Supplemental Information Table of Contents

SUPPLEMENTAL FIGURES 3

SUPPLEMENTAL EXPERIMENTAL PROCEDURES 15

Surgical and Implant Procedures 15

Recording Procedures 16

Behavioral Task 17

Behavioral Data Analysis 19

Saccade Entropy 19

Correlation between Cost and Saccade Pattern Selection 19

Alternative Definitions of Cost 20

Collinearity Diagnostics 21

Final Eye Position Control 22

Unit Data Analysis 22

Striatal Subtype Classification 23

Changes in Activity Patterns of Striatal Neurons across Learning 24

Regression Model Categories in G4 and Y4 25

Alternative Regression Models 25

IQR Correlation with Sessions in G4 and Y4 26

Depth Controls in Y9 26

IQR Correlation with Entropy in G4 and Y4 27

Separating Effects of Entropy and Session in Correlations with IQR 28

SUPPLEMENTAL REFERENCES 30

SUPPLEMENTAL FIGURES

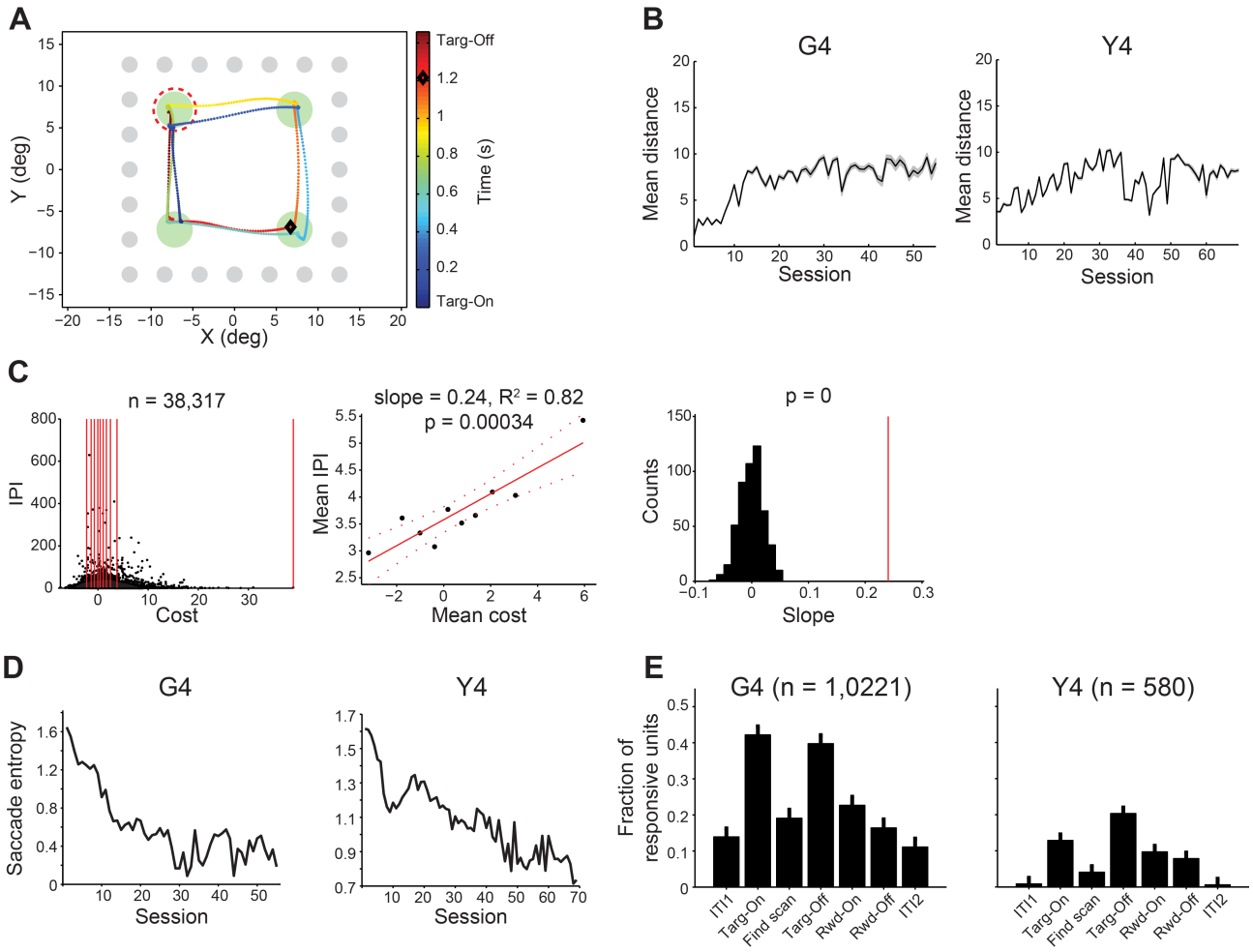

Figure S1. Behavior and Neural Correlates of Behavior in the 4-Target Task, Related to Figure 1
(**A-E**) Same as **Figure 1A** middle, **1C**, **1D**, **1E**, and **1F**, respectively, but for the 4-target task.

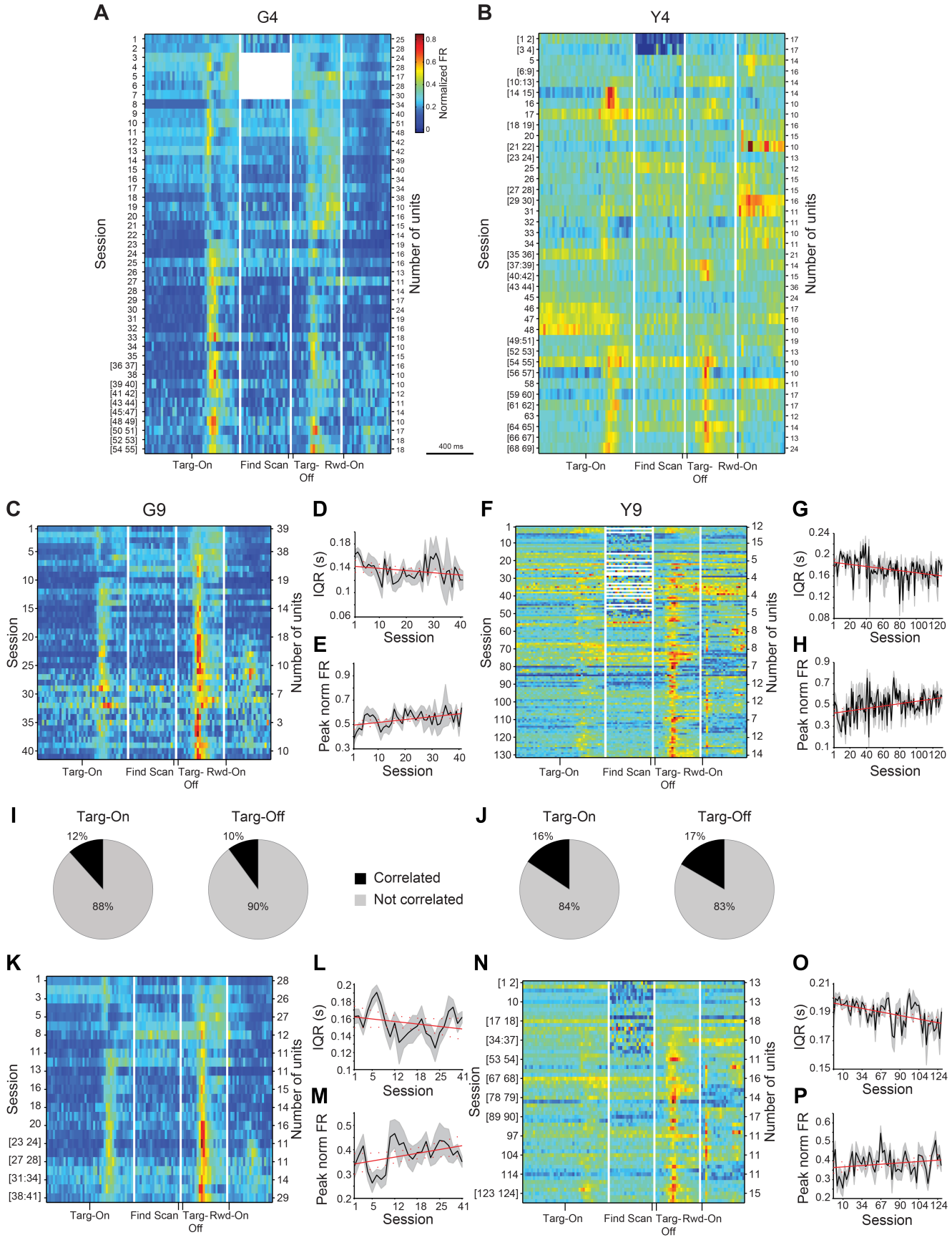

Figure S2. Changes in Activity Patterns of Striatal Neurons across Learning, Related to Figure 2

(**A** and **B**) Activity of all units recorded in monkeys G (**A**) and Y (**B**) across training on the 4-target task, plotted as in **Figure 2**. White sections during 0.4 s before Find Scan indicate that there were no trials in those sessions with sufficient time between Targ-On and Find Scan to display activity that did not overlap with the Targ-On window (−0.4 to 0.4 s).
(**C** and **F**) Activity of all units across training on the 9-target task in monkeys G (**C**) and Y (**F**), plotted as in **A** and **B**, except that units are not binned across sessions, and each row represents a single session.
(**D** and **G**) Response sharpness in the Targ-Off window, as measured by IQR and calculated for single sessions, increases across sessions in monkey G (**D**) and monkey Y (**G**).
(**E** and **H**) Decreases in the peak firing of all units across sessions in the Targ-Off window for monkeys G (**E**) and Y (**H**). The peak firing was calculated for single sessions as the mean normalized firing rate in the center two quartiles around the median firing time. The mean value in each bin is plotted across sessions (black; ± SEM, gray). The regression line (red) and confidence intervals (red dashed) are also plotted across sessions.
(**I** and **J**) Percentage of units with correlation between spike times and eye movement times in the 400 ms after Targ-On (left of pair) and in the 400 ms after Targ-Off (right of pair), recorded in G9 (**I**) and Y9 (**J**) across sessions.
(**K** and **N**) Activity of units not correlated with eye movements in Targ-On or Targ-Off windows across training in G9 (**K**) and Y9 (**N**), plotted as in **A** and **B**, binned across sessions such that each bin had ≥ 10 units. Session numbers are representative of the binning on every 2nd (**K**) or 5th (**N**) line.
(**L, M, O** and **P**) Same as **D**, **E**, **G and H** but during the Targ-On window and binned across sessions and only for those units not correlated with eye movements (same data as in **K** and **N**).

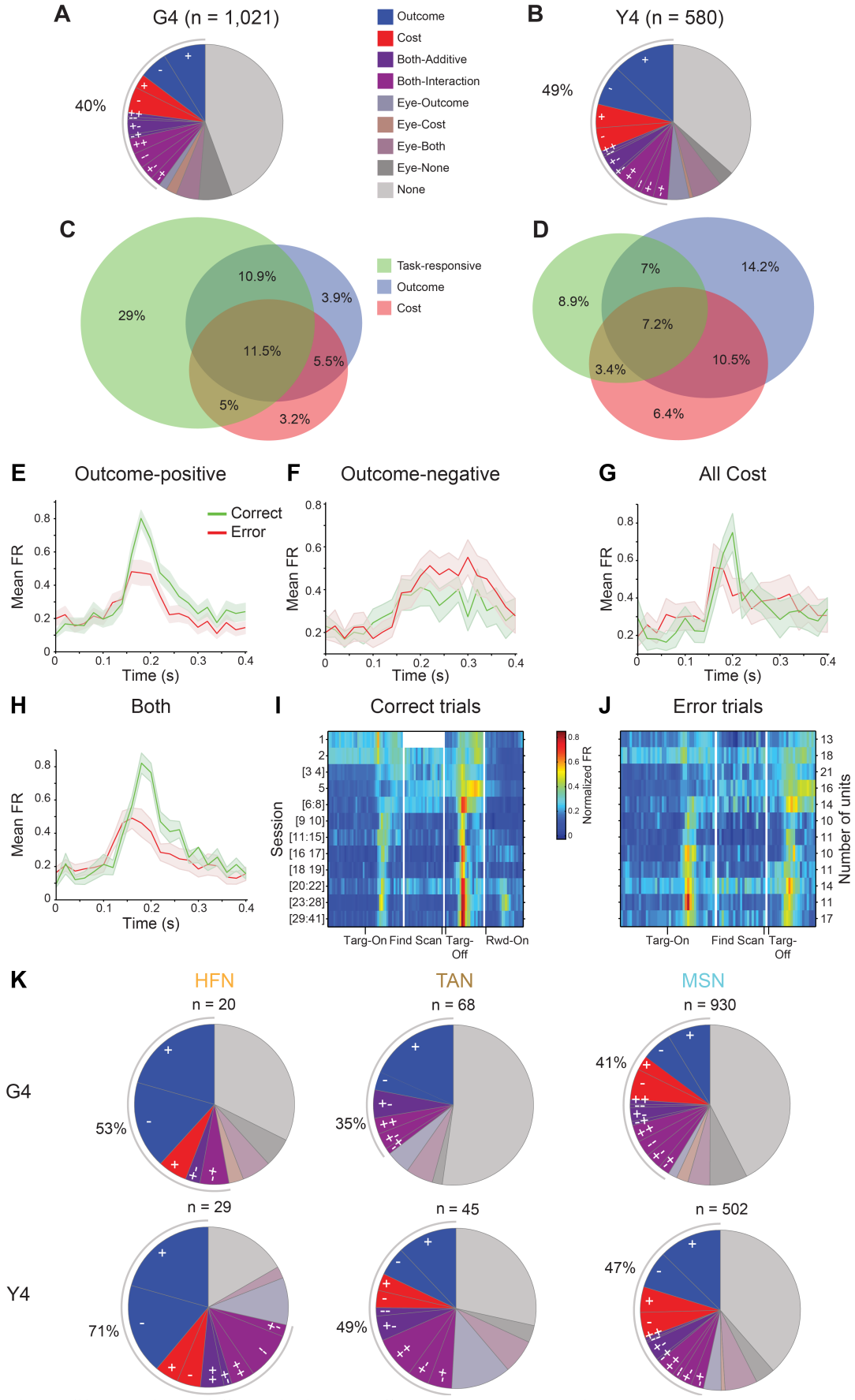

Figure S3. Fractions of CN Units in Each Category, Related to Figure 5

(**A** and **B**) Percentage of units in each category across sessions for G4 (**A**) and Y4 (**B**), plotted as in **Figures 3A** and **3B**.

(**C** and **D**) Venn diagrams of number of each unit type, plotted as in **Figures 3C** and **3D**.
(**E-H**) Mean (± SEM) firing rate in the 400 ms after Targ-Off for both correct (green) and error (red) trials (20 ms bins) at the end of training (last session bin containing ≥ 10 units) in G9 in all three unit types: Outcome units with positive (E) and negative (F) correlations, Cost units (**G**), and Both units (**H**).
(**I** and **J**) Activity of all Both units in G9 across sessions plotted as in **Figures 2A, 2B, S2A** and **S2B** for correct (**I**) and error (**J**) trials. Note that error trials do not contain the Rwd-On event, and therefore this event is omitted from the plot.
(**K**) HFN, TAN, and MSN classification in G4 (top) and Y4 (bottom). Pie charts are displayed as in **A** and **B**, but separately for putative HFNs (left), TANs (middle) and MSNs (right).

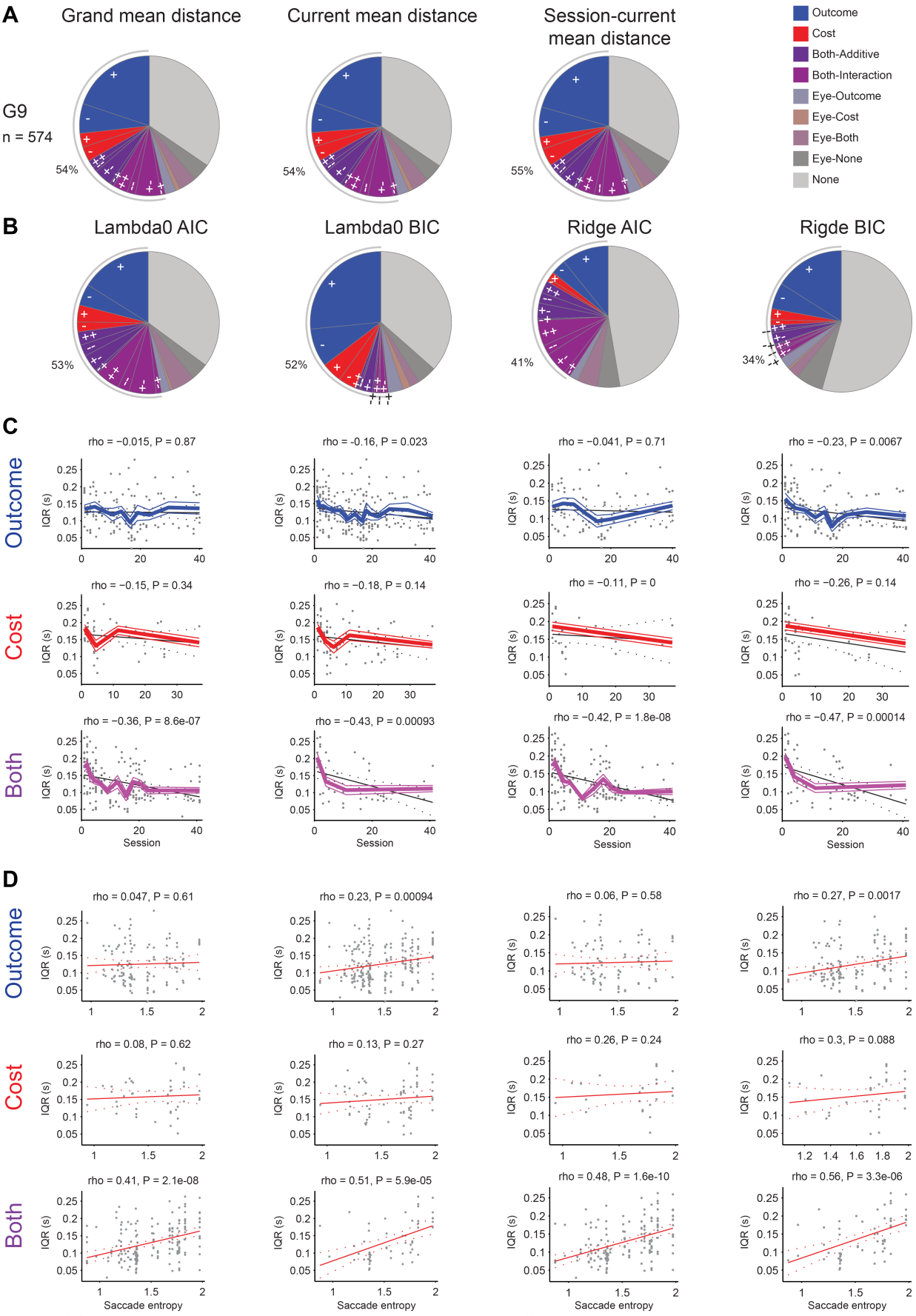

Figure S4. Results of Analyses from Alternate Regressions in G9, Related to Figure 7

(**A**) Proportions of CN units in each category, plotted as in **Figures 5** and **S3**. Units were classified using three different definitions of "cost" in separate stepwise regressions: trial distance minus mean distance across all trials on all sessions (left), trial distance minus mean distance on all trials (including previous sessions) up to that trial (middle), or trial distance minus mean distance on trials in the current session up to that trial (right).
(**B**) Proportions of unit types classified using four different regression types on the same data (using the standard definition of cost): lambda0 AIC, lambda0 BIC, ridge AIC and ridge BIC (from left to right).
(**C**) Correlations between IQR and the amount of training as in **Figures 7C-7H**. Columns correspond to those in **B**.
(**D**) Correlations between IQR and saccade entropy as in with the same rows as in **C** and the same columns as in **B**. Each point on the plot represents one unit recorded in one session. Red lines represent the correlation line fit (solid) and the 95% confidence intervals (dashed). Correlation coefficient and p-value of the correlation are indicated for each plot.

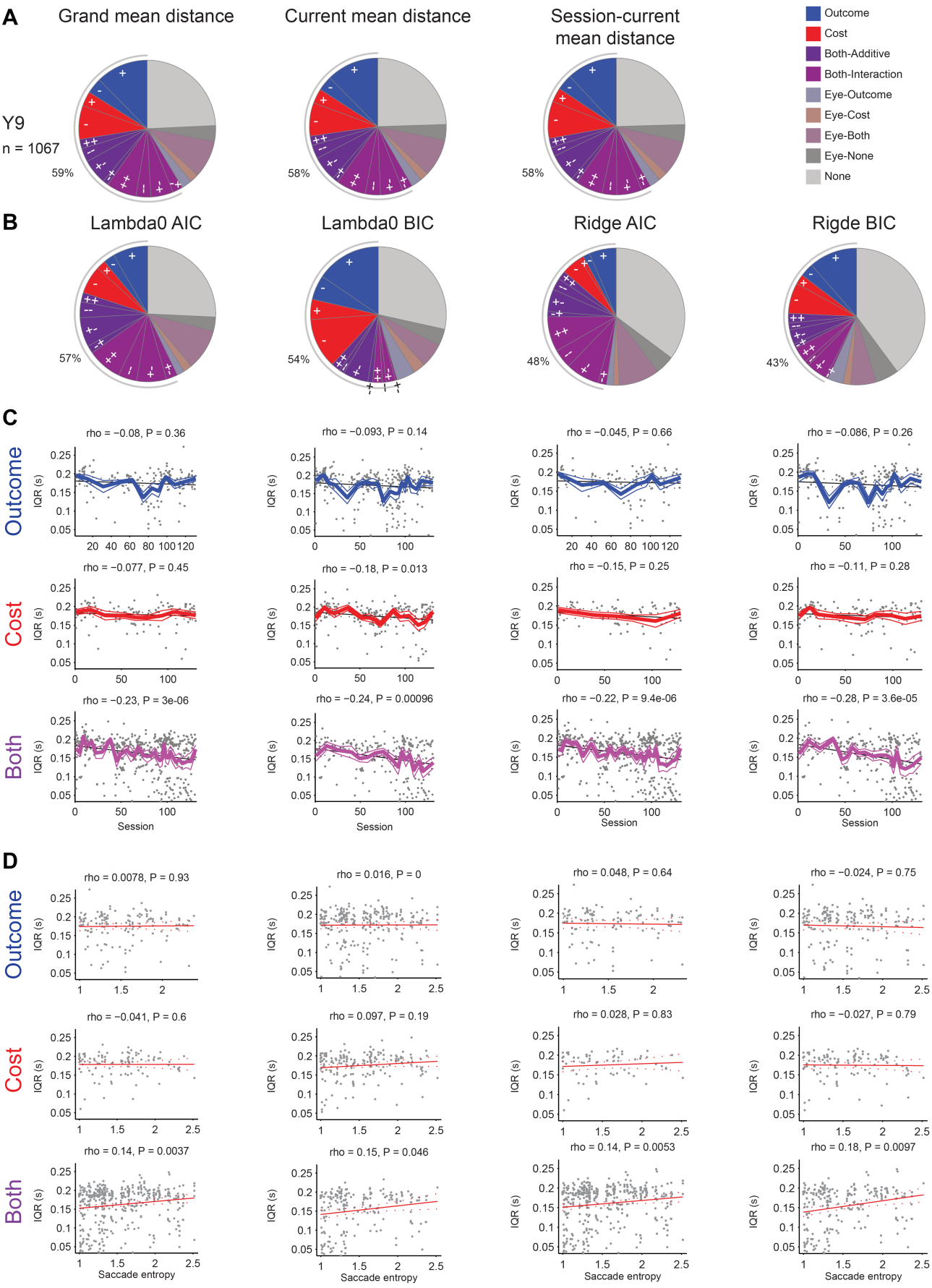

Figure S5. Results of Analyses from Alternate Regressions in Y9, Related to Figure 7
Results are plotted as in **Figure S4**.

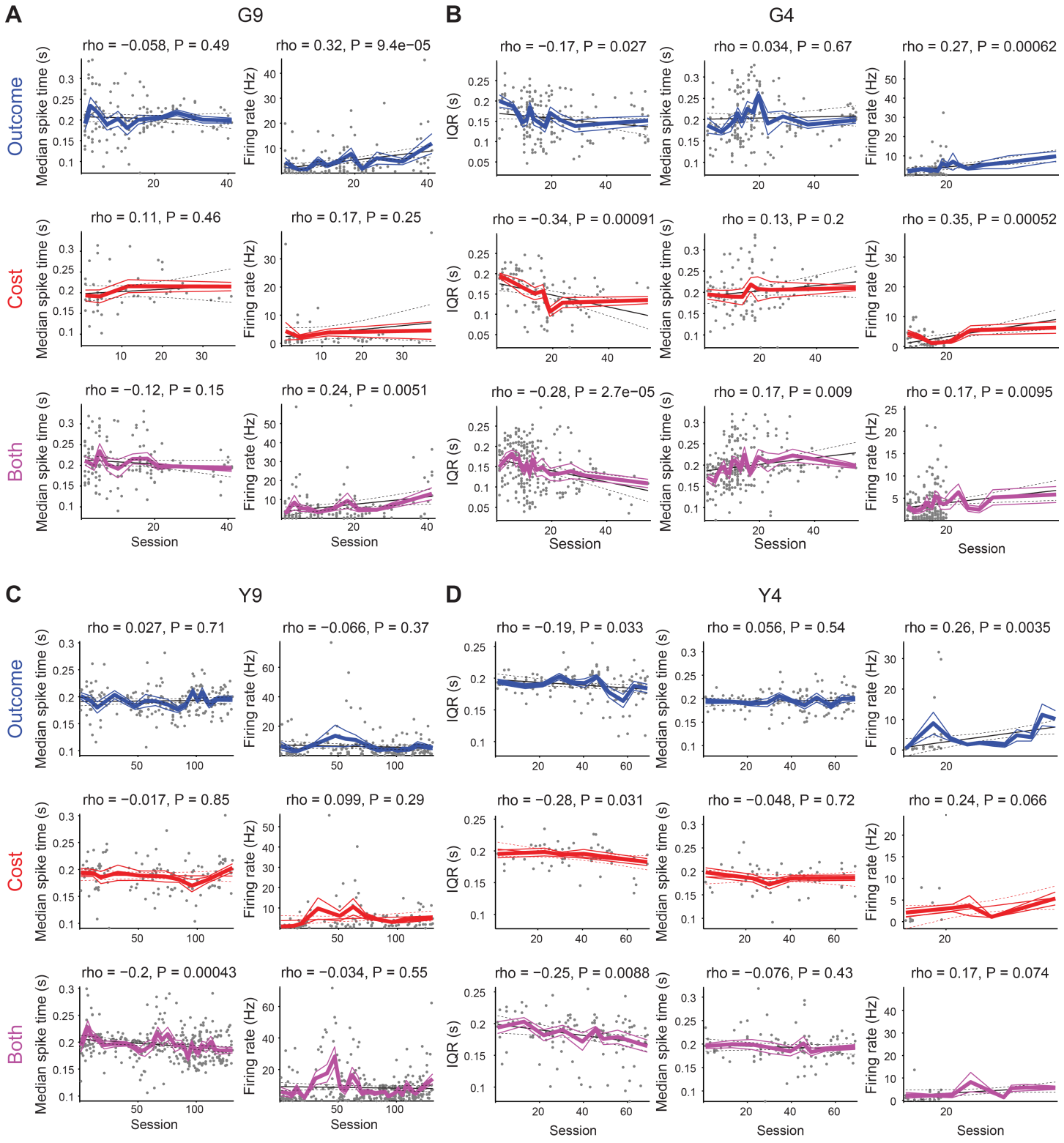

Figure S6. Correlations between Unit Activity and the Amount of Training, Related to Figure 7
IQR (left in **B** and **D**), median spike time (left in **A** and **C**, middle in **B** and **D**) and firing rate (right) calculated for Outcome (top), Cost (middle), and Both (bottom) units during the Targ-Off window are plotted (mean ± SEM) for G9 (**A**), G4 (**B**), Y9 (**C**) and Y4 (**D**). Each point on the plot represents one unit recorded in one session. Black lines represent the correlation line fit (solid) and the 95% confidence intervals (dashed). Correlation coefficient and p-value of the correlation are indicated for each plot.

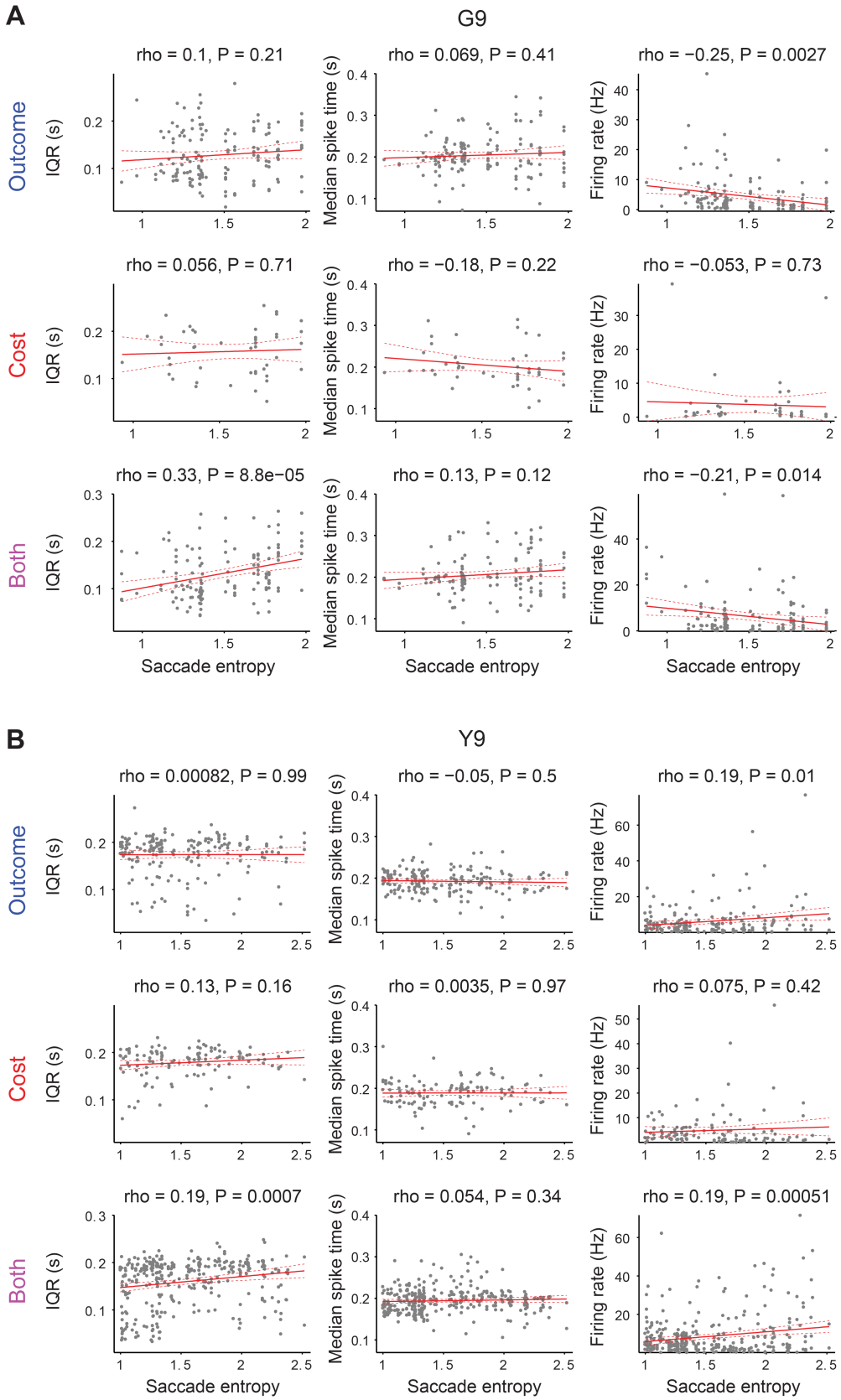

Figure S7. Correlations between Unit Activity and Saccade Entropy, Related to Figure 8
IQRs (left), median spike time (middle) and firing rate (right) of Outcome (top), Cost (middle), and Both (bottom) units measured during the Targ-Off window are plotted in relation to entropy in saccades for the performance of the 9-target task by monkeys G (**A**) and Y (**B**). Each point on the plot represents one unit recorded in one session. Red lines represent the correlation line fit (solid) and the 95% confidence intervals (dashed). Correlation coefficient and p-value of the correlation are indicated for each plot.

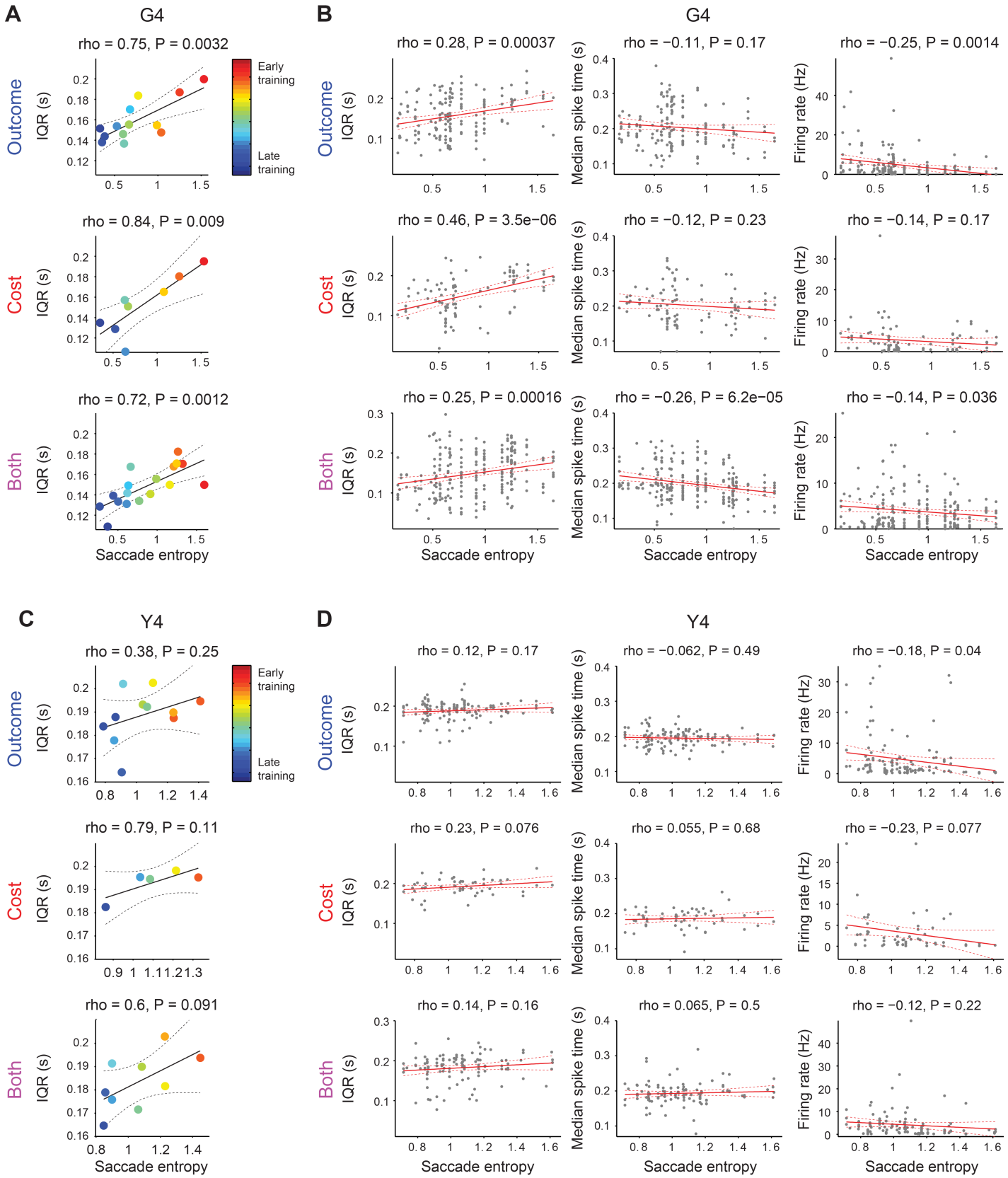

Figure S8. Correlations between Unit Activity and Saccade Entropy in 4-Target Task, Related to Figure 8
(**A** and **C**) Saccade entropy plotted relative to IQRs of Outcome (top), Cost (middle) and Both (bottom) units across sessions of the 4-target task with monkeys G (**A**) and Y (**C**), binned such that there are at least 10 units in each bin as in **Figure 8**.

(**B** and **D**) IQRs (left), median spike time (middle) and firing rate (right) of Outcome (top), Cost (middle) and Both (bottom) units measured during the Targ-Off window are plotted in relation to entropy in saccades in G4 (**B**) and Y4 (**D**), plotted as in **Figure S7**.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Across all sessions, we isolated 1,021 units from the caudate nucleus (CN) of monkey G in the 4-target task (G4) and 580 units from the CN of monkey Y in the 4-target task (Y4). CN units in these two task conditions replicated all the major findings from the 9-target task. In the 9-target task, 574 units from the CN of monkey G (G9) and 1,067 units from the CN of monkey Y (Y9) were isolated for study.

Surgical and Implant Procedures

Prior to any surgical procedures, structural MRIs (T1 and T2 weighted images, 1.5 or 3.0 Tesla; Siemens) were obtained to aid in the planning and placement of the recording chamber and head-fixation post with accompanying straps. All surgical procedures were performed under sterile conditions on deeply anesthetized monkeys placed in a standard stereotaxic apparatus. For monkey Y, a mold of the skull was made (VP Mix; Henry Schein) to aid in the shaping of the plastic chamber (Delrin, DuPont; Specialty Machining) that was surgically implanted on each animal. Each chamber was centered mediolaterally on the midline of the brain (with 20 mm on either side) and at approximately 22 mm (monkey G) or 30 mm (monkey Y) anterior to the interaural line. Monkey Y also received a post-operative structural MRI to confirm the placement of the recording chamber. Further detail on surgical procedures and a detailed explanation of the chronic recording method was previously published (Feingold et al., 2012), and thus a brief summary of the recording procedures will be provided here.

The plastic recording chamber accommodated a 30 × 40 mm grid with holes spaced at 1 mm center to center. These holes accommodated custom-built screw microdrives (Specialty Machining) that could each carry three, six or nine independently moveable electrodes (1–2 MΩ

at 1 kHz, 110-130 mm long, 125 μm shank, ~3 μm diameter tip; Frederick Haer). The microdrives could be configured to space electrodes as close as 1 mm apart and to record simultaneously from many brain regions, and their placements could be reconfigured for each chronic implant. Before the implantation of the chronic electrodes, electrophysiological mapping by the use of microstimulation (trains of 24-64 250 μs wide biphasic pulses, 333 Hz, 10-150 μA) was completed to determine the location of several key landmarks such as the frontal eye fields (FEF), primary motor cortex, and the depth of the dorsal surface of the CN.

Recording Procedures

Each monkey had a total of two chronic implants. Monkey G's first implant remained in place for 140 days, consisted of 72 independently moveable electrodes and targeted bilaterally the CN, the FEF, and the PFC. Recordings from the second implant in Monkey G were not analyzed and not presented here because the monkey was trained simultaneously on the 16-target and 25-target versions of the free-viewing scan task and was not successful in attaining a suitable level of performance to warrant analysis. Both of the chronic implants (166 and 366 days, respectively) in monkey Y contained 96 electrodes and also bilaterally targeted the CN, FEF, and PFC.

The recording of all data (behavioral and neural) began on the first day of training after electrodes had been implanted. During daily recording sessions, the monkey was seated in a chair approximately 50 cm from an LCD screen. Eye position was monitored by an infrared eye tracking system (500 Hz; SR Research Ltd.). A "hot" mirror that passed visual light and reflected infrared light to the eye tracking camera above was placed at a 45° angle in front of the monkey. Custom behavioral control software was designed in Delphi III. All signals were recorded with a Cheetah Data Acquisition system (Neuralynx, Inc.). Spike waveforms were digitized at 32 kHz

(filtered 600-6,000 Hz). Local field potentials were filtered at 1-475 Hz, and eye movements were recorded at 2 kHz.

After all experiments were completed, electrolytic lesions (10 μA DC for 10 s) were made at 4-6 sites to mark the locations in the brain relative to the grid holding the electrodes. Monkeys were perfused with fixative, and the brains were cut in the transverse plane on a sliding microtome (40-60 μm thick) and were stained for Nissl substance to permit detection of the electrode tracks. Anatomical tracing and 3-dimensional reconstruction was performed using Neurolucida (MicroBrightField).

Behavioral Task

The two monkeys were experimentally naïve prior to the first day of exposure to the free-viewing scan task, as they were not exposed to any explicit task training (on any task) prior to the first day of recording. They were only trained to be transferred from their home cage to the primate chair in the laboratory and to sit quietly in the chair with their head stabilized prior to the first day of neural recording. The free-viewing scan task and the related data analyses have been previously described in detail (Desrochers et al., 2010); therefore, a summary is provided below.

The trial began with a gray grid of dots displayed (**Figure 1A**, first panel). This variable Start Delay period lasted 1-2 s (mean 1.5 s), and while the gray dots were displayed, there were no behavioral requirements placed on the monkey. The green target grid then replaced the inner gray targets at the Targ-On event (**Figures 1A**, second panel, and **1B**). This Scan Delay period was short during initial training sessions to allow the monkey to first associate looking at the green targets with getting reward, and then was, over the course of ~10 sessions, gradually lengthened to be 1-2 s in duration (mean 1.5 s). During the Scan Delay, the only requirement

imposed was that the monkey's gaze not leave the area defined by the green grid. After the variable Delay Scan, at a time (Find Scan; **Figures 1A** and **1B**, black diamond) not signaled in any way to the monkey, a pseudorandomly chosen target became baited with reward. As long as the monkey's eyes remained within the green target grid, the monkey was free to scan the targets for as the remainder of the session, generally 3-4 hour long; typically each monkey captured the baited target by passing its gaze through it (fixation was not required) in less than a second (Reward Scan; mean 0.8 s). Immediately thereafter, the green target grid turned off (Targ-Off), and it was replaced by the gray target grid (**Figure 1A**, third panel). There were no behavioral requirements placed on the monkey for the remainder of the trial. After a variable Reward Delay (0.4-0.8 s, mean 0.6 s), reward was delivered for a constant duration across all trials (0.2 s juice reward for monkey G and 0.25 s food slush reward for monkey Y).

No particular sequence of saccades or fixations was required, and it was not required that the monkey's gaze land on each target, though the design of the task made it advantageous to do so. During the Reward Scan period, revisiting a previously scanned target did not lead to reward. Thus, the optimal strategy would be to scan each target exactly once. Because the onset of the Reward Scan was unpredictable and there was no evidence of different behaviors in the Delay Scan and Reward Scan periods, we assumed that the monkeys treated the entire period as the reward period. This assumption necessarily means that the optimal scanning pattern was a loop pattern in which each target was visited once before returning to the original target (see Desrochers et al., 2010 for further discussion).

Each recording session consisted of approximately 1,000 correct trials. The task was divided into blocks such that each of the targets in the grid would be chosen as the baited target approximately equal numbers of times. Monkeys first acquired the 4-target task, and in later

sessions, the monkeys performed the 4-target task and 9-target task in separate blocks during the same recording session.

Behavioral Data Analysis

One Y4 session and one Y9 session were excluded due to data loss. Session blocks with fewer than 55% rewarded trials were included only if performance on other blocks indicated the monkey was sufficiently motivated to perform the task, and session blocks with fewer than 40 rewarded trials were not included in analyses (~4% excluded overall). All analyses were done in Matlab.

Saccade Entropy

For each session, the probability of the monkey fixating each target in the green target grid following a fixation of each of the targets in the target grid was calculated to allow formation of the transition probabilities for that session. The entropy of the transition probabilities for each session (q) was defined as:

$$E = -\sum_i q_i \sum_j q_{ij} \log_2 q_{ij}$$

where q_i is the probability of observing target i, and q_{ij} is the probability of observing target j followed by target i.

Correlation between Cost and Saccade Pattern Selection

Previously we found that if in a trial in which a most frequent pattern was performed the distance was shorter than that in the previous trial, then the monkey tended to repeat the same pattern

again in fewer trials. The converse was also true; if the distance was longer, then the monkey waited a larger number of trials before again performing the pattern (Desrochers et. al., 2010).

We performed this same analysis using the current measure of cost (performance prediction error, PPE): the difference from the mean distance. Cost in the trial in which a stereotyped saccade pattern was executed was positively correlated with the number of trials that were completed before another stereotyped saccade pattern was formed (inter-pattern interval, IPI; **Figures 1D** and **S1C**). We found stronger correlations between cost and IPI than we previously found between the change in distance in two subsequent trials and IPI for all the most frequent 4-target patterns (previous: rho = 0.564, p = 0.089, slope = 0.01, boot p = 0.01; current: rho = 0.91, $p < 0.001$, slope = 0.24, boot p = 0) and all the most frequent 9-target patterns (previous: rho = 0.766, p = 0.010, slope = 0.25, boot $p < 0.002$; current: rho = 0.98, $p < 0.001$, slope = 0.33, boot p = 0). This difference was highly significant for 4-target patterns (IQR × distance type $F_{1,16} = 48.5$, p = 0), and although the slope and rho values were numerically larger for 9-target patterns with mean distance, this increase was not statistically significant (IQR × distance type $F_{1,16} = 0.47$). These results suggest that the cost (PPE) introduced in the current study was also correlated with the shifting of stereotyped saccade patterns on a trial-by-trial basis.

Alternative Definitions of Cost

To determine whether variations in the calculation of cost would alter the distributions of units classified as Outcome, Cost, or Both units, we calculated cost in one of three ways: 1) trial distance minus mean distance across all trials on all sessions, 2) trial distance minus mean distance on all trials (including previous sessions) up to that trial, and 3) trial distance minus

mean distance on trials in the current session up to that trial. There was no statistical difference in any of the three distributions from the distribution of units obtained from the original stepwise regression for either monkey (Kolmogorov-Smirnov test p's > 0.9, **Figures S4A** and **S5A**). These results suggest that different variants of the cost variable itself would be capable of driving the REINFORCE algorithm, as units that correlated with these cost variants were distributed nearly identically across the population recorded. The robustness of cost as an explanatory variable is also supported by the fact that variations in its calculation did not alter the finding that populations of units in the striatum represent this variable.

Collinearity Diagnostics

To examine whether the cost and outcome variables that we chose were too similar to construct the regressor, inducing a multicollinearity problem, we performed Belsley collinearity diagnostics on the regressors that we used for each session. In these diagnostics, the regressors are converted to condition indices (CIs) that identify the number and strength of any near dependencies in the design matrix. CIs less than 30 are generally accepted as the set of variables that do not produce a collinearity problem. All CIs were less than 18 as indicated in the table below:

	Minimum	Median	Maximum
G4	1.91	2.67	17.41
G9	1.96	2.62	3.85
Y4	2.27	2.74	4.50
Y9	1.79	2.62	5.60

These diagnostics suggested that even though the regressors were correlated, they did not present a multicollinearity problem for the regression, adding further legitimacy to the stepwise regression procedure that we performed.

Final Eye Position Control

To verify that there was no change in the eye position at Targ-Off that could account for changes observed in the neural activity, we performed the following control. For each trial, we assigned the final eye position to one of 9 two-dimensional position bins that tiled the green target grid space and were centered on the 9 targets that composed the green scan target grid. We then examined the distribution of those positions across sessions by taking the median of the eye position distribution and the IQR of the eye position distribution for each session. We tested whether there were changes from early to late sessions by comparing those values in the first 20 sessions versus the last 20 sessions (G9 has 41 total sessions). There was no difference between median last eye positions and no difference in the IQR (as a measure of variability) of the last eye position between early and late sessions (t_{39}'s < 2, p's > 0.05). These findings indicate that the distribution of final eye positions and the variability of those positions did not change from early to late sessions. Therefore it is highly unlikely that the changes we observed in the neural IQR could be due to changes in final eye position or final eye position variability.

Unit Data Analysis

Units were separated from noise manually and with templates using Offline Sorter (Plexon, Inc.). A unit was considered to have a significant response in a 0.4 s window adjacent to the event if the firing rate in four or more consecutive 20 ms bins were greater than two standard deviations above the mean firing rate, taken across the entire trial.

Striatal Subtype Classification

We used established methods (Thorn and Graybiel, 2014) to classify units as HFNs, TANs or MSNs. HFNs were classified according to three parameters: the peak full width at half-maximum (peak FWHM), the valley full width at half-maximum (valley FWHM), and the firing rate calculated as the total spikes in the recording session divided by the total time in the recording session. We then applied k-means clustering minimizing the squared Euclidean distance between two clusters. The cluster with high firing rates and low peak FWHM and valley FWHM was classified as containing points representing HFNs (**Figure 6A**).

TANs were classified based on three parameters: post-spike suppression, percent of spikes with a long inter-spike interval (ISI), and percent of spikes in bursts. Post-spike suppression is a measure of the refractory period (Schmitzer-Torbert and Redish, 2008), computed by constructing the 0-1 s autocorrelation histogram (1 ms bins), smoothing with a 50-point Hamming window, and calculating the time at which firing rate first exceeded the mean firing rate over the whole 1 s window. For classification purposes, a "long" ISI was considered to be > 0.5 s. A "burst" was considered to be any time two or more spikes occurred within 10 ms of each other. We applied squared Euclidean minimization k-means clustering on these three parameters with two clusters. The units in the cluster with long post-spike suppressions, with low percentages of spikes occurring in bursts, and with relatively small percentages of short ISI (most of them had relatively high percentages of long ISIs) were classified as TANs (**Figure 6B**). Those units that were not classified as HFNs or TANs were classified as MSNs. There was no overlap among the groups.

Changes in Activity Patterns of Striatal Neurons across Learning

The activity of striatal units, binned such that there were ≥ 10 units per bin as in **Figure 2A** and **B**, in the Targ-Off window, gradually changed across sessions. There was a significant decrease in the IQR for both G9 (rho = −0.0007, $p < 0.03$; **Figure 2C**) and Y9 (rho = −0.0003, $p < 0.001$; **Figure 2E**). To further support the gradual change in activity, we also examined the mean firing rate around the peak. The median spike time was designated the peak time and the mean firing was calculated during the center two quartiles, i.e., the quartiles on either side of the peak firing time. We found a significant increase across sessions (G9: rho = 0.003, $p < 0.01$; Y9: rho = 0.002, $p < 0.001$; **Figures 2D** and **F**).

We also examined the activity of striatal neurons when the activity binned by single sessions. Similar changes in neural activity were observed in the Targ-Off window (**Figures S2C** and **F**) with decreases in IQR (G9: rho = −0.0004, $p < 0.09$, **Figure S2D**; Y9: rho = −0.0002, $p < 0.001$, **Figure S2G**) and significant increases in peak firing (G9: rho = 0.002, $p < 0.01$, **Figure S2E**; Y9: rho = 0.001, $p < 0.001$, **Figure S2H**).

To examine how these changes in activity patterns across sessions were influenced by eye-movement related units, we first determined the percentage of neurons that had significant correlations with activity that showed correlations with eye movement (Pearson's, $p < 0.05$) in the Targ-On window (400 ms after Targ-On). We found very similar relative fractions of units correlated with eye movements at Targ-On as for Targ-Off (G9: 12% Targ-On, 10% Targ-Off, **Figure S2I**; Y9: 16% Targ-On, 17% Targ-Off, **Figure S2J**). Approximately 25% of the units with the activity correlated with eye movements in Targ-On also had activity correlated with eye movements in Targ-Off. We then examined the neural activity patterns across sessions of only those units with activity that was not correlated with eye movements in Targ-On or Targ-Off.

Although the pattern of activity changes across sessions was slightly weaker due to lower numbers of units, eliminating those movement-related units did not eliminate the phasic task-bracketing activity after Targ-On and after Targ-Off (**Figure S2K** and **N**). In the Targ-On window, the same pattern of decreasing IQR (G9: rho = −0.0006, $p < 0.2$, **Figure S2L**; Y9: rho = −0.0003, $p < 0.001$, **Figure S2O**) and increasing peak firing (G9: rho = 0.003, $p < 0.05$, **Figure S2M**; Y9: rho = 0.0007, $p < 0.1$, **Figure S2P**) was also observed across sessions.

Regression Model Categories in G4 and Y4

In the 4-target task, CN units in differing fractions, as in the 9-target task, again significantly represented task events (G4: $F_{6,378} = 27$, Y4: $F_{6,469} = 17$; p's < 0.0001; **Figure S1E**). Approximately 50% of units across sessions were task-responsive (TR), and the Targ-Off period was preferentially represented (post-hoc Tukey test $p < 0.05$; G4 Targ-On and Targ-Off greatest, Y4 Targ-Off greatest). Using stepwise linear regression, we found very similar fractions of units that responded to Cost, Outcome or Both (**Figures S3A** and **S3B**), and these fractions did not change across training sessions. TR units overlapped with both Outcome units and Cost units (**Figures S3C** and **S3D**). Dividing the units into HFNs, TANs, and MSNs did not alter the distributions of these Outcome, Cost and Both unit types (**Figure S3K**).

Alternative Regression Models

To examine whether the results obtained from the stepwise regression were robust to the difference in model selection methodology, we compared the results of the stepwise regression analysis with those derived from the All Possible Subset regressions using Akaike information criterion (AIC) and Bayesian information criterion (BIC). We further performed Ridge

regression to resolve potential multicollinearity that the regressors might contain. For each unit, four different regressions were performed (lambda0 AIC, lambda0 BIC, ridge AIC and ridge BIC). Lambda0 AIC and BIC are the usual AIC/BIC model selection procedures used without introducing ridge regression. Ridge AIC and Ridge BIC are the ridge regressions with the optimal lambda derived from the AIC and the BIC, respectively. To derive the optimal parameter for each ridge regression, we used AIC and BIC criteria using the effective degrees of freedom: $EDF_\lambda = \mathrm{tr}[X(X'X + \lambda D)^{-1}X']$, where X consisted of regression variables, D was a diagonal matrix, and λ was a regularization parameter. For each procedure, we performed the model selection among the following four models: outcome, cost, outcome + cost, and outcome + cost + outcome × cost. All models also included a constant term.

IQR Correlation with Sessions in G4 and Y4

G4 and Y4 CN units also showed changes across learning (**Figures S2A** and **B**). The IQRs of the Both type units were highly significantly correlated with session in both G4 ($R = -0.28$, $p < 0.0001$; **Figure S6B**) and Y4 ($R = -0.25$, $p < 0.01$; **Figure S6D**). Correlations with the other unit types (Outcome and Cost) for G4 and Y4 were reliable (R's ≤ -0.17, p's < 0.03), but the greatest significance levels were reached within the Both unit type (R's ≤ -0.25, p's < 0.009). Again these changes were specific to the IQR and were not accompanied by consistent changes in either median spike time or firing rate (**Figure S6B** and **S6D**).

Depth Controls in Y9

A key internal control for the effect of depth on IQR is that all three unit categories (Outcome, Cost and Both) were found at all depths throughout the CN (**Figure 4**). There were also no

significant differences among the depths of the electrodes across sessions for the three unit categories (session × category, G9: $F_{49,184} = 1.4$, $p > 0.05$, Y9: $F_{15,332} = 0.99$, $p > 0.5$). If it were the case that the decreases in IQR found in Both units were solely due to increases in the depth of the electrodes, then the IQR should also decrease in Outcome and Cost units; however, we found no such decreases in Outcome and Cost unit numbers.

As an additional control, because the data recorded from monkey Y was composed of recordings from two chronic implants (only one in monkey G), it was possible to compare the IQR of Both units recorded in the last two sessions of the first implant (Y9 sessions 7 and 8, n = 6, mean depth from grid surface = 45.5 mm) to the IQR of Both units recorded in the second implant (later in training), but at the same relative depth (45.5 ± 0.2 mm, n = 13, mean session = 58). We found that the IQR of the units recorded later in training was significantly lower than those recorded earlier in training, even though they were at the same depth ($t_{17} = 2.3$, $p < 0.05$).

IQR Correlation with Entropy in G4 and Y4

Similar to as was found in the 9-target task, IQR correlated strongly with entropy (**Figure S8**) in G4 (rho = 0.25, $p < 0.001$), but was marginally related in Y4 (rho = 0.14, $p = 0.16$). Though there were correlations between IQR and entropy with other unit types (Outcome and Cost) in G4, there were not for Y4, as was found for Y9 (**Figure S7**). We attribute these differences found in the 4-target task to the smaller range of possible behaviors and, consequently, a smaller range of entropy values. It could be that there is not enough richness in the 4-target behavior to draw conclusions about these correlations. Regardless, there were no consistent relationships with median spike time or firing rate in G4 and Y4, as was found for G9 and Y9 (**Figure S8**).

Separating Effects of Entropy and Session in Correlations with IQR

To attempt to separate the effects of behavioral entropy and session in correlations with IQR, we first asked whether the IQR was better correlated with behavioral entropy or with session number. We thus fitted the IQR over sessions using two different models, one with session and one with behavioral entropy as the behavioral factor. We then examined the quality of the model fit. For G9 Both units, the model fit slightly better for behavioral entropy (residual sum of squares, RSS = 0.33, log likelihood = −6.00) than for session (RSS = 0.34, log likelihood = −5.97). However, for Y9 Both units, the opposite was true, with the model for session (RSS = 0.72, log likelihood = −6.08) fitting slightly better than the model containing behavioral entropy (RSS = 0.75, log likelihood = −6.04). Thus, the IQR was similarly correlated to the entropy and session number across the two monkeys.

Then we examined whether the correlation between IQR and entropy would still exist after we subtracted the effect of session number. To accomplish this analysis, we first correlated IQR with session and, separately, behavioral entropy with session. We then performed a third correlation with the residuals of the first two models and examined the relationship between the residuals. For G9 Both units, we found a significant correlation between the residuals (rho = 0.50, $p < 0.05$). However, the same was not true for Y9 Both units (rho = −0.24, $p > 0.3$).

Therefore, although the behavioral entropy is one of the best parameters to represent the monkey's progress in learning, the session number itself is also correlated with learning. While there may be some evidence for an independent correlation between behavioral entropy and IQR, we conclude that session is one possible learning variable that both behavioral entropy and IQR change in concert with. We note further that the individual differences in results for the two

monkeys could reflect different strategies taken by them, and there are likely other hidden variables.

SUPPLEMENTAL REFERENCES

Desrochers, T.M., Jin, D.Z., Goodman, N.D., and Graybiel, A.M. (2010). Optimal habits can develop spontaneously through sensitivity to local cost. Proc. Natl. Acad. Sci. U. S. A. *107*, 20512–20517.

Feingold, J., Desrochers, T.M., Fujii, N., Harlan, R., Tierney, P.L., Shimazu, H., Amemori, K.-I., and Graybiel, A.M. (2012). A system for recording neural activity chronically and simultaneously from multiple cortical and subcortical regions in nonhuman primates. J. Neurophysiol. *107*, 1979–1995.

Schmitzer-Torbert, N., and Redish, A. (2008). Task-dependent encoding of space and events by striatal neurons is dependent on neural subtype. Neuroscience *153*, 349–360.

Thorn, C.A., and Graybiel, A.M. (2014). Differential entrainment and learning-related dynamics of spike and local field potential activity in the sensorimotor and associative striatum. J. Neurosci. *34*, 2845–2859.