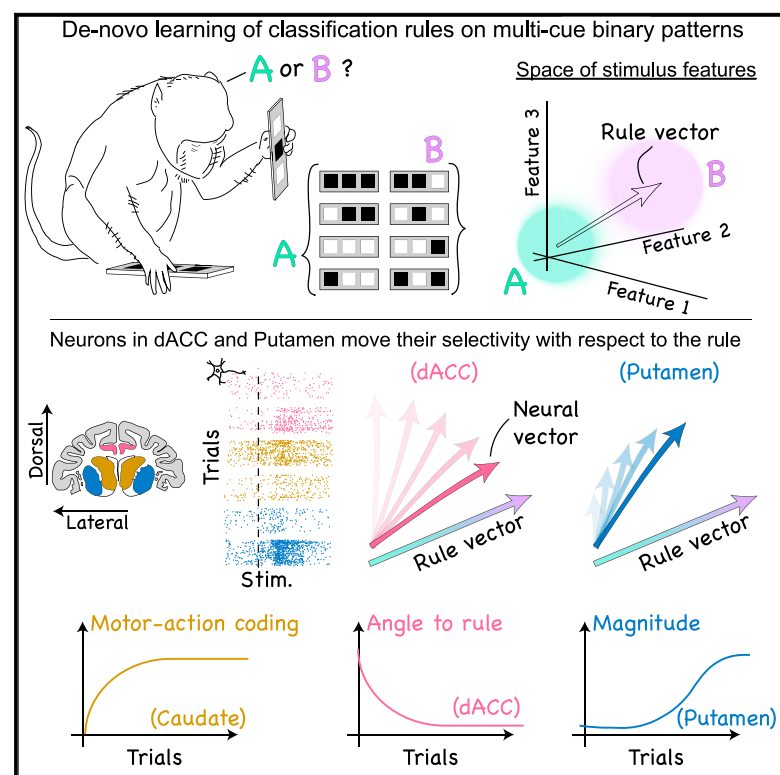


# The geometry of neuronal representations during rule learning reveals complementary roles of cingulate cortex and putamen

## Graphical Abstract



## Authors

Yarden Cohen, Elad Schneidman, Rony Paz

## Correspondence

ycohen1@mgh.harvard.edu (Y.C.), elad.schneidman@weizmann.ac.il (E.S.), rony.paz@weizmann.ac.il (R.P.)

## In Brief

Rule learning is highly evolved in primates. Cohen et al. record single neurons during *de novo* acquisition of classification rules and develop a geometrical framework to track dynamics of neural representations. They find complementary roles of a policy search in dACC neurons, followed by a confidence increase in the putamen, and further predict overnight retention.

## Highlights

- A geometrical framework reveals single-neuron dynamics during *de novo* rule learning
- Neurons in the dACC reflect search for the new rule-based behavioral policy
- Neurons in the putamen also reflect confidence increase that follows the policy search
- The neural representation predicts overnight retention and generalization of rules

Article

# The geometry of neuronal representations during rule learning reveals complementary roles of cingulate cortex and putamen

Yarden Cohen,<sup>1,2,\*</sup> Elad Schneidman,<sup>1,3,\*</sup> and Rony Paz<sup>1,3,4,\*</sup>

<sup>1</sup>Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel

<sup>2</sup>Present address: Massachusetts General Hospital, Boston, MA 02114, USA

<sup>3</sup>These authors contributed equally

<sup>4</sup>Lead contact

\*Correspondence: [ycohen1@mgh.harvard.edu](mailto:ycohen1@mgh.harvard.edu) (Y.C.), [elad.schneidman@weizmann.ac.il](mailto:elad.schneidman@weizmann.ac.il) (E.S.), [rony.paz@weizmann.ac.il](mailto:rony.paz@weizmann.ac.il) (R.P.)

<https://doi.org/10.1016/j.neuron.2020.12.027>

## SUMMARY

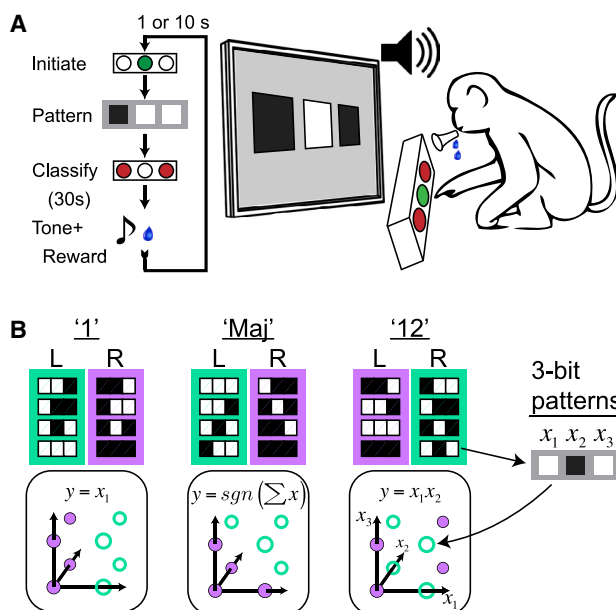
Learning new rules and adopting novel behavioral policies is a prominent adaptive behavior of primates. We studied the dynamics of single neurons in the dorsal anterior cingulate cortex and putamen of monkeys while they learned new classification tasks every few days over a fixed set of multi-cue patterns. Representing the rules and the neuronal selectivity as vectors in the space spanned by a set of stimulus features allowed us to characterize neuronal dynamics in geometrical terms. We found that neurons in the cingulate cortex mainly rotated toward the rule, implying a policy search, whereas neurons in the putamen showed a magnitude increase that followed the rotation of cortical neurons, implying strengthening of confidence for the newly acquired rule-based policy. Further, the neural representation at the end of a session predicted next-day behavior, reflecting overnight retention. The novel framework for characterization of neural dynamics suggests complementing roles for the putamen and the anterior cingulate cortex.

## INTRODUCTION

Learning to classify multi-cue stimuli is an adaptive behavior required by animals on a daily basis. To do so, primates acquire novel rules and apply new behavior policies based on combinations of stimulus cues. Accordingly, such tasks have been commonly used to explore learning strategies in humans under normal conditions (Gluck et al., 2002; Goodman et al., 2008; Lagnado et al., 2006; Nosofsky et al., 1992; Shepard et al., 1961) as well as pathological ones (Meeter et al., 2008; Shohamy et al., 2008; Speekenbrink et al., 2010; Stuss et al., 2000). While typically the complexity of classification rules was measured by the average performance over subjects (Feldman, 2000), more recent work has shown that the performance of individuals may vary widely, but can be predicted accurately using models that rely on a combination of stimulus features with a prior assigned to each individual subject (Cohen and Schneidman, 2013). In non-human-primates, studies of rule-based classification have ascribed complementary roles to the striatum and regions of the prefrontal cortex (PFC) (Balleine et al., 2007; Seger and Miller, 2010). Specifically, in paradigms that require classification of stimuli into distinct categories, individual neurons in the PFC acquire a preference for one category over the others (Cromer et al., 2010; Freedman and Assad, 2016; Freedman et al., 2003; Gold and Shadlen, 2007; Histed et al., 2009; Kim and Shadlen, 1999; Muhammad et al., 2006; Wallis et al., 2001).

Within the PFC, the dorsal anterior cingulate cortex (dACC) projects widely to striatal regions (Averbeck et al., 2014; Heilbronner et al., 2016; Ongür and Price, 2000) and is involved in several cognitive functions that contribute to the learning process. Neurons in the dACC represent attention, reflect actions that lead to reward, signal outcomes of previous trials, and form integrated representations of task structure (Chudasama et al., 2013; Haroush and Williams, 2015; Hayden and Platt, 2010; Heilbronner and Hayden, 2016; Koling et al., 2016; Lee et al., 2007; Mansouri et al., 2009; Rudebeck et al., 2008; Rushworth and Behrens, 2008; Saez et al., 2015; Seo and Lee, 2007, 2009; Wallis and Kennerley, 2011). The striatum, in turn, receives wide projections from the dACC and plays a role in choosing actions and providing reinforcement signals that may help to direct learning (Averbeck and Costa, 2017; Graybiel and Grafton, 2015; Jin and Costa, 2015; Kim and Hikosaka, 2013; Lau and Glimcher, 2008; Merchant et al., 1997; Seger, 2008; Seo et al., 2012; Williams and Eskandar, 2006).

Most studies have characterized neural correlates of final task representation, perception, and recognition, mainly following extensive training. Less is known about how single neurons form and change their preferences during *de novo* learning of such tasks. Rule-based classification can take several forms (Seger and Miller, 2010), and key studies showed that individual neurons represent behavior when animals learn to assign



**Figure 1. Trial-by-trial learning of classification rules**

(A) Behavioral paradigm: pressing and holding the middle button initiates a new trial. After a 3-bit pattern (black/white squares) appears on the screen, the monkey has 30 s to classify it with the left or right button. A fluid reward follows a correct choice, and a short timeout follows an incorrect choice.

(B) A schematic of the rule-based classification. Top: labeling all eight 3-bit patterns using the correct response (pressing the left or right buttons) divides the patterns into two categories (truth table marked by green and purple boxes). Bottom: representing the patterns as variables,  $[x_1, x_2, x_3] \in \{-1, 1\}^3$ , allows describing truth tables by Boolean expressions that define classification rules; namely, the relationship between patterns and correct responses. Different rules are presented graphically in the 3D space defined by individual bits and coloring each pattern to match its label. Shown are the one-bit rule (1), where the decision is based only on the identity of the left bit; the two-bit-rule (12), where the decision is based on a XOR of the left and middle bits; and the majority rule, where the decision is based on summing all 3 bits.

different outcome probability or value (Padoa-Schioppa and Assad, 2006; Yang and Shadlen, 2007), when they acquire arbitrary stimulus-motor associations (Brasted and Wise, 2004; Buch et al., 2006; Mitz et al., 1991), or when they switch contingencies between the rules (Buckley et al., 2009; Wallis et al., 2001).

Here we studied neural dynamics in the dACC and the striatum of monkeys during learning of visual rule-based classification tasks. Rules were changed every few days, which allowed us to characterize *de novo* learning and its neural correlates. We analyze the learning and neural activity in the space spanned by features of the stimuli, enabling a geometric interpretation of learning and describing complementing roles for the dACC and the striatum.

## RESULTS

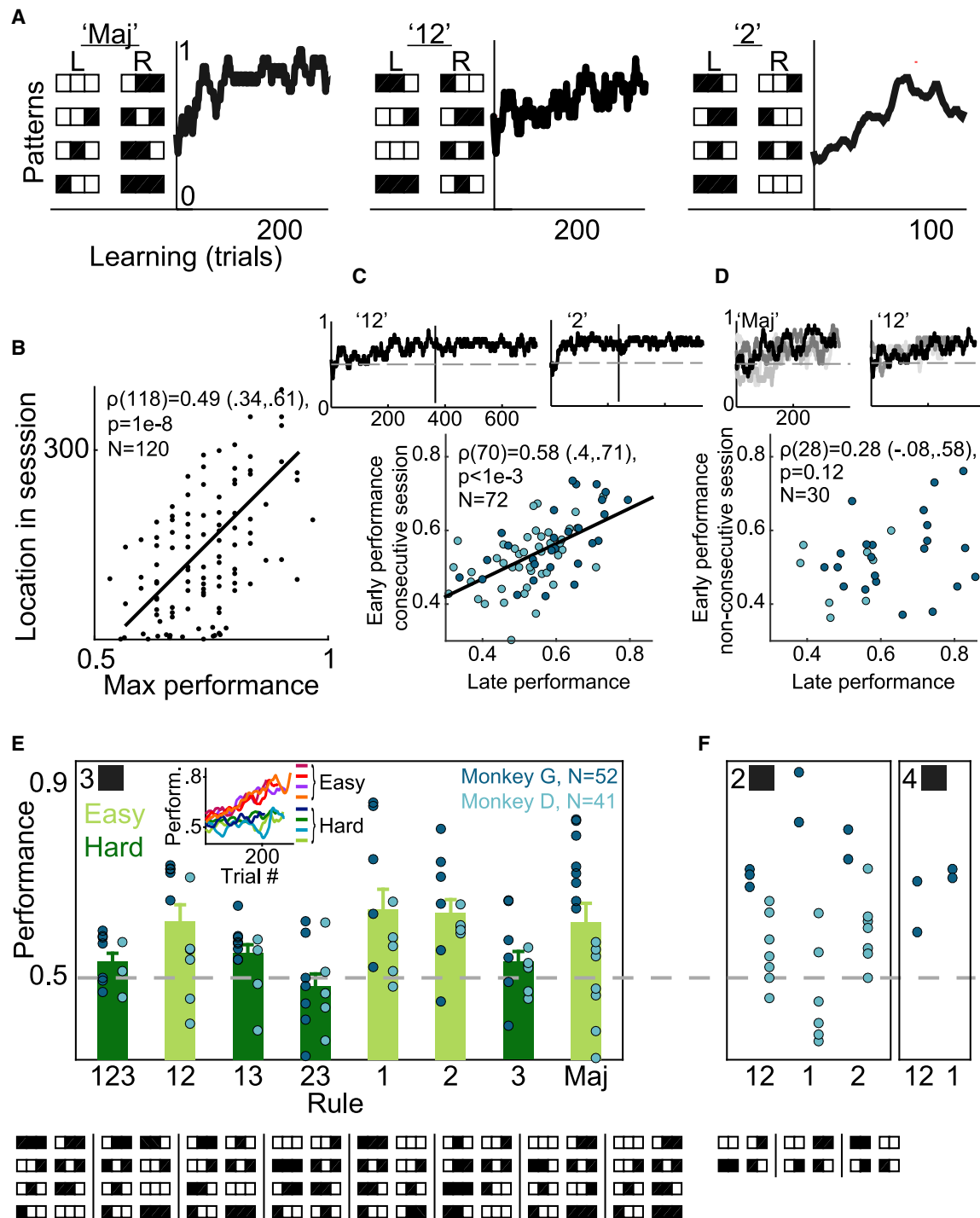
Two monkeys (*Macaca fascicularis*) performed trial-by-trial classification of visual patterns composed of black and white squares (“bits”; Figure 1A). In each session, the set of eight 3-bit patterns was presented in pseudo-random order, and the

monkeys had to learn by trial and error which of two response buttons to press when presented with each pattern. The relationship between the eight patterns and the correct responses (“labels”) defines a classification rule (Figure 1B). Of the 256 ( $2^8$ ) possible deterministic rules for 3 bits, we chose seven rules in which the label was determined according to single, pairwise, or triple-wise dependencies between the bits in the pattern. For example, in rule 1, the monkeys had to learn to choose the left/right button when the left bit was white/black, and in rule 12 when the left and middle bits were equal/different (Figure 1B). All seven rules are unbiased so that for 4 patterns the correct label is left, and for the other 4 it is right. Moreover, these rules are independent of each other in the sense that perfect learning of one rule would result in chance performance on all other rules. We also included the majority rule (which is also unbiased; “Maj” in Figure 1B), in which the label of a pattern depends on the majority color among the bits. Each monkey learned the eight rules over 4 weeks, each rule for typically 2–4 sessions on consecutive days. We then repeated the same set of rules for another 4 weeks (second and third cycles; Figure S1A). The monkeys did not experience any of these rules in the training prior to the recordings.

## Learning of classification rules

Both monkeys exhibited within-session learning (Figure 2A) and continuous performance improvement throughout long sessions (Figure 2B; Pearson  $r(118) = 0.49(0.34, 0.61)$ ,  $p < 1e-8$ ) as well as next-day retention of the learned rule (Figure 2C; Pearson  $r(70) = 0.58(0.4, 0.71)$ ,  $p < 1e-3$ ). Neither monkey showed retention benefits when presented with the same classification rules in the second cycle of the rules (Figure 2D; Pearson  $r(28) = 0.28(-0.08, 0.58)$ ,  $p = 0.12$ ). Both monkeys demonstrated learning of each of the rules in some sessions, but their performance levels varied across sessions and rules (Figures 2E and S1B). This diversity was also apparent in additional sessions in which the monkeys were presented with 2-bit rules and 4-bit ones (Figure 2F). This individual diversity is akin to that of human subjects who exhibited large variability across individuals and rules in a similar task (Cohen and Schneidman, 2013).

The notion that the monkeys learned the rules rather than the label of individual patterns was supported by several lines of evidence. First, in pure stimulus-response associative learning, the correct response to each stimulus would be acquired independent of other stimuli, and the performance on particular patterns would be the same under different rules. In contrast, we found that pattern-specific error rates show dependence on the specific rule (Figure S1C; Kruskal-Wallis test,  $\chi^2(7, 366) = 89$ ,  $p < 1e-15$ ) and were correlated with performance on other patterns (Figure S1D; Kruskal-Wallis test,  $\chi^2(7, 440) > 88$ ,  $p < 1e-10$ ). Second, classification of the visually salient patterns showed variability across rules and dependence on the specific rule (Figure S1E; Kruskal-Wallis test,  $24 < \chi^2(3, 36) < 33$ ,  $p < 1e-4$ ). Third, rule learning was supported by the consistency across learning of different rules (Figures S1G and S1H). Fourth, monkeys could even learn 4-bit rules for which memorization is extremely unlikely (Figure 2F; there are 65,536 different deterministic rules for the 16 stimulus patterns). Fifth, performance for specific patterns deteriorated following a change of the



**Figure 2. Rule-dependent learning**

(A) Learning curves show within-session improvement. Shown are individual learning curves from 3 example days, each with a different rule (monkey G, two left examples; monkey D, right example). Also shown is the underlying truth table of the rule (patterns are stacked by category).

(B) Continuous longer learning within the session is beneficial. Shown is the location within the session of the maximum performance plotted against the maximum performance (best 30 consecutive trials) for all sessions.

(C) Overnight retention. Shown are two examples of learning curves on consecutive days, with vertical lines separating the sessions. Over all sessions, there was a significant correlation between performance at the end of a session and performance at the beginning of the next-day session with the same rule (monkey D: light blue,  $r = 0.48$ ,  $p < 0.001$ ; monkey G: dark blue,  $r = 0.61$ ,  $p < 0.001$ ; both:  $r = 0.58$ ,  $p < 0.001$ ).

(legend continued on next page)

rule, even when the new rule did not require learning new associations for these patterns (Figures S1I–S1N). Finally, a simple strategy based on performance (win-stay, lose-switch) poorly described answer sequences (Figure S1O; Wilcoxon’s signed-rank test,  $p < 1e-15$ ). The evidence suggests that monkeys did not use simple memorization of stimulus-response associations as the main strategy and therefore supports the notion of rule learning.

We conclude that both monkeys learned the rules in individual sessions, with some rules easier to learn than others. The complexity and richness of the task led to error rates that reflect a more natural learning setting compared with over-trained animals and therefore enable comparison of neural responses between easy and hard rules as well as between sessions with low and high performance.

### Neurons in the dACC and putamen represent the learned rule

We recorded single-unit activity during all sessions of the 3-bit rules in the dACC (Brodmann area 24), caudate, and putamen (Figures 3A and 3B; 543,115,114 neurons recorded in the dACC, caudate, and putamen in 3-bit rule sessions). First we identified neurons with a stimulus-evoked spiking pattern that responded to the stimuli but with no preference for rule/category (stimulus-specific; Figure 3C, right). Supporting the notion that neurons represent rules rather than stimulus-response associations, we found that only 3% of neurons exhibited selectivity for a specific pattern (not different than chance level,  $p > 0.1$ , binomial tests; Figure S2A).

We then identified neurons that were not rule selective early on but changed during learning to distinguish between the two category labels (category-specific; Figure 3C, left and center; Figures S2B–S2E). To identify correlates of successful learning, we separated the sessions of easy and hard rules based on the average success rate for each rule (Figures 2E, S1P, and S1Q). This separation of easy versus hard rules is also evident in the learning curves (Figure 2E, inset; Figure S1R) and captures the largest performance variance (Figure S1S). We found that, at the end of the session, the dACC and putamen had significantly more category-specific neurons for easy rules compared with hard rules (Figure 3D; dACC: 21% versus 12.5%, Pu: 22% versus 8% for easy versus hard, binomial comparison z-test; dACC:  $z = 2.43$ ,  $p < 0.01$ ; Pu:  $z = 2.12$ ,  $p < 0.02$ ). This was also the case when considering individual sessions and separating them based on high versus low performance (Figures S2F and S2G; dACC:  $z = 2.046$ ,  $p < 0.021$ ; Pu:  $z = 1.79$ ,  $p < 0.04$ ; and for each monkey separately; Figures S2F and S2G). Moreover,

there was a monotonic relationship between performance and the proportion of neurons showing correct classification (Figure 3E).

Because the correct label of a stimulus and the choice of the monkey become correlated in successful learning, we used error trials to distinguish between them. In sessions of easy rules, more neurons became selective for the category than for the choice in the dACC and putamen (Figure 3F). In contrast, significantly more neurons became choice selective in the caudate (binomial comparison z-test for easy and hard rules; dACC: 13.3% versus 5.48%,  $z = 2.88$ ,  $p < 0.003$ ; Cd: 9.57% versus 1.86%,  $z = 1.72$ ,  $p < 0.05$ ; Pu: 9.95% versus 1.55%,  $z = 2$ ,  $p < 0.03$ ), with similar proportions of neurons being category specific in easy and hard rules (Figure 3D). Therefore, the modulation of single-neuron activity indicates that the dACC and putamen mainly represent the learned classification, whereas the caudate mainly represents the choice. Therefore, in the following analyses, we mainly focus on the dACC and putamen.

### A geometric representation to track neural dynamics during learning

We next examined how the modulation of neural activity develops during learning. To characterize the neural representations during rule acquisition, we represented each stimulus pattern using a complete set of statistically independent binary features (Figure 4A). Using this representation, we can then describe each of the rules as a linear weighted combination of these features. In other words, each rule can be described as a vector in this feature space, which we call the rule vector (Equation 4). We then express the selectivity of individual neurons using this set of basis features, which we call the neural vector (Equation 5). This gives a natural way to follow the neuron’s preference over time by computing the trajectory of its neural vector (using a rolling regression; Figures 4B and S3A–S3D).

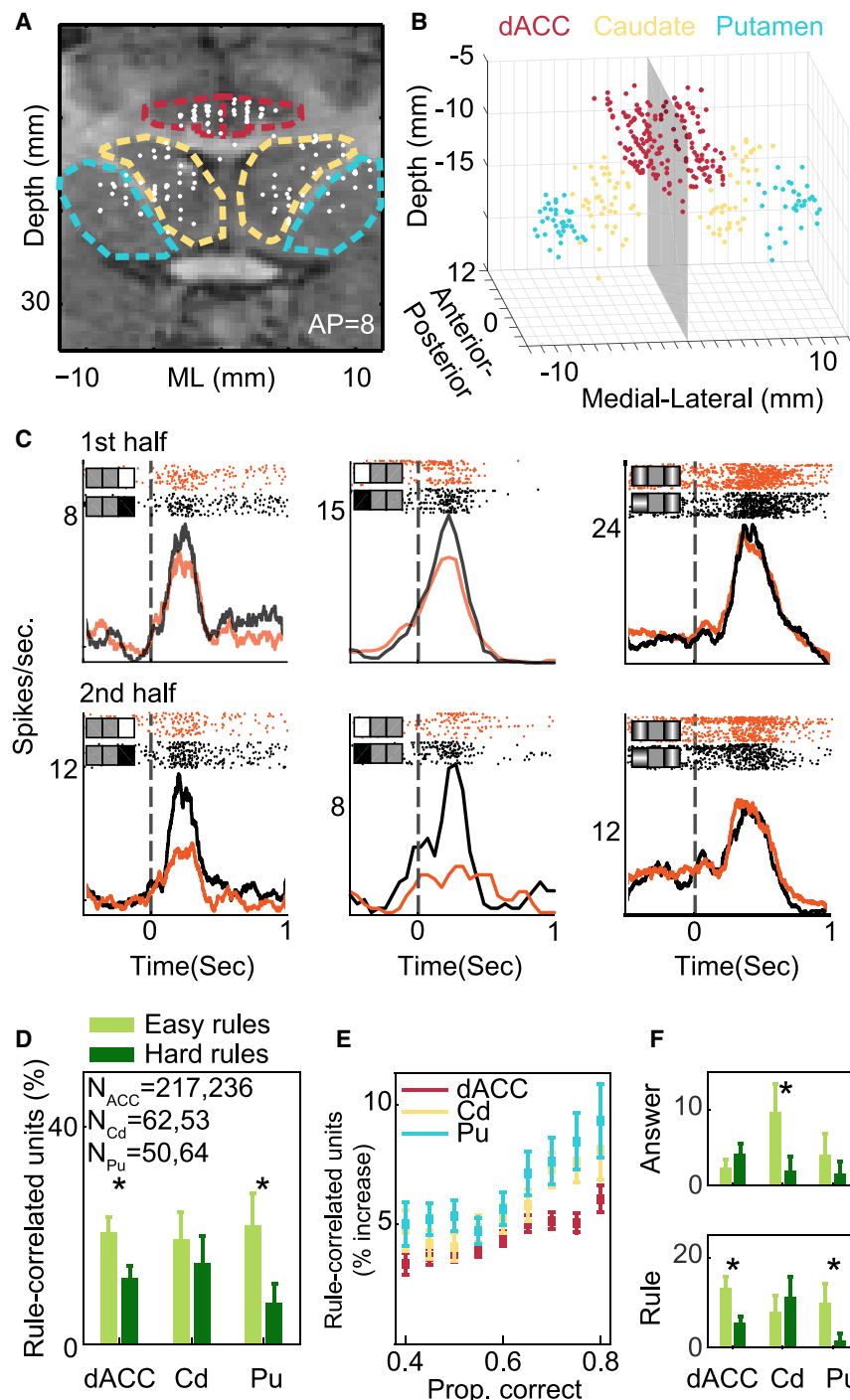
Importantly, because the neural vector is described in terms of movement in the space of features and by its relation to the rule vector, we can compare neural dynamics across sessions of different rules using the magnitude of the neural vector (vector magnitude) and its angle relative to the rule imposed in the session (angle to rule). The geometric projection of the neural vector onto the rule vector is equivalent to the correlation between neural activity and the categories determined by the rule (Lemma 3 corollary 2). This projection of the neural vector on the rule vector can increase by two independent mechanisms: changes in the vector magnitude (Figure 4B, left), suggesting an increase in confidence (Figure S3E), or by rotating toward the rule vector

(D) No retention between rule repetitions after several weeks. Use of eight different rules that use the same stimulus patterns over 4 weeks or more means that it would be very hard to memorize the different rules. Shown are two examples of repeats of the same rules over multiple sessions (scales of gray), with more than 4 weeks between repetitions. Over all sessions, there was no relationship between performance at the end of learning a rule and performance at the beginning of next repetition of the same rule (monkey D: light blue,  $r = 0.23$ ,  $p > 0.5$ ; monkey G: dark blue,  $r = 0.21$ ,  $p > 0.3$ ; both:  $r = 0.12$ ,  $p > 0.2$ ).

(E) Performance in all 3-bit rules. Shown is average performance in the last quarter of each session for all rules and all sessions for both monkeys (monkey D, light blue; monkey G, dark blue). For each rule, bars and error bars show the mean and SE over animals and sessions, classified into easy and hard rules (in easy rules, performance is significantly above chance level on average). The inset shows average learning curves for all rules, further justifying separation into easy and hard rules. Truth tables for all rules are shown below.

(F) Performance in 2-bit and in 4-bit rules (neural activity was recorded only during 3-bit rules). See also Figure S1.

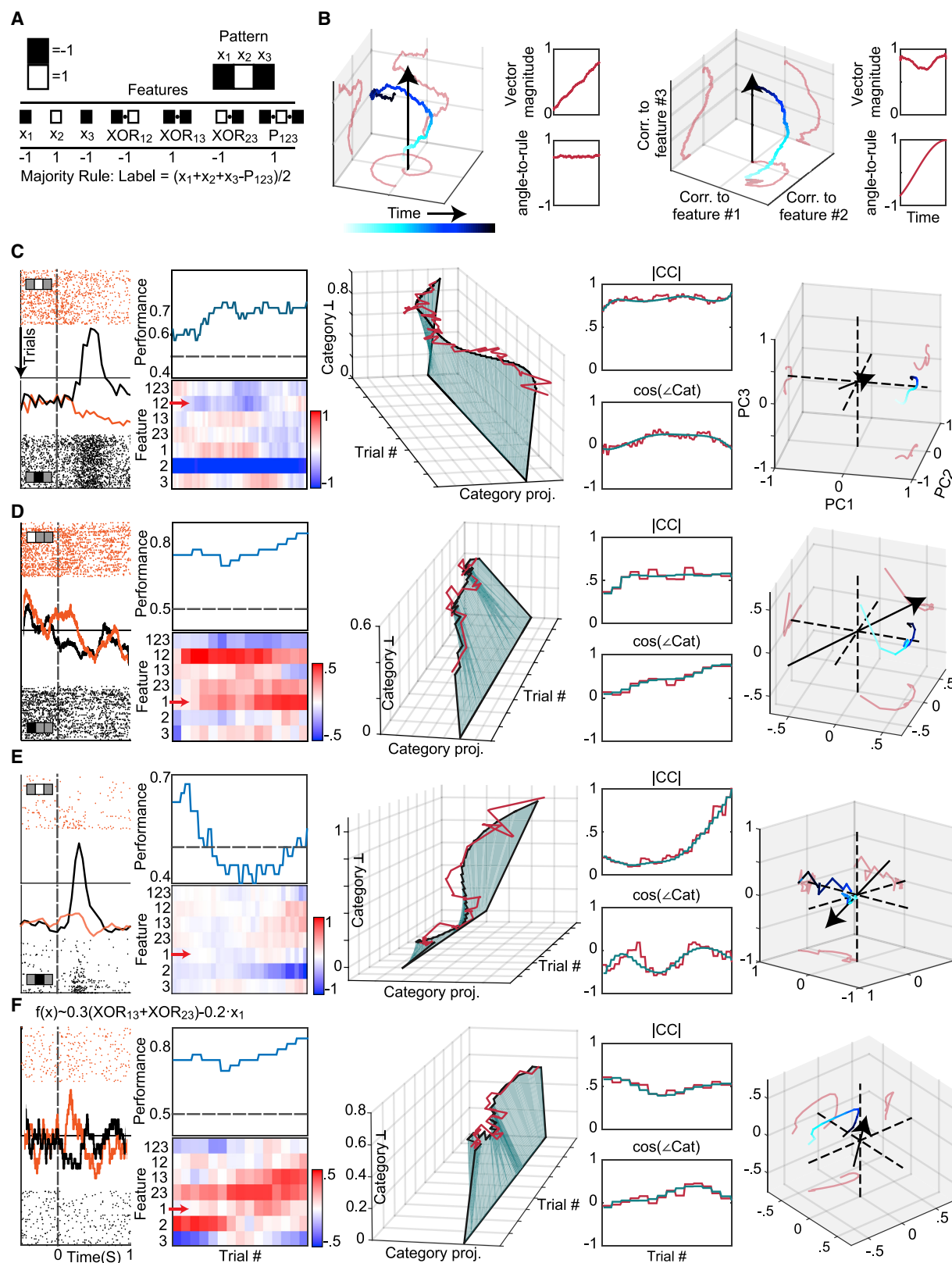




(smaller angle to rule; Figure 4B, right), suggesting a policy change. Choosing a different set of features to represent the rules and neural preferences would simply imply that this other basis is a linear combination of the current set of features, and so our analysis of the dynamics and our findings are independent of the specific set of features we used.

This geometrical representation of rules and neural responses revealed several types of neuronal dynamics: first, neurons that

(Figure 4E); and fourth, neurons that demonstrated changes in several features in parallel and, as a result, changes in rotation and magnitude (Figure 4F). These more complex trajectories can eventually lead to significant agreement with the rule vector (Figures S3F–S3J). Indeed, the two processes were not mutually exclusive, and many neurons with vector magnitude changes also exhibited angle-to-rule changes (Figures S5A and S5B).



(legend on next page)

### Differential representations of confidence and policy in the putamen and the dACC

To examine whether the two components of learning are represented differently in the two regions, we separately examined the increase in the magnitude of the neural vector and its rotation toward the rule vector (angle to rule). In the dACC and putamen, more neurons rotated toward the rule during learning of easy rules, resulting in having a smaller angle to rule (Figures 5A and 5B; binomial z-test comparing easy with hard rules; dACC: 40% versus 25.5%; putamen: 49% versus 34%; dACC:  $z = 3.32$ ,  $p < 0.001$ ; Pu:  $z = 1.68$ ,  $p < 0.05$ ). This rotation progressed gradually along the session as performance improved (Figures 5A and 5B, right panels). Similar findings were obtained when dividing sessions into high- and low-performance ones independent of the specific rule (Figures S4A–S4I). Moreover, the rotation across all dACC neurons correlated with the increase in performance and with performance at the end of the session for putamen neurons (Figure 5C).

Conversely, changes in vector magnitude between easy and hard rules were not observed in the dACC (Figure 5D; binomial z-test; dACC: 38% versus 34%,  $p > 0.1$ ) but did occur in the putamen, where more neurons increased their vector magnitude in easy rules (Figure 5E; binomial z-test; putamen: 47% versus 31%,  $z = 1.82$ ,  $p < 0.04$ ; Figures S4C and S4D: similar findings for high and low performance; Figure 5F: magnitude across all putamen neurons but not the dACC, correlated with performance). These changes became more prominent toward the end of the sessions (Figure 5E, right panels), suggesting strengthening of the recently acquired policy, and were accompanied by a reduction in response time (Figures S4J–S4L), further supporting the notion that they represent confidence.

In contrast to the angle and magnitude, there was no difference between easy and hard rules in terms of the representation of the actual choice in the dACC and putamen (Figures S4M and S4N). Complementing this and further supporting the conclusion that the dACC and putamen mainly represent learning, whereas

the caudate mainly represents the choice (Figure 3F), caudate neurons represented the actual choice differentially in easy versus hard rules (Figure S4O) but not for angle to rule or vector magnitude (Figures S4P–S4R).

These findings indicate that the putamen and dACC reflect a process of policy search for the correct rule; but although the dACC represents change of policy, the putamen also represents strengthening and confidence gain that accompany successful learning.

### Neuronal policy follows behavioral change

We first wanted to find out whether the representation depicted by the neural vectors shows similar dynamics as that of the behavior. We found that a large fraction of neurons in both regions showed dynamics of angle to rule that were correlated with the learning curve (Figures 6A and 6B; 32% of dACC neurons and 37% of putamen neurons, Pearson,  $p < 0.01$ ). Similarly, a large fraction of neurons changed their vector magnitude in correlation with performance (Figures 6C and 6D; 19% of dACC neurons and 18% of putamen neurons, Pearson,  $p < 0.01$ ; Figures S5A and S5B).

We hypothesized that a sequence of successful trials should lead to rotation of the neural vector; namely, adopting a more accurate policy. To test this, we measured the rotation by taking the derivative of the angle to rule and found that more neurons followed the behavior in easy rules than in hard ones when using a lag of four trials (Figure 6E; binomial z-test; dACC:  $z = 2.24$ ,  $p < 0.02$ ; Pu:  $z = 2.03$ ,  $p < 0.03$ ). The difference between easy and hard rules was smaller for lags of 8 trials as well as for zero lag (Figure 6E, bottom; Figure S5C). This suggests that, when a search for answers succeeds or fails beyond average, it is followed by a neural vector rotation. To test this more directly, we examined the distribution of angle to rule after a sequence of successful trials and found the angles to be smaller compared with the overall distribution (Figure 6F; both regions:  $p < 0.005$ ; putamen:  $p < 0.05$ ; dACC:  $p < 0.02$ ; Wilcoxon's rank-sum tests). The fact that the rotation

### Figure 4. A geometrical representation of neural activity reveals learning dynamics

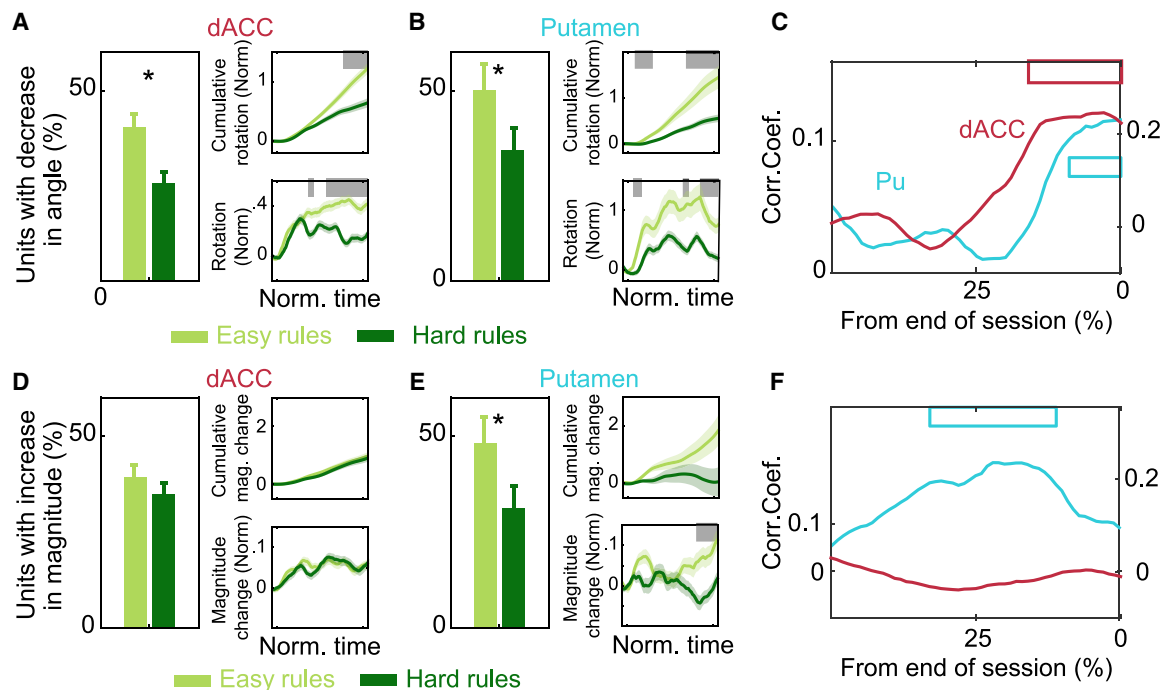
(A) The 7 features that form a spanning linear basis for the space in which all 3-bit rules reside. Shown is an example of the representation of the pattern  $\vec{x} = \blacksquare \square \blacksquare = (-1, 1, -1)$  by the set of binary features: the three first-order ones ( $\{\mathbf{x}_i\}_{i=1}^3$ ), three second-order ones ( $\{\mathbf{XOR}_{ij} = \mathbf{x}_i \cdot \mathbf{x}_j\}_{i < j=1}^3$ ), and a third-order one ( $\mathbf{P}_{123} = \mathbf{x}_1 \cdot \mathbf{x}_2 \cdot \mathbf{x}_3$ ). Also shown is a representation of the majority rule as a combination of these features (bottom).

(B) A schematic demonstrating that, in this feature space, the change in a neuron's response, the neural vector, modulates its projection on the rule vector (shown by the black arrow). An increase in the projection can result from two processes: a trajectory change that increases the neural vector magnitude (left) and/or neural vector rotation toward the direction of the rule vector—the angle to rule (right). The trajectory is color-coded to reflect time, and the red curves show the projections of the trajectory on the axes. Insets separate the dynamics of vector magnitude (top) and angle to rule (bottom). For presentation only, the neural trajectory is plotted in the space of the three main principal axes (PCs), but for all actual analyses, the neural vectors were computed in the complete 7D feature space.

(C–F) Examples of the neural vector dynamics of individual neurons. Each row shows the following for one neuron from left to right: (1) Raster plot aligned to stimulus onset, with trials ordered from top to bottom, divided by the two categories describing the neuron's preference at the end of the session (top, orange; bottom, black). Each raster is for the 4 patterns that together form a category. (2) Top: the behavioral learning curve (20 trials running average). Bottom: the regression correlation coefficients (blue–red color bar) for each feature in the basis (y axis) along all trials (x axis) construct the dynamic neural vector (Equation 5). Red arrow marks the feature that defined the correct label in the specific session. (3) The norm projections of neural vectors (red) on the category (x axis) and its orthogonal subspace (category  $\perp$ , z axis) over time (y axis). The blue surface connects (0,0) to the smoothed dynamics to visualize rotation and extension. (4) The neural vector magnitude (top) and angle to category (bottom) plotted over time. Red, raw data; blue, smoothed data. (5) The high-dimensional (7D) neural trajectory projected on the three main PCs and curve-fitted, color-coded for time progress in the session (light  $\rightarrow$  dark blue). The black arrow is the rule vector in that session, and the red curves are the projections of the trajectory on the axes. (C) shows a neuron that does not change its selectivity during learning; (D) shows a neuron that rotates toward the correct rule by learning the right feature; (E) shows a neuron that extends its vector toward a “wrong” feature; (F) shows a neuron that exhibits a complex relationship of vector extension and rotation.

See also Figure S3.





**Figure 5. Differential representation in the dACC and putamen**

(A and B) The proportion of neurons that significantly decreased their angle to rule during the session was different for easy rules versus hard rules in the dACC (A) and putamen (B) (error bars show SE,  $p < 0.05$  in both, binomial z-test). The right panels show the population-average angle change (bottom, SE in shaded color) and the normalized cumulative change (top). Sessions were time warped for averaging. Gray bars indicate significant difference between easy and hard rules ( $p < 0.05$ , bootstrap).

(C) Correlation coefficient between performance change and rotation for all dACC neurons and between performance at the end of the session and angle to rule for all putamen neurons. Bars mark the range of significant correlations (Pearson,  $p < 0.05$ ).

(D and E) The same as in (A) and (B) for change in vector magnitude. In contrast to rotation, only the putamen showed a significant increase in vector magnitude in easy versus hard rules ( $p < 0.05$ , binomial z-test).

(F) Correlation coefficient between performance change and vector magnitude for all dACC neurons and between performance at the end of the session and vector magnitude for all putamen neurons. Bars mark the range of significant correlations (Pearson,  $p < 0.05$ ).

See also Figure S4.

was directed toward the rule (smaller angles) indicates that the change is not only due to reward. These findings show that adopting a new correct policy by rotation of the neural vector toward the rule vector is more likely to occur after a sequence of successful responses.

### Representation of confidence in the putamen follows policy representation in the dACC

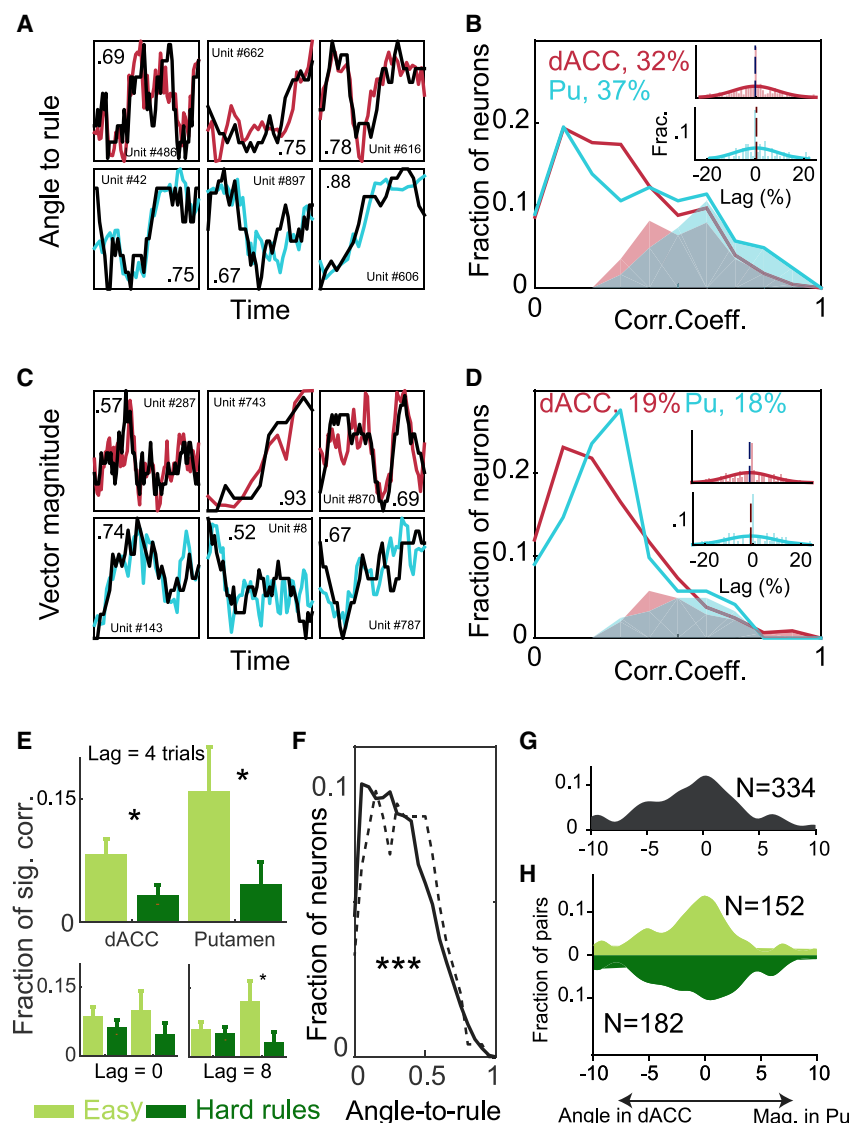
To examine temporal differences between the dACC and putamen, we also studied neurons that were recorded simultaneously from the two regions. Using pairs of neurons from the two areas that were recorded in the same session, we computed the correlation between all combinations of vector magnitude and angle to rule, across regions and within a region, and using different temporal lags. There was no directional temporal preference when comparing vector magnitude or angle to rule across regions or within a region or when taking angle to rule in the putamen and vector magnitude in the dACC (Figure S5D;  $p > 0.1$  for all, Wilcoxon's signed-rank tests). In contrast, only the lags between vector magnitude in the putamen and angle to rule in the dACC showed a clear directionality, with the putamen

magnitude following dACC rotation (Figure 6G; Wilcoxon signed-rank test,  $z = -2.81$ ,  $p < 0.005$ , mean lag  $-2.8 \pm 0.96$  trials). Moreover, this was more dominant in neural pairs recorded during easy rules (Figure 6H, top; Wilcoxon signed-rank test,  $z = -2.64$ ,  $p < 0.01$ , mean lag  $-3.56 \pm 1.36$  trials).

These results are in agreement with previous findings of information transfer in corticostriatal loops, showing that the interaction between the two regions occurs over a short time window. The findings also suggest that when a series of successful trials is followed by neural rotation toward the rule, the rotation in the dACC is followed by magnitude extension in the putamen.

### Neural policy at the end of a session predicts next-day behavior

If the neural vector at the end of a session reflects learning a policy during that session, then it might predict early performance in the session of the next day. This would also point to a retention mechanism. To examine this, we projected each neuron's neural vector at the end of a session onto the rule vector of the following day and computed the correlation between these projections and the behavior from the next day. Indeed, neural vectors in



**Figure 6. Differential learning dynamics in the putamen and dACC**

(A) Change in angle to rule during a session overlaid with the performance behavior (black), showing a highly similar temporal pattern. Three dACC neurons (top row, red) and three putamen neurons (bottom row, cyan), all with significant correlation between neural dynamics and performance ( $p < 0.01$  for all, Pearson).

(B) Histograms of correlation coefficients for all neurons, with the shaded area marking neurons exhibiting significant correlations at  $p < 0.01$ . Insets show the distribution of trial lags between neural dynamics and behavior, with the mean lag not different than zero ( $p > 0.1$ ,  $t$  tests).

(C) The same as (A) for dynamics in vector magnitude.

(D) The same as (B) for dynamics in vector magnitude.

(E) Local shifts in angle to rule follow the behavior in more neurons when comparing easy with hard rules, but only in a lag of 4 trials (main panel) and not in zero- or 8-trial lags (bottom insets). Error bars show SE.

(F) The distribution of angle to rule (absolute cosine value) at the end of all strips of 10 successful trials (dashed line) compared with the distribution of all angles (solid line) ( $p < 0.005$ , Wilcoxon rank-sum test).

(G) Distribution of optimal lags for all simultaneously recorded pairs of neurons between vector magnitude of the putamen neuron and angle to rule of the dACC neuron. The mean is significantly below zero, indicating that changes in angle in dACC neurons preceded changes in magnitude in putamen neurons.

(H) The same as (G) but separately for easy and hard rules.

See also [Figure S5](#).

## DISCUSSION

We studied the dynamics of neural representations in the dACC and the striatum while monkeys learned to classify multi-cue patterns. We developed a novel mathematical

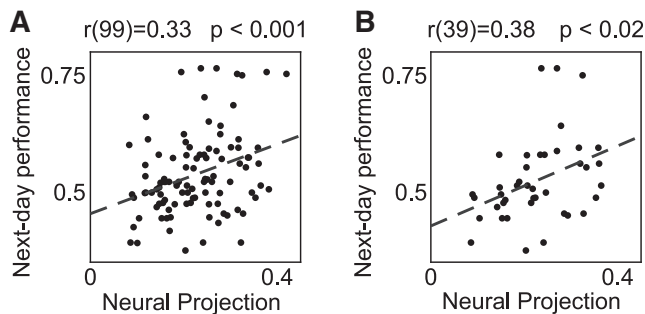
approach to represent neural selectivity and its properties for a rich set of rules that changed every few days. To characterize the neural dynamics during such *de novo* learning, we represented each rule in the space formed by a spanning set of the stimulus features, which enabled us to explore the information carried by single neurons as a vector in that space. The angle between the neural vector and the rule vector reflects learning-related policy and change in strategy, and the magnitude of the neural vector reflects changes in neural confidence. We then used these two geometric traits as a framework for exploring the dynamics during learning and across rules.

This approach revealed a dissociation in functionality; neurons in the dACC mostly rotated to decrease their angle to rule (changed their strategy), whereas neurons in the putamen mostly changed their activity to increase the magnitude of the neural vector, reflecting confidence and reinforcement of the correct strategy (Graybiel and Grafton, 2015). In line with this interpretation and the role of the striatum in reinforcement, rotation

the putamen predicted next-day early performance (Figure 7A; Pearson's  $r(99) = 0.33(0.15, 0.5)$ ,  $p < 0.0007$ ; no correlation was found when reversing the days as a control). Therefore, the closer the neural vector was at the end of one day to the rule used the next day, the better the initial performance.

We further tested whether the neural vectors also bias learning in sessions when a new rule was presented the next day and again found a significant correlation for putamen neurons (Figure 7B; Pearson's  $r(39) = 0.38(0.08, 0.62)$ ,  $p < 0.014$ ; no correlation when reversing days as a control). Together with the finding that the angle to rule across putamen neurons correlated with performance at the end of the session (Figure 5C), this suggests an overnight retention of policy in single neurons.

These results establish the effectiveness of using the geometric framework for identifying functional representations in neural activity, further showing that the policy represented by the angle to rule at the end of daily learning affects and biases performance the next day.



**Figure 7. Neural representation at the end of a session predicts next-day performance**

(A) The projection of each putamen neuron's policy (angle to rule) at the end of the session onto the next-day rule (rule vector) against the mean performance at the beginning of the next-day session (the maximal neural vector projection is calculated over the last 25% of the session, and performance is averaged over the first 25% of the next-day session), showing a positive correlation ( $r = 0.33$ ,  $p < 0.001$ , black regression line).

(B) Similar to (A) but when the rule changed overnight, again showing a positive correlation ( $r = 0.38$ ,  $p < 0.02$ ).

changes in the dACC were followed by magnitude extensions in the putamen. In addition, we found that neurons in the putamen were related to overnight retention because their representation at the end of a day predicted next-day behavior. Because the classification task was complex and learning usually continued over sessions, this suggests that the striatum represents the policy used for retention within striatal populations or by transfer to other regions (e.g., as in a consolidation processes). This interpretation further supports previous studies showing that the striatum maintains intermediate representations, potentially via sustained activity (Deffains et al., 2016), and by studies showing that the striatum supports learning under spaced conditions that require reinforcement and memory (Doll et al., 2015; Wimmer et al., 2018).

The geometrical framework we present here allows identification of neurons that retain stable representation and ones that change during learning (Chen et al., 2001; Genovesio et al., 2005; Sadtlir et al., 2014). Interestingly, these changes were observed during successful and unsuccessful learning sessions, suggesting that single-neuron dynamics reflect a mixture of the search for a new policy and memory for previously acquired rules. However, although most neurons change by rotating toward the correct rule, some change in other directions. One possible interpretation is consistent with the common view that behavioral decisions rely on averaging of large neural populations where the misguided neurons are overruled. Alternatively, this could suggest a strategy of maintaining representations of other rules but suppressing their effect on the decision (Klavir et al., 2012), which would be a beneficial memory for future (different) tasks. Our approach would enable examination of such memories and their relative strength (e.g., by characterizing the rules to which the neurons that do not rotate point). It could also enable prediction of behavioral biases by examining the distance of a new rule from the preferred direction of the set of observed neurons. The notion is directly related to the ability to generalize from one rule to another and could explain why

some rules are learned better when presented in succession whereas others are not (Oby et al., 2019). This is in line with the idea that similarity in the space of neural activity defines a metric on the stimulus or action space (Bernardi et al., 2020; Ganmor et al., 2015; Pryluk and Paz, 2019; Resnik and Paz, 2015; Tkačik et al., 2013).

In contrast to the putamen and the dACC, caudate neurons did not show substantial representation of the learned rule and instead reflected the instantaneous choice. This was not a motor-related activity per se because it was stronger for easy rules than for hard ones. Because in the current task the objective value of stimuli is constant, the difference observed between rule types might suggest that caudate neurons are sensitive to the value of applying the chosen behavioral policy (internal, even when wrong) or with representation of subjective value (Cai et al., 2011). This is also in line with findings showing that, in visuomotor associations, the dorsal striatum represents the value of actions (Lau and Glimcher, 2008) and follows the choice representation observed in the lateral PFC (Seo et al., 2012) but can also learn without the LPFC (Minamimoto et al., 2010). Our observations therefore support the notion that, in rule learning, caudate activity reflects value-based action selection (Cai et al., 2011; Desrochers et al., 2015; Kim and Hikosaka, 2013; Williams and Eskandar, 2006; Yanike and Ferrera, 2014).

Our analyses were focused on single-neuron activity, demonstrating that we can track dynamics of representations even at the single-cell level. This complements the recent emphasis on neural ensembles to decode behavior from mixed representations (Golub et al., 2018; Grewe et al., 2017; Gründemann et al., 2019; Karpas et al., 2019; Levy et al., 2019; Mante et al., 2013; Maoz et al., 2020; Rigotti et al., 2013). The readout of a neural ensemble can improve performance because of a change in individual neuron properties or because of a change in weights given to each neuron by a readout node. Our results offer a window into single-neuron properties and dissect task-relevant dynamics and therefore provide novel insights into the mechanisms involved in the changes seen in population dynamics during learning (Figure S7). Our results here show categorical properties after learning is complete (Figure S6), in agreement with recent studies of prefrontal neurons (Hirokawa et al., 2019; Onken et al., 2019). In addition, here we describe the dynamics that led them to develop these final representations.

Overall, here we developed a new computational framework to examine dynamics of neural representations during *de novo* learning and used it to shed light on two complementing roles that are necessary to accomplish any learning of classification rules: acquisition of policy and strengthening of confidence. We show that the former is more represented in the anterior cingulate cortex, whereas the latter is more represented by the putamen. Future work is required to understand how abnormalities in extended cingulate networks that enable rich learning in primates (Pryluk et al., 2019) can result in maladaptive processes that would lead to applying incorrect rules and, in extreme cases, even lead to psychopathologies (Averbeck and Chafee, 2016; Dunsmoor and Paz, 2015; Lee, 2013; Likhtik and Paz, 2015).

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Materials and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Ethics declaration
  - Subjects
- **METHOD DETAILS**
  - Animal training
  - Experiment sessions
  - Neural recordings
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Behavior analysis
  - Neural activity analysis

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.neuron.2020.12.027>.

## ACKNOWLEDGMENTS

We thank Yosef Shohat for animal training, welfare, and experimental procedures; Dr. Yoav Kfir for scientific consulting; Dr. Eilat Kahana for help with medical and surgical procedures; Dr. Edna Furman-Haran and Fanny Attar for MRI procedures; and the members of the Paz and Schneidman labs. E.S. was supported by European Research Council grant 311238, Israel Science Foundation grant 1629/12, a CRCNS grant, and Simons Collaboration on the Global Brain grant 542997, as well as research support from Martin Kushner Schnur and Mr. and Mrs. Lawrence Feis. R.P. was supported by Israel Science Foundation grant ISF #2352/19 and European Research Council grant ERC-2016-CoG #724910.

## AUTHOR CONTRIBUTIONS

Y.C., E.S., and R.P. conceived and designed the study. Y.C. performed the experiments and analyses. Y.C., E.S., and R.P. wrote the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 25, 2020

Revised: November 7, 2020

Accepted: December 30, 2020

Published: January 22, 2021

## REFERENCES

- Averbeck, B.B., and Chafee, M.V. (2016). Using model systems to understand errant plasticity mechanisms in psychiatric disorders. *Nat. Neurosci.* **19**, 1418–1425.
- Averbeck, B.B., and Costa, V.D. (2017). Motivational neural circuits underlying reinforcement learning. *Nat. Neurosci.* **20**, 505–512.
- Averbeck, B.B., Lehman, J., Jacobson, M., and Haber, S.N. (2014). Estimates of projection overlap and zones of convergence within frontal-striatal circuits. *J. Neurosci.* **34**, 9497–9505.
- Balleine, B.W., Delgado, M.R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* **27**, 8161–8165.
- Bernardi, S., Benna, M.K., Rigotti, M., Munuera, J., Fusi, S., and Salzman, C.D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell* **183**, 954–967.e21.
- Brasted, P.J., and Wise, S.P. (2004). Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. *Eur. J. Neurosci.* **19**, 721–740.
- Buch, E.R., Brasted, P.J., and Wise, S.P. (2006). Comparison of population activity in the dorsal premotor cortex and putamen during the learning of arbitrary visuomotor mappings. *Exp. Brain Res.* **169**, 69–84.
- 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.
- Cai, X., Kim, S., and Lee, D. (2011). Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* **69**, 170–182.
- Chen, N.H., White, I.M., and Wise, S.P. (2001). Neuronal activity in dorsomedial frontal cortex and prefrontal cortex reflecting irrelevant stimulus dimensions. *Exp. Brain Res.* **139**, 116–119.
- 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.
- Cohen, Y., and Schneidman, E. (2013). High-order feature-based mixture models of classification learning predict individual learning curves and enable personalized teaching. *Proc. Natl. Acad. Sci. USA* **110**, 684–689.
- Cromer, J.A., Roy, J.E., and Miller, E.K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* **66**, 796–807.
- Deffains, M., Iskhakova, L., Katabi, S., Haber, S.N., Israel, Z., and Bergman, H. (2016). Subthalamic, not striatal, activity correlates with basal ganglia downstream activity in normal and parkinsonian monkeys. *eLife* **5**, e16443.
- Desrochers, T.M., Amemori, K., and Graybiel, A.M. (2015). Habit Learning by Naïve Macaques Is Marked by Response Sharpening of Striatal Neurons Representing the Cost and Outcome of Acquired Action Sequences. *Neuron* **87**, 853–868.
- Doll, B.B., Shohamy, D., and Daw, N.D. (2015). Multiple memory systems as substrates for multiple decision systems. *Neurobiol. Learn. Mem.* **117**, 4–13.
- Dunsmoor, J.E., and Paz, R. (2015). Fear Generalization and Anxiety: Behavioral and Neural Mechanisms. *Biol. Psychiatry* **78**, 336–343.
- Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature* **407**, 630–633.
- Freedman, D.J., and Assad, J.A. (2016). Neuronal Mechanisms of Visual Categorization: An Abstract View on Decision Making. *Annu. Rev. Neurosci.* **39**, 129–147.
- Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J. Neurosci.* **23**, 5235–5246.
- Ganmor, E., Segev, R., and Schneidman, E. (2015). A thesaurus for a neural population code. *eLife* **4**, e06134.
- Genovesio, A., Brasted, P.J., Mitz, A.R., and Wise, S.P. (2005). Prefrontal cortex activity related to abstract response strategies. *Neuron* **47**, 307–320.
- Gluck, M.A., Shohamy, D., and Myers, C. (2002). How do people solve the “weather prediction” task?: individual variability in strategies for probabilistic category learning. *Learn. Mem.* **9**, 408–418.
- Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574.
- Golub, M.D., Sadtler, P.T., Oby, E.R., Quick, K.M., Ryu, S.I., Tyler-Kabara, E.C., Batista, A.P., Chase, S.M., and Yu, B.M. (2018). Learning by neural reassociation. *Nat. Neurosci.* **21**, 607–616.



- Goodman, N.D., Tenenbaum, J.B., Feldman, J., and Griffiths, T.L. (2008). A rational analysis of rule-based concept learning. *Cogn. Sci.* 32, 108–154.
- Graybiel, A.M., and Grafton, S.T. (2015). The striatum: where skills and habits meet. *Cold Spring Harb. Perspect. Biol.* 7, a021691.
- Grewe, B.F., Gründemann, J., Kitch, L.J., Lecoq, J.A., Parker, J.G., Marshall, J.D., Larkin, M.C., Jercog, P.E., Grenier, F., Li, J.Z., et al. (2017). Neural ensemble dynamics underlying a long-term associative memory. *Nature* 543, 670–675.
- Gründemann, J., Bitterman, Y., Lu, T., Krabbe, S., Grewe, B.F., Schnitzer, M.J., and Lüthi, A. (2019). Amygdala ensembles encode behavioral states. *Science* 364, eaav8736.
- Haroush, K., and Williams, Z.M. (2015). Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell* 160, 1233–1245.
- Hayden, B.Y., and Platt, M.L. (2010). Neurons in anterior cingulate cortex multiplex information about reward and action. *J. Neurosci.* 30, 3339–3346.
- Heilbronner, S.R., and Hayden, B.Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annu. Rev. Neurosci.* 39, 149–170.
- Heilbronner, S.R., Rodriguez-Romaguera, J., Quirk, G.J., Groenewegen, H.J., and Haber, S.N. (2016). Circuit-Based Corticostriatal Homologies Between Rat and Primate. *Biol. Psychiatry* 80, 509–521.
- Hirokawa, J., Vaughan, A., Masset, P., Ott, T., and Kepecs, A. (2019). Frontal cortex neuron types categorically encode single decision variables. *Nature* 576, 446–451.
- Histed, M.H., Pasupathy, A., and Miller, E.K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63, 244–253.
- Jin, X., and Costa, R.M. (2015). Shaping action sequences in basal ganglia circuits. *Curr. Opin. Neurobiol.* 33, 188–196.
- Karpas, E., Maoz, O., Kiani, R., and Schneidman, E. (2019). Strongly correlated spatiotemporal encoding and simple decoding in the prefrontal cortex. *bioRxiv*. <https://doi.org/10.1101/693192>.
- Kim, H.F., and Hikosaka, O. (2013). Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron* 79, 1001–1010.
- Kim, J.N., and Shadlen, M.N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat. Neurosci.* 2, 176–185.
- Klavir, O., Genuit-Gabai, R., and Paz, R. (2012). Low-frequency stimulation depresses the primate anterior-cingulate-cortex and prevents spontaneous recovery of aversive memories. *J. Neurosci.* 32, 8589–8597.
- 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.
- Lagnado, D.A., Newell, B.R., Kahan, S., and Shanks, D.R. (2006). Insight and strategy in multiple-cue learning. *J. Exp. Psychol. Gen.* 135, 162–183.
- Lau, B., and Glimcher, P.W. (2008). Value representations in the primate striatum during matching behavior. *Neuron* 58, 451–463.
- Lee, D. (2013). Decision making: from neuroscience to psychiatry. *Neuron* 78, 233–248.
- Lee, D., Rushworth, M.F., Walton, M.E., Watanabe, M., and Sakagami, M. (2007). Functional specialization of the primate frontal cortex during decision making. *J. Neurosci.* 27, 8170–8173.
- Levy, D.R., Tamir, T., Kaufman, M., Parabucki, A., Weissbrod, A., Schneidman, E., and Yizhar, O. (2019). Dynamics of social representation in the mouse prefrontal cortex. *Nat. Neurosci.* 22, 2013–2022.
- Likhtik, E., and Paz, R. (2015). Amygdala-prefrontal interactions in (mal)adaptive learning. *Trends Neurosci.* 38, 158–166.
- Mansouri, F.A., Tanaka, K., and Buckley, M.J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nat. Rev. Neurosci.* 10, 141–152.
- Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 503, 78–84.
- Maoz, O., Tkačik, G., Esteki, M.S., Kiani, R., and Schneidman, E. (2020). Learning probabilistic neural representations with randomly connected circuits. *Proc. Natl. Acad. Sci. USA* 117, 25066–25073.
- Meeter, M., Radics, G., Myers, C.E., Gluck, M.A., and Hopkins, R.O. (2008). Probabilistic categorization: how do normal participants and amnesic patients do it? *Neurosci. Biobehav. Rev.* 32, 237–248.
- Merchant, H., Zainos, A., Hernández, A., Salinas, E., and Romo, R. (1997). Functional properties of primate putamen neurons during the categorization of tactile stimuli. *J. Neurophysiol.* 77, 1132–1154.
- Minamimoto, T., Saunders, R.C., and Richmond, B.J. (2010). Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron* 66, 501–507.
- Mitz, A.R., Godschalk, M., and Wise, S.P. (1991). Learning-dependent neuronal activity in the premotor cortex: activity during the acquisition of conditional motor associations. *J. Neurosci.* 11, 1855–1872.
- Muhammad, R., Wallis, J.D., and Miller, E.K. (2006). A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. *J. Cogn. Neurosci.* 18, 974–989.
- Nosofsky, R.M., Kruschke, J.K., and McKinley, S.C. (1992). Combining exemplar-based category representations and connectionist learning rules. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 211–233.
- Oby, E.R., Golub, M.D., Hennig, J.A., Degenhart, A.D., Tyler-Kabara, E.C., Yu, B.M., Chase, S.M., and Batista, A.P. (2019). New neural activity patterns emerge with long-term learning. *Proc. Natl. Acad. Sci. USA* 116, 15210–15215.
- Ongür, D., and Price, J.L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* 10, 206–219.
- Onken, A., Xie, J., Panzeri, S., and Padoa-Schioppa, C. (2019). Categorical encoding of decision variables in orbitofrontal cortex. *PLoS Comput. Biol.* 15, e1006667.
- Padoa-Schioppa, C., and Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226.
- Pryluk, R., and Paz, R. (2019). Learning outside the box. *Proc. Natl. Acad. Sci. USA* 116, 15316–15318.
- Pryluk, R., Kfir, Y., Gelbard-Sagiv, H., Fried, I., and Paz, R. (2019). A Tradeoff in the Neural Code across Regions and Species. *Cell* 176, 597–609.e18.
- Resnik, J., and Paz, R. (2015). Fear generalization in the primate amygdala. *Nat. Neurosci.* 18, 188–190.
- Rigotti, M., Barak, O., Warden, M.R., Wang, X.J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590.
- Rudebeck, P.H., Behrens, T.E., Kennerley, S.W., Baxter, M.G., Buckley, M.J., Walton, M.E., and Rushworth, M.F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J. Neurosci.* 28, 13775–13785.
- Rushworth, M.F., and Behrens, T.E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* 11, 389–397.
- Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., and Batista, A.P. (2014). Neural constraints on learning. *Nature* 512, 423–426.
- Saez, A., Rigotti, M., Ostojic, S., Fusi, S., and Salzman, C.D. (2015). Abstract Context Representations in Primate Amygdala and Prefrontal Cortex. *Neuron* 87, 869–881.
- Seger, C.A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neurosci. Biobehav. Rev.* 32, 265–278.
- Seger, C.A., and Miller, E.K. (2010). Category learning in the brain. *Annu. Rev. Neurosci.* 33, 203–219.
- Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* 27, 8366–8377.



- Seo, H., and Lee, D. (2009). Behavioral and neural changes after gains and losses of conditioned reinforcers. *J. Neurosci.* 29, 3627–3641.
- Seo, M., Lee, E., and Averbeck, B.B. (2012). Action selection and action value in frontal-striatal circuits. *Neuron* 74, 947–960.
- Shepard, R.N., Hovland, C.I., and Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychol. Monogr.* 75, 1–42.
- Shohamy, D., Myers, C.E., Kalanithi, J., and Gluck, M.A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neurosci. Biobehav. Rev.* 32, 219–236.
- Speekenbrink, M., Lagnado, D.A., Wilkinson, L., Jahanshahi, M., and Shanks, D.R. (2010). Models of probabilistic category learning in Parkinson's disease: Strategy use and the effects of L-dopa. *J. Math. Psychol.* 54, 123–136.
- Stuss, D.T., Levine, B., Alexander, M.P., Hong, J., Palumbo, C., Hamer, L., Murphy, K.J., and Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia* 38, 388–402.
- Tkačik, G., Granot-Atedgi, E., Segev, R., and Schneidman, E. (2013). Retinal metric: a stimulus distance measure derived from population neural responses. *Phys. Rev. Lett.* 110, 058104.
- van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 2579–2605.
- Wallis, J.D., and Kennerley, S.W. (2011). Contrasting reward signals in the orbitofrontal cortex and anterior cingulate cortex. *Ann. N Y Acad. Sci.* 1239, 33–42.
- Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956.
- Williams, Z.M., and Eskandar, E.N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat. Neurosci.* 9, 562–568.
- Wimmer, G.E., Li, J.K., Gorgolewski, K.J., and Poldrack, R.A. (2018). Reward Learning over Weeks Versus Minutes Increases the Neural Representation of Value in the Human Brain. *J. Neurosci.* 38, 7649–7666.
- Yang, T., and Shadlen, M.N. (2007). Probabilistic reasoning by neurons. *Nature* 447, 1075–1080.
- Yanike, M., and Ferrera, V.P. (2014). Representation of outcome risk and action in the anterior caudate nucleus. *J. Neurosci.* 34, 3279–3290.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Critical commercial assays</b>		
0.6–1.2 M $\Omega$ glass coated tungsten electrodes	Alpha Omega	N/A
metal guide (outer diameter: 0.51 mm, inner diameter: 0.41 mm)	Cadence	Gauge 25xxtw
Head-tower and electrode-positioning-system	Alpha Omega	N/A
Recording chamber	Alpha Omega	N/A
<b>Deposited data</b>		
Raw electrophysiology and behavior data	This manuscript	Upon request
<b>Experimental models: organisms/strains</b>		
Macaca Fascicularis monkeys	N/A	N/A
<b>Software and algorithms</b>		
Alpha Lab SNR	Alpha Omega	N/A
Offline sorter	Plexon	N/A
MATLAB	Mathworks	N/A
In-house developed code	This manuscript	<a href="https://github.com/yardencsGitHub/CohenSchneidmanPaz2020_Code">https://github.com/yardencsGitHub/CohenSchneidmanPaz2020_Code</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to Yarden Cohen ([ycohen1@mg.harvard.edu](mailto:ycohen1@mg.harvard.edu)), Elad Schneidman ([elad.schneidman@weizmann.ac.il](mailto:elad.schneidman@weizmann.ac.il)), or Rony Paz ([rony.paz@weizmann.ac.il](mailto:rony.paz@weizmann.ac.il))

#### Materials availability

This study did not generate new unique reagents.

#### Materials and code availability

Data will be supplied upon request. All custom-made code in this manuscript is publicly available in the Github repository [https://github.com/yardencsGitHub/CohenSchneidmanPaz2020\\_Code](https://github.com/yardencsGitHub/CohenSchneidmanPaz2020_Code).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Ethics declaration

All surgical and experimental procedures were approved and conducted in accordance with the regulations of the Weizmann Institute Animal Care and Use Committee, following National Institutes of Health regulations and with accreditation from the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC).

#### Subjects

Two healthy male monkeys (Monkeys G and D, macaca fascicularis, 4–6kg, ages 3–5) were chosen from our colony and participated in the experiment. The animals were housed in pairs in cages according to EU and NIH standards, in rooms with controlled temperature, humidity and daylight-cycle. The cages included toys, hideouts, and installments in different heights allowing the monkeys to perform their natural social behaviors. When not participating in behavior training or experiments, the animals had unlimited access to food and water. Animals participating in training and experiments were under a fluid restriction regime that always exceeded the minimal daily consumption of 20 mL/kg body weight (see Guidelines for Use of Fluid Regulation for Nonhuman Primates in Biomedical Research published by AAALAC).

The animals did not participate in other experimental procedures prior to this study.

## METHOD DETAILS

### Animal training

Two male monkeys (Monkeys G and D, macaca fascicularis, 4–6 kg) participated in the experiment. Before data collection, each monkey went through a training phase that acquainted it with all the task components and their sequence in a learning session. Both monkeys were trained similarly. They first learned rules with 2-bit patterns to understand the concept of the task. Then, immediately before the surgery, they experienced three rules with 3-bit patterns so that they would not be surprised when they see 3-bits for the first time during recordings. Importantly, these were 3 different rules than the 8 rules tested during recordings and shown in the manuscript (there are 256 possible for assigning 8 patterns into 2 categories). Other than that, no training was done. Therefore, in the recording sessions the monkeys were familiar with the concept of patterns and classification (Figure 1A). However, the monkeys did not experience any of the rules reported in this manuscript before the electrophysiological recordings began.

### Experiment sessions

The monkeys learned to classify binary patterns of  $N = 3$  squares. In each session, the entire set of  $2^N$  possible patterns was presented. The order of patterns was generated by concatenating full sets of randomly ordered  $2^N$  patterns. This process ensured that all patterns appear with the same temporal frequency and that no choice of behavioral rule, apart from the correct one, is beneficial in large portions of the session. For compactness we refer to the rules by their constituent squares. So, for example, in rule ‘3’ the label is determined by the color of the 3<sup>rd</sup> square and in rule ‘12’ the label is determined by the XOR of squares 1 and 2. See Figures 1, 2, and S1 and main text for the list of rules used in this study and during recordings. Each classification rule was replaced every few days and repeated after about a month and after a full cycle of the 8 rules was presented.

### Neural recordings

#### Surgery

A craniotomy was performed under deep anesthesia and aseptic conditions and a recording chamber (27x27mm) was implanted above the midline and anterior commissure to allow daily electrodes insertion. The chamber’s positioning was done according to MRI calculated coordinates with respect to the identified bone structure around the ear canals and eye sockets. Still images were taken during the surgery to record the location of the chamber, the head holder and the screws on the skull for easier extraction process.

After surgery the monkeys were treated with analgesics (Buprenorphine) and antibiotics (Rocephin, Baytril). The monkeys were allowed to recover for 1–2 weeks before the first head restraining in the setup. The fluid consumption regime was gradually reinstated starting two weeks after surgery.

#### MRI-Based Electrode Positioning

Anatomical MRI scans were acquired before, during, and after the recording period. Images were acquired on a 3-Tesla MRI scanner: (MAGNETOM Trio, Siemens) with a CP knee coil (Siemens). A T1-weighted, three-dimensional gradient-echo (MPRAGE) pulse sequence was acquired with a repetition time of 2,500 ms, an inversion time of 1,100 ms, an echo time of 3.36 ms, an 8 flip angle, and two averages. Images were acquired in the sagittal plane, 192 × 192 matrix, and 0.63 mm resolution. The first scan was performed before surgery and used to align and refine anatomical maps for each individual animal (relative location of the dACC and the striatum, and anatomical markers such as the interaural line and the anterior commissure; confirmed using atlas). We used this scan to guide the positioning of the chamber on the skull at the surgery. After surgery, we performed another scan with 2–4 electrodes directed toward the dACC, Putamen and caudate. The regions’ depth was calculated from the dura surface and the plane of the top of the chamber. We assessed estimation of electrode tip locations and comparison to the MRI image with < 1 mm accuracy (mean = 0.5 mm).

#### Mapping recording regions

During the first week of electrode insertions we performed a mapping procedure to identify the depth of cell bodies in prominent recording regions. During that week no behavior recordings were made and the fluid restriction was gradually reinstated.

Additionally, with every electrode insertion during the experiment we recorded the depths of cell bodies and were able to reconstruct the boundaries of our regions of interest.

#### Electrophysiology

The monkeys were seated in a dark room and each day, up to six microelectrodes (0.6–1.2 MΩ glass coated tungsten, Alpha Omega) were lowered inside a metal guide (Gauge 25xxtw, outer diameter: 0.51 mm, inner diameter: 0.41 mm, Cadence) into the brain using a head-tower and electrode-positioning-system (Alpha-Omega). The guide was lowered to penetrate and cross the dura and stopped once in the superficial layer of the cortex. The electrodes were then moved independently further into either the dACC, Caudate, or Putamen. Electrode signals were pre-amplified, 0.3 Hz–6 kHz band-pass filtered, and sampled at 44 kHz; and online spike sorting was performed using a template-based algorithm (Alpha Lab SNR, Alpha Omega). We allowed 15–30 minutes for the tissue and signal to stabilize before starting acquisition and behavioral protocol. At the end of the recording period, offline spike sorting was further performed for all sessions to improve unit isolation (offline sorter, Plexon).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Behavior analysis

#### Performance

Each learning session results in a series of correct and incorrect answers,  $\{y_t\}_{t=1:T} \in \{0, 1\}^T$ ,  $T$  being the number of trials. To measure learning behavior and account for erratic tendencies we took the following steps:

- 1) To avoid the behavioral decline that may bias performance at the end of the sessions we disregarded up to the last 10% of the session if it contained only wrong answers. On average we ended up ignoring  $\sim 1\%$  or 2-3 trials in each session.
- 2) We define performance at the end of the session by averaging correct and incorrect answers in the last quarter of the session,  $P_{end} = \langle y_t \rangle_{t \in [\frac{3}{4}T \rightarrow T]}$ . The confidence level for rejecting the null hypothesis of chance performance follows the regularized incomplete beta function,  $I_{1-p}\left(\frac{T}{4} - k, 1 + k\right)$ , where  $\frac{T}{4}$  is the number of trials in the last quarter of the session and  $k$  is the number of correct answers during that segment (Figure S1B).
- 3) Identically to  $P_{end}$ , we define  $P_{start}$  as the mean performance during the first quarter of the session.
- 4) In Figure 2B, we define maximal performance as the best mean performance in 30 consecutive trials,  $P_{max} = \max_{\tau \leq T-29} \langle y_t \rangle_{t \in [\tau \rightarrow \tau+29]}$ .
- 5) In Figure S1 we use pattern-specific performance. The order of pattern presentation, randomized batches containing all  $2^N$  patterns, guaranteed that the sequences of pattern-specific presentations were perfectly-interleaved – allowing for the comparison of pattern-specific errors conditioned on prior presentations of the same pattern (Figure S1C) or other patterns (Figure S1D). Similarly, the comparisons of pattern-specific learning curves is temporally-aligned between rules (Figure S1E) and with the general performance (Figures S1I–S1N).

#### Sessions of easy and hard rules

We label rules according to the monkeys' ability or inability to recurrently achieve high performance in learning those rules. Thus, given that a subset of rules is labeled 'easy' and another subset is labeled 'hard', we computed the amount of variance, within

the set of  $P_{end}$ 's that the labeling explains. The  $R^2$  value is:  $R^2 = 1 - \frac{\sum_i (P_{end}^i - \mu^i)^2}{\sum_i (P_{end}^i - \mu)^2}$  where,  $\mu = \langle P_{end} \rangle$  is the mean end-performance

of all the sessions and  $\mu^i$  are the mean end performances of sessions of either 'easy' or 'hard' rules (Figure 2E, Figure S1P–S).

#### Sessions of high and low performance

We also labeled individual sessions as 'high' or 'low' performance independent of the rules to make sure our results are robust (Figures S2F–G, S4A–D). High and low performances are defined as  $P_{end}$  values above or below the median and are calculated for each monkey separately. This ensures no bias due to the differences between animals when contrasting sessions of high and low performance.

#### Rule repetition effects

Each classification rule was used for 1-5 consecutive days and repeated after about a month. We compare the monkeys' performance at the end of a session ( $P_{end}(n)$  with 'n' standing for the n'th session of the rule) to the mean performance in the first quarter of the following session of the same rule,  $P_{start}(n+1)$ . The comparison is made by calculating the Pearson correlation between the  $P_{end}$ 's and  $P_{start}$ 's.

We examined two distinct cases:

1. Taking sessions only from consecutive days, we calculate the correlation,  $\rho_{across}$ , of the across-days learning (Figure 2C).
2. Taking only sessions from the end of a consecutive sequence and the beginning of the following sequence, we calculate the correlation,  $\rho_{recall}$ , of the monkeys' ability to recall rules they encountered a month before (Figure 2D).

#### Testing for pattern-specific memorization

We first consider a memorization scheme in which subjects perfectly learn a list of correct pattern-label pairs but do not generalize. The acquisition of such memorized patterns can be an all-or-none event, which means that after a certain pattern-label pair was memorized it will dictate choice behavior. Alternatively, we consider a gradual probabilistic association strengthening process for the observed patterns, which leaves room for errors. Our data rules out the case of all-or-none memorization: Figures S1C and S1D demonstrate that our subjects frequently made mistakes on specific patterns even after they were labeled correctly in previous presentations.

We can also rule out memorization of the types mentioned above as the sole mechanism. Relying on memorization alone would mean that all rules on patterns of three bits would be learned in the same rate. Figure S1R clearly shows that this is not the case, and that rule identity plays a key role. Even finer memorization aspects, such as pattern-specific acquisition rates, can also be ruled out from our data. Figure S1E shows that subjects learned the labeling of the same pattern under different rules at different rates. Finally, pattern-response pairs can be influencing each other during learning. Figure S1D complements Figure S1C showing that the rule identity impacts this influence as well.

To hone in the general rule-based behavior we add [Figures S1I–S1N](#) to show specific cases of pattern specific performance deterioration accompanying general performance increase. The examples in [Figures S1I–S1N](#). Specifically highlight learning sessions in which a pattern-specific performance was high at the end of one session and decreased following a rule switch. Importantly, the rule switch did not require changing the learned response to the specific pattern (as 4 out of 8 patterns did not change response category in the rule switch). The pattern-specific performance deterioration during a general performance improvement is not expected in learning by stimulus-response association and is a hallmark of rule-based behavior.

### Behavior stability tests

**Feature-based behavior stability.** We want to summarize how consistent were the monkeys in a single number for each session ([Figure S1F](#)). This is done with answers from the last 1/4 of each session. We define as a consistency measure the mean (across patterns) distance of the logistic classifier (fitted to answers in the last 1/4 session) from the chance (0.5) answer.

Namely, if the monkeys adopt a feature based consistent policy at the last quarter of each session, then we can fit their sequence of answers with:

$$P(y = 1|x; \vec{\alpha}, \gamma) = \frac{1}{1 + \exp\left(-\gamma - \sum_{\mu} \alpha_{\mu} f_{\mu}(x)\right)} \quad [\text{Equation 1}]$$

where  $x$  are the presented patterns,  $f_{\mu}(x)$  are the features, and  $\vec{\alpha}, \gamma$  are fitted to maximize the likelihood of the answers.

Fitting a features-based classifier overcomes the sparsity of specific pattern presentations. Specifically, whereas each pattern repeats on average every 8 trials, features are defined in all trials. Calculating per pattern consistency also imposes the assumption that the monkeys can tell all patterns apart from each other. The classifier's way doesn't make any assumption beyond a features based behavior policy.

The consistency measure is thus the mean distance from 0.5. or,

$$\text{Consistency} = \frac{1}{8} \sum_x |P(y = 1|x; \vec{\alpha}, \gamma) - 0.5| \quad [\text{Equation 2}]$$

The chance level for a completely unbiased classifier is  $\frac{\sum_{n=0}^8 \left| \frac{n}{8} - 0.5 \right| \cdot \binom{n}{8}}{\sum_{n=0}^8 \binom{n}{8}} = 0.1367$

**Across sessions stability.** Next, we want to check if the monkeys were stable across sessions. Namely, regardless of performance, how similar is the features-based behavior at the last quarter of different sessions. Or, how similar is the behavior when learning the same rule in different sessions.

To avoid interference by behavioral confidence (the inconsistency that drives  $P(y|x)$  in Equation 1 close to 0.5 and can be impacted by session-specific motivation) on the policy-related across-trial variability (The conceptual inconsistency that separates the logistic classifiers fitted to different sessions) we threshold the classifiers,  $c(x) = [P(y|x) > 0.5]$ , and for each rule compare all pairs of sessions. The thresholding guarantees that pairs of sessions will be considered similar if the average behavior policy was similar. [Figure S1G](#) shows the mean (and SE in error bars) of the across session similarity score:

$$S_{\text{rule}} = \frac{1}{N_{\text{pairs}}} \sum_{i < j} \frac{1}{8} \sum_x [c_i(x) \cdot c_j(x) + (1 - c_i(x)) \cdot (1 - c_j(x))] \quad [\text{Equation 3}]$$

where  $i, j$  are different sessions of the same rule and  $N_{\text{pairs}}$  is the number of pairs of sessions with the same rule. In this analysis we pool all sessions of each rule including sessions that repeat the same rule in consecutive days or months apart.

### Behavior similarity to the rules

To estimate how similar is the monkeys' behavior to the rules they learn we repeat the calculation in Equation 3 but replace one of the classifiers ( $c_j$ ) and subtract 0.5 to shift the mean expected overlap to 0. The, above chance level, results are presented in [Figure S1H](#).

### Testing for performance-based strategy

Animal behavior could potentially obey a local performance-based strategy called win→stay, lose→switch. This strategy is observed in animal studies and suggests that animals will repeat a choice that led to reward and switch a choice that didn't. To test if the monkeys significantly relied on such a strategy, we simulated answer sequences following this strategy for all the learning sessions in our experiments. We then compared the true answers, given by the monkeys, to the simulated sequences. Any above-chance (50%) agreement would indicate that the monkey might be using this strategy. However, we find that the agreement with the win-stay, lose-switch strategy is below chance for nearly all sessions. [Figure S1O](#) shows distributions of per-session agreement for simulations initiated in the left or right choice (the only free parameter). 93.5% of sessions had below-chance agreement and the median agreement was significantly lower than chance (Wilcoxon signed rank test,  $p < 1e-15$ )



## Neural activity analysis

### Single-neuron responses

We expect the learning-relevant neural activity to be influenced by both trial-by-trial variations, such as changing behavior and stimulus identity, and by slower processes, namely learning. Studying the learning related dynamics, we are interested in the single unit neural activity that correlates to such inherently variable computational primitives. Namely, we seek a measure of the spiking activity that communicates the variations across trials. Accordingly, for every neuron we examine the spikes in the 500ms following the stimulus onset and bin them into 5x100ms segments to obtain sensitivity to temporal effects in addition to the spike count. The result is a 5-vector of spike counts from each trial,  $\vec{V}(t = 1 \dots T) \in \mathbb{R}^5 \times \mathbb{N}$ . In this representation, the component of largest across-trials variance is  $\vec{v}^* = \arg \max_{\|\vec{V}\|=1, \vec{V} \in \mathbb{R}^5} \text{Var}(\vec{V} \cdot \vec{V})$ . We then project each 5-dimensional vector on this principal component and get a single number

from each trial,  $r(t) = (\vec{V}(t) - \langle \vec{V} \rangle) \cdot \vec{v}^*$ . This number scores the spiking patterns of the neuron with respect to its most prominent fluctuations or change. Importantly, several unrelated processes may contribute to the across-trials variability and in choosing the projection,  $r(t)$ , as the representation of stimulus neural response, we tune to the largest source of variability, regardless of its nature (Figures S3A–S3C).

### Pattern-specific neurons

We devised a criterion for exemplar preference based on a neuron's firing rate in the 500mSec after the pattern presentation. We build a table of all responses of the neuron to each of the patterns. Next, we compare the sets of responses to each pair of patterns using a rank-sum test for equal medians and treat the distribution of responses as different using a threshold at  $p < 0.05$ . A neuron is pattern-specific if the distribution of responses to only one pattern is different from all the others (Figure S2A)

### Feature-based representations

For a pattern,  $\vec{x}$ , we chose a basis of features that are polynomials of the variables  $x_1, x_2, x_3$  that take the values  $\pm 1$ . Features,  $f_{ijk}(x) = x_1^i x_2^j x_3^k$  differ in the polynomial degrees ( $i, j, k = 0, 1$ ) and satisfy, when averaging across all patterns,  $\langle f \rangle = 0$ ,  $\langle f^2 \rangle = 1$ .

**Lemma 1:** Classification rules factor in this basis.

**Proof 1:** Any classification rule can be described by a function,  $y(\vec{x})$ , assigning the label  $y \in \{1, -1\}$  to every pattern  $\vec{x}$ . Define

$$\gamma = \frac{1}{8} \sum_{\vec{s}} y(\vec{s}) \text{ and } \forall l \in \{i, j, k\}, \alpha_l = \frac{1}{8} \sum_{\vec{s}} y(\vec{s}) f_l(\vec{s}) \text{ and construct:}$$

$$f(\vec{x}) = \gamma + \sum_l \alpha_l f_l(\vec{x}) \quad [\text{Equation 4}]$$

It is straightforward to show that  $\sum_l f_l(\vec{s}) f_l(\vec{x}) = \begin{cases} 7 & \text{if } \vec{x} = \vec{s} \\ -1 & \text{otherwise} \end{cases}$  and therefore  $f(\vec{x}) = y(\vec{x})$ .

Also, note that for the 8 rules we used  $\gamma = 0$  and, by construction,  $\vec{\alpha}$  is a unique definition of the rule as a vector in features' space (i.e., the 'rule vector').

**Lemma 2:** Different features in this base have zero covariance.

**Proof 2:** Let  $f_1$  and  $f_2$  be features in this set and without losing generality assume that they differ in the polynomial degree of  $x_1$  s.t.  $f_2$  doesn't contain  $x_1$ .  $f_1$  and  $f_2$  can still have overlapping cues like the case of  $f_1 = x_1 x_2 x_3$  and  $f_2 = x_2 x_3$ .

$$\text{Cov}(f_1, f_2) = \sum_{x_1, x_2, x_3 = \pm 1} f_1 \cdot f_2 = \sum_{x_1 = \pm 1} x_1 \sum_{x_2, x_3 = \pm 1} \frac{f_1}{x_1} \cdot f_2 = 0$$

where  $\frac{f_1}{x_1}$  denotes the function  $f_1$  without its  $x_1$  component

**Lemma 3:** Correlation projections factor in this basis.

**Proof 3:** The corollary of Lemma 2 is that any combination of basis functions,  $f = \sum_{l \in \{i, j, k\}} a_l f_l$  such that  $\sum_l a_l^2 = 1$  will have  $\langle f \rangle = 0$  and  $\langle f^2 \rangle = 1$  (because  $\langle f_l \cdot f_j \rangle = 0, \forall l \neq j$ ). Which leads to  $\text{Var}(f) = 1$ .

For the scalar neural response  $r$ , the correlation  $C(f, r) = \frac{\langle (f - \langle f \rangle) (r - \langle r \rangle) \rangle}{\sqrt{\text{Var}(f) \text{Var}(r)}} = \sum_l a_l \frac{\langle (f_l - \langle f_l \rangle) (r - \langle r \rangle) \rangle}{\sqrt{\text{Var}(f_l) \text{Var}(r)}} = \sum_l a_l C(f_l, r)$  because  $\text{Var}(f_l) = 1$  and  $f_l = 0 \forall l$ .

If we describe the correlations to basis features as a vector (the 'neural vector'),

$$\vec{C} = (C(f_1, r), C(f_2, r), \dots) \quad [\text{Equation 5}]$$

then  $C(f, r)$  is the projection  $\vec{C} \cdot \vec{a}$ .

### Corollary 1:

If we measure a neuron's correlations to the features separately, then the unit vector,  $\hat{a}$ , that maximizes  $\sum_i a_i C(f_i, r)$  will give us the feature  $f = \sum_{i \in \{i,j,k\}} a_i f_i$  that maximizes  $C(f, r)$  and can be thought of as the neuron's preferred feature.

### Corollary 2:

If  $\hat{a}$  is a vector holding the coefficient for a rule that we chose in advance,  $f = \sum_{i \in \{i,j,k\}} a_i f_i$  (e.g., the one being learned), then the projection of the neural correlations,  $\vec{C}$ , on  $\vec{a}$ ,  $\vec{C} \cdot \vec{a} = \sum_i a_i C(f_i, r)$  is indeed the neuron's rule-correlation  $C(f, r)$  (Figure 4B, Figure S3C,D).

### Dynamics of representations

To study the dynamics of task related neural correlates we divided each session to partially overlapping windows (40 trials segments with 4 trials jumps). For each neuron, calculating the correlation between its spiking patterns,  $r(t)$ , following stimulus onset, and the stimulus features,  $\vec{f}(\vec{x})$ , (as well as to the correct category and the monkey's future answer) yields a set of correlation coefficients,  $CC_i(t) = \text{corr}(r(\tau \in w_t), f_i(\vec{x}(\tau \in w_t)))$ , for each regression window  $w_t$ . These rolling regression coefficients were used to calculate the following measures:

### Comparing representation between conditions

To judge whether neurons show rule selectivity during a certain segment of the session (Figure 3D) we test the fraction of regression windows within that segment, that exhibit significant rule correlation (Pearson,  $p < 0.05$ ). This test is done comparatively between sessions of different conditions, and we set a criterion of 10% to declare a neuron as showing rule selectivity during the segment. If there are more than 5 neurons meeting each condition (2 conditions, e.g., easy and hard rules) we use the binomial comparison

z statistic,  $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \cdot (\frac{1}{n_1} + \frac{1}{n_2})}}$  with  $\hat{p}_1, \hat{p}_2$  the measured success rate in two populations of sizes  $n_1, n_2$  and  $\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$ .

### Comparing rule versus answer representations

The correlations of a neuron's activity to the rule (the correct label in each trial) and to the animal's answers have a mutual component (the spiking pattern). The rule and answer variables are also interrelated via the performance level. To compare neuron's correlations to the answer and rule and account for these dependencies we use Williams's t-statistic for correlated

correlations,  $t = \frac{(C_{12} - C_{13}) \cdot [(n - 3) \cdot (1 + C_{23}) / 2D_3]^{\frac{1}{2}}}{\sqrt{1 + \frac{(n - 3) \cdot (C_{12} + C_{13})^2 \cdot (1 - C_{23})^2}{8(n - 1)D_3}}}$ , where  $C_{12}$  is the category correlation,  $C_{13}$  is the answer correlation, and  $C_{23}$

is the correlation between answers and categories.  $n$  is the number of trials and  $D_3$  is the determinant of the sample correlations matrix. The statistic is compared to the t - distribution with  $n - 3$  degrees of freedom (Figure 3F).

### Categorical properties of neural responses

To check if neurons correlated to the category or the animals' answers show categorical response properties we follow the method described in Onken et al. (2019). Briefly, we use two task variables, the correct category and the animals' action in each trial to define four task condition. For each neuron and each regression window we calculate the mean response in each of the conditions yielding a 4-vector. These vectors are normalized to unit length and assigned with cluster identity using the spherical k-means algorithm. Figure S6 presents vectors both as 3d plots using their 3 leading principal components and also as 2d plots using t-SNE (van der Maaten and Hinton, 2008).

### Relating neural correlates to performance

To relate any regression measure and performance within a group of neurons we take the following steps:

1. For every 40-trials-long regression window we calculate the mean performance.
2. Given a performance level, we collect all the regression windows with performance within 0.15 of that level and calculate the mean and standard error of the measures of interest (Figure 3E).

### Angle-to-rule and vector-magnitude

Given a basis of visual features there is a unique spanning of the classification rule in each session. For each regression window we define the angle-to-rule as the angle between the vector of correlation coefficients to visual features (the *neural-vector*) and the vector that represents the rule (the *rule-vector*). Similarly, we define the features' correlation magnitude as the norm ( $L_2$ ) of the correlation coefficients vector.

When presenting the learning related dependence of these geometrical variables over time, we smooth the curves with a running window with length of 10 percent of regression windows in a session (Figures 4, 5A, 5B, 5D, 5E, insets, and S3 and S4)

### Correlation to rule and answer

In Figures S4M–S4O and S4R, we present similar analyses of representation dynamics (as in Figures 5A, 5B, 5D, and 5E) but instead of the geometric measures we show correlations to the categories determined by the rule and correlations with the monkeys physical answers (their right or left choices in each trial)

### Session-length standardization

Several calculations require the comparison or grouping of segments from relative session fractions and/or location. To enable this, we standardized the regression measures from each session to a fixed length of 100 bins. This means that all rolling regressions were stretched to the same length, because none of the original analyses exceeded 100 regression windows.

### Changing angle-to-rule or vector-magnitude

To quantify neurons that decreased angle-to-rule or increased the vector-magnitude (Figures 5A, 5B, 5D, and 5E), we compare regression windows in the first 15%-segment of the sessions to regression windows in the last 15%-segment of each session with a 1-tailed t test. In Figures 5A, 5B, 5D, and 5E, the fractions of cells that passed the test are compared with a 2-tailed binomial z test.

Figures S4A–S4D replicates these contrasts when dividing sessions into high and low performance.

### Fractional change in angle and magnitude

We calculate the fractional difference of neural-vector's angle-to-rule and magnitude from the average baseline values in the initial 15% of regression windows (Figure 5). The resulting traces are smoothed with a 10% window and significant difference between sessions of easy and hard rules is determined with bootstrapping – shuffling the easy/hard label 10,000 times and checking if the correct labeling surpasses the required confidence level (95%).

### Interpreting neural-vector magnitude as confidence

Confidence in a binary decision is defined by the decision boundary. A decision boundary can be defined using a sigmoid expression  $y = \frac{1}{1 + \exp(-\alpha r)}$  where  $y$  is the choice probability,  $r$  is the mean-subtracted activity of a neuron, and  $\alpha$  is a scalar. In stochastic behavior, the trial by trial decisions can be modeled as a flip of a biased coin with probabilities  $y$  and  $1-y$  to press either one response button or the other. In a low confidence state these decision probabilities are close to 0.5. Decisions in a high confidence state are reliable and these probabilities are close to 0 or 1.

If the neural activity depends on just one feature,  $f_i$ , from our basis then we can express it as  $r = af_i + \xi$  using a scalar coefficient 'a' and a zero-mean error  $\xi$  ( $r, f_i, \xi$  change from trial to trial).

The correlation between the neural activity and the feature will take the form:

$$\text{Corr}(r, f_i) = \frac{a}{\sqrt{a^2 + \langle \xi^2 \rangle}}$$

This correlation coefficient will approach 1 (–1) as 'a' approaches  $\infty$  ( $-\infty$ ). This means that the strength of the correlation has a monotonous relation to the 'steepness' of the sigmoid  $y(r)$ . This relation is depicted for example Figure S3E.

In general, if the neural activity is spanned by a mixture of our entire basis of features as  $r = \sum_{\mu} a_{\mu} f_{\mu} + \xi$  and  $b = \sum_{\nu} b_{\nu} f_{\nu}$  is a direction in

our features space ( $|\vec{b}| = \sum_{\nu} b_{\nu}^2 = 1$ ) then we get  $\text{Corr}(r, b) = \frac{\vec{a} \cdot \vec{b}}{\sqrt{|\vec{a}|^2 + \langle \xi^2 \rangle}}$  where ' $\vec{a}$ ' and ' $\vec{b}$ ' are vectors. This correlation is maxi-

mized if ' $\vec{b}$ ' is in the direction of ' $\vec{a}$ ' and boils down to the same scalar expression as above (replacing the scalar 'a' with the magnitude  $|\vec{a}|$ ). This means that the decision boundary is defined by the direction of ' $\vec{a}$ ' and its steepness is defined by the magnitude of ' $\vec{a}$ '.

### Relating change in reaction time and change in vector magnitudes

The behavioral reaction time is defined as the time interval from the moment the animals release the trial onset button to the moment they press one of the answer buttons. This measure cleans the initial processing time required to release the trial onset button after the stimulus onset. The remaining motion time is used here as a potential indication of the animals' confidence.

Our task did not train the animals to respond quickly and the reaction times distribute broadly. We used the median in each regression window to avoid outlier trials and define the change in reaction time as the difference between the mean at the first and last quarters of the session.

We then compute the correlation across neurons between the change in reaction time and the fractional change in vector magnitudes (Figure S4J–L using the same parameters as Figures 5C and 5F) separately for the sessions with high and low performance as defined above. We use Fisher's transformation,  $\hat{r} = \frac{1}{2} \log \left[ \frac{1+r}{1-r} \right]$  and the z-test for:

$$z = \frac{\hat{r}_{\text{high}} - \hat{r}_{\text{low}}}{s}$$

where  $s = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$ , to rule out the hypothesis of equal correlation coefficients.

### Graded relation between neural correlates and learning behavior

In the section ‘Changing angle-to-rule or vector-magnitude’ above we described contrasts between neurons recorded in groups of sessions (easy versus hard rules in Figures 5A, 5B, 5D, and 5E and high versus low performance in Figures S4A–S4D). In addition to these categorical contrasts, we use Pearson correlations across neurons to estimate the graded relation between performance measures and neural correlates. We correlate behavior measures ( $P_{end}$ ,  $P_{end} - P_{start}$  defined above and normalized for each monkey separately) with the neurons’ angle-to-rule and vector magnitude (Figures 5C and 5F). The neural correlates are averages in regression windows in baseline and segment durations similar to Figures S4A–S4D.

### Optimal lags between time series

Given two time series, e.g., the angle-to-rule of a dACC neuron and the simultaneously-recorded vector-magnitude of a Putamen neuron, we find the shift that maximizes their Pearson correlation. Only pairs with significant correlation in the optimal lag contribute (as in Figures 6E, 6G, and 6H).

### Relating neural-vector to next-day behavior

To examine if the learning-related change in the neural-vectors indicate a real shift in the monkeys’ preferred policy (Figure 7), we tested if the neurons’ preferred feature combination (i.e., their neural-vector) predicts the monkeys’ behavior early in the following day. For each neuron we calculated the neural-vector in the last 20% of the sessions’ regression windows. We then calculated the projection of these neural-vectors on the subsequent day’s rule and used the maximal value as the predictor of the next day’s performance. In Figure 7A we calculate the Pearson correlations between these neural projections and the mean performance in the early fraction of the next day’s session across the neural population. In Figure 7B we repeat the same calculation but only take cases in which the rule was changed between the current and next day.

### Behavior decoding from neural ensemble activity states

The analysis of neural ensemble responses are descriptions of neural computations that are conceptually-different from the framework developed in this manuscript:

- (1) The space of feature correlations has different dimensionality and units from the space of ensemble activity patterns. In the former, axes are correlation coefficients and the dimension equals the number of features. In the latter, axes are the activity of individual neurons and the dimension equals the number of neurons.
- (2) The correlation coefficients in our framework are calculated across trials in rolling regression windows, not in individual trials, whereas population activity vectors can be measured per-trial.
- (3) Conceptually, our approach first extracts the task-correlated dynamics of single neurons whereas ensemble patterns are insensitive to the task and require further assumptions to ‘decode’ behavior.

To relate these frameworks, we consider an ensemble of neurons jointly-recorded in a sequence of learning steps,  $t$ , during one regression window,  $W$ . Each neuron, indexed by  $i = 1..N$ , has the response  $r_i(t)$  in trial  $t$ . For simplicity we consider the responses to be centered and linearly-dependent on the visual features s.t.  $r_i(t) = \sum_{\mu} \alpha_i^{\mu} f_{\mu}(x(t)) + \xi_i(t)$ , where  $x(t)$  is the pattern shown in trial  $t$  and  $\xi_i$  is a zero-mean Gaussian random variable with variance  $\sigma_i^2$ .

Fisher’s linear discriminant is a commonly-used population level code to describe policy-based behavior. In the linear discriminant formalism, the per-trial vectors of neural responses,  $\{\vec{r}(t)\}_{t \in W} \in \mathbb{R}^N$ , are treated as points in the space of ensemble activity patterns (with dimension = number of neurons) and used to decode a binary behavior condition (the correct category or the animals’ choice). This is done by finding a direction,  $\vec{\beta} \in \mathbb{R}^N$ , defining the optimal plane separating ensemble activity in one condition from the other. In this framework  $\vec{\beta}$  is identical to the set of coefficients that maximize  $CC\left(\sum_{i=1}^N \beta_i r_i, \sum_s a^s f^s(x)\right)$ , the correlation coefficient (across trials) between the projection of the neural ensemble on the direction  $\vec{\beta}$  and the behavior condition,  $\sum_s a^s f^s(x) + \text{constant}$ .

Calculating this correlation coefficient and using the properties of our basis of features we get:

$$CC = \frac{\sum_{i=1}^N \beta_i (\vec{\alpha}^i \cdot \hat{a})}{\sqrt{\sum_{i,j=1}^N \beta_i \beta_j (\vec{\alpha}^i \cdot \vec{\alpha}^j + \sigma_{ij})}} \quad [\text{Equation 6}]$$

This expression is composed from interpretable and illuminating parts.

The term  $(\vec{\alpha}^i \cdot \hat{a})$  is the projection of the  $i$ ’th neuron’s preferred visual feature on the direction defined by the behavior condition in feature space  $(\hat{a})$ . As already shown in our methods section this is equivalent to the correlation of the  $i$ ’th neuron with the behavior condition.

The term in the denominator is independent of the behavior condition and separates to terms identifiable as ‘signal correlation’  $(\vec{\alpha}^i \cdot \vec{\alpha}^j)$  and terms identified as ‘noise correlation’  $(\sigma_{ij} = \langle \xi_i \xi_j \rangle)$  between pairs of neurons.

An assumption of conditional independence between neurons is implied in most analyses of ensemble activity frameworks (like many forms of dimensionality reduction) and amounts to setting  $\sigma_{ij} = 0 \ \forall i \neq j$ . Under this assumption, the conclusion from Equation 6

is that the coefficients  $\vec{\beta}$  will optimize the behavior decoding if they give large weights to neurons that increase their correlation to the behavior condition.

To verify this intuition, we fit one logistic classifier (equation below) to the correct category and another to the behavior, the actual answer provided by the animals, in each rolling regression window. For each neuron we calculate the Pearson correlations between the dynamics of its  $\beta$  coefficient and the dynamics of its correlation to the category (or answer) and summarize the results in [Figure S7D](#).

A logistic classifier models the probability of behavior state ‘y = 1’ given the vector of neural activity  $\vec{r}$ :

$$Prob(y = 1 | \vec{r}) = \frac{1}{1 + \exp[-\vec{\beta} \cdot \vec{r} + \gamma]}$$

Fitting is done by maximizing the likelihood in each regression window separately.