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

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

Yarden Cohen[1,2], Elad Schneidman[1,#], Rony Paz[1,#,‡]

[1]Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel
[2]Current address: Massachusetts General Hospital, Boston, MA 02114, USA
[#] These authors contributed equally
[‡] Lead contact

Correspondence should be addressed to: Y.C. (ycohen1@mgh.harvard.edu), R.P. (rony.paz@weizmann.ac.il), or E.S. (elad.schneidman@weizmann.ac.il)

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 the 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 towards the rule, implying a policy search, whereas neurons in the putamen showed 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 from 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 previous studies have typically used the average performance of subjects to define rule complexity (Feldman, 2000), it has been shown that individual learning curves can be predicted 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 preference to 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 anterior-cingulate-cortex (dACC) projects widely to striatal regions (Averbeck et al., 2014; Heilbronner et al., 2016; Ongur 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 outcome of previous trials, and form integrated representations of task structure (Chudasama et al., 2012; Haroush and Williams, 2015; Hayden and Platt, 2010; Heilbronner and Hayden, 2016; Kolling 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; Seo and Lee, 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. Therefore, 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 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', Fig.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 relation between the eight patterns and the correct responses (`labels') defines a classification rule (Fig. 1B). Out 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 if the left bit is white/black, and in rule '12' if the left and middle bits are equal/different (Fig.1B). All seven rules are unbiased, such 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 Fig.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 (2^{nd} and 3^{rd} cycles, Fig. 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 (Fig.2A) and continuous performance improvement throughout long sessions (Fig.2B, Pearson r(118)=0.49(0.34,0.61), p<1e-8), as well as next-day retention of the learned rule (Fig.2C, Pearson r(70)=0.58(0.4,0.71), p<1e-3). Neither monkey showed retention benefit when presented with the same classification rules in the second cycle of the rules (Fig.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 (Fig. 2E, Fig. S1B). This diversity was also apparent in additional sessions in which the monkeys were presented with 2-bit rules and 4-bit ones (Fig. 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 independently 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 (Fig. S1C, Kruskal-Wallis test,$\chi^2(7,366) = 89$, p < 1e-15), and were correlated with performance on other patterns (Fig. S1D, Kruskal-Wallis test ,$\chi^2(7,440) \quad > 88$, p < 1e-10). Second, the classification of the visually salient patterns showed variability across rules and dependence on the specific rule (Fig. 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 (Fig. S1G-H). Fourth, monkeys could even learn 4-bit rules for which memorization is extremely unlikely (Fig.2F; note that there are 65,536 different deterministic rules for the 16 stimulus patterns). Fifth, performance for specific patterns deteriorated following a change of the rule, even when the new rule did not require learning new associations for these patterns (Fig. S1I-N). Finally, a simple strategy based on performance (win-stay, lose-switch) poorly

described answer sequences (Fig. S1O, Wilcoxon's signed rank test, $p < 1e\text{-}15$). Together, the evidence suggest that monkeys did not use simple memorization of stimulus-response associations as the main strategy, and therefore support the notion of rule-learning.

We conclude that both monkeys learned the rules in individual sessions with some rules being easier to learn than others. The complexity and richness of the task led to error rates that reflect a more natural learning setting compared to 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 (Fig.3A-B, 543,115,114 neurons recorded in dACC, Caudate, and Putamen in 3-bit rule sessions). First, we identified neurons with stimulus-evoked spiking pattern that responded to the stimuli, but with no preference for rule/category (*stimulus-specific,* Fig.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, Fig. 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*, Fig.3C, left and middle, Fig. S2B-E). To identify correlates of successful learning, we separated the sessions of *easy* and *hard* rules based on the average success rate for each rule (Fig.2E, Fig. S1P-Q). This separation for easy vs. hard rules is also evident in the learning curves (Fig.2E inset; Fig. S1R) and captures the largest performance variance (Fig. S1S). We found that at the end of the session, both the dACC and the Putamen had significantly more *category-specific* neurons for *easy* rules compared to *hard* rules (Fig.3D, dACC: 21% vs. 12.5%, Pu: 22% vs 8% for *easy* vs. *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 vs. low performance (Fig. S2F-G, dACC: $z=2.046$, $p<0.021$, Pu: $z=1.79$, $p<0.04$; and for each monkey separately, Fig.S2F-G). Moreover, there was a monotonic relation between performance and the proportion of neurons showing correct classification (Fig.3E).

Since 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 (Fig.3F). In contrast, significantly more neurons became choice-selective in the Caudate (binomial comparison z-test for easy and hard rules, dACC: 13.3% vs. 5.48%, $z = 2.88$, $p<0.003$, Cd: 9.57% vs 1.86%, $z=1.72$, $p<0.05$, Pu: 9.95% vs. 1.55%, $z=2$, $p<0.03$), with similar proportions of neurons being *category-specific* in *easy* and *hard* rules (Fig.3D). Therefore, the modulation of single neuron activity indicates that dACC and Putamen mainly represent the learned classification, whereas the Caudate mainly represent the choice. We therefore focus in the following analyses mainly on the dACC and the Putamen.

A geometric representation to track neural dynamics during learning

We next examined how the modulation of neural activity develops during the learning. To characterize the neural representations during rule acquisition, we represented each stimulus pattern using a complete set of statistically independent binary features (Fig.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 term the *rule-vector* (Eq 4). We then express the selectivity of individual neurons using this set of basis features, which we term the *neural-vector* (Eq 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, Fig. S3A-D, Fig.4B).

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 (Fig.4B, left) – suggesting an increase in confidence (Fig. S3E), or by rotating towards the *rule-vector* (smaller angle-to-rule, Fig.4B, right) - suggesting a policy change. We note that 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 remained selective for a feature regardless of the performance improvement (Fig.4C); Second, neurons that rotated towards the rule during learning (Fig.4D. Fig. S2B-E); Third, neurons that increased their neural-vector magnitude but with no regard to the rule of that session (Fig.4E); Fourth, neurons that demonstrated changes in several features in parallel, and as a result, changes in both rotation and magnitude (Fig.4F). These more complex trajectories can eventually lead to significant agreement with the *rule-vector* (Fig.S3F-J). Indeed, the two processes were not mutually exclusive and many neurons with vector-magnitude changes also exhibited angle-to-rule changes (Fig. S5A-B).

Differential representations of confidence and policy in the Putamen and the dACC

To examine if the two components of learning are differently represented in the two regions, we examined the increase in the magnitude of the neural-vector and its rotation towards the rule-vector (angle-to-rule) separately. In both the dACC and the Putamen, more neurons rotated towards the rule during learning of easy rules, resulting in having a smaller angle-to-rule (Fig.5A,B, binomial z-test comparing *easy* to *hard* rules, dACC: 40% vs 25.5%, Putamen: 49% vs 34%, dACC: $z=3.32$, $p<0.001$, Pu: $z=1.68$, $p<0.05$). This rotation gradually progressed along the session as performance improved (Fig.5A-B right panels). Similar findings were obtained when dividing sessions into high and low performance ones, independent of the specific rule (Fig.S4A-I). 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 (Fig.5C).

Conversely, changes in vector-magnitude between easy and hard rules were not observed in the dACC (Fig.5D, binomial z-test. dACC: 38% vs 34%, $p>0.1$), but did occur in the Putamen, where

more neurons increased their vector-magnitude in *easy* rules (Fig.5E, binomial z-test. Putamen: 47% vs 31%, z=1.82, p<0.04; Fig.S4C-D: similar findings for high and low performance; Fig.5F: magnitude across all Putamen neurons, but not dACC, correlated with performance). These changes became more prominent towards the end of the sessions (Fig.5E, right-panels), suggesting strengthening of the recently acquired policy, and were accompanied by a reduction in response time (Fig S4J-L), 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 both the dACC and the Putamen (Fig.S4M-N). Complementing this and further supporting the conclusion that the dACC and the Putamen mainly represent learning whereas the Caudate mainly represent the choice (Fig.3F) - Caudate neurons represented the actual-choice differentially in *easy vs. hard* rules (Fig.S4O), but not for angle-to-rule or vector-magnitude (Fig.S4P-R).

These findings indicate that both the Putamen and dACC reflect a process of policy search for the correct rule; but whereas the dACC represent change of policy, the Putamen also represents strengthening and confidence gain that accompanies successful learning.

Neuronal Policy follows behavioral change

We first asked whether the representation depicted by the neural-vectors shows similar dynamics to 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 (Fig.6A-B, 32% of the dACC neurons and 37% of the Putamen neurons, Pearson, p<0.01). Similarly, a large fraction of neurons changed their vector-magnitude in correlation with the performance (Fig.6C-D, 19% of the dACC neurons and 18% of the Putamen neurons, Pearson, p<0.01, Fig. S5A-B).

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 (Fig.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 (Fig.6E bottom; Fig.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 to the overall distribution (Fig.6F, both regions: p<0.005, Putamen: p<0.05; dACC: p<0.02, Wilcoxon's rank-sum tests). The fact that the rotation was directed towards the rule (smaller angles) indicates that the change is not due only to reward. Together, these findings show that adopting a new correct policy – by rotation of the neural-vector towards 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 the Putamen, we also studied neurons that were simultaneously recorded 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 either vector-magnitude or angle-to-rule across regions, or within a region, or when taking angle-to-rule in Putamen and vector-magnitude in dACC (Fig.S5D, $p>0.1$ for all, Wilcoxon's signed rank tests). In contrast, only the lags between vector-magnitude in Putamen and angle-to-rule in dACC showed a clear directionality, with the Putamen magnitude following dACC rotation (Fig.6G, Wilcoxon's signed-rank test, $z=-2.81$, $p<0.005$; mean lag -2.8 ± 0.96 trials). Moreover, this was more dominant in neural pairs recorded during *easy* rules (Fig.6H. top, Wilcoxon's signed-rank test, $z=-2.64$, $p<0.01$, Mean lag -3.56 ± 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 once a series of successful trials is followed by neural rotation towards 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 the Putamen predicted next-day early performance (Fig.7A, Pearson's $r(99) = 0.33$ (0.15,0.5), $p<0.0007$; no correlation was found when reversing the days as control). Therefore, the closer the *neural-vector* was at the end of one day to the rule used in the next day, the better was the initial performance.

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

Discussion

We studied dynamics of neural representations in the dorsal-anterior-cingulate-cortex (dACC) and the striatum while monkeys learn to classify multi-cue patterns. We developed a novel mathematical approach to represent the 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, namely 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 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 as 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 that is used for retention, either 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 both reinforcement and memory (Doll et al., 2015; Wimmer et al., 2018).

The geometrical framework we presented here allows for the identification of neurons that retain a stable representation and ones that change during learning (Chen et al., 2001; Genovesio et al., 2005; Sadtler et al., 2014). Interestingly, these changes were observed during both 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. Yet, while most neurons change by rotating towards 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 over-ruled. 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 beneficial memory for future (different) tasks. Our approach would enable examination of such memories and their relative strength (e.g., by characterizing the rules that the neurons that do not rotate point to). 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 do 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; Tkacik 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, as it was stronger for easy rules than for hard ones. Since 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 if 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), 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; Grundemann 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 on the 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 (Fig.S7). Our results here show categorical properties after learning is complete (Fig.S6), in agreement with recent studies in prefrontal neurons (Hirokawa et al., 2019; Onken et al., 2019). In addition, we further describe here the dynamics that led them to develop these final representations.

Overall, we developed here 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).

Acknowledgments: We thank Yosef Shohat for animal training, welfare and experimental procedures; Dr. Yoav Kfir for scientific consult, 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 a European Research Council Grant 311238, an Israel Science Foundation Grant 1629/12, a CRCNS grant, 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.

Main figure titles and legends

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 seconds to classify it with the left or the right button. A fluid reward follows a correct choice and a short timeout follows an incorrect choice.

B. A scheme of the rule-based classification. Top: labeling all eight 3-bit patterns by the correct response (pressing the **L**eft or **R**ight buttons) divides the patterns into two categories (a truth table marked by green and purple boxes). Bottom: representing the patterns as variables, $[\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}] \in \{-\mathbf{1}, \mathbf{1}\}^{\boldsymbol{3}}$, allows describing truth tables by Boolean expressions that define classification rules, namely 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: one-bit-rule ('1') where the decision is based only on the identity of the left bit; two-bit-rule ('12') where decision is based on a XOR of left and middle bits, and the Majority-rule where decision is based on summing all 3 bits.

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). Shown also is the underlying truth table of the rule (patterns are stacked by category).

B. Continuous longer learning within-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 in 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).

D. No retention between rule repetitions after several weeks. The use of eight different rules that use the same stimulus patterns over four weeks or more means 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. Average performance in the last quarter of each session for all rules and all sessions, for both monkeys (D: light blue; G: dark blue). For each rule, bars and error bars show the mean and SE over animals and sessions, and classified into 'easy' and 'hard' rules (in 'easy' rules performance is significantly above chance-level on average). Inset shows average learning curves for all rules, further justifying the 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 Fig. S1.

Figure 3. Single-neurons represent learning of rule-based classification.

A. Recording locations projected on a coronal MRI section. Recording locations span ~12mm in the anterior-posterior axis (hence some locations seem to be outside the regions of interest). Red marks borderline of the dACC, Yellow for Caudate and Cyan for Putamen.

B. 3-D reconstruction of all recording locations. Gray plane is the midline, anterior-posterior zero is at the anterior-commissure and depth is measured from the dura surface.

C. Rule-learning in neurons. Spiking patterns are divided by the category label (orange and black rasters and PSTHs). Spike times are aligned to the stimulus onset (dashed line). The order of trials is top to bottom. Trials are divided into the first half (upper row) and second half (bottom row) of the sessions. The two left neurons differentiated between categories in the second half significantly more than in the first half, whereas the right neuron did not differentiate categories.

D. Fraction of single-units whose firing pattern correlated with the rule at the end of learning (Mean and SE). Sessions are divided according to easy vs. hard rules, with more neurons signaling correct classification in easy rules in the dACC and the Putamen ($p<0.05$ for both, binomial tests), but not in the Caudate.

E. Instantaneous performance parallels neural-based categorization. The fraction of neural segments with significant rule representation increases with the mean performance (calculated by rolling regression windows of 40 trials in steps of 4 trials, averaged over sessions).

F. Percent of neurons with stronger correlation to the rule (bottom) and with stronger correlation to the actual choice (top), over both successful and error trials (William's test), and comparing easy vs. hard rules. The dACC and the Putamen are significantly correlated more with the rule and not with the actual choice, whereas caudate neurons are significantly correlated with the choice only and not with the rule (significant comparisons are $p<0.05$, binomial tests).

See also Fig. S2.

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 for the representation of the pattern $\vec{x}$ =■□■= $(-\mathbf{1}, \mathbf{1}, -\mathbf{1})$ by the set of binary features: the three first order ones, $\{\boldsymbol{x_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{3}}$, three second order ones, $\left\{\boldsymbol{XOR_{ij}} = \boldsymbol{x_i} \cdot \boldsymbol{x_j}\right\}_{\boldsymbol{i<j=1}}^{\boldsymbol{3}}$, and a third order one, $\boldsymbol{P_{123}} = \boldsymbol{x_1} \cdot \boldsymbol{x_2} \cdot \boldsymbol{x_3}$. Also Shown is the 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, modulate 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 towards 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 (PCA), yet 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 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 of the rule (top-orange and bottom-black). Namely 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* (Eq 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 ⊥, 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 principal axes (PCA) and curve fitted, color-coded for time progress in the session (light → 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. A neuron that does not change its selectivity during learning.

D. A neuron that rotates towards the correct category by learning the right feature.

E. A neuron that extends its vector towards a 'wrong' feature.

F. A neuron that exhibits a complex relationship of vector extension and rotation.

See also Fig. S3.

Figure 5. Differential representation in the dACC and the Putamen.

A,B. Proportion of neurons that significantly decreased their angle-to-rule during the session was different in easy rules vs. hard rules in both the dACC (**A**) and the Putamen (**B**) ($p<0.05$ in both, binomial z-test). 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,E. Same as in (**A,B**) for change in vector magnitude. In contrast to rotation, only the Putamen showed significant increase in the vector-magnitude in easy vs. 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 Fig. S4.

Figure 6. Differential learning dynamics in the Putamen and the 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 (upper row, red) and three Putamen neurons (lower 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 shaded area marking neurons exhibiting significant correlations at $p<0.01$. Both regions contained a high number of neurons with dynamics similar to behavior ($p<0.01$ for both, χ^2). Inset 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. Same as **A** for dynamics in vector magnitude.

D. Same as **B** for dynamics in vector magnitude.

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

F. The distribution of angle-to-rule (absolute cosine value) at the end of all strips of 10 successful trials (dashed line) compared to the distribution of all angles (solid line) ($p<0.005$, Wilcoxon's 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. Same as **G** but separately for easy and hard rules.

See also Fig. S5

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 in 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, showing again a positive correlation ($r=0.38$, $p<0.02$).

STAR Methods

RESOURCE AVAILABILITY

Lead Contact

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

Materials Availability

Materials Availability Statement

- 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.

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, 4-6kg) 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 (Fig.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 3rd square and in rule '12' the label is determined by the XOR of squares 1 and 2. See Figs.1-2, Fig. 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 x 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 <1mm accuracy (mean=0.5mm).

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 ~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, $P_{end} = \langle y_t \rangle_{t\in\left[\frac{3}{4}T\to T\right]}$. 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 (Fig. S1B).

3) Identically to P_{end}, we define P_{start} as the mean performance during the first quarter of the session.

4) In Fig. 2B, we define maximal performance as the best mean performance in 30 consecutive trials, $P_{max} = \max_{\tau\le T-29} \langle y_t \rangle_{t\in[\tau\to\tau+29]}$.

5) In Fig.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 (Fig. S1C) or other patterns (Fig. S1D). Similarly, the comparisons of pattern-specific learning curves is temporally-aligned between rules (Fig. S1E) and with the general performance (Fig. S1I-N).

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: $R^2 = 1 - \frac{\sum_i \left(P_{end}^i - \mu^i\right)^2}{\sum_i \left(P_{end}^i - \mu\right)^2}$ where, $\mu = \langle P_{end} \rangle$ is the mean end-performance of all the sessions and μ^i are the mean end performances of sessions of either 'easy' or 'hard' rules (Fig.2E, Fig. 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 (Fig.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, ρ_{across}, of the across-days learning (Fig. 2C).
2. Taking only sessions from the end of a consecutive sequence and the beginning of the following sequence, we calculate the correlation, ρ_{recall}, of the monkeys' ability to recall rules they encountered a month before (Fig. 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: Fig. S1C and Fig. 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. Fig 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. Fig.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. Fig. S1D complements Fig. S1C showing that the rule identity impacts this influence as well.

To hone in the general rule-based behavior we add Fig. S1I-N to show specific cases of pattern specific performance deterioration accompanying general performance increase. The examples in Fig. S1I-N. 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 (Fig.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:

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

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,

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

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 Eq. 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. Fig.S1G shows the mean (and SE in error bars) of the across session similarity score:

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

where i, j are different sessions of the same rule and N_{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 Eq. 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 Fig.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 than 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. Fig. 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 $\overrightarrow{v^*} = arg \max_{\|\vec{v}\|=1, \vec{v} \in \mathbb{R}^5} Var\left(\vec{V} \cdot \vec{v}\right)$. We then project each 5-dimensional vector on this principal component and get a single number from each trial, $r(t) = \left(\vec{V}(t) - \langle \vec{V} \rangle\right) \cdot \overrightarrow{v^*}$. 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 (Fig. S3A-C).

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 (Fig.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 ±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 $\gamma = \frac{1}{8}\sum_{\vec{s}} y(\vec{s})$ and $\forall I \in \{i, j, k\}, \alpha_I = \frac{1}{8}\sum_{\vec{s}} y(\vec{s})\, f_I(\vec{s})$ and construct:

$$f(\vec{x}) = \gamma + \sum_I \alpha_I f_I(\vec{x}) \qquad \text{[Eq 4]}$$

It is straightforward to show that $\sum_I f_I(\vec{s}) f_I(\vec{x}) = \begin{cases} 7 \; if \; \vec{x} = \vec{s} \\ -1 \; otherwise \end{cases}$ 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 $f_2 = x_2 x_3$.

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

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, $f = \sum_{I \in \{i,j,k\}} a_I f_I$ such that $\sum_I a_I^2 = 1$ will have $\langle f \rangle = 0$ and $\langle f^2 \rangle = 1$ (because $\langle f_I \cdot f_J \rangle = 0, \forall_{I \neq J}$). Which leads to $Var(f) = 1$.

For the scalar neural response r, the correlation $C(f, r) = \frac{\langle (f - \langle f \rangle)\,(r - \langle r \rangle) \rangle}{\sqrt{Var(f) Var(r)}} = \sum_I a_I \frac{\langle (f_I - \langle f_I \rangle)\,(r - \langle r \rangle) \rangle}{\sqrt{Var(f_I) Var(n)}} = \sum_I a_I C(f_I, r)$ because $Var(f_I) = 1$ and $\langle f_I \rangle = 0 \; \forall I$.

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

$$\boldsymbol{C} = (C(f_1, r), C(f_2, r), ..) \qquad \text{[Eq 5]}$$

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

Corollary 1:

If we measure a neuron's correlations to the features separately, then the unit vector, $\boldsymbol{a}$, 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 $\boldsymbol{a}$ 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, $\boldsymbol{C}$, on $\boldsymbol{a}$,$\boldsymbol{C}\cdot\boldsymbol{a} = \sum_I a_I C(f_I, r)$ is indeed the neuron's rule-correlation $C(f,r)$ (Fig.4B, Fig. 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, $CC_i(t) = corr\left(r(\tau \in w_t), f_i\left(\vec{x}(\tau \in w_t)\right)\right)$, 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 (Fig.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, $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ 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 vs. 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 William'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 (Fig.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 describes in Onken, A., Xie, J., Panzeri, S., and Padoa-Schioppa, C. (2019). Categorical encoding of decision variables in orbitofrontal cortex. PLOS Computational Biology *15*, e1006667. 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. Fig S6 presents vectors both as 3d plots using their 3 leading principal components and also as 2d plots using t-SNE (Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research *9*, 2579–2605.)

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 (Fig. 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 (Fig.4, Fig.5A,B,D,E insets, Fig.S3,4)

Correlation to rule and answer

In Fig. S4M-O,R. we present similar analyses of representation dynamics (as in Fig. 5A,B,D,E) 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 (Fig.5A,B,D,E), 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 Fig. 5A,B,D,E, the fractions of cells that passed the test are compared with a 2-tailed binomial z test.

Fig. S4A-D 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 (Fig.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 α 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 ξ (r,f_I,ξ change from trial to trial).

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

$$Corr(r, f_I) = \frac{a}{\sqrt{a^2 + <\xi^2>}}$$

This correlation coefficient will approach 1 (-1) as 'a' approaches ∞(-∞). 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 Fig 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 ($|\boldsymbol{b}| = \sum_\nu b_\nu^2 = 1$) then we get $Corr(r, b) = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{\sqrt{|\boldsymbol{a}|^2 + <\xi^2>}}$ where '**a**' and '**b**' are vectors. This correlation is maximized if '**b**' is in the direction of '**a**' and boils down to the same scalar expression as above (replacing the scalar 'a' with the magnitude |**a**|). This means that the decision boundary is defined by the direction of '**a**' and its steepness is defined by the magnitude of '**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 (Fig. S4J-L using the same parameters as Fig. 5C,F) 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:

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

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 vs. hard rules in Fig. 5A,B,D,E and high vs. low performance in Fig. S4A-D). 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 (Fig. 5C,F). The neural correlates are averages in regression windows in baseline and segment durations similar to Fig. S4A-D.

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 Fig.6E,G,H).

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 (Fig. 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 Fig.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 Fig.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 ξ_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, $\{\vec{r}(t)\}_{t\in W} \in R^N$, 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, $\vec{\beta} \in R^N$, 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 $CC(\sum_{i=1}^{N}\beta_i r_i, \sum_s a^s f^s(x))$, the correlation coefficient (across trials) between the projection of the neural ensemble on the direction $\vec{\beta}$ and the behavior condition, $\sum_s a^s f^s(x) + constant$.

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

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

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' ($\sigma_{ij} =< \xi_i\xi_j >$) 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 Eq. 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 β coefficient and the dynamics of its correlation to the category (or answer) and summarize the results in Fig. S7D.

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

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

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

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. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 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.

Brasted, P.J., and Wise, S.P. (2004). Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. The European Journal of Neuroscience 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. Experimental Brain Research 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. Experimental Brain Research 139, 116-119.

Chudasama, Y., Daniels, T.E., Gorrin, D.P., Rhodes, S.E., Rudebeck, P.H., and Murray, E.A. (2012). The Role of the Anterior Cingulate Cortex in Choices based on Reward Value and Reward Contingency. Cereb Cortex.

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 U S A 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.

Desrochers, T.M., Amemori, K., and Graybiel, A.M. (2015). Habit Learning by Naive 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.

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. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 23, 5235-5246.

Ganmor, E., Segev, R., and Schneidman, E. (2015). A thesaurus for a neural population code. Elife 4.

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. Learning & Memory 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. Cognitive Science 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., Grundemann, 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.

Grundemann, J., Bitterman, Y., Lu, T., Krabbe, S., Grewe, B.F., Schnitzer, M.J., and Luthi, A. (2019). Amygdala ensembles encode behavioral states. Science 364.

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. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 30, 3339-3346.

Heilbronner, S.R., and Hayden, B.Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. Annual Review of Neuroscience, Vol 39 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.

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., Genud-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. Journal of Experimental Psychology General 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., Tkacik, G., Esteki, M.S., Kiani, R., and Schneidman, E. (2020). Learning probabilistic neural representations with randomly connected circuits. Proc Natl Acad Sci U S A 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? Neuroscience and Biobehavioral Reviews 32, 237-248.

Merchant, H., Zainos, A., Hernandez, 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. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 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 Cognitive Neurosci 18, 974-989.

Nosofsky, R.M., Kruschke, J.K., and McKinley, S.C. (1992). Combining exemplar-based category representations and connectionist learning rules. Journal of experimental psychology Learning, memory, and cognition 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 U S A 116, 15210-15215.

Ongur, 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 computational biology 15, e1006667.

Padoa-Schioppa, C., and Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. Nature 441, 223-226.

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 e518.

Pryluk, R., and Paz, R. (2019). Learning outside the box. Proc Natl Acad Sci U S A 116, 15316-15318.

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. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 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. Psychological Monographs: General and Applied 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. Journal of Mathematical Psychology 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.

Tkacik, G., Granot-Atedgi, E., Segev, R., and Schneidman, E. (2013). Retinal metric: a stimulus distance measure derived from population neural responses. Physical review letters 110, 058104.

Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature 411, 953-956.

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.

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.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Critical Commercial Assays		
0.6–1.2 MΩ glass coated tungsten electrodes	Alpha Omega	
metal guide (outer diameter: 0.51 mm, inner diameter: 0.41 mm)	Cadence	Gauge 25xxtw
Head-tower and electrode-positioning-system	Alpha Omega	
Recording chamber	Alpha Omega	
Deposited Data		
Raw electrophysiology and behavior data	This manuscript	Upon request
Experimental Models: Organisms/Strains		
Macaca Fascicularis monkeys		
Software and Algorithms		
Alpha Lab SNR	Alpha Omega	
Offline sorter	Plexon	
Matlab	Mathworks	
In-house developed code	This manuscript	https://github.com/yardencsGitHub/CohenSchneidmanPaz2020_Code
Other		

Figure 1

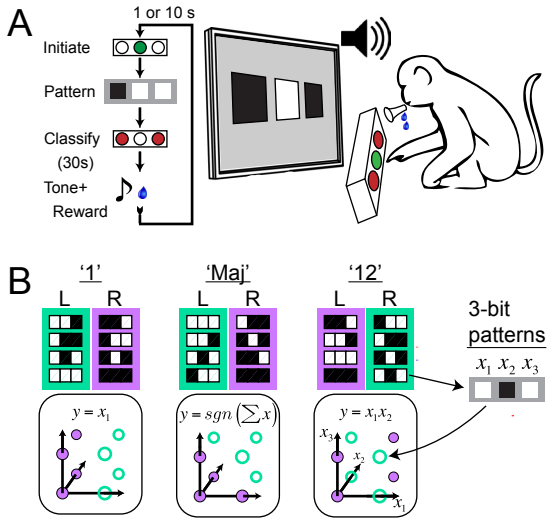

A
1 or 10 s
Initiate
Pattern
Classify
(30s)
Tone+
Reward
B
'1'
'Maj'
'12'
L
R
L
R
L
R
3-bit
patterns
$x_1\ x_2\ x_3$
$y = x_1$
$y = sgn\left(\sum x\right)$
$y = x_1 x_2$
x_3
x_2
x_1

Figure 2

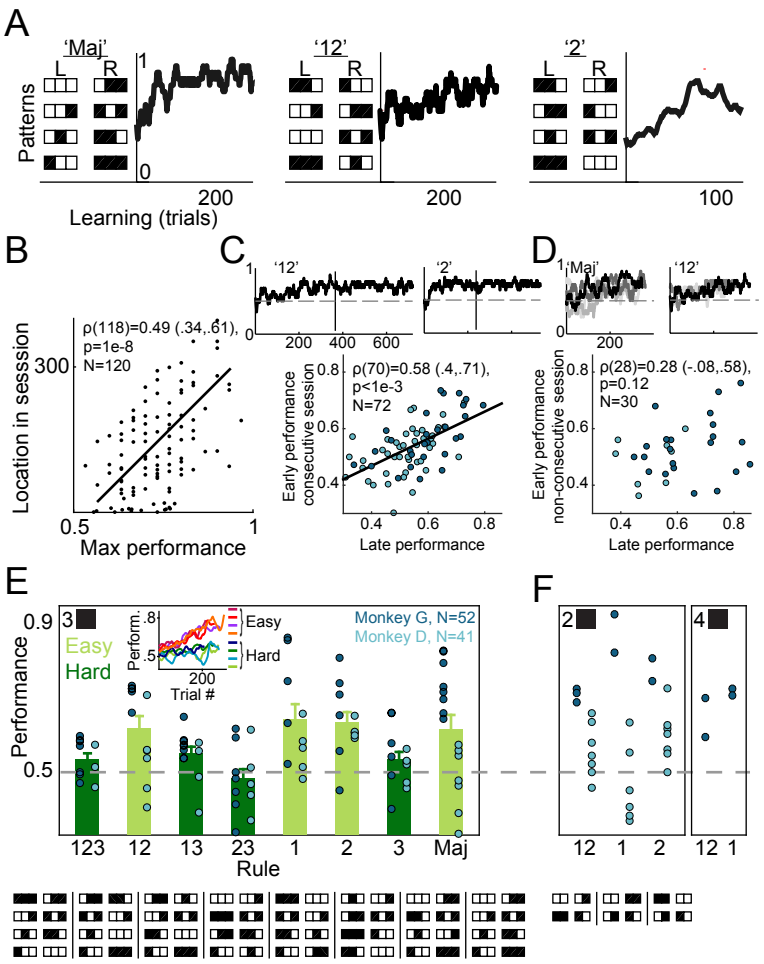


Figure 2
A
'Maj'
'12'
'2'
L
R
Patterns
1
0
200
200
100
Learning (trials)
B
C
D
'12'
'2'
'Maj'
'12'
ρ(118)=0.49 (.34,.61),
p=1e-8
N=120
Location in sesssion
300
0.5
1
Max performance
ρ(70)=0.58 (.4,.71),
p<1e-3
N=72
Early performance
consecutive session
0.4
0.6
0.8
Late performance
ρ(28)=0.28 (-.08,.58),
p=0.12
N=30
Early performance
non-consecutive session
E
F
Performance
0.9
0.5
3
Easy
Hard
Perform.
.8
.5
200
Trial #
Easy
Hard
Monkey G, N=52
Monkey D, N=41
123
12
13
23
1
2
3
Maj
Rule
2
4
12
1
2
12
1

Figure 3

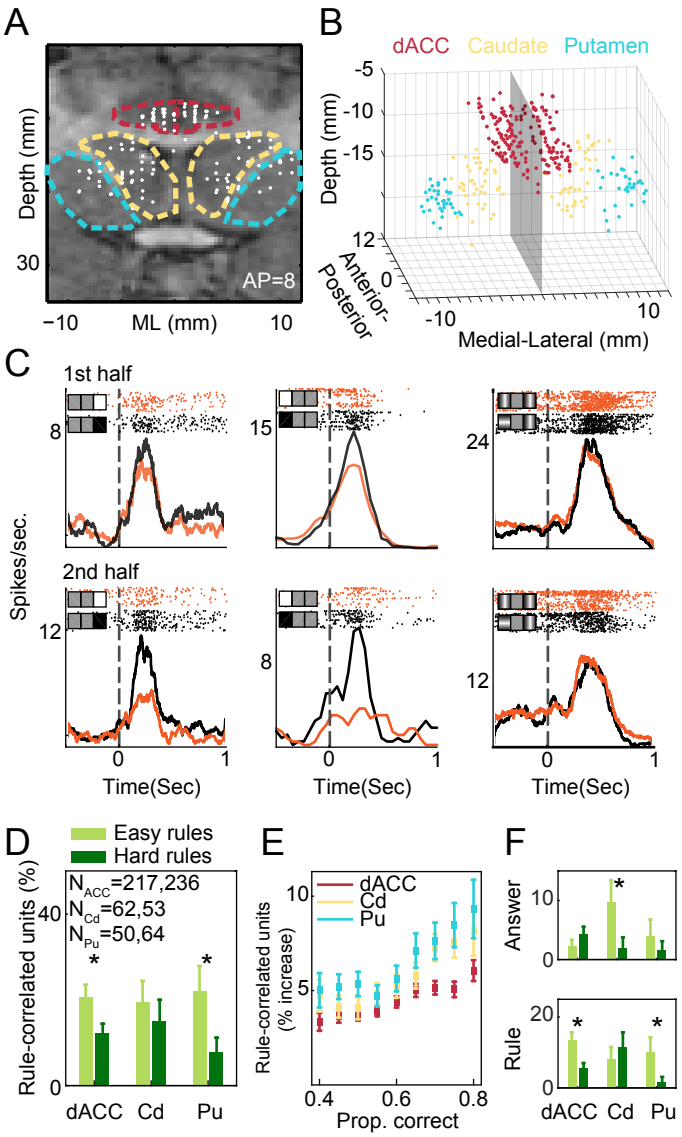

A
Depth (mm)
30
AP=8
−10
ML (mm)
10
B
dACC
Caudate
Putamen
-5
-10
-15
Depth (mm)
12
0
Anterior-
Posterior
-10
10
Medial-Lateral (mm)
C
1st half
8
15
24
Spikes/sec.
2nd half
12
8
12
0
1
Time(Sec)
Time(Sec)
Time(Sec)
D
Easy rules
Hard rules
N_ACC=217,236
N_Cd=62,53
N_Pu=50,64
Rule-correlated units (%)
40
0
dACC
Cd
Pu
E
dACC
Cd
Pu
10
5
Rule-correlated units
(% increase)
0.4
0.6
0.8
Prop. correct
F
Answer
10
0
Rule
20
0
dACC
Cd
Pu

Figure 4

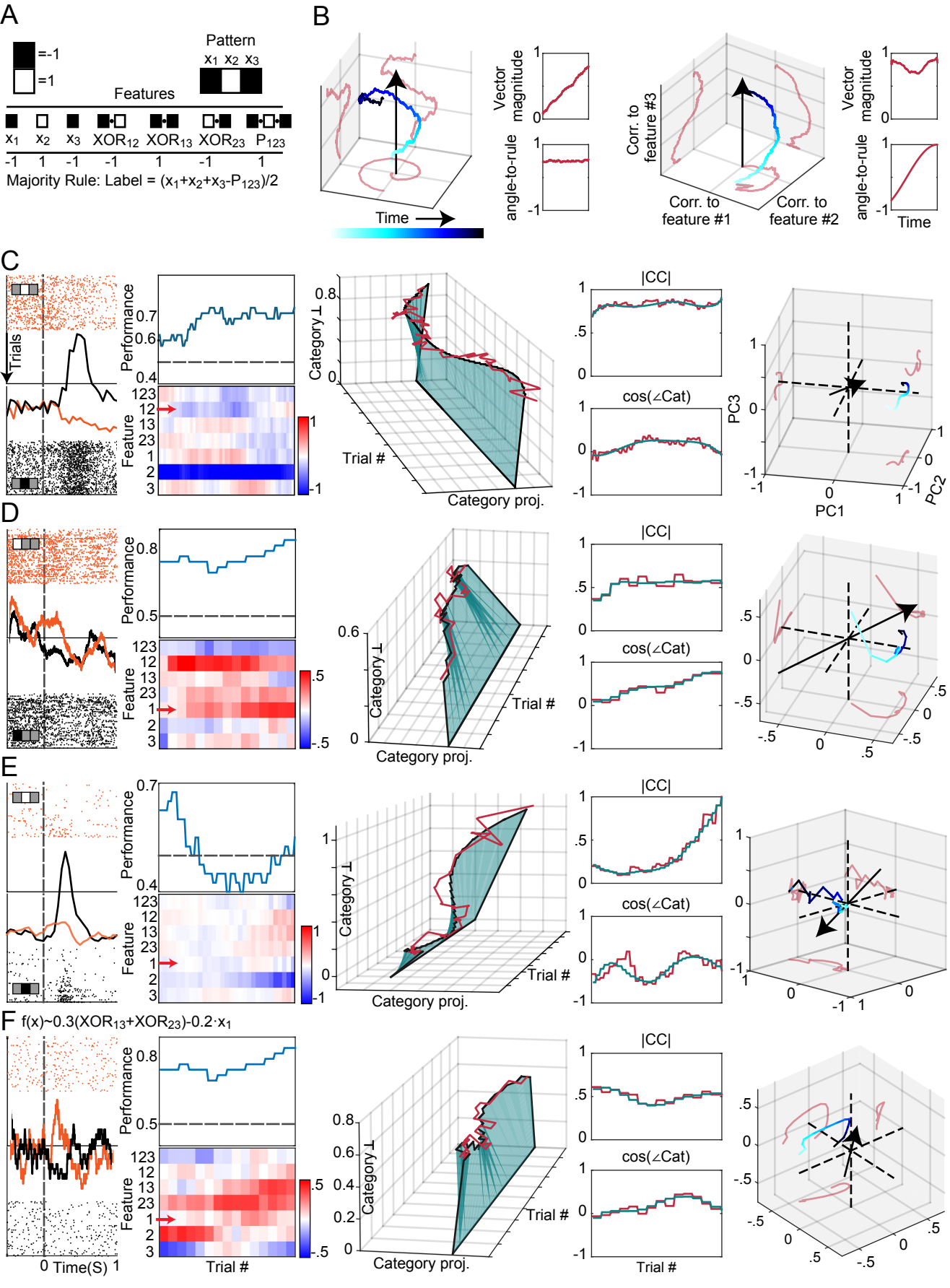

A
=-1
=1
Pattern
x1 x2 x3
Features
x1 x2 x3 XOR12 XOR13 XOR23 P123
-1 1 -1 -1 1 -1 1
Majority Rule: Label = (x1+x2+x3-P123)/2
B
Time
Vector magnitude
angle-to-rule
Corr. to feature #3
Corr. to feature #1
Corr. to feature #2
C
D
E
F
f(x)~0.3(XOR13+XOR23)-0.2·x1
Trials
Performance
Feature
Category ⊥
Trial #
Category proj.
|CC|
cos(∠Cat)
PC1
PC2
PC3
0 Time(S) 1

Figure 5

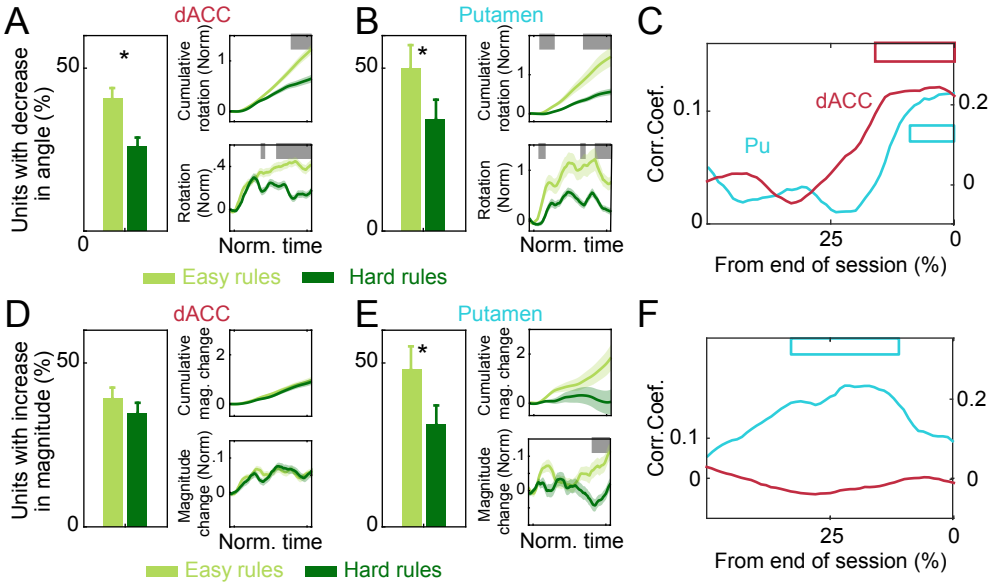

A
dACC
Units with decrease in angle (%)
50
0
*
Cumulative rotation (Norm)
1
0
Rotation (Norm)
.4
0
Norm. time
B
Putamen
50
0
*
Cumulative rotation (Norm)
1
0
Rotation (Norm)
1
0
Norm. time
C
Corr.Coef.
0.1
0
dACC
Pu
0.2
0
25
0
From end of session (%)
Easy rules
Hard rules
D
dACC
Units with increase in magnitude (%)
50
0
Cumulative mag. change
2
0
Magnitude change (Norm)
.1
0
Norm. time
E
Putamen
50
0
*
Cumulative mag. change
2
0
Magnitude change (Norm)
.1
0
Norm. time
F
Corr.Coef.
0.1
0
0.2
0
25
0
From end of session (%)
Easy rules
Hard rules

Figure 6

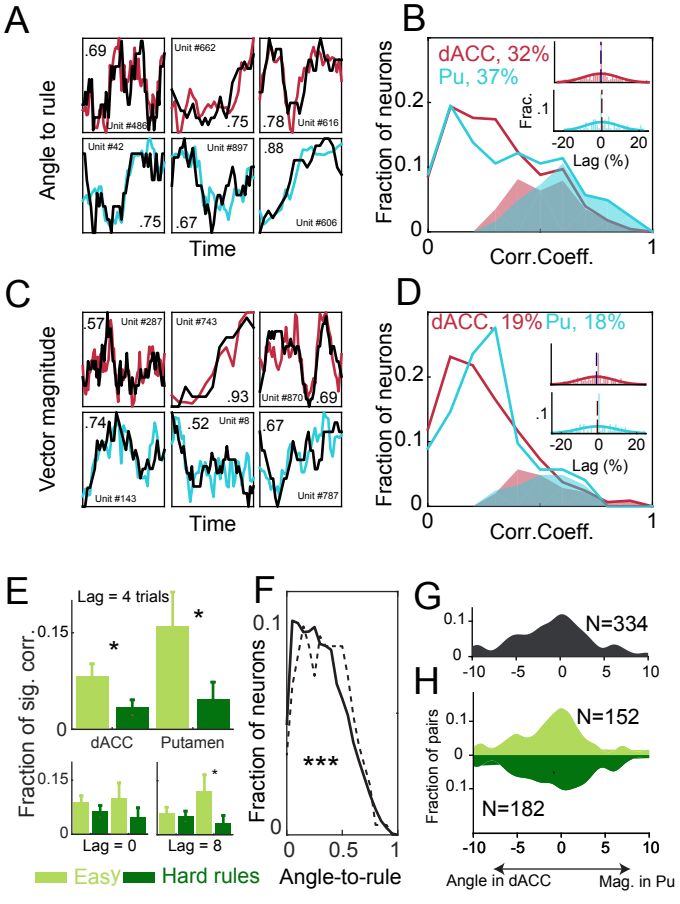
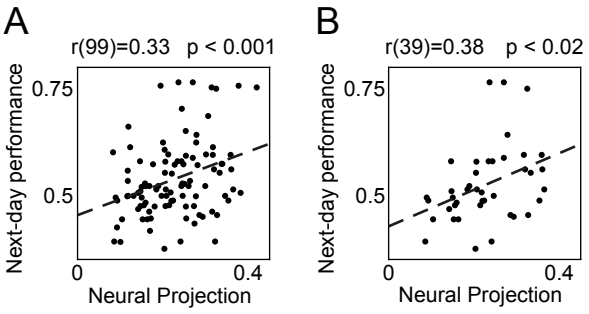

A
Angle to rule
Time
Unit #662
Unit #486
Unit #616
Unit #42
Unit #897
Unit #606
B
dACC, 32%
Pu, 37%
Fraction of neurons
Corr.Coeff.
Frac.
Lag (%)
C
Vector magnitude
Time
Unit #287
Unit #743
Unit #870
Unit #143
Unit #8
Unit #787
D
dACC, 19% Pu, 18%
Fraction of neurons
Corr.Coeff.
Lag (%)
E
Lag = 4 trials
Fraction of sig. corr.
dACC
Putamen
Lag = 0
Lag = 8
Easy
Hard rules
F
Fraction of neurons
Angle-to-rule
G
N=334
H
N=152
N=182
Fraction of pairs
Angle in dACC
Mag. in Pu

Figure 7


A
r(99)=0.33 p < 0.001
Next-day performance
0.75
0.5
0
0.4
Neural Projection
B
r(39)=0.38 p < 0.02
Next-day performance
0.75
0.5
0
0.4
Neural Projection

A

rule	novel	repeat	frac
'123'	5	5	0.5
'12'	5	6	0.45
'13'	5	7	0.41
'23'	5	7	0.41
'1'	4	5	0.44
'2'	3	7	0.3
'3'	4	8	0.33
'Maj'	5	9	0.35

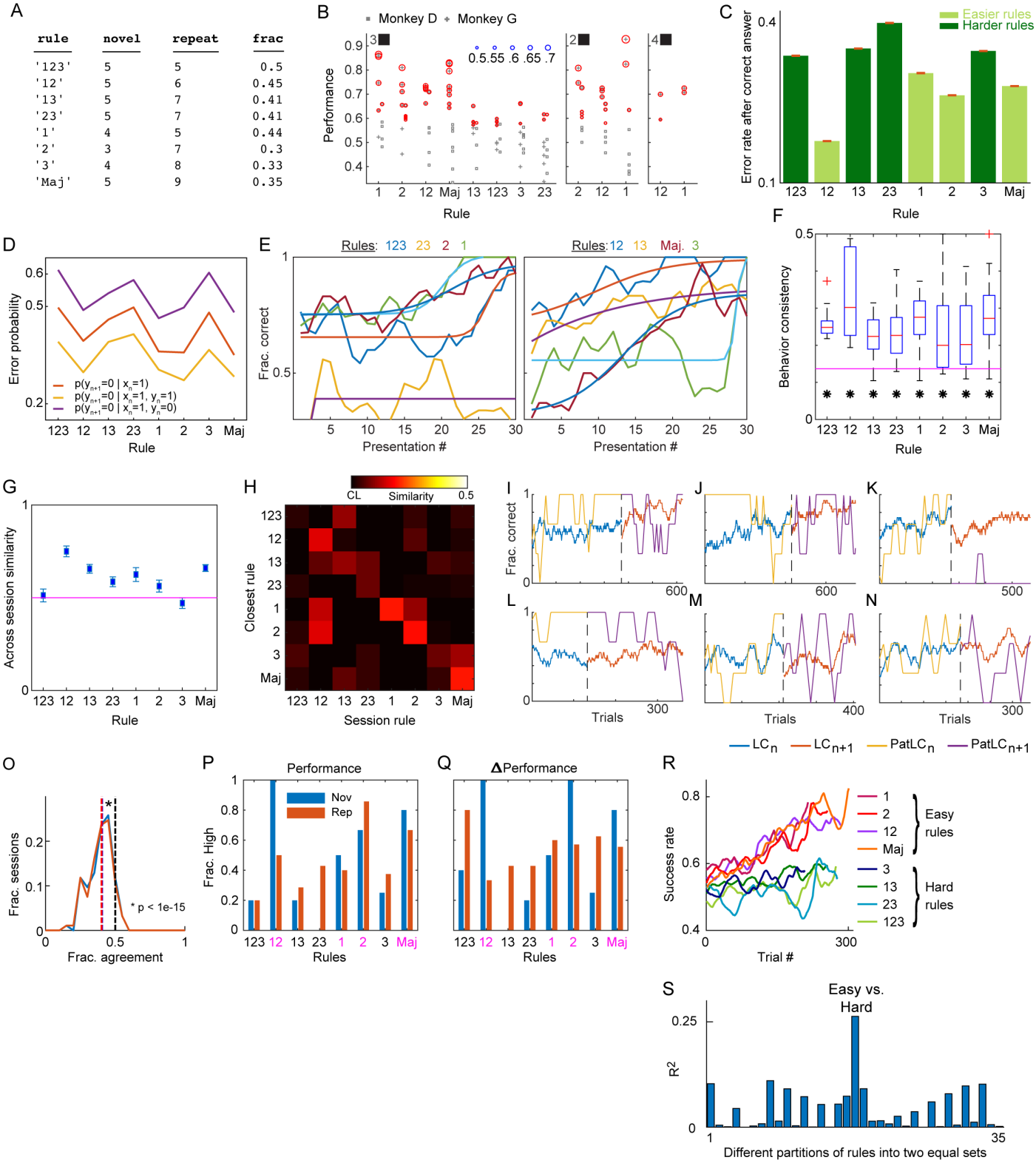


Fig. S1. Performance and rule-dependency (Related to main figure 2)

A. The number of sessions for each rule when it was different from the rule in the previous day-session (novel) or when it is the same rule (repeat).

B. **Performance**. Identical to Fig. 2E but with ordering the rules according to mean performance (left to right). Shown is the performance at the last quarter of each session (P_{end}), and the red circles show the confidence level. Both animals had at least one successful session in each rule, but four rules (left) had many more successful sessions than the other four (right).

C. **Error probability after making a correct classification**. For each rule shown is the probability of making an error on a pattern after choosing correctly in the previous presentation of that same pattern. Error bars depict SE (red). The strong dependency on the specific rule suggests the monkeys did not memorize specific patterns after a correct classification ($\chi^2(7)$=88.97, p < 1e-15, Kruskal-Wallis test).

D. **Error probability in different stimulus-response associations show rule-dependent coupling.** For each rule (x-axis) shown is the probability of making an error in classifying a pattern 'y' after correctly-classifying a different pattern 'x' in the previous presentation of 'x' (red line). In yellow/purple shown is the same error probability conditioned also on correctly/incorrectly classifying 'y' in the previous presentation. The latter calculation shows the global effect of learning – fewer errors after correct choices. All probabilities have strong dependency on the rule being learned (Kruskal-Wallis test, $\chi^2 > 88$, p<1e-10 for all), demonstrating a strong coupling between stimulus-response pairs as expected in rule-based learning.

E. **Acquisition rates for 'salient' patterns**. The most salient pattern, '000', was labeled 'right' and 'left' equally in the set of rules (4+4 rules). Shown are the learning curves (colored lines + sigmoidal fits), the proportion of correct labeling (y-axis, smoothed across days with a running window of 4 presentations) as a function of pattern presentations (x-axis). The two panels are for the 4 rules in which 000 is labeled 'left' (left panel) and for the 4 it was labeled 'right' (right panel). There was variability across rules in how the visually salient pattern was learned, showing that patterns that draw more attention are not learned faster or similarly across different rules, and providing more evidence against a simple memorization process (Kruskal-Wallis test, $24 < \chi^2(3,36) < 33$, 1e-7 < p < 1e-4 for all).

F-H. **Behavior consistency**. **F.** Within-session consistency scores (y-axis, methods) for all sessions with a single rule (x-axis). All rules showed above-chance level consistency (t-test, p<0.05 for all, chance-level in magenta). **G.** Between-session pairwise consistency score across sessions with the same rule. **H**. A classifier fitted to the last 25% of each session in which a certain rule was learned (x-axis) and compared to the other rules (y-axis, see methods). Color scale is from chance level (CL) to the maximal possible classification value (0.5).

I-N. Learning disrupts previously acquired stimulus-response associations. Blue and red lines show overall performance in consecutive sessions (LC_n,LC_{n+1}). The underlying rule is switched between the sessions (dashed line) but in these rules 4 out of 8 patterns did not change their correct stimulus-response association. The yellow and purple lines show the performance in one of these 4 patterns ($PatLC_n$,$PatLC_{n+1}$). As seen and in contrast to learning by independent stimulus-response pairing, rule-based learning implies that pattern-specific performance can deteriorate while the general performance improves (**A-D** for monkey G. **E,F** for monkey D). This happens even when the association-specific performance reaches 100% (**A,E**).

O. Win→Stay, Lose→Shift strategy poorly describes behavior. We simulated win→stay, lose→shift answer sequences, using the same pattern presentations and rules as in the actual sessions. Shown are the histograms of the overlap between the simulated sessions with the actual behavior (Red and blue for simulations starting with the left and right choices). Colored dashed lines show the median values, both smaller than chance level (0.5, black dashed line, Wilcoxon's signed ranked test, p<1e-15).

P. The fraction of sessions in which performance was high, separating sessions with novel rules (blue) from sessions that repeat the rule from the previous day (red). Easy rules (magenta) are the highest 4 rules in de-novo learning.

Q. The fraction of sessions in which change-in-performance within a session was high, separating sessions with novel rules (blue) from sessions that repeat the rule from the previous day (red). Here as well, easy rules (magenta) are the highest 4 rules in de-novo learning.

R. Average learning curves also separate easy and hard rules.

S. Performance variance explained by partitioning rules into two equal sets of four rules. The Easy-Hard partition explains significantly more than all other partitions.

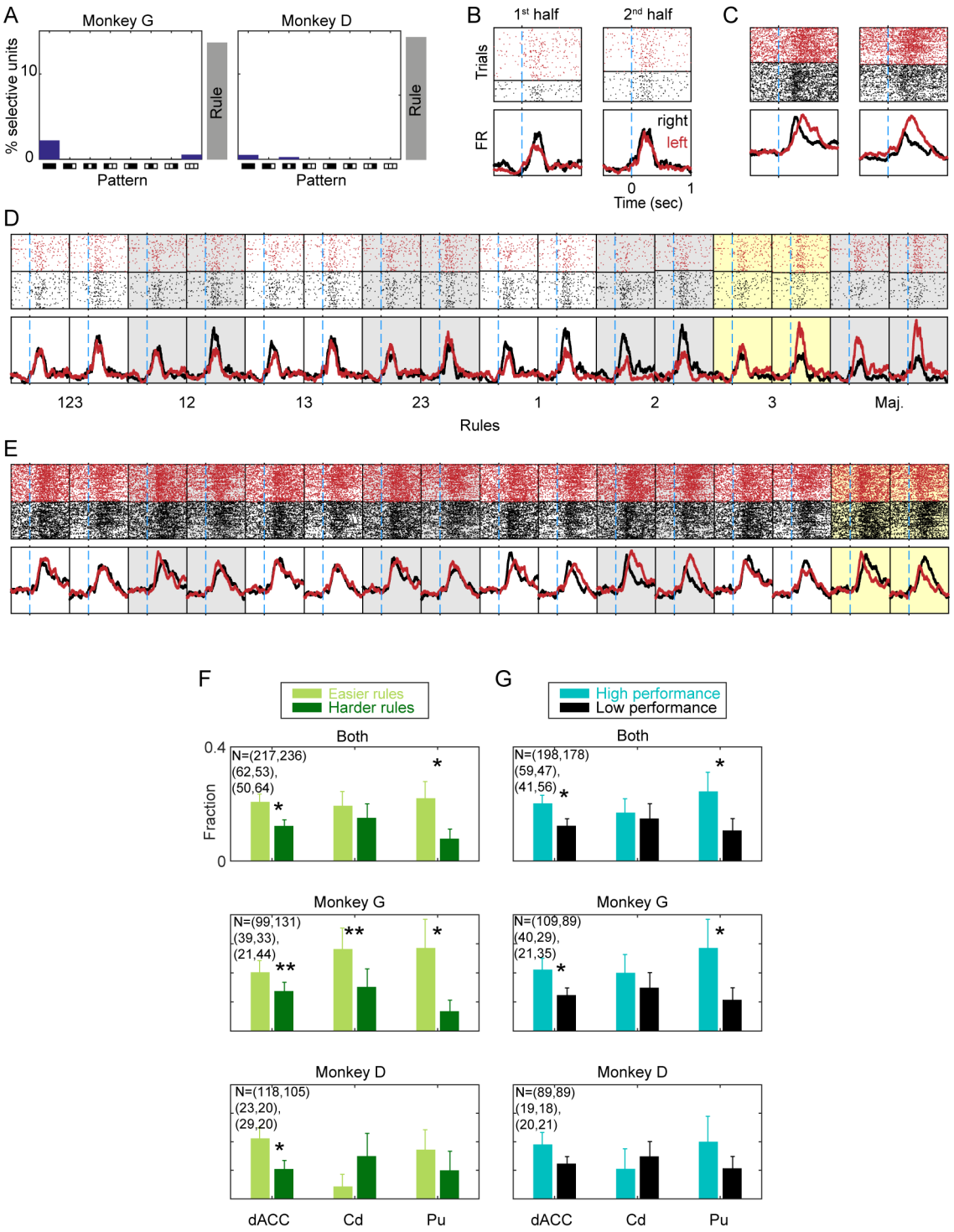


Fig. S2. Rule-representation in single-units (Related to main figure 3)

A. **Pattern selective neurons**. The percent of neurons that were selective for a single pattern (bars, y-axis) plotted for all patterns (x-axis). Gray bars on the right indicate the percent of rule correlated neurons. Left: monkey G. Right: monkey D. Very few neurons were selective for a single-patterns, suggesting against single stimulus-response learning.

B-E. Rasters and PSTHs from single neurons showing changes in representing the animals' choice, the correct rule, and all other rules. Spikes aligned to stimulus onset (t=0, x-axis, light blue dashed line). Trials (y-axis) are stacked according to the animal left or right choice (red, black dots in the top panel and curves in the bottom panel mark spike times and firing rates accordingly). Calculations are made separately in the first and second halves of the session (left - right panels).

B. A neuron that does not differentiate the animal's choice in the 2nd half of the session.

C. A neuron that learns to differentiate the animal's choice.

D. The same neuron (and data) as in panel **B,** separated by the labels of the 8 rules in the experiment (x-axis labels). Gray and white shadings distinguish the different rules and yellow shading marks the actual rule being learned.

E. similar to panel **D** but for the neuron in **C**.

F,G. Rule-representation differ in sessions of high/low performance, easy/hard rules in both monkeys.

Fraction of neurons that exhibited significant rule correlation during the last third of the session (Bars+SE), plotted for the 3 regions, and shown for both animals combined (top), separately for monkey G (middle) and monkey D (bottom).

F. Separating sessions by easy-hard rules. A-top (similar to Fig.3D): dACC: $z=2.43$, $p<0.01$, Pu: $z=2.12$, $p<0.02$; Monkey G: dACC: $z=1.305$, $p<0.1$, Cd: $z=1.3275$, $p<0.1$, Pu: $z=2.37$, $p<0.01$; Monkey D: dACC: $z=2.17$, $p<0.02$.

G. Separating sessions by performance, independent of the rule. Both: dACC: $z=2.046$, $p<0.021$, Pu: $z=1.79$, $p<0.04$; Monkey G: dACC: $z=1.85$, $p<0.04$, Pu: $z=1.97$, $p<0.03$**.**

Binomial z-test (*=$p<0.05$, **=$p<0.1$).

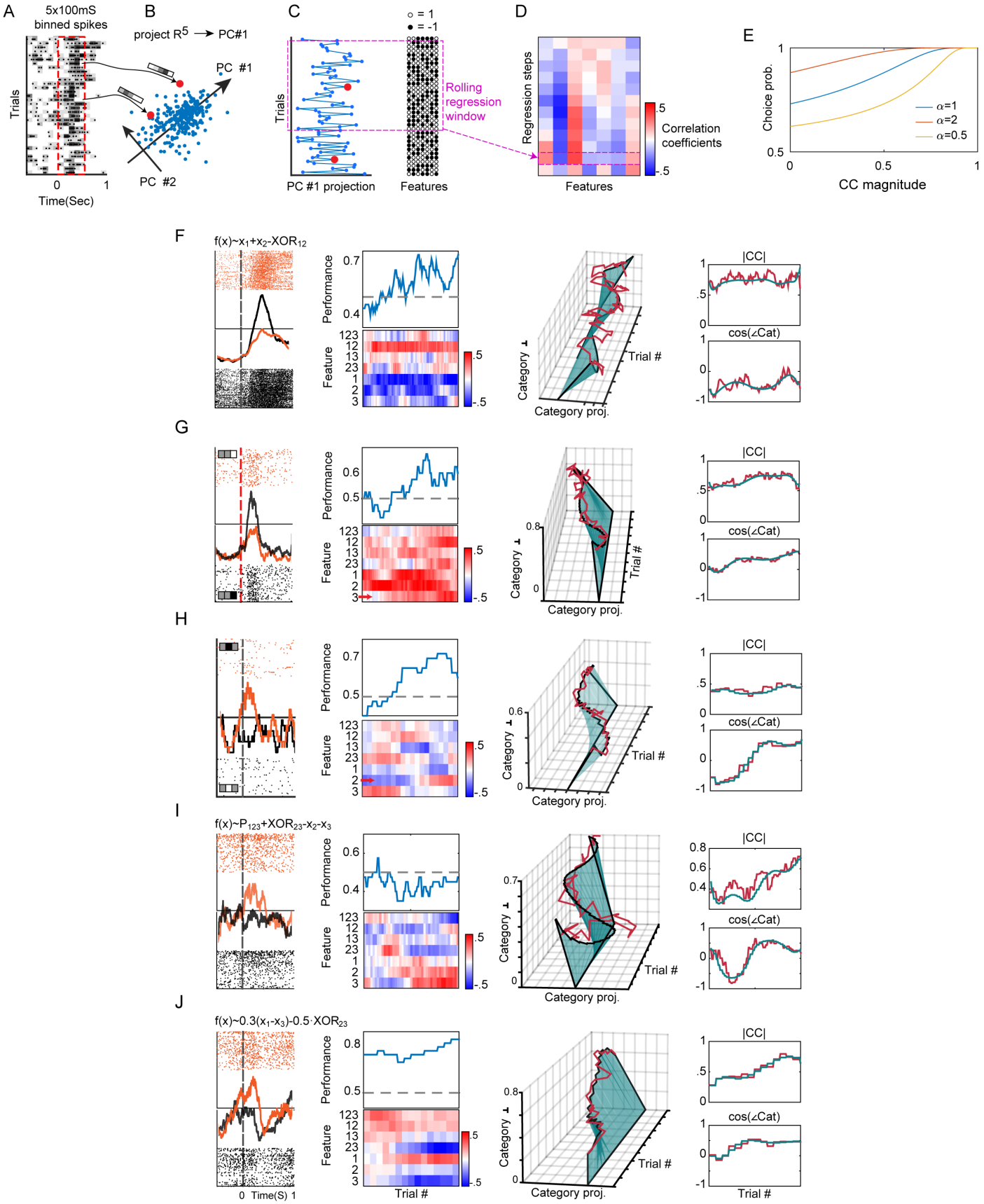


Fig .S3. Single neurons' dynamic feature representation (Related to main figure 4)

A. Spikes (dots) during 500 msec after the stimulus presentation (x-axis, dashed red frame) are counted in 5 x 100 msec bins (gray scale).

B. The 5-vectors of spike counts from all trials are projected on the principal component explaining the most variance across trials. (arrows and red dots show specific examples).

C. The resulting neural 1-d stimulus response, x-axis, is correlated across trials (y-axis) with stimulus visual features (7-vectors of black and white circles) in a rolling regression window (purple frame).

D. The resulting vector of 7 correlation coefficients spans the neuron's visual feature preference during that regression window and its dynamics across regression steps (*neural-vector*, y-axis).

E. Neural vector magnitude translates to behavioral confidence: confidence is defined as the choice probability of a decision based on neural activity (see methods 'Interpreting neural-vector magnitude as confidence'). As the magnitude of the features correlation vector (x-axis) increases, so does the choice probability (y-axis), indicating an increasing confidence.

F-J. Examples of learning related neural dynamics. Same format as in Fig.4, showing neurons with stable feature selectivity, high dimensional rotation, magnitude increase, and complex trajectories.

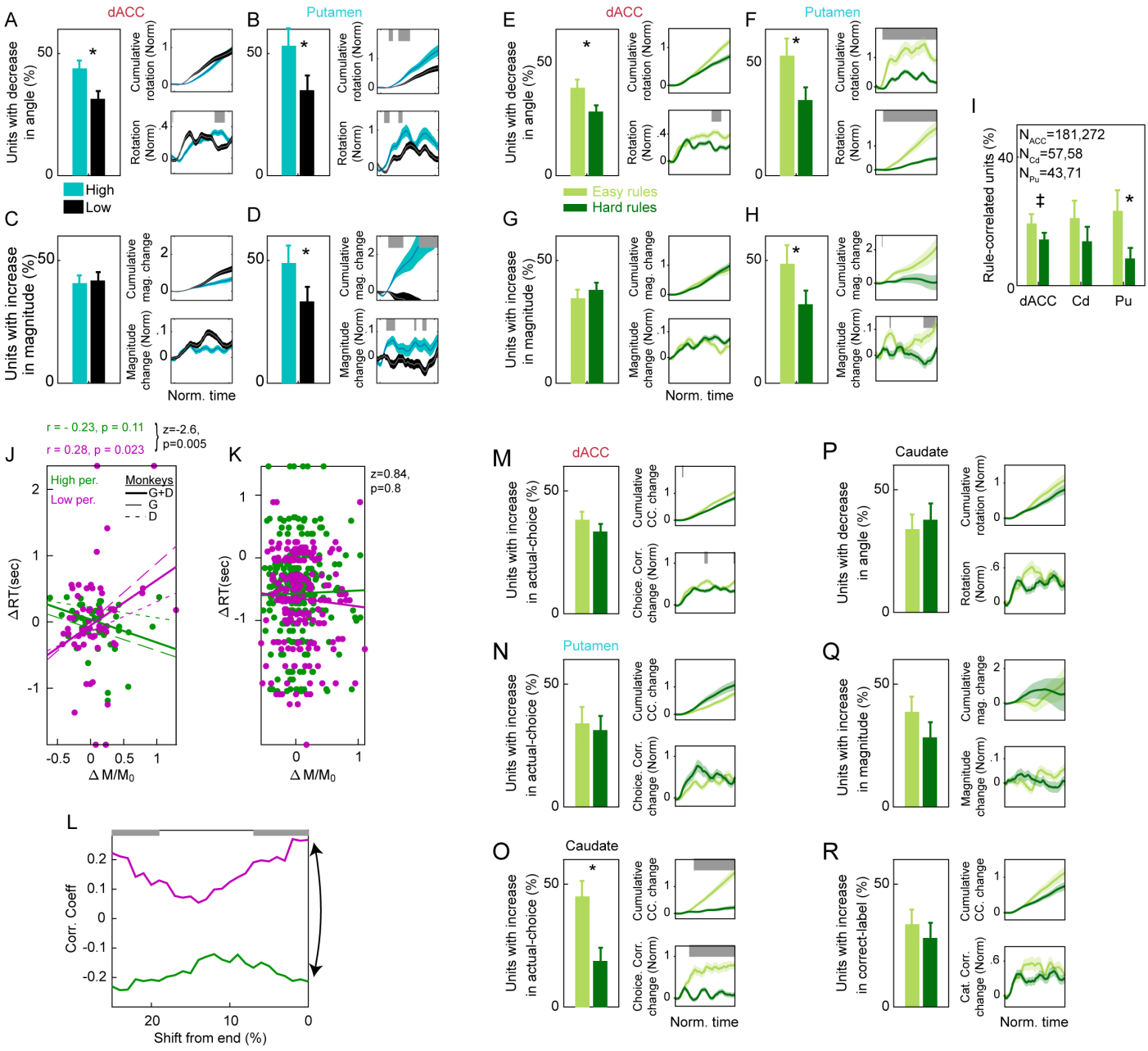


Fig .S4. Differential representation dynamics in the dACC, the Putamen, and the Caudate (Related to main figure 5)

A-D. Dividing sessions by performance. Replicating the results of main Fig. 5 but when dividing sessions into high and low performance independent of the rule type (i.e. instead of dividing session according to easy and hard rules, as in Fig. 5). High – Low performance was determined by above – below the median performance for each animal.

E-I. Redefining the Majority rule as hard for monkey D replicates the results. E-H. Same presentation as in Fig. 5 and replicating the results. **I.** Same presentation as in Fig. 2D and replicating the results.

J. Correlations between change in neural-vector magnitude and reaction-time are significantly different between sessions of high and low performance (Z-test of Fisher transformed Pearson correlation coefficients, Z=-2.6,-2.05,-1.6, p=0.005, p=0.02, p=0.055 for both monkeys, monkey G, monkey D)

K. Same as in **J** for dACC neurons.

L. The correlation across neurons between change in reaction time and change in magnitude, shown for different offsets from the end of the session (x-axis, % of session duration). Green and purple lines show

the coefficients for high and low performance sessions. Gray bars mark significant differences (Fisher corrected measure, methods, p<0.05). Arrow marks the contrast in panel **J**.

M-R. Neural representation in the Caudate reflect choice. M,N, The dACC (**M**) and Putamen (**N**) do not show a differential change in representing the actual choice. **O.** Caudate neurons show a rule-dependent (easy vs. hard) increase in representing the animals' actual-choice. **P-R.** Caudate neurons do not show a differential (easy vs. hard) decrease in angle-to-rule (**P**), or an increase in magnitude (**Q**), or in representation of the correct-label (**R**).

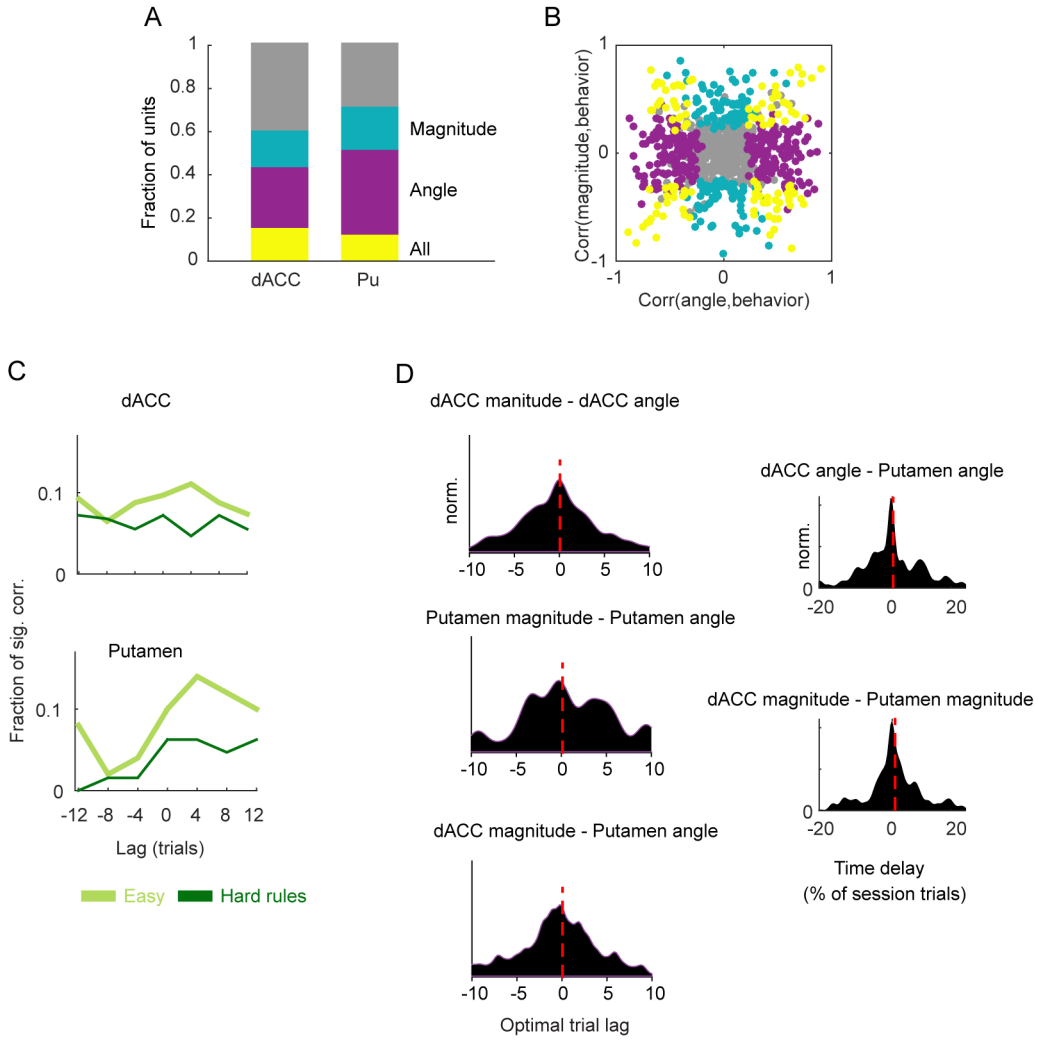


Fig. S5. **Temporal alignment and similarity of dynamic geometric representations and learning curves (Related to main figure 6)**

A. Enumeration of the significant (Pearson, p<0.05) correlations to all variable combinations in the different regions (x-axis). The colored bars show the fraction of units with significant correlations between the learning curve and both angle-to-rule and correlations vector-magnitudes (yellow), only angle-to-rule (purple), or only vector-magnitudes (turquoise).

B. Comparing the magnitude correlation (y-axis) to the angle-to-rule correlation (x-axis) for all neurons. Color-coding as in panel (**A**)

C. Fraction of significant correlations between the change in rotation (the derivative of the angle-to-rule) and the behavioral curve, computed for different lags between the two measures. As shown in main Fig. 6E, more neurons followed the behavior (positive lags) in *easy* than in *hard* rules.

D. For all simultaneously recorded pairs, we computed the optimal lag between the vector-magnitude and the angle-to-rule, for all possible combinations of within and between regions. All options shown above were not different than zero (t-test, p>0.1 for all). Only lags between vector-magnitude in Putamen and angle-to-rule in dACC were different than zero with the Putamen magnitude following dACC rotation (shown in Fig. 6G).

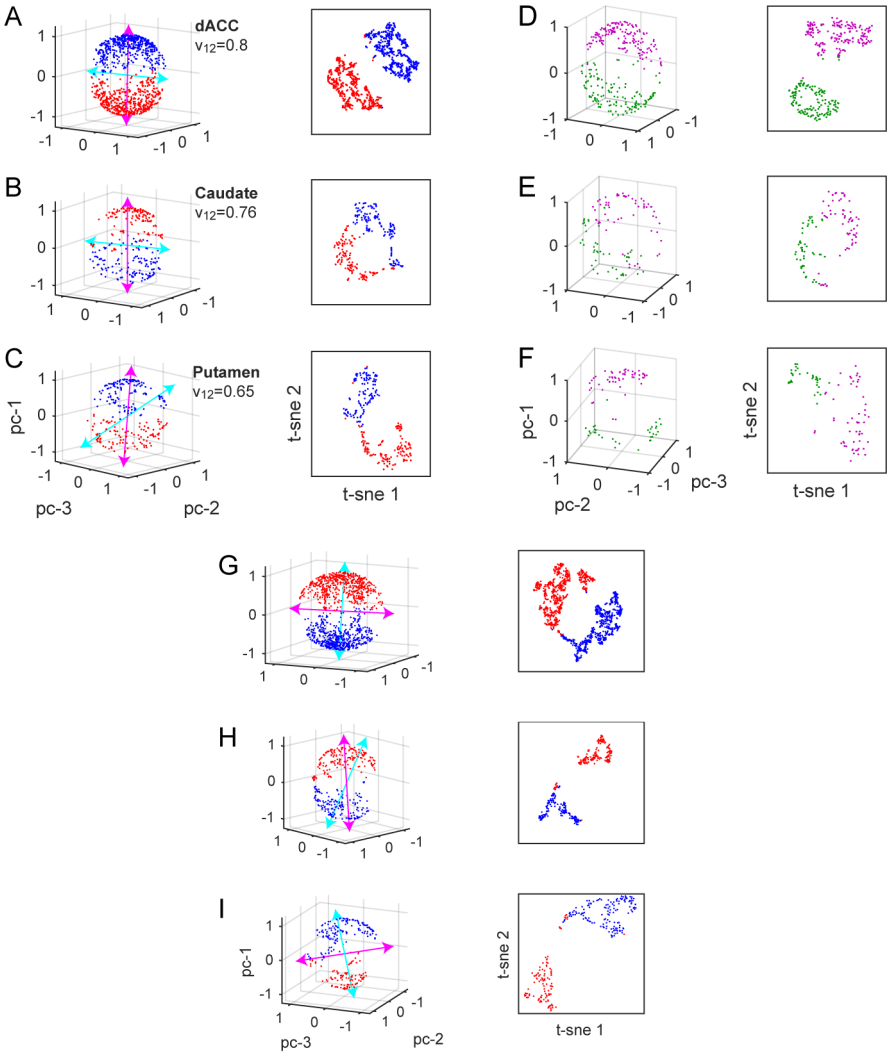


Fig. S6. Single-unit responses exhibit categorical properties (Related to Discussion)

Four task conditions defined by the correct category label and the animals' answers (left or right) are used to cluster single neuron responses. Dots mark mean responses in 40 trial regression windows normalized to unit length. The left panels show responses from category-correlated regression windows, projected on their 3 main principal components. Magenta/cyan arrows show the direction of the category/answer variables. Right panels show the same data projected on 2 t-SNE components to reveal categorical separation.

A-C. Neurons in the dACC showed the lowest dimensionality (V_{12} is the fraction of variance captured by the first 2 PC's).

D-F. Same as **A-C** but only for windows in which ANOVA tests showed significant dependence of the neural response on the task condition.

G-I. Same as **A-C** for regression windows with significant answer correlations. This reveals stronger categorical properties in the striatum and weaker in the dACC.

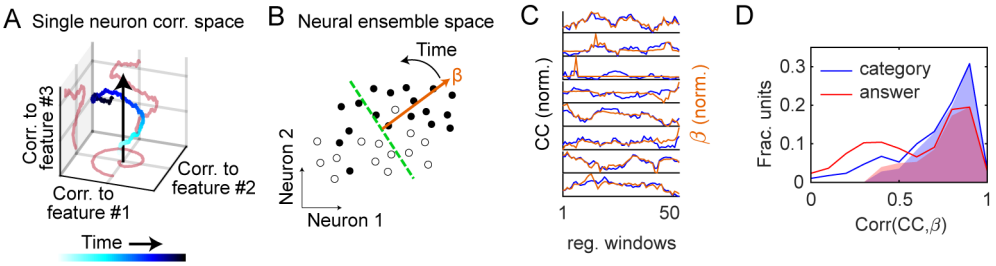


Fig. S7. Correspondence between feature-based correlations and population-level decoding of ensemble activity (Related to Discussion)

A. In our study the single neuron dynamics is a trajectory in the space of correlations to visual features (axes). A behavior policy is represented as a direction in this space (arrow). In successful learning the trajectory moves to increase the projection on the policy (magnitude increase and/or rotation towards the rule).

B. The behavior policy can be decoded from the activity of multiple neurons. The space of ensemble activity is defined by axes matching the activity if individual neurons and the response in each trial is a point in this space. A linear decoder, fitted to the behavior, defines a separating plane (dashed line) in this space perpendicular to the vector $\vec{\beta}$. To learn, the plane is rotated to achieve optimal separation by the desired policy (full vs. empty circles).

C. Example of 8 simultaneously recorded neurons. For each neuron (insets) the normalized category correlation (z-scored, blue) is overlaid on the β (z-scored, orange) for that neuron computed by a category classifier fitted in the same regression window (x-axis) to all 8 neurons. The traces expose the similarity of the two approaches and also points of disagreement.

D. Under assumptions of linear visual response of single-neurons and conditional independence between neurons, we find that the two approaches shown in A,B, can be related. The histograms show correlation coefficients between the single neuron category dynamics and the dynamics of the weight β fitted to neurons in decoding the category from the population activity. The shaded areas mark correlations with $p<0.001$, amounting to 81% (or 62% for answers) of all neurons.