter remained largely unchanged in ACC and was unaffected by stimulation in
432 striatum (**Figure 6K**,**L**).
433
434 **ACC and striatum uses different neural codes for tracking errors and uncertainty**
435 What are possible neurophysiological sources for the opposite effects of microstimulation on
436 learning performance and model mechanisms in ACC (impaired learning, AE and RL), and the
437 striatum (improved learning, AE and RL)? To address this question, we recorded neuronal activity
438 with linear 64 channel silicon probes in subject F in ACC (737 recording sites) and in the striatum
439 (835 recording sites) at those anatomical coordinates where stimulation had altered behavior
440 (**Figure S7**). We fitted the behavioral performance with the *AE-SF-WM model* and evaluated
441 whether activity in the ACC and the striatum differently encoded the learning variables of the
442 model during the choice period (0.3-0.7 s following fixation onset) at which electrical stimulation
443 was administered in the previous experiment. We found that neuronal activity in the ACC and the
444 striatum showed different patterns of neural encoding of (*1*) choice probabilities, (*2*) object values,
445 (*3*) trial outcomes and (*4*) error history (**Figure 7**). Firstly, there were more recording sites in the
446 ACC that showed negative correlations of neural activity with choice outcomes during the choice,
447 reflecting elevated activity that predicted that the forthcoming choice will be incorrect
448 (unrewarded), while in the striatum there were more positively correlated sites with stronger firing
449 during correct choices (**Figure 7A-C**; two-prop Z-test for neg. versus pos. correlations with
450 correct/incorrect outcome during the 0.3-0.7 s in ACC: Two prop. Z-stat, -3.75, p = 0.00017; in
451 striatum: Two prop. Z-stat, 3.1417, p = 0.00168; two-prop. Z-test for a difference in proportion of
452 ACC vs. striatum for pos. correlations: Z-stat, -3.85, p = 0.00012; for neg. correl. Z-stat, 5.9501,

p = 2.6798e-09). A similar pattern was evident for trial-by-trial correlations of firing with choice probability (**Figure 7D-F**) and with expected stimulus values (**Figure 7G-I**), both were being

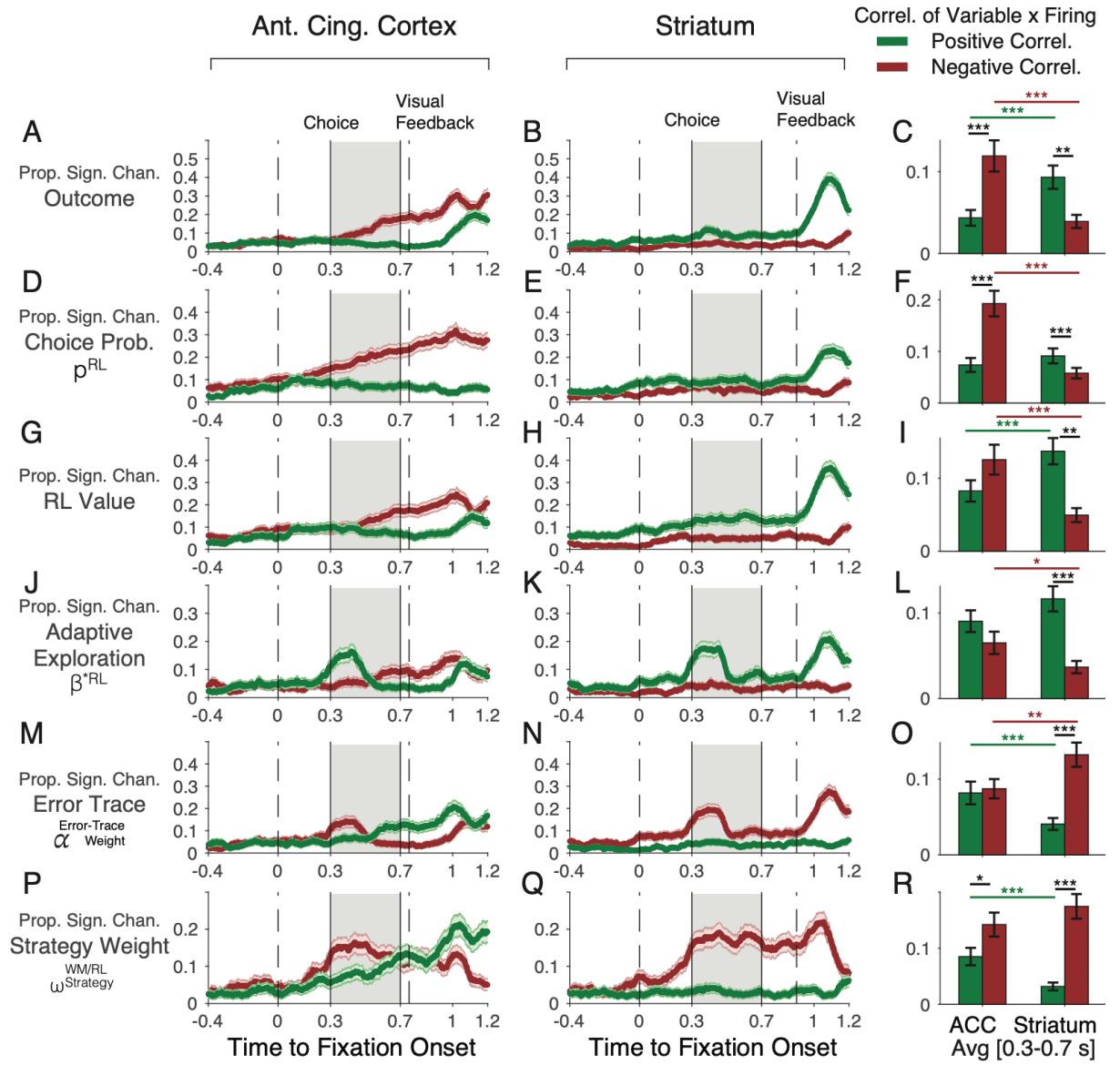


**Figure 7. Encoding of behavioral and latent model variables with neuronal activity differs in the ACC and striatum.** (**A,B**) The proportion of recording channels with multiunit activity in the ACC (*A*) and striatum (*B*) significantly correlating trial-by-trial with outcomes (0=unrewarded,1=rewarded) relative to the onset of fixating the chosen object. Green lines denote the proportion for channels with positive correlations, red lines denote proportions of channels with negative correlations. Grey shaded area is the time period when the subjects committed to choosing the fixated object (at 0.3 s) and when the choice is registered (0.7 s). This choice period was used for electrical stimulation. MUA activity was smoothed with a 0.2 gaussian window prior to the correlation with the variables. Correlations were performed every 25 ms. (**C**) Average proportion of channels with significant positive (green) and negative (red) correlations in ACC (*left*) and the striatum (*right*). Error bars are SDs, horizontal bars and stars denote proportional difference significance level. (**D-R**) Same format as *A-C* for neural activity correlations with choice probability of the chosen object (*D-F*), the expected values of the chosen object (*G-I*), the adaptive exploration-exploitation variable $\beta^{*RL}$ (*J-L*), the error-trace weight variable (*M-O*), and trial-by-trial fluctuations with the WM/RL strategy weight (*P-R*).

encoded by a larger neural population with negative than positive firing correlations in the ACC, and by a larger population with positive than negative correlations in the striatum (for choice probability: two-prop Z-test for neg. versus pos. correlations in ACC: Two prop. Z-stat, -4.7513, p = 2.0212e-06; in striatum: Two prop. Z-stat, 1.8402, p = 0.06574; two-prop. Z-test for a difference in proportion of ACC vs. striatum for pos. correlations: Z-stat, -1.2605, p = 0.20749; for neg. correl. Z-stat, 8.1773, p = 2.2204e-16; for the expected value of chosen objects: two-prop Z-test for neg. versus pos. correlations in ACC: Two prop. Z-stat, -1.9074, p = 0.05647; in striatum: Two prop. Z-stat, 4.3509, p = 1.3558e-05; two-prop. Z-test for a difference in proportion of ACC vs. striatum for pos. correlations: Z-stat, -3.4253, p = 0.00061; for neg. correl. Z-stat, 5.3876, p = 7.1387e-08) (**Figure 7F**,**I**).

For other variables, the coding in ACC and striatum was only partially different. The ACC and the striatum encoded the adaptive exploration parameter value in neuronal populations showing positive activity correlation, reflecting increased neural firing when there was more exploitation than exploration, i.e. when the adaptive exploration variable value was larger. The positive correlation was indistinguishable between areas (two-prop. Z-test for a difference in proportion of ACC vs. striatum for positive correlations: Z-stat, -1.6938, p = 0.0903) (**Figure 7J-L**). But in ACC, an additional population with negative firing correlations emerged during the choice period that was absent in the striatum (two-prop. Z-test for a difference in proportion of ACC vs. Striatum for neg. correl. Z-stat, 2.5747, p = 0.010032) (**Figure 7J-L**). These negatively correlated sites fired stronger when the choice was based more on exploring options than exploiting values, i.e. when the adaptive exploration variable value was lower. Similarly, a separate neuronal population in ACC was evident for the encoding of the error-trace that reflected how previous errors are

accumulated to adjust values of the adaptive exploration variable. During the early choice period ACC and the striatum had a neuronal population with increased activity when errors accumulated, i.e. they were negatively correlated with the error-trace variable (two-prop Z-test for neg. versus pos. correlations with error-trace in ACC: Two prop. Z-stat, -0.27286, p = 0.78496; in striatum: Two prop. Z-stat, -4.7002, p = 2.5997e-06; two-prop. Z-test for a difference in proportion of ACC vs. striatum for neg. correl. Z-stat, -2.823, p = 0.0047575) (**Figure 7M-O**). But only the ACC had a separate neuronal population that increased firing with increasing error-trace values, evident in a positive firing correlation starting ~0.5 s during the choice (two-prop. Z-test for a difference in proportion of ACC vs. striatum for pos. correlations: Z-stat, 3.416, p = 0.00063552) (**Figure 7M)**.

Similar to the error-trace correlations, the ACC and the striatum showed more often negative firing correlations with the strategy weight ($w^{\frac{WM}{RL}Strategy}$), reflecting higher activity when the RL strategy dominated, i.e. when $w^{\frac{WM}{RL}Strategy}$ had a smaller value (two-prop Z-test for neg. versus pos. correlations with the WM/RL strategy weight in ACC: Two prop. Z-stat, -2.4396, p = 0.014702; in striatum: Two prop. Z-stat, -6.7735, p = 1.2571e-11; two-prop. Z-test for neg. correl. Z-stat, -1.7512, p = 0.079914) (**Figure 7P-R**). In addition to this negative correlation a neural population gradually emerged in ACC during the choice period with activity correlating positively with the strategy weight and which was absent in the striatum (two-prop. Z-test for a difference in proportion of ACC vs. striatum for pos. correlations: Z-stat, 4.5459, p = 5.4707e-0) (**Figure 7P,Q**).

## Discussion

We have shown that microstimulating the ACC and the striatum modulated the speed of learning the value of object features when there were multiple distracting features present (3D condition).

501 Microstimulation in ACC and the striatum caused largely opposite behavioral effects. In the ACC
502 stimulation impaired learning (**Figure 2C**), and this impairment (*1*) was more pronounced with
503 higher motivational saliency when gains were higher (**Figure 2E**,**G**), (*2*) was more apparent
504 following correct trials (**Figure 4A**,**C**; **S4K**,**M**), (*3*) was linked to a difficulty in sustaining correct
505 responses (**Figure S6E**), and (*4*) gradually emerged with increasing stimulation trials (**Figure**
506 **S6A**). In the striatum, stimulation improved learning (**Figure 2C**), and this improvement (*1*) did
507 not increase with higher incentives (**Figure 2D-G**), (*2*) was more apparent following correct trials
508 (**Figure 4A**,**C**; **S4O**,**Q**), and (*3*) was evident already following the first stimulation trials (**Figure**
509 **S6C**,**D**). Beyond these overall effects, there was heterogeneity of the stimulation effects, but for
510 each area, these overall effects were reliably evident in about a third of the stimulation sessions in
511 ACC (monkey F / I: 29% / 42% of sessions) and in the striatum (monkey F / I: 33% / 38% of
512 sessions) (**Figure 3**).
513
514 Taken together, these results document a dissociation of stimulation-induced changes of adaptive
515 goal-directed behavior in the ACC versus the striatum. This dissociation became evident when the
516 behavioral goal involved learning feature values in a task environment with multiple distracting
517 features (3D condition). The impairment (in ACC) and improvement (in the striatum) of learning
518 in this multidimensional environment was closely associated with a weakening (in ACC) and
519 strengthening (in the striatum) of adaptive reinforcement learning mechanisms, while cognitive
520 learning strategies such as selective forgetting and working memory were not consistently affected.
521 The differences in behavioral stimulation effects were consistent with a more heterogeneous
522 neuronal encoding of choice probabilities, values, error history, and adaptive exploration in the
523 ACC when compared with the striatum (**Figure 7**). While both areas encoded these variables with

524 firing rate increases (positive correlations of variables and firing), the ACC contained larger
525 neuronal subpopulations than the striatum that showed negative firing correlations with choice
526 probabilities, values, and exploration, while having separate neuronal populations that increased
527 firing when errors accumulated (**Figure 7D**,**G**,**J**,**M**). This finding suggests that an activating
528 electrical stimulation protocol (*see* below) will likely have increased neuronal signaling in the
529 ACC linked to higher choice uncertainty, lower valuation of the chosen target stimulus, more
530 exploratory choices and larger error traces. In contrast, activating the striatum will have more likely
531 increased activity of neurons who would otherwise increase their firing when the choice was made
532 with larger certainty (high choice probability, **Figure 7E**), to an object with higher value (higher
533 value of the chosen stimulus, **Figure 7H**), and more exploitative choices (positive correlation with
534 the adaptive exploration parameter, **Figure 7K**). These differential effects of stimulation are
535 putative as they were collected at the same anatomical coordinates as the stimulation experiments,
536 but not at precisely the same spot as they were not collected in the same experimental sessions.
537 Despite this limitation, the consistency of differential stimulation induced behavioral effects, and
538 the differential neuronal encoding provides a parsimonious framework for understanding our key
539 results.
540
541 **Electrical microstimulation biases excitation of the local circuit output.** An assumption for
542 interpreting the main findings of our study is that the brief 0.3 s microstimulation pulse trains used
543 in our experiment activated the neural circuit. Previous studies support this assumption, showing
544 that biphasic ~50 μA microstimulation pulse trains activate neurons within a ~200-300 μm radius
545 and increase net excitation across distances that can exceed 1500 μm, depending on the spread of
546 excitatory horizontal projections [52-55]. Microstimulation protocols with similar parameters (60-300

547 ms long, 20-100 μA pulses) applied in the frontal eye field and lateral intraparietal area bias
548 saccades towards target stimuli overlapping with the peripheral response field of the stimulation
549 site [56-60]. Similarly ~50μA microstimulation pulse trains have different effects in more mediodorsal
550 frontal areas such as the supplementary eye field where microstimulation elicits eye movements
551 at higher thresholds, can slow-down overt saccadic eye movements to target stimuli in difficult
552 tasks, speed-up eye movements in simple tasks, and reduce the correct selection of sequential
553 targets [61-63]. These effects suggest that microstimulation in the supplementary eye field can
554 enhance a net inhibitory effect of it on other saccade generating structures. These results suggest
555 that stimulation in the non-eye movement related ACC (our stimulation sites were anterior to the
556 eye movement fields in the mid-cingulate [33], **Figure 1D**) likely excites the local circuit, which is
557 associated with an inhibitory effect on prefrontal projection targets [64] and a net excitatory effect
558 on the striatum [35].
559
560 **ACC provides error monitoring and directed exploration over exploitation**
561 Our stimulation effects provide evidence that the ACC is causally involved in the monitoring of
562 errors and outcome uncertainty and in the directed exploration of visual objects to reduce outcome
563 uncertainty during learning. Behavioral learning is governed by latent variables that are not
564 explicitly observable but need to be estimated from the sequence of behavioral choices. We
565 modeled various latent variables implementing learning mechanisms and cognitive strategies and
566 found microstimulation affected similar mechanisms in ACC and striatum but in opposite ways
567 (**Figure 5C-H**). ACC stimulation reduced weighting of recent errors ($\alpha^{\text{ErrorTrace}}$). In the model
568 lower weighting of error increases exploration (lower $\beta^{*RL}$ values). Consequently, the choices of
569 both monkeys during ACC stimulation blocks reflected higher exploration levels (**Figure 6I**),

570 more choice uncertainty (**Figure 6E**), and more outcome uncertainty as evident in higher
571 prediction errors across the learning block in stimulation than in sham blocks (**Figure 6G**). The
572 increases of uncertainty and exploration levels are consistent with the result that ACC stimulation
573 induced a difficulty in sustaining corrected responses (**Figure S6E**), an effect that has also been
574 found after extensive lesions of the ACC [11].
575
576 These results provide causal evidence that one overarching function of the ACC is to evaluate
577 outcome uncertainty and to use this information to guide information seeking, i.e. to determine
578 how to explore options to reduce uncertainty [9,65,66]. In our study, these functions were affected in
579 the 3D condition in which uncertainty is reflected in the number of features that could be associated
580 with reward. In the presence of this larger feature space, subjects do not only need to know when
581 to explore versus exploit, but also which specific feature should be credited for a choice outcome.
582 Such a feature specific credit assignment is encoded in neurons in the ACC during feature learning
583 [29], and these neuronal representations of feature-specific prediction errors provide critical
584 information to guide exploration towards those features that result in positive prediction errors and
585 to avoid choosing features whose choice results in negative prediction errors.
586
587 Taken together, our modeling results suggests that the ACC contributes to an adaptive learning
588 process by tracking the history of recent prediction errors to adjust the degree of exploration over
589 exploitation that subjects apply in subsequent trials. This adjustment of exploration may be either
590 un-directed, i.e. modifying random exploration, or directional, i.e. modifying which features to
591 explore in future trials. Our results are more consistent with the ACC supporting directed
592 exploration, because stimulation did not induce random exploration in the 2D or 1D condition but

specifically altered exploration in the 3D condition when high feature uncertainty calls upon exploratory guidance. In this 3D environment learning was impaired most prominently following gains compared to losses – irrespective of whether stimulation was administered on correct or error trials - suggesting that assigning positive credit to specific features for exploratory guidance was disrupted more than using loss to avoid features that were unrewarded. These conclusions resonate with a hypothesis that posits the ACC acts together with the ventrolateral prefrontal cortex to guide information seeking by detecting the sources of uncertainty and directing exploratory sampling towards objects to reduce the uncertainty [65]. This hypothesis also clarifies that the role of the ACC for monitoring sources of uncertainty requires interactions with the ventrolateral prefrontal cortex to update those representations that guide information seeking, attention, and choices, consistent with lesion studies [67,68], and with recording studies showing neuronal tuning to attended object features [69,70] and attention-dependent spiking interactions between the ACC and the ventrolateral prefrontal cortex [35,71].

**Causal facilitation of adaptive reinforcement learning with striatum stimulation**

In the striatum, the microstimulation effects on behavior and on the learning mechanisms of the learning model were opposite in sign compared to the ACC. Striatum stimulation enhanced exploiting expected values over exploration (higher $\beta^{*RL}$), and improved the adaptive reinforcement learning mechanisms either by enhanced weighting of recent errors ($\alpha^{\text{ErrorTrace}}$, in one subject) (**Figure 5E,F**), or by an enhanced exploration weighting ($\omega^{Exploration}$, in the other subject) (**Figure 5D**). These changes were reflected in reduced prediction errors (**Figure 6H**), enhanced certainty (choice probability) with which choices were made (**Figure 6F**), and more exploitation of value expectations (**Figure 6J**). These findings provide evidence that the striatum

616 causally supports adaptive reinforcement learning. This conclusion is consistent with suggestions
617 that the striatum is situated in fronto-striatal networks to adjust learning processes about goals
618 versus values [10,72], and about predictive versus selective value associations [1]. However, causal
619 evidence for a role in adjusting reinforcement learning during goal-directed behavior has remained
620 sparse. Our findings indicate that striatal microstimulation enhances the weighting of prediction
621 errors, which will improve value updating and reduced uncertainty about the most valuable feature
622 in future trials. Support for this scenario comes from multiple sources. Previous studies have found
623 that striatal neurons that encode feature-specific prediction errors begin to encode the value of
624 reward predicting stimuli when learning completes [29], and that subpopulations of striatal neurons
625 fire stronger the less uncertain the reward-value of attended stimuli [28,29]. These findings were
626 confirmed in our neural recordings that revealed larger populations of neurons encoding choice
627 probability and the value of chosen stimuli with firing rate increases (i.e. via positive correlations
628 of firing with variables) than with firing decreases (**Figure 7E,H**). In addition, a prior stimulation
629 study has documented that stimulation of the caudate nucleus when subjects process a single
630 stimulus on the screen can enhance the value of that stimulus, effectively enhancing the likelihood
631 it will be chosen in future choices [31]. Our results extend these findings by suggesting that
632 stimulation does not enhance the value of all features of a processed object, but that stimulation
633 selectively enhances the reward-relevant target feature particularly when there were two
634 distracting features of the multidimensional objects that need to be filtered during learning. The
635 specificity of the stimulation effect to improve learning in the 3D condition is consistent with the
636 overarching function of the striatum as implementing an adaptive filtering of cortical inputs. This
637 filtering function of the striatum has largely been inferred for the motor domain, but recent studies
638 have started documenting that similar adaptive filtering and attentional selection is evident also

639 for non-motor tasks that involve the learning and choices of visual features and objects [28,35,42,73,74].

640

641 **ACC mediates cognitive control at high motivational saliency.** We observed the most apparent
642 learning deficits with ACC stimulation when two conditions were met: when the motivational
643 salience of a choice was high (high gains or losses) and when the credit assignment about feature
644 values was difficult (in the 3D condition). This pattern of result is consistent with a study causing
645 a more prolonged disruption of the ACC [17] and shows that the ACC fundus multiplexes
646 motivational and cognitive functions, which is consistent with two major accounts of ACC
647 function: Consistent with a motivational effort-control account [75], the ACC causally supported
648 learning when the motivational saliency of choices is high by either anticipating a high
649 motivational benefit (gaining 3 vs 2 token) or (in one subject) a high potential punishment (loss of
650 3 vs 1 token). Consistent with a cognitive account, the ACC causally guided searching for choice
651 options when outcome uncertainty was high (3D versus 1D or 2D) [19,76]. Importantly, the
652 multiplexing of motivation and cognitive variables is consistent with the ACC's rich direct
653 anatomical connectivity to subcortical sites that directly affect motivation, including dopaminergic
654 neurons, and that mediate attentiveness and cognitive control, including cholinergic and
655 norepinephrinergic nuclei [77,78]. This specific anatomical connectivity has been used as a key
656 argument for suggesting that the ACC adaptively boost cognitive control when task demands are
657 high [79]. Our results provide direct causal support for this 'boosting' function of the ACC.
658 Importantly, our modeling results suggest that the proposed theoretical function of the ACC to
659 'boost cognitive control' [80] may be realized by modulating exploration when a difficult task
660 requires motivational support [17].

661

662 Taken together, the present study reveals a causal influence of the ACC and the striatum to support
663 adaptive reinforcement learning. The behavioral stimulation effects were traced back to the ACC's
664 and striatum's role in the ongoing estimation of the reinforcement learning values and their
665 uncertainty given the history of recent experiences and to the feature-specific credit assignment
666 that is needed to optimize attention and decision making in a changing environment. These results
667 and conclusions may prove important for developing brain computer interfaces designed to
668 enhance cognitive control and support cognitive flexibility.
669
670
671 **Methods**
672 **Ethics Statement.** All animal and experimental procedures were in accordance with the National
673 Institutes of Health Guide for the Care and Use of Laboratory Animals, the Society for
674 Neuroscience Guidelines and Policies, and approved by the Vanderbilt University Institutional
675 Animal Care and Use Committee (M1700198-01).
676
677 **Experimental Procedures.** Two adult male macaque monkeys performed the experiments
678 (monkey F, 13 years of age, 11.3 kg, and monkey I, 11 years of age, 11.0 kg). Behavior and
679 electrical stimulation triggers were controlled using the Multi-Task Unified Suite for Experiments
680 (M-USE) [81]. The animals were trained on the feature-reward rule learning task in the cage-based
681 touchscreen Kiosk Testing Station [82] and stimulation experiments proceeded in sound attenuating
682 experimental booths with subject's head position fixed, facing a 21'' LCD screen at a distance of
683 63 cm from their eyes to the screen center. For each stimulation experiment, a tungsten stimulation
684 electrode (FHC Inc., Bowdoin, ME) was lowered through a guide tube at a pre-determined location

in a custom recording/stimulation chamber implanted over the left hemisphere. The stimulation electrodes were lowered into the ACC or the striatum and stopped once sustained spiking activity was noticeable and constant.

**Task paradigm.** The feature-reward rule learning task required learning through trial-and-error which feature of multidimensional objects is associated with reward. Objects were 3D rendered Quaddles [83] that could vary features in 1,2, or 3 feature dimensions within a block (**Figure 1A-C**). The feature dimensions were body shape, arm type, color, and body pattern with 9-11 feature per dimensions (e.g. oblong, pyramidal, spherical, etc. for body shapes). In each block, only one feature value was associated with token reward. The feature that was rewarded, i.e. the feature-reward rule, stayed constant for 30-50 trials, then switched randomly to another feature in a new block. Monkeys initiated a trial by fixating a black circle for 0.5 s, which triggered after a 0.3 s delay the presentation of three objects randomly positioned at three of the four corners of a virtual square-grid spanning 10.5 cm with 5° eccentricity relative to the screen center. Objects extended over ~3cm on the screen corresponding to ~2.5° degrees of visual angle. The animals had up to 5 s to choose one object by maintaining gaze at an object for >0.7 s Choosing the correct object was followed by a yellow halo around the stimulus as visual feedback, an auditory tone, and either 2 or 3 tokens (green circles) added to the token bar (*G2* and *G3* conditions). Choosing an object without the rewarded target feature was followed by a blue halo around the selected objects, a lower-pitched auditory feedback, and for the loss conditions, the presentation of one or three grey 'loss' token that traveled to the token bar where one or three already obtained tokens were removed. In different blocks either 1 or 3 tokens were lost after incorrect choices (*L1* and *L3* conditions). The timing of the feedback was identical for gain and loss conditions. In each session,

708 monkeys were presented with up to 36 separate learning blocks, each with a unique feature-reward
709 rule.
710
711 Across all 48 experimental sessions, monkeys F/I completed on average 33.5 / 36 blocks during
712 ACC stimulation sessions and 32.0 / 35.8 blocks during striatum stimulation sessions. For each
713 experimental session, a unique set of objects was defined by randomly selecting three dimensions
714 and three features per dimension (e.g., 3 body shapes: oblong, pyramidal and ellipsoid; 3 arm
715 types: upward pointy, straight blunt, downward flared; 3 patterns: checkerboard, horizontal striped,
716 vertical sinusoidal) [83]. Based on this feature set three different task conditions were defined: One
717 task condition contained objects that differed in features of only one dimension, while features of
718 the other dimensions were identical; for example, the object body shapes were oblong, pyramidal,
719 and ellipsoid, but all objects had blunt straight arms and uniform grey color ('1D' condition). A
720 second task condition defined objects that varied features in two dimensions ('2D' condition), and
721 a third task condition defined objects that varied in features of three dimensions ('3D' condition).
722 Learning which feature is rewarded is systematically more demanding with objects varying in
723 more feature dimensions [17].
724

725 **Experimental design.** Each session was composed of 36 feature-reward rule blocks that pseudo-
726 varied four motivational conditions and three cognitive demand conditions. Cognitive demand
727 varied the number of object features that varied from trial to trial with features varying in one, two,
728 or three dimensions (conditions 1D, 2D, or 3D). The motivational conditions varied the amount of
729 tokens gained for correct choices to be 2 or 3 (G2 and G3 conditions) and the amount of tokens
730 subtracted (lost) from incorrect choices were either -1 or -3 (L1 and L3 conditions) as introduced

731 in prior studies [84]. Subjects had to earn 8 tokens to completely fill a token bar on top of the screen.
732 Each block began with an asset of 3 tokens. This starting asset ensured that subjects lost tokens
733 already after the first incorrect trial in a block and thereby recognized the block's loss condition.
734 Across the 36 blocks all combinations of motivational conditions (+2/-1, +3/-1, +2/-3, +3/-3) and
735 cognitive conditions (1D, 2D, 3D) were pseudo randomly assigned to blocks in equal number.

736 Each session contained 15 blocks with electrical stimulation, 15 sham stimulation blocks, and 6
737 initial blocks without stimulation. A session was designated either a *StimReinf+* or a *StimReinf-*
738 session in which *Sham* and one of the stimulation conditions (*StimReinf+* or *StimReinf-*) alternated
739 starting after the sixth block. A session ended either when all 36 blocks were completed, or when
740 a monkey stopped initiating trials after 90 minutes.

741 **Monitoring and analysis of gaze.** Gaze was monitored with the binocular infrared eye-tracker at
742 600 Hz sampling rate (Spectrum, Tobii). Gaze was calibrated before the task began with a 9-point
743 eye-tracker calibration routine. Object fixations were reconstructed using a robust gaze
744 classification algorithm described in [85] . The duration of object fixations was analyzed for objects
745 fixated prior to the last fixation onto the object that the subjects choose by maintaining fixation for
746 >0.3 s onto that object.
747

748 **Electrical microstimulation**. Electrical stimulation was delivered with the Intan Stimulation
749 Controller (RHS 2000, Intan Technologies, LLC), starting 300 ms after monkeys continuously
750 fixated an object and lasted for 300 ms. Choice fixations lasted >700 ms ensuring that the electrical
751 stimulation was administered only while the chosen object was fixated. Monkey's fixation
752 durations had a bimodal distribution with the fixations indicating the sampling of information prior

753 to making a choice lasting 150-200 ms on average in the 1D, 2D, and 3D conditions, and the last
754 fixation that monkeys used to indicate a choice lasting on average 800 ms (see Suppl. Figure S7
755 in [17]). The 300 ms lasting 200 Hz electrical stimulation pulse train consisted of 60 biphasic 50 μA
756 pulses, with a 0.2 ms cathodal current injection followed immediately by a 0.2 ms anodal current,
757 and a 5 ms interval between onsets of these 0.4 ms lasting pulses (*see* inset in **Figure 1A**). Prior to
758 starting microstimulation in an experimental session, we measured the impedance using the built-
759 in Intan stimulation controller's *electrode impedance measurement*, resulting in an average
760 impedance of 94.45 and 85.06 $k\Omega$ for stimulation sessions of Monkeys F and I. These values are
761 as expected given the 50kΩ – 100kΩ impedance of the stimulation electrode (tungsten
762 microelectrode UEWLFDSEAN1E, 100 mm length; FHC Inc., Bowdoin ME).
763
764 **Stimulation-triggered multi-unit activity and its relation to learning performance**. Prior to
765 starting stimulation within an experimental session, we ensured through visual and auditory
766 inspection of neural activity that the stimulation electrode was in the grey matter at sites in which
767 neurons showed spiking activity. Offline, we quantified how strong multi-unit activity at the
768 stimulation electrode was modulated by electrical stimulation and tested whether the modulation
769 of stimulation-triggered multi-unit activation (MUA) was predictive of behavioral effects of
770 stimulation. To this end, we extracted continuous multi-unit activity using a 0.5-3 kHz bandpass
771 filter, followed by rectification, and 200 Hz low-pass filtering of the recorded signal before and
772 after the 0.3 s stimulation period. We removed outlier trials reflective of artifacts with the inter-
773 quartile method and for each session normalized the data to a -0.3-0 s pre-stimulation baseline
774 period. To quantify the stimulation triggered effect, we extracted the maximum response within a
775 0.65-1.3 s post-stimulation window in the stimulation and sham conditions, and calculated a *MUA*

776 *Index* as the normalized difference (*Stim* – *Sham*)/(*Stim* + *Sham*) of that maximum response. The
777 *MUA Index* ranges from -1 to +1 with positive values reflecting stronger MUA activity in the
778 stimulation condition compared to the sham condition. For the same sessions we also quantified
779 the learning speed (trials-to-criterion) and calculated a *Performance Index* (*Stim* – *Sham*)/(*Stim* +
780 *Sham*). We statistically analyzed the relation of the MUA Index and the Performance Index for
781 those experimental conditions in which stimulation affected learning using Pearson correlations.
782
783 **Reconstruction of anatomical stimulation locations**. We reconstructed the recording and
784 stimulation sites in the ACC and striatum using the software 3D slicer (http://www.slicer.org,
785 version 4.11) (**Figure 1D**,**E**). For each monkey, we first co-registered MRI's taken when the
786 recording chamber was implanted, with CT scans that were taken after the chamber was implanted.
787 Then, we segmented the recording chamber and reformatted the image space to reference the
788 chamber space. We recorded the x,y,z (z being the depth of the electrode) coordinate the
789 stimulation electrode tips relative to the chamber to reproduce the penetration depth for each
790 stimulation electrode. The reconstructed electrode tip site was labeled on the MRI images, where
791 monkey F's stimulation sites were overlayed onto monkey I's image to have a unified image of
792 stimulation sites.
793
794 **Electrophysiological recording experiment**. Following the stimulation experiment, we
795 performed electrophysiological recordings in thirteen experimental sessions in the ACC and the
796 striatum at those anatomical coordinates used during the stimulation experiment. Recordings were
797 performed in monkey F performing the same experimental task as during the stimulation
798 experiment. In these recording-only sessions neural activity was recorded with 64-channel linear

799 silicon probes (Neuronexus, 50 μm interelectrode distance) using Open-Ephys (open-ephys.org)
800 and the Intan Recording System (RHD2000, Intan Technologies, LLC). Neuronal wideband
801 recordings were notch filtered (60, 120, and 180 Hz) with a filter bandwidth of 2 Hz to remove
802 power noise and its harmonics and re-referenced using median referencing across channels.
803 Continuous multi-unit activity (MUA) was quantified by 750-5000 Hz bandpass filtering,
804 rectifying, 300 Hz low pass filtering, and down-sampling to 1000 Hz. Individual trials were
805 extracted from trial initiation until the token and reward feedback was given with additional 0.5 s
806 padding. Trials were aligned to the time of the decision, and MUA smoothed using a gaussian
807 window of 0.2 s. We then used the mean activity during the baseline period 0.5 s at the beginning
808 of the trial and prior to the onset of the stimuli to detrend the data channel-wise. We then
809 determined whether a recording channel showed task modulated neuronal activity with a
810 regression model that predicted MUA activity every 25 ms using as factors trial outcome (correct
811 vs error), reward outcomes (gain and loss of tokens), learning status (before vs. after trial-to-
812 criterion was reached), stimulus dimensionality (1-,2-, and 3 dimensional objects), and previous
813 trial outcome (correct vs error). We removed channels that showed no task modulation for any of
814 these task variables in at least three consecutive time windows which affected on average 0.5
815 channels per probe.
816
817 For the neuronal analysis, we aligned the MUA activity to the onset of the fixation of the object
818 the subject chose on that trial by continuing fixating it for 0.7 s and smoothed it with a 200ms
819 gaussian kernel (**Figure 7**). We then calculated every 25 ms the correlation of MUA activity across
820 trials with the outcomes (error vs. correct) and with trial-by-trial changes of the model-estimated
821 choice-probability, prediction error, expected object value of the chosen object, adaptive

822 exploration value, error-trace, and the strategy weight (see Figure 7 and Methods). For each
823 variable we quantified the proportion of MUA channels that shows positive versus negative
824 correlation values at a $p<0.05$ significance level. We compared the average proportion of positive
825 versus negatively correlations in the 0.3 – 0.7 s time window for each variable within each area
826 and between areas (ACC versus Striatum) with a 2-sample Z test for proportions (**Figure**
827 **7C**,**F**,**I**,**L**,**O**,**R**). The time window overlaps with the 0.3-0.6 s time window that was used for
828 electrical microstimulation.
829
830 **Behavioral analysis of learning and plateau performance**. We estimated the efficiency of
831 learning by calculating the trials required to reach criterion performance as the first trial within a
832 block at which animals reached 70% accuracy over the following 12 trials. To compare the trials-
833 to-criterion between conditions we used two-way Student's T tests with FDR corrected (alpha =
834 0.05) p-values. If monkeys did not reach this learning criterion during a block, we calculated the
835 trial at which the learning criterion would have been reached by using a linear regression of the
836 last 12 trials in the block to find the predicted trials needed to reach criterion. We counted these
837 blocks as being unlearned to determine that the proportion of learned blocks did not differentiate
838 the experimental conditions (**Figure S1A**). For the condition with *StimReinf-* stimulation, we
839 excluded blocks with less than 2 active stimulation trials (i.e. with less than 2 errors). We applied
840 two-way Student's T tests with FDR corrected (alpha = 0.05) p-values to compare plateau accuracy
841 of the animals. Plateau performance was indexed as the average accuracy across trials once the
842 learning criterion was reached (**Figure S1B**).
843

844 **Analysis of Variance (ANOVA) on experimental conditions**. N-way ANOVAs were performed
845 to determine if there were significant stimulation effects on learning (measured by trials-to-
846 criterion for the block) between four experimental factors: dimensionality (*Dim*) with three levels
847 (1D, 2D, and 3D); *Gain* context with 2 levels (Gain 2, Gain 3); *Loss* context with 2 levels (Loss 1,
848 Loss 3), and stimulation (*Stim*) contexts 3 levels (Sham, *StimReinf*+, *StimReinf*-). We reported the
849 significant ($p < 0.05$) interaction effects as either 2-way (stimulation condition x dimensionality)
850 or 3-way (stimulation condition x dimensionality x gain/loss) effects.
851
852 **Identifying sessions with individually significant behavioral stimulation effects.** We identified
853 sessions in which stimulation impaired or improved learning relative to sham using Fisher's Exact
854 test comparing stimulation versus sham blocks in the 3D condition (**Figure 3**). We computed an
855 index to facilitate comparison of the trials-to-criterion (learning speed) across sessions as *Index* =
856 *(Stim – Sham)/(Stim-Sham)* with values ranging from +1 to -1, with positive values indexing slower
857 learning in the stimulation than sham blocks and negative values indexing faster learning in the
858 stimulation than sham blocks.
859
860 **Linear mixed effect models for block-level analysis of behavioral metrics.** We used linear
861 mixed effects models (LME) to evaluate the influence of experimental parameters on several
862 behavioral metrics, such as 1) trials-to-criterion, labeled 'LP' for learning point; choice reaction
863 times 2) during learning (prior to reaching the learning point), and 3) after learning; the number of
864 fixations prior to choosing an object 4) during learning and 5) after learning; and 6) the plateau
865 accuracy when learned completed. We tested stimulation effects on behavior with four
866 experimental factors: feature dimension (*Dim*) with three levels (1D, 2D, and 3D); *Gain* context

867 with 2 levels (Gain 2, Gain 3); *Loss* context with 2 levels (Loss 1, Loss 3), and stimulation (*Stim*)
868 contexts with 2 or 3 levels (Sham, *StimReinf*+, *StimReinf*-). We used two other factors as random
869 effects: the feature dimension of the target feature (*Feat*) with 4 levels (color, pattern, arm, and
870 shape) and subject (*Monkey*) factor with 2 levels (F, I). The LME is formalized as $Metric = Dim +$
871 $Gain + Loss + Stim + (1|Feat) + (1|Monkey) + b + \varepsilon$. We extended this LME to test for interactions
872 of *Stim2* with *Dim* , *Gain* , and *Loss:* $Metric = Dim + Gain + Loss + Dim \times Stim + Gain \times Stim$
873 $+ Loss \times Stim + (1|Feat) + (1|Monkey) + b + \varepsilon$.
874
875 **Analysis of choice reaction times.** Stimulation might have an influence on the reaction time from
876 display onset to the time when an object is fixated. We analyzed choice reaction times separately
877 for correct trials in which the subjects made only one saccade and chose the first object they fixated
878 and trials in which they fixated a second and third object (**Figure S2A**,**B**). The distribution of
879 choice reaction times was determined separately for the sham, *StimReinf*+, and *StimReinf*-
880 conditions from the first block in which stimulation began (the sixth block onwards). Choice
881 reaction time distributions for trials with one, two, and three object fixations were compared with
882 a two-way Student's t-test with FDR corrected (alpha = 0.05) p-values.
883
884 **Analysis of fixational sampling of objects.** To test whether stimulation altered fixational
885 sampling of objects we analyzed the number of exploratory fixations of objects prior to the last
886 choice-fixation during learning (prior to reaching learning criterion) and when learning completed
887 for the sham and stimulation conditions, comparing them using permutation tests with FDR
888 corrected (alpha = 0.05) p-values (**Figure S1E,F**). In addition, we compared the duration of stable
889 fixations on objects prior to the last fixation of an object that indicated it was chosen (**Figure S2A-**

890 **B**). Fixational sampling duration distributions for trials with one, two, and three or more object
891 fixations were compared with a two-way Student's t-test with FDR corrected (alpha = 0.05) p-
892 values. The distribution of fixational sampling duration was done separately for the *sham*,
893 *StimReinf*+, and *StimReinf*- conditions from the first block in which stimulation began (the 7$^{th}$
894 block onwards).
895
896 **Temporal stability of stimulation effects on learning across sessions and within sessions.** To
897 discern whether the effect of stimulation on learning was stable over time within a session, we
898 fitted a regression to the average trials-to-criterion for sham and stimulation blocks per session and
899 compared the slope and intercept across conditions for each monkey, separately for the 1D, 2D,
900 and 3D conditions (**Figure S2C**,**D**,**G**,**H**). To test for temporal stability of the stimulation effects
901 within a session, we subdivided the number of stimulation blocks into three equally sized sets of
902 ten blocks (blocks 7-16; 17-26; 27-36) and calculated the average trials-to-criterion criteria
903 separately for the stimulation conditions and for the 1D, 2D, and 3D conditions (**Figure**
904 **S2E**,**F**,**I**,**J**). We used the two-way Student's T test with FDR corrected p-values (alpha = 0.05) to
905 test for significance between the time epochs within a session.
906
907 **Analysis of the behavioral adjustment following erroneous and correct choices**. We quantified
908 the accuracy (proportion correct choices) for those trials that were preceded by different types of
909 behaviors separately for the *Sham*, *StimReinf*+, and *StimReinf*- conditions. These analyses were
910 limited to trials during learning, i.e. prior to completion of learning and performed with data
911 combined from both monkeys for all sessions (**Figure 4**) and limited to the effective sessions with
912 individually significant behavioral stimulation effects (**Figure S4K-R**). We separately calculated

accuracy in trials relative to an erroneous or a correct choice. The main analysis was limited to the trials prior to completion of learning, but results were qualitatively similar when all trials were considered.

To compare stimulation effects on previous trial dependent performance, we used permutation tests. For the post-error and post-correct performance, we randomly shuffle, without replacement, the stimulation condition labels (sham, *StimReinf+*, *StimReinf-*) of the trial outcome for each $n^{th}$ trial (n=1-5 trails following the outcome), for 10000 iterations to establish a randomized sampling distribution of the mean difference between groups. We then calculated the probability of a finding a value greater than the observed mean difference per stimulation condition following an outcome; the null hypothesis being that accuracy following a given trial outcome is not different across stimulation conditions. The p-value calculated is the probability that the difference between the groups mean being greater than the observed mean for each $n^{th}$ trial. P-values were FDR corrected (alpha = 0.05) across post outcome trials.

**Accuracy relative to the number of stimulation trials.** We tested how accuracy changed with increasing number of stimulation trials by calculating the average performance across the subsequent 3 trials following the $1^{st}$, $2^{nd}$, $3^{rd}$ , … $n^{th}$ stimulation trial within stimulation blocks and sham-stimulation trials in sham blocks (**Figure S6A-D**). We compared conditions using permutation tests for each subject and the 1D, 2D, and 3D experimental conditions.

**Error-Correct ($EC_n$) analysis of trial effects on accuracy.** We tested if stimulation affected the ability to sustain correct or incorrect choices over trials (**Figure S6E-H**). We calculated the

accuracy in the trial following correct (C) and incorrect (E) sequences occurring in the block, where 'CEE' would correspond to a 'Correct, Incorrect, Incorrect' sequence the accuracy of the following trial. We calculated this across blocks and separately for the 1D, 2D, 3D conditions separately for the sham versus *StimReinf*+ and sham versus *StimReinf*- comparison. The choice outcome sequences used for the Sustained Correct analysis were: E, EC, ECC, ECCC, ECCC; and for the Sustained Error analysis: C, CE, CEE, CEEE, CEEEE. P-values were FDR corrected (alpha = 0.05) for permutation testing.

**Logistic regression model to compare trial-wise differences between stimulation and sham**

We used a generalized linear model (GLM) with a binomial distribution and logit link function to analyze binary choice outcomes across trials in sham and stimulation conditions. The model tested how trial number (*Trial*) following gains or losses, and stimulation condition (*Stim*) influenced response probabilities (correct = 1, incorrect =0), while also considering their interaction. Thus, the GLM was formalized as: $logit(P(Y)) = \beta 0 + \beta 1(Trial) + \beta 2(Stim) + \beta 3(Trial \times Stim) + \varepsilon$, where *P(Y)* is the probability of a response (1 or 0), and $\epsilon$ represents residual error. To assess the significance of stimulation effects, we examined the estimated coefficients (β1, β2, β3) and their corresponding p-values (Wald tests). A significant interaction (β3) would indicate that the influence of trial number on response probability differs between sham and stimulation conditions (interaction). This approach allows us to statistically test whether stimulation alters behavioral response patterns over time. For analyses involving trials following gain or loss, only the five subsequent trials were included in the model.

958 **Behavioral modeling combining reinforcement learning, working memory and adaptive**
959 **exploration.** We modeled the choices of the monkeys with a recently proposed reinforcement
960 learning model [23] that combines the classical mechanism of reinforcement learning (RL) of values
961 using reward prediction errors with (1) an adaptive exploration (AE) mechanism that adjusts the
962 degree of exploration relative to exploitation depending on the recent history of negative prediction
963 errors (when a reward was predicted but not obtained), (2) a selective forgetting (SF) mechanism
964 that decays the values of features from non-chosen (non-attended) objects, and (3) a working
965 memory (WM) mechanism that updates objects in a working memory after a rewarded choice. The
966 full model combining all three mechanisms is labeled the *AE-SF-WM model*, depicted in **Figure**
967 **5A**. It generates on every trial two hypothetical choices (choice probabilities) based on the RL and
968 the WM component and arrives at the final choice by weighting the expected feature values (RL)
969 and object values (WM) using a weighting parameter that reflects the relative success of each
970 mechanism of predicting rewarded choices in the past trials (**Figure 5A**). We refer to this
971 weighting parameter as the strategy weight as its value indicates whether the RL component or the
972 WM component is contributing more to a choice. There are several static model mechanisms as
973 well as two adaptive variables of the adaptive RL-WM model that vary trial-by-trial depending on
974 the performance of the subjects. These model mechanisms have been systematically explored and
975 validated to be important for adaptive behavior with multidimensional stimuli elsewhere [23]. In
976 brief, the reinforcement learning (RL) model updates expected feature values (V) using reward
977 prediction errors (RPEs) scaled by learning rates. The RL component is a traditional Rescorla
978 Wagner Q-learning model that estimates for each trial the expected Q value of objects. For each
979 of the three objects presented on a given trial, the object value is calculated as the sum of the values
980 of its features ($V^F$)**,** which are updated on every trial according to the reward prediction error (RPE)

981 equal to $(R_t - V_{i,t}^F)$, where R is the rewarded (1) or nonrewarded (0) outcome of the trials
982 multiplied by a learning rate $\eta_t$ resulting in the equation $V_{t+1}^F = V_{i,t}^F + \eta_t\ (R_t - V_{i,t}^F)$, here *i* indexes
983 the feature and *t* the trial. We evaluated models with two separate learning rates for error outcomes
984 ($\eta_t = \eta_{Loss}$), and for correct outcomes ($\eta_t = \eta_{Gain}$), similar to Ref. [86]. The RL component was
985 augmented with a selective forgetting (SF) mechanism by decaying the values of non-chosen (nc)
986 and not-presented (np) features by $\omega_{nc,t} = \omega_{np,t} = \omega^{\text{SelectiveForgetting}}$. This mechanism prioritizes
987 learning from chosen object features, which are also those that were attended when making a
988 choice, which makes the SF an attentional filtering mechanism that has been shown to be important
989 in previous studies [23,29,50]. On any given trial, each object has a probability of being chosen from
990 the RL component of the model according to a softmax function applied to the object values.
991 Specifically, the choice probability is calculated as $p_{i,t}^{RL} = \frac{\exp\left(\beta_t^{RL} \sum_{j \in O_i} V_{j,t}^F\right)}{\sum_k \exp\left(\beta_t^{RL} \sum_{j \in O_k} V_{j,t}^F\right)}$ , which uses a $\beta^{RL}$
992 parameter (the so called *inverse temperature* that we interpret as exploration/exploitation weight
993 and can vary from trial-to-trial in our model setting, see below) to weigh the sum of the values of
994 all features ($V^F$) of the presented set of objects ($O_i$). At low $\beta^{RL}$ the choice is exploratory and less
995 strictly determined by the value of objects, while at high values of $\beta^{RL}$ the choices follow the
996 highest valued object and thus are exploitative.
997
998 *Fast one-shot working memory.* Working memory of object values is implemented using static
999 parameters for the capacity and the decay of working memory content. An object will be
1000 represented in WM when it is chosen and rewarded, which is represented as an object's WM value
1001 ($V_{i,t+1}^{WM} = 1$) [3,21]. WM values $V^{WM}$ decay slowly by $decay^{WM}$ which allows its maintenance over a
1002 few trials as long as the same object is not chosen without being rewarded. When an object is

1003 chosen and not-rewarded, $V^{WM}$ is reset to the default value $\frac{1}{n_o}$, here $n_o$ is the number of presented

1004 objects. This rapid updating of the WM after rewarded and unrewarded choices of an object reflects

1005 a 'one-shot' learning mechanism that does not use reward prediction errors (**Figure 5A**). WM

1006 content contributes to the choice according to a $\beta^{WM}$ parameter determining the choice probability

1007 $p^{WM}$ of object *i* on trial *t* in WM through a softmax selection process $p_{i,t}^{WM} = \frac{\exp(\beta^{WM} V_{i,t}^{WM})}{\sum_j \exp(\beta^{WM} V_{j,t}^{WM})}$.

1008

1009 *Adaptive changes of exploration-exploitation balance.* We use an adaptive exploration (AE)

1010 mechanism that adjusts the balance of exploration and exploitation. In the RL model with the AE

1011 mechanism, a choice is predicted by a $\beta^{*RL}$ exploration parameter that is not fixed but varies over

1012 trials with the history of recent negative prediction errors. When subjects make successive errors

1013 (i.e. reward prediction errors ($\delta_t$) are negative) during learning a new feature-reward association

1014 the $\beta^{*RL}$ value is reduced to increase exploratory choices among objects [23,25]. In contrast, the $\beta^{*RL}$

1015 parameter is initially increased after correct trials (more exploitative) after which reward prediction

1016 errors start decreasing, hence the increase in $\beta^{*RL}$ attenuates. The above behavior of the time-

1017 varying $\beta^{*RL}$ is captured by weighting after each trial reward prediction errors with a magnitude

1018 that depends on its sign in terms of an auxiliary variable $\beta_t^*$ . Specifically, $\beta_{t+1}^* = \beta_t^* +$

1019 $(\alpha^{Fixed} max(\delta_t, 0) + \alpha^{ErrorTrace} min(\delta_t, 0))$, where $\alpha^{ErrorTrace}$ is a free parameter to scale the

1020 error-trace (the $\alpha^{Fixed}$ parameter is fixed, as only its relative size matters) and $\delta_t$ is the reward

1021 prediction error computed across all object features ($\delta_t = R_t - \frac{1}{\#(j \in O_i)} \sum_{j \in O_i} V_{j,t}^F$), here i is the

1022 index of the chosen object and t the trial. Both $\alpha^{ErrorTrace}$ and $\alpha^{Fixed}$ are negative, which means

1023 a positive RPE decreases $\beta_t^*$ and a negative RPE increases it. This prediction error trace is

1024 translated into an actual $\beta^{*RL}$ value by transforming it with a (negative) exploration factor

1025 $\omega^{\text{Exploration}}$ according to a sigmoid-like function $\beta_t^{*RL} = \frac{\beta_m}{1+exp\left(-\omega^{Exploration}(\beta_t^* - \omega_2)\right)}$. The $\omega_2$

1026 parameter was fixed to 0.5. When $\beta_t^*$ is positive, the denominator is larger than one, decreasing

1027 $\beta^{*RL}$, (with lower $\beta^{*RL}$ values causing more exploratory choices) whereas when it is negative the

1028 exponential is much smaller than one, hence $\beta^{*RL}$ is large (causing more exploitative choices).

1029 Taken together, successive errors lead to an increase in $\beta_t^*$, it thus more or less counts the errors,

1030 which leads via the sigmoid transformation to lower $\beta^{*RL}$ hence more exploration. Conversely,

1031 successive correct choices decrease $\beta_t^*$, which via the sigmoid transformation reduces $\beta^{*RL}$

1032 yielding more exploitation.

1033

1034 *Combining SF, AE and WM mechanisms using a strategy weight.* Combining the RL mechanisms

1035 of selective forgetting and adaptive exploration with the WM component was achieved with a trial-

1036 by-trial varying strategy weight which adjusted the relative weight of the WM choice probabilities

1037 relative to the RL choice probabilities for a final choice. Previous work has shown that learning is

1038 better predicted when both, the RL and the WM component, influence choosing objects during

1039 learning [23]. The relative influence of WM on choosing an object is determined by a weighting

1040 factor $w_t^{\frac{WM}{RL}Strategy}$ which combines the WM and RL choice probabilities into an integrated choice

1041 probability ($p^{IN}$), calculated as $p_{i,t}^{IN} = w_t^{\frac{WM}{RL}Strategy} p_{i,t}^{WM} + (1 - w_t^{WM/RL}) p_{i,t}^{RL}$. A

1042 higher $w_t^{\frac{WM}{RL}Strategy}$ corresponds to a more prominent role of the WM component in determining

1043 the choice. We refer to $w_t^{\frac{WM}{RL}Strategy}$ as learning *strategy weight*. The model increases the WM

1044 influence (by increasing $w_t^{\frac{WM}{RL}Strategy}$ ) when it has sufficient capacity to hold the value of objects

1045 and when its choice probability is high relative to the choice probability of the RL component,

which occurred transiently during the initial learning after a block switch when the expected values are only slowly adjusting while the working memory object value is updated immediately after a correct choice (**Figure 6K**,**L**).

**Selection of behavioral model.** To determine whether the AE, SF, and WM mechanisms contributed to predicting task performance of the monkeys we fit models with all combinations of these mechanisms to the choices of each of the subjects of the stimulation experiment and calculated the negative log-likelihood and the Akaike Information Criterion (AIC). We fit the model to choices of the sham and the stimulation conditions but separately for each monkey and each brain areas in order to derive the best-fitting model for each monkey and area but across all experimental conditions (**Figure 5B**). The fitting minimized the Negative Log Likelihood (NLL) using the MATLAB (The Mathworks, Inc.) function *fminsearch* followed by calls to *fmincon*.

After determining that the *AE-SF-WM model*, best described the overall data, we separately fitted the *AE-SF-WM model* to the sham and the stimulation condition for the 3D task performance of each monkey. To estimate the variability of the model parameters we used a bootstrap approach that fit the behavioral data of each condition with random subsets of 80% of the blocks from that condition, generated the sequence of choice with the bootstrapped model parameter values, and extracted the static and the trial-by-trial varying parameter values of from the model generated data. We visualize the mean and standard error of the mean of these parameters in **Figure 5C-H** and **Figure 6-L**.

**Acknowledgments**

This work was supported by the National Institute of Mental Health (R01MH123687). The funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of this manuscript.

**Author Contributions**

R.L.T., K.B.B and T.W. prepared the experiment and reconstructed the anatomical stimulation locations. R.L.T. and A.N. performed the stimulation experiments. R.L.T., C.G. and A.N. performed the electrophysiological experiments. C.G. analyzed the electrophysiological signals. R.L.T., T.W. and P.T. performed analysis, modeling and visualization. T.W. conceived the research and supervised the study. All authors contributed writing the paper.

**Data and code accessibility**

Data and custom programming code for analysis and modeling is available upon reasonable request.

**Figure 1. Task paradigm and stimulation locations**. (**A**) On each trial monkeys freely explored three stimuli until they fixated on one stimulus for longer than 0.3 s which indicated they committed choosing that stimulus. Electrical (or sham) stimulation was delivered 0.3-0.6 s after the onset of the fixation of the chosen stimulus. Stimulation was triggered either when the positive reinforced (rewarded) object was chosen (*StimReinf+)* or when a negative reinforced stimulus (resulting in loosing tokens) was chosen (*StimReinf-).* Stimulation was only delivered during learning new feature-reward associations, i.e. prior to reaching a 70% performance criterion. 0.75 s after the choice-fixation onset a halo behind the chosen object provided visual feedback (yellow/blue for correct/error choices), followed 0.5 s later by auditory feedback and the reveal of tokens gained or lost. (**B**) The task switched the rewarded feature (y-axis) in blocks of 30-50 trials. (**C**) Blocks randomly varied the motivational context from gaining +2 or +3 for choosing the object with the rewarded feature, and the loss of 1 or 3 tokens otherwise. Blocks also varied randomly

whether features from one, two, or three feature dimensions of the objects changed from trial to trial (1D, 2D, and 3D conditions). (**D**) Stimulation locations in the ACC for monkeys F / I (blue / red) in the reference frame of Procyk et al, 2016 (left) and shown on a coronal view (right). Coloring indicates the anterior-posterior distance to the interaural line. (**E**) Stimulation locations in the striatum (head of the caudate nucleus) for monkeys F / I (blue / red).

**Figure 2. Stimulation in anterior cingulate cortex impairs and in striatum improves learning in multidimensional (3D) conditions.** (**A**,**B**) Average Learning curves for blocks with stimulation (*StimReinf*+ and *StimReinf*- combined) and sham in the ACC (A) and the striatum (B) for each monkey in the 1D, 2D, and 3D condition. Vertical grey lines denote the trial at which performance reached criterion performance (first trial with 70% correct over subsequent 12 trials) in the 3D condition, which was *later* with stimulation vs sham in the ACC and *earlier* with stimulation vs sham in the striatum. (**C**) Average number of trials to reach 70% performance criterion (y-axis) for the 1D/2D/3D dimensionality conditions for the sham, *StimReinf*+, *StimReinf*- stimulation conditions in the ACC (left) and the striatum (right). Colored lines reflect both monkeys, grey lines show individual monkeys. Horizontal bars mark slower learning in *StimReinf*+ than sham condition for high dimensionality blocks. (**D-G**) Same format as C for 1D/2D/3D blocks where correct trials led to gains of +2 tokens (*D*), +3 tokens (*E*), and when erroneous choices led to a loss of 1 token (*F*) and 3 tokens (*G*). Error bars are SEM. Asterisks (*) indicate FDR-corrected $p < 0.05$ using t-tests with both monkeys pooled.

**Figure 3. The majority of stimulation sessions in the ACC impairs learning and in the striatum improves learning.** (**A**) The scatterplot shows for each experimental session in the ACC the difference in learning speed (trial-to-criterion) with stimulation vs sham in the 3D condition (*y-axis*) and the 1D condition (*x-axis*). Black outlined markers denote sessions with individually significant difference of stim vs. sham. Blue /Red data are from subject F/I. Errors bars are STD. Solid point. is mean. Histogram summarized the effects for 1D (upper) and 3D (rightmost) condition. Values <0 shows slower learning with stimulation than sham. (**B**) Pie charts summarizing the proportion of sessions in which stimulation improved (green) or impaired (yellow) learning for each subject. (**C**,**D**) Same format as (*A*,*C*) for experiments in the striatum.

**Figure 4. Stimulation modulates behavior following gains in the 3D condition.** (**A**) Average accuracy (*y-axis*) for trials before and after a correct trial that gained subjects tokens for blocks with stimulation in the ACC on correct trials (*StimReinf*+) and sham in the 3D condition for trials until the mean trials criterion in the ACC / *StimReinf*+ condition. *'s denote FDR-corrected $p<0.05$ (permutation test) significant differences of stim vs sham. The p-value at bottom refers to a significant coefficient using the Wald test after stimulation and sham conditions were fitted to a logistic regression model, produced from a generalized linear model with a binomial distribution and logit link function (see Methods). (**B**) Same as (*A)* for accuracy before and after incorrect trials that lost tokens. (**C**,**D**). Same format as (*A*,*B*) for experimental sessions with *StimReinf*- stimulation and sham. (**E-H**). Same format as (*A-D*) for experimental sessions in the striatum with *StimReinf*+ stimulation (*E*,*F*), and with *StimReinf*- stimulation (*G*,*H*).

**Figure 5. A triple-strategy model accounts for learning at high (3D) feature uncertainty.** (**A**) Overview of the triple-strategy model combining adaptive exploration / selective forgetting / working memory (AE-SF-WM). The model learns to make choices of objects by a combination of (i) working memory (WM) and by using expected feature values of an RL module, which uses (ii) an adaptive mechanism to adjust exploration during learning and (iii) a selective forgetting of values from non-chosen features. (**B**) The AE-SF-WM model best fit choices in the effective high (3D) feature-uncertainty condition across sessions and stimulation / sham conditions in the ACC (blue) and striatum (orange) for both monkeys (solid and dashed line). The lowest normalized Akaike Information Criterion corresponds to the best model. (**C,D**) The effect size (absolute value of t-statistics) and Bonferroni corrected p-values for the difference of parameter values of the model fit separately to the sham and stimulation conditions in the 3D condition in ACC (*C*) and in the striatum (*D*). The parameters are defined in Methods. (**E-H**) Change in fitting parameters between stimulation and sham condition for ACC (blue) and striatum (orange) for both monkeys (F dot-dashed; I solid line) for (E) the adaptive exploration-exploitation value $\beta^{*RL}$ , (F) the error trace weight $\alpha^{ErrorTrace}$, (G) the working memory softmax constant $\beta^{WM}$ and (H) the strategy weight $\omega^{\frac{WM}{RL}Strategy}$ . Stars in panel (C-H) denote standard significance levels.

**Figure 6. Stimulation has opposite effects in ACC and striatum on reinforcement learning and adaptive exploration.** (**A**,**B**) Visual summary of the model mechanisms with coloring reflecting modulation with microstimulation in the ACC (*A*) and the striatum (*B*). See Figure 5A for the full model. (**C**,**D**) Trial-by-trial changes in behavioral accuracy for each monkey in ACC (*C*) and the striatum (*D*) when monkeys learned choices in the 3D condition in sham blocks (black lines) and stimulation blocks (colored lines). Analysis used the seventeen ACC sessions and seventeen striatum sessions in which stimulation affected behavior. (**E-L**) Same format as (*A,B*) showing estimated values from the AE-SF-WM model for the choice probability of the RL component (*E,F*), the positive prediction errors (*G,H*), the adaptive exploration-exploitation value $\beta^{*RL}$ (*I,J*), and the strategy weight $\omega^{\frac{WM}{RL}Strategy}$ (*K,L*). Error shading is SD across fifty bootstrap model fits.

**Figure 7. Encoding of behavioral and latent model variables with neuronal activity differs in the ACC and striatum.** (**A,B**) The proportion of recording channels with multiunit activity in the ACC (*A*) and striatum (*B*) significantly correlating trial-by-trial with outcomes (0=unrewarded,1=rewarded) relative to the onset of fixating the chosen object. Green lines denote the proportion for channels with positive correlations, red lines denote proportions of channels with negative correlations. Grey shaded area is the time period when the subjects committed to choosing the fixated object (at 0.3 s) and when the choice is registered (0.7 s). This choice period was used for electrical stimulation. MUA activity was smoothed with a 0.2 gaussian window prior to the correlation with the variables. Correlations were performed every 25 ms. (**C**) Average proportion of channels with significant positive (green) and negative (red) correlations in ACC (*left*) and the striatum (*right*). Error bars are SDs, horizontal bars and stars denote proportional difference significance level. (**D-R**) Same format as *A-C* for neural activity correlations with choice probability of the chosen object (*D-F*), the expected values of the chosen object (*G-I*), the

adaptive exploration-exploitation variable $\beta^{*RL}$ (*J-L*), the error-trace weight variable (*M-O*), and trial-by-trial fluctuations with the WM/RL strategy weight (*P-R*).

## References


1. Soltani, A., and Koechlin, E. (2022). Computational models of adaptive behavior and prefrontal cortex. Neuropsychopharmacology *47*, 58-71. 10.1038/s41386-021-01123-1.
2. Rmus, M., McDougle, S.D., and Collins, A.G.E. (2021). The role of executive function in shaping reinforcement learning. Curr Opin Behav Sci *38*, 66-73. 10.1016/j.cobeha.2020.10.003.
3. Collins, A.G.E., Ciullo, B., Frank, M.J., and Badre, D. (2017). Working Memory Load Strengthens Reward Prediction Errors. J Neurosci *37*, 4332-4342. 10.1523/JNEUROSCI.2700-16.2017.
4. Yoo, A.H., and Collins, A.G.E. (2022). How Working Memory and Reinforcement Learning Are Intertwined: A Cognitive, Neural, and Computational Perspective. J Cognitive Neurosci *34*, 551-568. 10.1162/jocn_a_01808.
5. Farashahi, S., Donahue, C.H., Khorsand, P., Seo, H., Lee, D., and Soltani, A. (2017). Metaplasticity as a Neural Substrate for Adaptive Learning and Choice under Uncertainty. Neuron *94*, 401-414 e406. 10.1016/j.neuron.2017.03.044.
6. Viejo, G., Girard, B., Procyk, E., and Khamassi, M. (2018). Adaptive coordination of working-memory and reinforcement learning in non-human primates performing a trial-and-error problem solving task. Behav Brain Res *355*, 76-89. 10.1016/j.bbr.2017.09.030.
7. Grossman, C.D., Bari, B.A., and Cohen, J.Y. (2022). Serotonin neurons modulate learning rate through uncertainty. Curr Biol *32*, 586-599 e587. 10.1016/j.cub.2021.12.006.

8. Behrens, T.E., Woolrich, M.W., Walton, M.E., and Rushworth, M.F. (2007). Learning the value of information in an uncertain world. Nat Neurosci *10*, 1214-1221.
9. White, J.K., Bromberg-Martin, E.S., Heilbronner, S.R., Zhang, K., Pai, J., Haber, S.N., and Monosov, I.E. (2019). A neural network for information seeking. Nat Commun *10*, 5168. 10.1038/s41467-019-13135-z.
10. Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci *8*, 1704-1711.
11. Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., and Rushworth, M.F. (2006). Optimal decision making and the anterior cingulate cortex. Nat Neurosci *9*, 940-947.
12. Fouragnan, E.F., Chau, B.K.H., Folloni, D., Kolling, N., Verhagen, L., Klein-Flugge, M., Tankelevitch, L., Papageorgiou, G.K., Aubry, J.F., Sallet, J., and Rushworth, M.F.S. (2019). The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. Nat Neurosci *22*, 797-808. 10.1038/s41593-019-0375-6.
13. Amiez, C., Joseph, J.P., and Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. Cereb Cortex *16*, 1040-1055. 10.1093/cercor/bhj046.
14. Buckley, M.J., Mansouri, F.A., Hoda, H., Mahboubi, M., Browning, P.G., Kwok, S.C., Phillips, A., and Tanaka, K. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science *325*, 52-58.
15. Chudasama, Y., Daniels, T.E., Gorrin, D.P., Rhodes, S.E., Rudebeck, P.H., and Murray, E.A. (2013). The role of the anterior cingulate cortex in choices based on reward value and reward contingency. Cereb Cortex *23*, 2884-2898. 10.1093/cercor/bhs266.
16. Mansouri, F.A., Buckley, M.J., Fehring, D.J., and Tanaka, K. (2020). The Role of Primate Prefrontal Cortex in Bias and Shift Between Visual Dimensions. Cereb Cortex *30*, 85-99. 10.1093/cercor/bhz072.
17. Banaie Boroujeni, K., Sigona, M.K., Treuting, R.L., Manuel, T.J., Caskey, C.F., and Womelsdorf, T. (2022). Anterior cingulate cortex causally supports flexible learning under motivationally challenging and cognitively demanding conditions. PLoS Biol *20*, e3001785. 10.1371/journal.pbio.3001785.
18. Klein-Flugge, M.C., Bongioanni, A., and Rushworth, M.F.S. (2022). Medial and orbital frontal cortex in decision-making and flexible behavior. Neuron *110*, 2743-2770. 10.1016/j.neuron.2022.05.022.
19. Kolling, N., Wittmann, M.K., Behrens, T.E., Boorman, E.D., Mars, R.B., and Rushworth, M.F. (2016). Value, search, persistence and model updating in anterior cingulate cortex. Nat Neurosci *19*, 1280-1285. 10.1038/nn.4382.
20. Walton, M.E., and Mars, R.B. (2007). Probing human and monkey anterior cingulate cortex in variable environments. Cogn Affect Behav Neurosci *7*, 413-422. 10.3758/cabn.7.4.413.
21. Collins, A.G., and Frank, M.J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci *35*, 1024-1035. 10.1111/j.1460-9568.2011.07980.x.
22. Voytek, B., and Knight, R.T. (2010). Prefrontal cortex and basal ganglia contributions to visual working memory. Proc Natl Acad Sci U S A *107*, 18167-18172. 10.1073/pnas.1007277107.
23. Womelsdorf, T., Watson, M.R., and Tiesinga, P. (2021). Learning at Variable Attentional Load Requires Cooperation of Working Memory, Meta-learning, and Attention-augmented Reinforcement Learning. J Cogn Neurosci *34*, 79-107. 10.1162/jocn_a_01780.
24. Viejo, G., Khamassi, M., Brovelli, A., and Girard, B. (2015). Modeling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning. Front Behav Neurosci *9*, 225. 10.3389/fnbeh.2015.00225.
25. Khamassi, M., Enel, P., Dominey, P.F., and Procyk, E. (2013). Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. Progress in brain research *202*, 441-464. 10.1016/B978-0-444-62604-2.00022-8.

26. Soltani, A., and Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. Nat Rev Neurosci *20*, 635-644. 10.1038/s41583-019-0180-y.
27. Khamassi, M., Quilodran, R., Enel, P., Dominey, P.F., and Procyk, E. (2015). Behavioral Regulation and the Modulation of Information Coding in the Lateral Prefrontal and Cingulate Cortex. Cereb Cortex *25*, 3197-3218. 10.1093/cercor/bhu114.
28. Banaie Boroujeni, K., Oemisch, M., Hassani, S.A., and Womelsdorf, T. (2020). Fast spiking interneuron activity in primate striatum tracks learning of attention cues. Proc Natl Acad Sci U S A *117*, 18049-18058. 10.1073/pnas.2001348117.
29. Oemisch, M., Westendorff, S., Azimi, M., Hassani, S.A., Ardid, S., Tiesinga, P., and Womelsdorf, T. (2019). Feature-specific prediction errors and surprise across macaque fronto-striatal circuits. Nature Communications *10*.
30. Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F., and Peters, J. (2020). Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. Elife *9*. 10.7554/eLife.51260.
31. Santacruz, S.R., Rich, E.L., Wallis, J.D., and Carmena, J.M. (2017). Caudate Microstimulation Increases Value of Specific Choices. Curr Biol *27*, 3375-3383 e3373. 10.1016/j.cub.2017.09.051.
32. Williams, Z.M., and Eskandar, E.N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. Nat Neurosci *9*, 562-568. 10.1038/nn1662.
33. Procyk, E., Wilson, C.R., Stoll, F.M., Faraut, M.C., Petrides, M., and Amiez, C. (2016). Midcingulate Motor Map and Feedback Detection: Converging Data from Humans and Monkeys. Cereb Cortex *26*, 467-476. 10.1093/cercor/bhu213.
34. Hutchison, R.M., Womelsdorf, T., Gati, J.S., Leung, L.S., Menon, R.S., and Everling, S. (2012). Resting-state connectivity identifies distinct functional networks in macaque cingulate cortex. Cerebral cortex *22*, 1294-1308. 10.1093/cercor/bhr181.
35. Banaie Boroujeni, K., and Womelsdorf, T. (2023). Routing states transition during oscillatory bursts and attentional selection. Neuron *111*, 2929-2944 e2911. 10.1016/j.neuron.2023.06.012.
36. Kaping, D., Vinck, M., Hutchison, R.M., Everling, S., and Womelsdorf, T. (2011). Specific contributions of ventromedial, anterior cingulate, and lateral prefrontal cortex for attentional selection and stimulus valuation. PLoS Biol *9*, e1001224. 10.1371/journal.pbio.1001224.
37. Kennerley, S.W., and Wallis, J.D. (2009). Reward-dependent modulation of working memory in lateral prefrontal cortex. J Neurosci *29*, 3259-3270.
38. Kennerley, S.W., Behrens, T.E., and Wallis, J.D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nature neuroscience *14*, 1581-1589. 10.1038/nn.2961.
39. Kawai, T., Yamada, H., Sato, N., Takada, M., and Matsumoto, M. (2019). Preferential Representation of Past Outcome Information and Future Choice Behavior by Putative Inhibitory Interneurons Rather Than Putative Pyramidal Neurons in the Primate Dorsal Anterior Cingulate Cortex. Cerebral Cortex *29*, 2339-2352. 10.1093/cercor/bhy103.
40. Quilodran, R., Rothe, M., and Procyk, E. (2008). Behavioral shifts and action valuation in the anterior cingulate cortex. Neuron *57*, 314-325. 10.1016/j.neuron.2007.11.031.
41. Hayden, B.Y., Heilbronner, S.R., Pearson, J.M., and Platt, M.L. (2011). Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J Neurosci *31*, 4178-4187.
42. Lau, B., and Glimcher, P.W. (2008). Value representations in the primate striatum during matching behavior. Neuron *58*, 451-463. 10.1016/j.neuron.2008.02.021.
43. Peng, X., Trambaiolli, L.R., Choi, E.Y., Lehman, J.F., Linn, G., Russ, B.E., Schroeder, C.E., Haber, S.N., and Liu, H. (2024). Cross-species striatal hubs: Linking anatomy to resting-state connectivity. Neuroimage *301*, 120866. 10.1016/j.neuroimage.2024.120866.
44. Tang, W., Jbabdi, S., Zhu, Z., Cottaar, M., Grisot, G., Lehman, J.F., Yendiki, A., and Haber, S.N. (2019). A connectional hub in the rostral anterior cingulate cortex links areas of emotion and cognitive control. Elife *8*. 10.7554/eLife.43761.

45. Borra, E., Biancheri, D., Rizzo, M., Leonardi, F., and Luppino, G. (2022). Crossed Corticostriatal Projections in the Macaque Brain. J Neurosci *42*, 7060-7076. 10.1523/JNEUROSCI.0071-22.2022.
46. Voloh, B., Oemisch, M., and Womelsdorf, T. (2020). Phase of firing coding of learning variables across the fronto-striatal network during feature-based learning. Nature Communications *11*.
47. Amemori, K., and Graybiel, A.M. (2012). Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. Nature neuroscience *15*, 776-785. 10.1038/nn.3088.
48. Amemori, K.I., Amemori, S., Gibson, D.J., and Graybiel, A.M. (2018). Striatal Microstimulation Induces Persistent and Repetitive Negative Decision-Making Predicted by Striatal Beta-Band Oscillation. Neuron *99*, 829-841 e826. 10.1016/j.neuron.2018.07.022.
49. Cox, J., and Witten, I.B. (2019). Striatal circuits for reward learning and decision-making. Nat Rev Neurosci *20*, 482-494. 10.1038/s41583-019-0189-2.
50. Niv, Y., Daniel, R., Geana, A., Gershman, S.J., Leong, Y.C., Radulescu, A., and Wilson, R.C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. J Neurosci *35*, 8145-8157. 10.1523/JNEUROSCI.2978-14.2015.
51. Khamassi, M., Wilson, C.R.E., Rothé, M., Quilodran, R., Dominey, P.F., and Procyk, E. (2011). Meta-Learning, Cognitive Control, and Physiological Interactions between Medial and Lateral Prefrontal Cortex. In Neural Basis of Motivational and Cognitive Control, R.B. Mars, J. Sallet, M.F.S. Rushworth, and N. Yeung, eds. (MIT Press), pp. 351.
52. Butovas, S., and Schwarz, C. (2003). Spatiotemporal effects of microstimulation in rat neocortex: a parametric study using multielectrode recordings. J Neurophysiol *90*, 3024-3039. 10.1152/jn.00245.2003.
53. Seidemann, E., Arieli, A., Grinvald, A., and Slovin, H. (2002). Dynamics of depolarization and hyperpolarization in the frontal cortex and saccade goal. Science *295*, 862-865. 10.1126/science.1066641.
54. Tehovnik, E.J. (1996). Electrical stimulation of neural tissue to evoke behavioral responses. J Neurosci Methods *65*, 1-17. 10.1016/0165-0270(95)00131-x.
55. Card, N.S., and Gharbawie, O.A. (2020). Principles of Intrinsic Motor Cortex Connectivity in Primates. J Neurosci *40*, 4348-4362. 10.1523/JNEUROSCI.0003-20.2020.
56. Rakhshan, M., Schafer, R.J., Moore, T., and Soltani, A. (2025). Neural Mechanisms Underlying Robust Target Selection in Response to Microstimulation of the Oculomotor System. J Neurosci *45*. 10.1523/JNEUROSCI.2356-23.2024.
57. Gold, J.I., and Shadlen, M.N. (2000). Representation of a perceptual decision in developing oculomotor commands. Nature *404*, 390-394. 10.1038/35006062.
58. Hanks, T.D., Ditterich, J., and Shadlen, M.N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. Nat Neurosci *9*, 682-689. 10.1038/nn1683.
59. Murd, C., Moisa, M., Grueschow, M., Polania, R., and Ruff, C.C. (2020). Causal contributions of human frontal eye fields to distinct aspects of decision formation. Sci Rep *10*, 7317. 10.1038/s41598-020-64064-7.
60. Schafer, R.J., and Moore, T. (2007). Attention governs action in the primate frontal eye field. Neuron *56*, 541-551. 10.1016/j.neuron.2007.09.029.
61. Berdyyeva, T.K., and Olson, C.R. (2014). Intracortical microstimulation of supplementary eye field impairs ability of monkeys to make serially ordered saccades. J Neurophysiol *111*, 1529-1540. 10.1152/jn.00503.2013.
62. Histed, M.H., and Miller, E.K. (2006). Microstimulation of frontal cortex can reorder a remembered spatial sequence. PLoS Biol *4*, e134. 10.1371/journal.pbio.0040134.
63. Stuphorn, V., and Schall, J.D. (2006). Executive control of countermanding saccades by the supplementary eye field. Nat Neurosci *9*, 925-931. 10.1038/nn1714.
64. Nacher, V., Hassani, S.A., and Womelsdorf, T. (2019). Asymmetric effective connectivity between primate anterior cingulate and lateral prefrontal cortex revealed by electrical microstimulation. Brain Struct Funct *224*, 779-793. 10.1007/s00429-018-1806-y.

65. Monosov, I.E., and Rushworth, M.F.S. (2022). Interactions between ventrolateral prefrontal and anterior cingulate cortex during learning and behavioural change. Neuropsychopharmacology *47*, 196-210. 10.1038/s41386-021-01079-2.
66. Alexander, W.H., and Brown, J.W. (2011). Medial prefrontal cortex as an action-outcome predictor. Nature neuroscience *14*, 1338-1344. 10.1038/nn.2921.
67. Rudebeck, P.H., Saunders, R.C., Lundgren, D.A., and Murray, E.A. (2017). Specialized Representations of Value in the Orbital and Ventrolateral Prefrontal Cortex: Desirability versus Availability of Outcomes. Neuron *95*, 1208-1220 e1205. 10.1016/j.neuron.2017.07.042.
68. Murray, E.A., and Rudebeck, P.H. (2018). Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. Nat Rev Neurosci *19*, 404-417. 10.1038/s41583-018-0013-4.
69. Bichot, N.P., Heard, M.T., DeGennaro, E.M., and Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. Neuron *88*, 832-844. 10.1016/j.neuron.2015.10.001.
70. Bichot, N.P., Xu, R., Ghadooshahy, A., Williams, M.L., and Desimone, R. (2019). The role of prefrontal cortex in the control of feature attention in area V4. Nat Commun *10*, 5727. 10.1038/s41467-019-13761-7.
71. Oemisch, M., Westendorff, S., Everling, S., and Womelsdorf, T. (2015). Interareal Spike-Train Correlations of Anterior Cingulate and Dorsal Prefrontal Cortex during Attention Shifts. J Neurosci *35*, 13076-13089. 10.1523/JNEUROSCI.1262-15.2015.
72. van der Meer, M., Kurth-Nelson, Z., and Redish, A.D. (2012). Information processing in decision-making systems. Neuroscientist *18*, 342-359. 10.1177/1073858411435128.
73. Hikosaka, O., Ghazizadeh, A., Griggs, W., and Amita, H. (2018). Parallel basal ganglia circuits for decision making. J Neural Transm (Vienna) *125*, 515-529. 10.1007/s00702-017-1691-1.
74. Arcizet, F., and Krauzlis, R.J. (2018). Covert spatial selection in primate basal ganglia. PLoS Biol *16*, e2005930. 10.1371/journal.pbio.2005930.
75. Shenhav, A., and Botvinick, M.M. (2013). Motivated action: new light on prefrontal-neuromodulatory circuits. Current biology : CB *23*, R161-163. 10.1016/j.cub.2012.12.028.
76. Hunt, L.T., Malalasekera, W.M.N., de Berker, A.O., Miranda, B., Farmer, S.F., Behrens, T.E.J., and Kennerley, S.W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. Nature Neuroscience *21*, 1471-+. 10.1038/s41593-018-0239-5.
77. Barbas, H. (2000). Complementary roles of prefrontal cortical regions in cognition, memory, and emotion in primates. Adv Neurol *84*, 87-110.
78. Barbas, H., and Zikopoulos, B. (2007). The prefrontal cortex and flexible behavior. Neuroscientist *13*, 532-545. 13/5/532.
79. Silvetti, M., Lasaponara, S., Daddaoua, N., Horan, M., and Gottlieb, J. (2023). A Reinforcement Meta-Learning framework of executive function and information demand. Neural Netw *157*, 103-113. 10.1016/j.neunet.2022.10.004.
80. Silvetti, M., Vassena, E., Abrahamse, E., and Verguts, T. (2018). Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. PLoS Comput Biol *14*, e1006370. 10.1371/journal.pcbi.1006370.
81. Watson, M.R., Traczewski, N., Dunghana, S., Boroujeni, K.B., Neumann, A., Wen, X., and Womelsdorf, T. (2023). A Multi-task Platform for Profiling Cognitive and Motivational Constructs in Humans and Nonhuman Primates. bioRxiv. 10.1101/2023.11.09.566422.
82. Womelsdorf, T., Thomas, C., Neumann, A., Watson, M.R., Banaie Boroujeni, K., Hassani, S.A., Parker, J., and Hoffman, K.L. (2021). A Kiosk Station for the Assessment of Multiple Cognitive Domains and Cognitive Enrichment of Monkeys. Front Behav Neurosci *15*, 721069. 10.3389/fnbeh.2021.721069.
83. Watson, M.R., Voloh, B., Naghizadeh, M., and Womelsdorf, T. (2019). Quaddles: A multidimensional 3-D object set with parametrically controlled and customizable features. Behav Res Methods *51*, 2522-2532. 10.3758/s13428-018-1097-5.

84. Boroujeni, K.B., Watson, M., and Womelsdorf, T. (2022). Gains and Losses Affect Learning Differentially at Low and High Attentional Load. J Cogn Neurosci *34*, 1952-1971. 10.1162/jocn_a_01885.
85. Voloh, B., Watson, M.R., Konig, S., and Womelsdorf, T. (2020). MAD saccade: statistically robust saccade threshold estimation via the median absolute deviation. J Eye Mov Res *12*. 10.16910/jemr.12.8.3.
86. Frank, M.J., Seeberger, L.C., and O'Reilly R, C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science *306*, 1940-1943. 10.1126/science.1102941.