Title

A Comparison of Rapid Rule-Learning Strategies in Humans and Monkeys.

Journal

The Journal of Neuroscience, 44(28)

Authors

Goudar, Vishwa

Kim, Jeong-Woo

Liu, Yue

et al.

Publication Date

2024-07-10

DOI

10.1523/JNEUROSCI.0231-23.2024

Peer reviewed

Behavioral/Cognitive

A Comparison of Rapid Rule-Learning Strategies in Humans and Monkeys

Vishwa Goudar,[1]* Jeong-Woo Kim,[1]* Yue Liu,[1] Adam J. O. Dede,[2] Michael J. Jutras,[2] Ivan Skelin,[3,4] Michael Ruvalcaba,[5] William Chang,[5] Bhargavi Ram,[3,4] Adrienne L. Fairhall,[2] Jack J. Lin,[3,4] Robert T. Knight,[5,6] Elizabeth A. Buffalo,[2,7] and Xiao-Jing Wang[1]

[1]Center for Neural Science, New York University, New York 10003, [2]Department of Physiology and Biophysics, University of Washington, Seattle, Washington 98195, [3]Department of Neurology, University of California, Davis, California 95616, [4]The Center for Mind and Brain, University of California, Davis, California 95616, [5]Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, [6]Department of Psychology, University of California, Berkeley, California 94720, and [7]Washington Primate Research Center, University of Washington, Seattle, Washington 98195

Interspecies comparisons are key to deriving an understanding of the behavioral and neural correlates of human cognition from animal models. We perform a detailed comparison of the strategies of female macaque monkeys to male and female humans on a variant of the Wisconsin Card Sorting Test (WCST), a widely studied and applied task that provides a multiattribute measure of cognitive function and depends on the frontal lobe. WCST performance requires the inference of a rule change given ambiguous feedback. We found that well-trained monkeys infer new rules three times more slowly than minimally instructed humans. Input-dependent hidden Markov model–generalized linear models were fit to their choices, revealing hidden states akin to feature-based attention in both species. Decision processes resembled a win–stay, lose–shift strategy with interspecies similarities as well as key differences. Monkeys and humans both test multiple rule hypotheses over a series of rule-search trials and perform inference-like computations to exclude candidate choice options. We quantitatively show that perseveration, random exploration, and poor sensitivity to negative feedback account for the slower task-switching performance in monkeys.

Significance Statement

Advances in training and recording from animal models support the study of increasingly complex behaviors in nonhumans. Before interpreting their neural computations as human-like, we must first ascertain whether their computational algorithms are human-like. We compared rapid rule-learning strategies of macaque monkeys and humans on a Wisconsin Card Sorting Test variant and found that monkeys are 3–4 times slower than humans at learning new rules. Model fits to choice behavior revealed that both species use qualitatively similar exploration strategies with different decision criteria. These differences produced distinct errors in monkeys that are similar to those observed in humans with prefrontal deficits. Our results generate detailed neural hypotheses and highlight the need for systematic interspecies behavioral and neural comparisons.

Introduction

Animal models are essential for mechanistic investigations of the circuit underpinnings of complex computation. However, any extrapolation of the findings to an understanding of human cognition relies on an interspecies overlap of the computational and neurocognitive means used to carry out complex tasks (Melloni et al., 2019; Birch et al., 2020). The prefrontal cortex, which plays an essential role in higher cognitive functions (Fuster, 2015), has more neurons in humans compared with nonhuman primates in absolute terms (Gabi et al., 2016). Therefore, it has been suggested that this evolutionary increase may underlie superior human cognitive abilities (Deacon, 1997; Herculano-Houzel, 2009; Gabi et al., 2016). However, anatomical studies disagree on cross-primate differences in the relative size of the prefrontal cortex, with some finding that it is disproportionately enlarged in humans compared with that in macaque monkeys (Passingham

Received Feb. 7, 2023; revised May 28, 2024; accepted May 31, 2024.

Author contributions: V.G., A.J.O.D., A.L.F., J.J.L., R.T.K., E.A.B., and X-J.W. designed research; A.J.O.D., M.J.J., I.S., M.R., W.C., B.R., and A.L.F. performed research; V.G., J.-W.K., Y.L., and A.J.O.D. analyzed data; V.G., J-W.K., Y.L., A.J.O.D., A.L.F., J.J.L., and X.-J.W. wrote the paper.

We thank S. Ahmad, N. Germanos, M.J.J., K. Morrisroe, C.I. O'Leary, and S. Schleufer for their roles in animal care and training. We also thank S.W. Linderman for generously sharing his code and advice on fitting HMM-GLM models to data and I.R. Stone, J. Ferre, and E.Y. Walker for their fruitful discussions. This work was supported by the National Institute of Health U-19 program Grant No. 5U19NS107609-03, R01 Grant No. R01MH062349, the Office of Naval Research Grant No. N00014-17-1-2041 (to X-J.W.), and the National Institutes of Health Office of Research Infrastructure Programs under Award Number P510D010425 (to E.A.B.).

*V.G. and J.-W.K. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Xiao-Jing Wang at xjwang@nyu.edu.

https://doi.org/10.1523/JNEUROSCI.0231-23.2024

and Smaers, 2014; Donahue et al., 2018) and others showing that it is not (Semendeferi et al., 2002; Barton and Venditti, 2013). This is further confounded by differences in primate prefrontal cell density (Semendeferi et al., 2011; Gabi et al., 2016). However, this debate can benefit from a clear characterization of interprimate differences in prefrontal-dependent cognitive functions. For these reasons, an interpretation of findings from animals demands rigorous comparisons between cognitive computations in humans and nonhuman animals.

Toward this end, we compared the behavioral strategies of macaque monkey and humans on the same task: a rule-switching task inspired by the Wisconsin Card Sorting Test (WCST) which is widely used to evaluate the cognitive functions involved in abstract thinking, rule search, cognitive set shifting, and the effective use of feedback (Grant and Berg, 1948; Kopp et al., 2021). The WCST has long been employed in the study of prefrontal function and dysfunction (Milner, 1963; Passingham, 1972; Drewe, 1974; Nelson, 1976; Gold et al., 1997), lending support to the presence of abstract thinking and computation in the monkey brain (Mansouri et al., 2020). In this task, subjects must match a test object to reference objects, each composed of multiple visual features, based on a hidden rule. Feedback indicates whether the match was correct, but does not unambiguously reveal the rule identity. The rule identity must be inferred from the collective outcome of multiple trials. Additionally, the rule undergoes uncued changes across trial blocks, requiring detection of and adaptation to such changes based solely on positive or negative feedback.

In our version of the task, subjects must select an object and receive feedback contingent on whether it contains a specific feature (the hidden rule). Each object has three feature dimensions (color, pattern, and shape), with one of four feature values per dimension, defining 12 possible rules (Fig. 1*a*). Simply learning the value of each of the 64 individual object–reward associations is inefficient; each object is a conjunction of multiple features, and learning the value of one object does not generalize to the others. Instead, identifying object features and learning their value facilitate generalization and consequently are more efficient. The problem of attributing binary feedback when there are many features (high-dimensional environments), referred to as the "curse-of-dimensionality," is effectively resolved through such abstract reasoning (Gershman and Niv, 2010; Wilson and Niv, 2012; Niv et al., 2015). Yet, tracking and updating the value of the 12 features impose prohibitive working-memory demands and is computationally daunting. Conversely, selectively attending to and evaluating one feature at a time are inefficient as they discard relevant feedback regarding other features. The strategy that individuals use to address this trade-off between computational complexity and information efficiency remains to be elucidated.

We found that monkeys were 3–4 times slower than humans at identifying new rules. To understand how the two species handle the task's complexity and to explain this performance difference, we fit a hypothesis-free behavioral model that predicts upcoming choices based on choices and their outcomes on previous trials. The best-fit models developed hidden states that aligned with a feature-based attention strategy wherein some visual features are selectively examined over others while making a choice and attributing feedback. The decision process recovered by the model revealed each species' rule-learning strategy. While similar to the win–stay, lose–shift (WSLS) strategy, these strategies deviated from it in important ways. First, both species explore more than one feature at a time. Second, they perform inference-like computations—their attentional state toward one feature can change based on the outcome of choosing another. We further identified distinct stages of rule learning, which revealed three key reasons for the lower performance in monkeys: (1) following a rule switch, monkeys perseverate on the previous rule more than humans; (2) monkeys occasionally make random choices that do not involve any of the features under exploration, even after finding the rule, which delays expression of the learnt rule; and (3) poorer attention to negative feedback in monkeys particularly when they simultaneously explore the rule and nonrule features poses a credit assignment challenge which delays learning.

Materials and Methods

Task description

Human and monkey subjects were tested on a rule-switching task whose design was inspired by the WCST. On each trial, they were simultaneously presented with an array of four objects on a computer screen. Each object was comprised of a stimulus feature from each of three stimulus dimensions: color, pattern, and shape (Fig. 1*a*), e.g., a blue polka-dotted triangle. For each trial, these 4 objects were chosen from a pool of 64 unique objects, each containing a possible combination of individual features from each of the three dimensions, such that there was no feature overlap between them. Accordingly, for each possible array, all four features of each dimension appeared on the screen, but the combination of features represented by each individual object varied across trials. Within a single rule-learning block of trials, one color, pattern, or shape was designated as the target, resulting in 12 possible rules. The identity of this rule was not cued, but had to be learned by trial and error, based on the feedback received at the end of each trial. Upon meeting a rule-learning criterion for the current rule, the rule feature changed on the next trial in an uncued manner, initiating a new rule-learning block. This rule shift could be either intradimensional, where the dimension of the new rule feature matched that of the previous rule feature (e.g., changing from triangle to square), or extradimensional, where the dimensions of the old and new rule features did not match (e.g., changing from triangle to yellow).

Experimental design

Monkeys. All procedures were carried out in accordance with the National Institutes of Health guidelines and were approved by the University of Washington Institutional Animal Care and Use Committee. Subjects were four adult female rhesus monkeys (*Macaca mulatta*) with mean age of 12.5 ± 2.5 years and mean weight of 7.5 ± 0.6 kg at the start of the experiment. All subjects were experimentally naive when acquired. Prior to training on the rule-switching task, three subjects had been trained on two working-memory tasks (delayed match-to-sample task and oculomotor delayed response task), and the fourth subject had been trained on only one of these tasks (delayed match-to-sample task).

The structure of the task was as follows. A monkey initiated each trial by fixating on a white cross (0.5°) at the center of a computer screen. Following 500 ms of successful fixation, the cross disappeared and was replaced by an array of four objects. During the self-paced decision epoch that followed, the monkey was free to explore the array of objects; her choice was signaled by maintaining her gaze within a 9 × 9° window centered on the object for 800 ms. The monkey received a food slurry reward over a 1.4 s duration for selecting the object that contained the rule feature. A timeout period (either 1 or 5 s) occurred on trials when the monkey did not choose the object containing the rule feature or when she did not make a choice within 4 s. The timeout period was reduced in two monkeys to increase the number of trials they completed per session, after it was confirmed that this reduction did not alter their trial-to-criterion performance. The feedback period was immediately followed by a 400 ms or 1 s intertrial interval (ITI). The larger ITI was used for one of the four monkeys to permit an examination of neural activity for the development and maintenance of neural representations of the rule during this interval. We classified a rule as learned either when

the monkey made 8 consecutive correct responses or when she made 16 correct responses in 20 trials or fewer.

Each monkey was trained on this task in four steps. Monkeys were advanced from one training step to the next after consistently being able to acquire a new rule (meeting criteria for consecutive correct responses) within 3 min or after consistently acquiring a certain number of rule shifts across an entire session (typically at least 20 shifts in a 90 min session). Once monkeys completed the final step, they were prepared for testing. The steps were as follows:

Step 1: Monkeys received training on the eye-tracker calibration task, which entailed fixating on a small (0.3°) square in various locations on the computer screen and releasing a touch-sensitive bar in response to a change in the square's color for reward. Next, they were trained to fixate on an image equivalent to those used in the final version of the task, for reward. Monkeys progressed through this step in one or two sessions.

Step 2: Monkeys were introduced to all four stimuli, where they were rewarded for fixating on the stimulus with the correct feature (e.g., red). In a given session, the dimension of the correct feature was constant (i.e., all rule shifts were intradimensional), and the four stimuli varied in only this target dimension (e.g., all targets were the same shape and pattern and varied only in their color). The dimension of the target features was varied across sessions (e.g., the target features were colors on one session and patterns on the next). Monkeys spent ∼20–25 sessions on this step.

Step 3: Monkeys were introduced to extradimensional shifts. Target features varied across rule blocks within a session by only two dimensions (e.g., in color and shape). In keeping with the previous step, the four stimuli did not vary along the nontarget dimension. The two target dimensions changed from one session to the next (e.g., the target dimensions were color and pattern in one session and pattern and shape in the next). Monkeys typically spent ∼25 sessions on this step.

Step 4: The target feature and stimuli varied in all three dimensions—this was the final version of the task.

The monkeys learned and solidified their rule-learning and set-shifting strategies for solving the task over the course of these training steps. Since each training phase involved a distinct, relatively notable advancement in complexity over the previous training step (e.g., the requirement to identify the correct rule among several possible rules or the introduction of extradimensional shifts), we observed progressive learning over the course of these training phases—monkey performance generally improved between the beginning and end of each phase (Extended Data Fig. 1-2). However, we did not see such a learning effect during the transition to the final version of the task (Step 3–Step 4), which we believe is due to the relatively smaller difference in complexity between these steps compared with earlier steps. The monkeys had already been introduced to extradimensional shifts between two dimensions per session in Step 3 and were easily able to generalize this to three extradimensional shifts in a single session when they were transitioned to Step 4.

Prior to testing, a titanium post was surgically affixed to each monkey to hold its head for eye-gaze tracking and a separate electrophysiological investigation. During testing, each monkey was head-fixed in a dimly illuminated room and positioned 60 cm away from a 19 in CRT monitor with a screen refresh rate of 120 Hz noninterlaced. The monitor had a resolution of 800 × 600 pixels, subtending 33° by 25° of the visual angle. Eye movements were recorded using a noninvasive infrared eye-tracking system (EyeLink 1000 Plus, SR Research). Stimuli were presented using an experimental control software (NIMH Cortex or NIMH MonkeyLogic). Calibration of the infrared eye-tracking system was accomplished using a nine-point manual calibration task. Following the calibration task, the monkey was tested on the rule-switching task.

The majority of caloric intake during testing was provided in the form of the food slurry reward. Monkeys were supplemented after testing each day with fruits and vegetables as well as monkey chow. Daily pretesting weights were taken to monitor weight, and the caloric intake was adjusted to maintain a vet-approved weight range based on sex and age. All animals had *ad libitum* access to drinking water.

The type of rule shift (intradimensional or extradimensional) was determined pseudorandomly and occurred with equal probability. Earlier studies in monkeys and humans performing the WCST suggest differential contributions of brain regions to rule learning following intradimensional versus extradimensional shifts (Rogers et al., 2000; Watson et al., 2006). The probability of these two conditions was set to be equal to facilitate a balanced comparison between them during electrophysiological investigation. A block consisted of all the trials from the initial rule shift to the final trial of criterion performance. We analyzed a total of 1,305 blocks in 81 recording sessions from Monkey B, 872 blocks in 29 recording sessions from Monkey C, 805 blocks in 29 recording sessions from Monkey S, and 224 blocks in 13 recording sessions from Monkey T.

Cross-sessional (Extended Data Fig. 1-3) and intersession (Extended Data Fig. 1-4) performance comparisons show no discernible trend during behavioral recordings at the testing stage, indicating that monkey rule-learning performance had plateaued. Therefore, an interspecies comparison based on data collected at this stage is fair, and we used every session of usable data recorded during testing in our analysis. We excluded trials when monkeys did not make a decision within the required 4 s duration (Extended Data Table 1-1) and trials from incomplete rule blocks (e.g., at the end of a session).

Humans. The studies involving human participants were reviewed and approved by the Institutional Review Board of University of California. All participants provided their written informed consent to participate in this study and received a small monetary compensation not linked to performance. We collected two datasets that differed in the task parameters. Subjects in Dataset 1 were four adult males and one adult female with mean age of 26.4 ± 4.1 years. Subjects in Dataset 2 were two adult males and three adult females with mean age of 20.2 ± 1.6 years. Subjects were brought into a room where they sat and completed a computer-adapted version of the rule-switching task on a recording laptop. Subjects in Dataset 1 received the following instructions: "In this experiment, you will see 4 cards on each trial. Each card has 3 unique features (color, pattern and shape). No feature is shown on more than one card, so you will see 12 different features on each trial (4 colors, 4 patterns, 4 shapes). The card containing the correct feature (1 out of 12 possible) will be correct choice. The correct feature might change during the task. The answer is given by pressing one of the four arrow keys that corresponds with the selected card position on the screen (up, down, left or right). You have 4 s to provide the answer, or the trial times out. The task goes on for 200 trials or about 15 min." Subjects in Dataset 2 received a more limited set of instructions, so that they would learn the task and trial structures from experience similar to the monkeys: "In this experiment, you will use one of the four arrow keys on each trial as a response. The 'correct' or 'incorrect' feedback will be provided following each choice. Your task is to maximize the number of correct responses."

Individual trials consisted of the following epochs: cross fixation (black cross displayed in the center of the screen on a gray background for 300 ms in Dataset 1 and 500 ms in Dataset 2); choice (four objects displayed on the screen at locations corresponding to up, down, left, or right positions for up to 4,000 ms); feedback (feedback message "correct" or "incorrect" displayed for 1,500 ms for subjects in Dataset 1 and feedback message "correct" displayed for 1,500 ms or "incorrect" displayed for 5,000 ms for subjects in Dataset 2); and ITI (gray screen for 1,000 ms). Subjects indicated their choice by pressing the arrow key on the laptop keyboard, corresponding to the chosen object's position on the screen. If the choice was not indicated within the 4,000 ms, the trial was considered timed-out. The learning criterion was defined as 5 consecutive correct trials or 8 correct out of the last 10 trials for Dataset 1. Dataset 2 paralleled the monkey criterion of 8 consecutive correct trials or 16 correct out of the last 20 trials. After reaching a learning criterion, the rule switched and a new rule-learning block began. The new rule was randomly determined, with the probability of intradimensional versus extradimensional rule shifts set to be consistent with the monkey experiments. Each participant completed five task sessions (300 trials/sessions for a total of 1,500 trials in Dataset 1 and 200 trials/session for a total of 1,000 trials in Dataset 2). This spanned between 107 and 138 blocks across the five subjects in Dataset 1 and between 98 and 110 blocks across the five subjects in Dataset 2.

WSLS agent. We simulated behavior using a WSLS strategy. The task structure (rule selection, learning criterion) for the WSLS agent was identical to that of the humans in Dataset 1, except for the trial structure—the agent's algorithm determined its choice immediately upon stimulus presentation. The WSLS algorithm (Extended Data Fig. 2-2*d*, left) holds a single feature as its target (the "persist" state) and deterministically chooses the object with that feature at each trial. All other features are in the "avoid" state. Positive feedback maintains the current feature in the persist state. Negative feedback demotes it to the avoid state and promotes a randomly selected feature from among the 11 others that were in the avoid state to the persist state. The agent completed 500 rule blocks.

Input–output hidden Markov model–generalized linear model (IOHMM-GLM) for the prediction of feature choices

Model design. The four objects presented during a trial consist of 12 visual features, $f \in \{1, \ldots, 12\}$. Assuming a feature-based mental representation, the model predicts the choice of each feature f at the next trial t. This choice is represented by $c_t^f \in \{0, 1\}$, where $c_t^f = 1$ indicates the f was part of the chosen object and $c_t^f = 0$ indicates it was not. Either choice can result in a reward or timeout for the trial, given by $r_t \in \{0, 1\}$. The choice–outcome history of f given the past l trials is denoted $h^f \in \{1, \ldots, 2^{2\ell}\}$. We refer to l as the lag, and it is a hyperparameter of the model. The value of h^f at trial t is given by the binary vector $(r_{t-1}, c_{t-1}, \ldots, r_{t-\ell}, c_{t-\ell})$ of size $2l$. Therefore, it can take on 2^{2l} possible values. In all our analyses, we choose a lag 1 ($l = 1$) model for further analysis. Such a model depends on a choice–outcome history that takes on one of four possible values at trial t, $(r_{t-1} = 0, c_{t-1} = 0)$, $(r_{t-1} = 1, c_{t-1} = 0)$, $(r_{t-1} = 0, c_{t-1} = 1)$ or $(r_{t-1} = 1, c_{t-1} = 1)$, which we refer to as NC^-, NC^+, C^-, and C^+, respectively.

The transformation of the choice–outcome history into a choice at trial t is mediated by discrete hidden states $s^f \in \{1, \ldots, K\}$ that determine the parameters of the transformation. The maximum number of states K is a second model hyperparameter. The transformation is modeled as a Bernoulli GLM as follows:

$$p(c_t = 1 | s_t = k, h_t) = \frac{1}{1 + \exp(-w_k^T h_t)}, \quad (1)$$

where the parameters $w_k \in \mathbb{R}^{1 \times 2^{2\ell}}$ are determined by the state $s_t = k$. We denote the set of parameters across all K states as $w \in \mathbb{R}^{K \times 2^{2\ell}}$.

Transitions between states also depend on the choice–outcome history and are modeled by multinomial logistic regression as follows:

$$p(s_{t+1} = k | s_t = j, h_{t+1}) = \frac{\exp(\log(P_{jk}) + u_{jk}^T h_{t+1})}{\sum_{k'=1}^{K} \exp(\log(P_{jk'}) + u_{jk'}^T h_{t+1})}, \quad (2)$$

where the parameters $P \in \mathbb{R}_+^{K \times K}$ and $u \in \mathbb{R}^{K \times K \times 2^{2\ell}}$ represent the bias or baseline transition probability and history weights. This model design is schematized in Figure 2*a*.

Finally, the probability distribution of initial states π is a model parameter that specifies the state at the first trial of a session.

Model fitting. We fit the parameter values for the choice GLM weights w, the baseline transition probability P, the transition GLM weights u, and the initial state distribution π to the choices of each subject. To avoid overfitting, the parameter values were shared across all features. In other words, all parameter values were the same for all 12 features. The likelihood of the data under a model is its probability subject to the model's parameters and inputs $p(c_{1..T} | w, P, u, \pi, h_{1..T})$, where T is the number of trials in the session. It is expressed in terms of these parameters as follows:

$$\begin{aligned} p(c_{1..T} | w, P, u, \pi, h_{1..T}) &= \sum_{s_{1..T}} p(c_{1..T}, s_{1..T} | w, P, u, \pi, h_{1..T}) \\ &= \sum_{s_{1..T}} p(s_1 | \pi) \left[\prod_{t=2}^{T} p(s_t | P, u, h_t) \right] \left[\prod_{t=1}^{T} p(c_t | w, s_t, h_t) \right], \end{aligned}$$

where the last two terms are given by Equations 1 and 2, respectively.

The model parameters were fit by minimizing $-log[\, p(c_{1..T} | w, P, u, \pi, h_{1..T})]$, i.e., the negative log-likelihood of the data, via gradient descent with the ADAM optimizer. The choice GLM weights for each of the k states were initialized to a single $2^{2\ell}$-dimensional vector drawn from a standard normal distribution. The baseline transition probability was initialized to the sum of a diagonal matrix with value $0.9I$ where I is the identity matrix and a random matrix with elements drawn from a uniform distribution in the interval (0, 0.05). The larger diagonal values enforce "stickiness" that bias transitions back into a state. The transition GLM weights were initialized to zero, and the initial state distribution was initialized to $1/K$ for each state k. For each subject and each pair of hyperparameters (l, K), the parameters were optimized over 10,000 iterations with fivefold cross-validation (Fig. 2*b*).

The best-fit model was sought for each subject and hyperparameter setting across 10 independent parameter initializations for the humans and across five initializations for the monkeys. Figure 2*b* shows the mean negative log-likelihood taken over all initializations and cross-validation folds. The best-fit model for each subject was selected for further analysis from the resulting 50 models for each human and 25 models for each monkey at hyperparameter values $l = 1$ and $K = 4$. We found that a majority of these models produced very similar choice and transition probabilities. However, fits to the WSLS agent varied much more. Since negative feedback immediately demoted features from the persist to avoid state, exploration of nonrule features typically lasted 1–2 trials. This likely makes it harder for the model fitting procedure to identify exploration and introduces more variability across fits.

Once the best-fit model is identified, the most-likely sequence of states, s^*, for each subject, session, and feature determined by the Viterbi algorithm (Viterbi, 1967; Fig. 2*d*). For each trial t and feature f, the algorithm performs a forward pass across all past trials and a backward pass across all future trials to determine the most-likely state of f at trial t that best explains past, present, and future history-dependent choices under the constraints of the model's parameters and the choice and transition probabilities they yield. Extended Data Figure 2-1*b* shows the cumulative distribution of the posterior probabilities ($p(s_t = s_t^* | c_{1..T}, h_{1..T})$) of these state estimates calculated for the Viterbi algorithm.

All model fits and the most-likely state determination were performed with the state space model Python package (Linderman et al., 2020).

Model extension for the prediction of object choices

We extended the feature choice prediction model described in the previous section to predict object choices at each trial t. Given the predicted choice probability ($p(c_t^{f_{ij}} | w, P, u, \pi, h_{1..t}^{f_{ij}})$) for each feature $f_{i,j}$, $i \in \{1, \ldots, 3\}$ in an object o_j, $j \in \{1, \ldots, 4\}$ presented at trial t, the model predicts which object is chosen at t. This transformation of predicted feature choice probabilities $p(f_{i,j})$ into object choice probabilities $p(o_j | p(f))$ is modeled by multinomial logistic regression as follows:

$$p(o_j | p(f)) = \frac{\exp\left[\sum_{i=1}^{3} v_{ij} \log(\, p(f_{i,j})) + b_j\right]}{\sum_{j'=1}^{4} \exp\left[\sum_{i=1}^{3} v_{ij'} \log(\, p(f_{i,j'})) + b_{j'}\right]}, \quad (3)$$

where the parameters $v \in \mathbb{R}^{3 \times 4}$ and $b \in \mathbb{R}^{1 \times 4}$ represent the feature choice probability weights and biases in selecting each object, respectively. These values were fit to the choices of each subject by minimizing the cross-entropy loss $-\sum_{t=1}^{T} \sum_{j=1}^{4} y_{j,t} \log(\, p(o_{j,t} | p(f)_t))$ where $y_{j,t} \in \{0, 1\}$ indicates whether object $o_{j,t}$ was chosen on trial t. Model fitting was performed via stochastic gradient descent with the ADAM optimizer implemented by the PyTorch Python package (Paszke et al., 2019). The parameter values for v and b were initialized from a uniform distribution in the interval $\left[-\frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}\right]$ and optimized until convergence with a maximum of 100,000 iterations. Cross-validation was performed with the same training and test sets used while training the feature choice prediction models (Extended Data Fig. 2-1*a*).

The accuracy of the object choice prediction model based on the best-fit feature choice prediction model with four states and lag 1 is shown in Figure 2*c*, left. We also fit a model to determine the chosen object in a similar fashion using the feature choice probabilities based

on their most-likely state estimates ($p(c_t^{f_{ij}}|s_t^{f_{ij}} = s_t^{*f_{ij}}, h_t^{f_{ij}})$) instead. The accuracy of this model is shown in Figure 2*c*, right.

Model analysis

The probability distribution of histories in each state (Extended Data Fig. 3-1*b*) is as follows:

$$p(h = i|s^* = j) = \frac{\sum_{f,t} \mathbb{1}(h_t^f = i, s_t^{*f} = j)}{\sum_{f,t} \mathbb{1}(s_t^{*f} = j)}, \quad (4)$$

where $\mathbb{1}$ is the indicator function and $\sum_{f,t}$ is a sum over features and trials. The state and history-dependent choice probability (Extended Data Fig. 3-1*a*) can be directly calculated from the model's parameters (Eq. 1) or empirically as follows:

$$p(c = 1|s^* = j, h = i) = \frac{\sum_{f,t} \mathbb{1}(c_t^f = 1, s_t^{*f} = j, h_t^f = i)}{\sum_{f,t} \mathbb{1}(s_t^{*f} = j, h_t^f = i)}. \quad (5)$$

The choice probability of a feature in each state (Fig. 3*a*) can be computed by utilizing Equations 4 and 5 or Equation 1 as follows:

$$p(c = 1|s^* = j) = \sum_{i\in\{1,\ldots,4\}} p(c = 1|s^* = j, h = i) \cdot p(h = i|s^* = j). \quad (6)$$

Similarly, the state transition probabilities (Extended Data Fig. 3-2) can be directly calculated from the model's parameters (Eq. 2) or empirically as follows:

$$p(s_{t+1}^* = k|s_t^* = j, h_{t+1} = i) = \frac{\sum_{f,t} \mathbb{1}(s_{t+1}^{*f} = k, s_t^{*f} = j, h_{t+1}^f = i)}{\sum_{f,t} \mathbb{1}(s_t^{*f} = j, h_{t+1}^f = i)}. \quad (7)$$

We approximated the decision process in each species (Extended Data Fig. 2-2*b*) from the state transition probability and the "reverse" state transition probability ($p(s_t^* = j|s_{t+1}^* = k, h_{t+1} = i)$). The latter helps in conditions where transitions into a state are typically rare. This quantity (Extended Data Fig. 3-3) is calculated empirically as follows:

$$p(s_t^* = j|s_{t+1}^* = k, h_{t+1} = i) = \frac{\sum_{f,t} \mathbb{1}(s_t^{*f} = j, s_{t+1}^{*f} = k, h_{t+1}^f = i)}{\sum_{f,t} \mathbb{1}(s_{t+1}^{*f} = k, h_{t+1}^f = i)}. \quad (8)$$

Trial categorization

Trials were categorized based on the identity of the rule feature and the most-likely state estimates for all 12 features as in Figure 5*a*. Since each trial is always designated to belong to one and only one category, the trial categories are mutually exclusive and exhaustive. For each rule block, this allows us to determine the number of trials spent in each category (Fig. 5*c*). Moreover, since the categories are mutually exclusive, we can explain summary statistics (mean and variance) of the block length for each subject in terms of statistics of their category lengths (Extended Data Fig. 5-1*a*) as follows:

$$\mathbb{E}[\text{block length}] = \sum_{\text{category c}} \mathbb{E}[\text{no. trials in category c}]$$

$$\text{Var}[\text{block length}] = \sum_{\text{category c}} \text{cov}[\text{no. trials in category c, block length}].$$

Interspecies comparison of category lengths

In Figure 7, the higher probability of continued exploration of nonrule features by monkeys during the rule-favored exploration category is attributed to poor (direct and indirect) negative feedback sensitivity (Fig. 7*c*,*d*). In addition, we attribute the higher probability of continued exploration in monkeys during rule-favored exploration trials compared with nonrule exploration trials to a higher prevalence of direct positive feedback during rule-favored exploration trials (Fig. 7*f*).

These determinations were made based on the following decomposition:

$$\begin{aligned} &p(s_{t+1}^* \in \text{explore}|s_t^* \in \text{explore}, (t, t+1) \subseteq \text{category c}) \\ &= \sum_{i\in\{1,\ldots,4\}} p(s_{t+1}^* \in \text{explore}, h = i|s_t^* \in \text{explore}, (t, t+1) \subseteq \text{category c}) \\ &= \sum_{i\in\{1,\ldots,4\}} [\,p(s_{t+1}^* \in \text{explore}|h = i, s_t^* \in \text{explore}, (t, t+1) \subseteq \text{category c}) \\ &\qquad \times p(h = i|s_t^* \in \text{explore}, (t, t+1) \subseteq \text{category c})]. \end{aligned}$$

The joint probability above is shown in Figure 7*f*, left, and in Extended Data Figure 7-1, and quantities resulting from its decomposition below are shown in Figure 7*f*, middle-right.

Statistical analyses

Between-species statistical comparisons of learning performance, inferred state occupancy, and state transition statistics were carried out using bootstrap tests with a t statistic.

Code accessibility. All training and analysis codes will be available at publication on GitHub (https://github.com/xjwanglab).

Results

Monkeys are slower rule learners than humans

We compared the ability of monkeys and humans to rapidly adapt to changes in task contingencies in a rule-switching task inspired by the WCST. On each trial, subjects were presented with four objects and received feedback upon selecting one of them (Fig. 1*a*, middle). Each object was composed of one unique feature from each of three dimensions—visual pattern, shape, and color (Fig. 1*a*, top). Each of the 12 possible features appeared in one of the objects on each trial, but object compositions changed across trials. On a given block of trials, subjects received positive feedback (monkeys, food reward; humans, the word "CORRECT" displayed on the screen) for selecting the object which contained the feature defined by the current hidden rule (e.g., red) and negative feedback (monkeys, timeout; humans, the word "INCORRECT" displayed on the screen) otherwise. After subjects demonstrated that they had learned the current rule by reaching criterion performance on the current block, a new block was initiated through an uncued switch to a randomly chosen new rule (Fig. 1*a*, bottom). Similarities and differences in task parameters and learning criteria across the monkey and two human datasets are outlined in Extended Data Tables 1-1 and 1-2. The second human dataset was collected to match the task parameters in the monkey dataset, enabling a better comparison between species (see Materials and Methods).

Remarkably, well-trained monkeys learned new rules within only tens of trials. Yet, they were over three times slower than humans (Fig. 1*b*; monkeys, 27.84 ± 2.92 trials (mean ± SEM); Human Dataset 1, 5.98 ± 0.52 trials; Human Dataset 2, 6.47 ± 0.19 trials). This learning deficit in monkeys was significant in comparison with that in subjects in Human Dataset 2 (bootstrap test with t statistic, $p < 0.01$), as well as in comparison with subjects in Human Dataset 1 following a correction for the difference in the learning criterion between the two datasets (Extended Data Fig. 1-1; monkeys, 20.61 ± 1.52 trials; Human Dataset 1, 5.98 ± 0.52 trials; bootstrap test with t statistic; $p < 0.005$). We then sought to explain the interspecies computational differences that produce this rule learning slowing in monkeys. Specifically, we focused on inferring individuals' rule-learning strategies from behavior and identifying the species differences that contribute to the learning speed difference. One possible strategy, WSLS, is widely reported during rule learning in

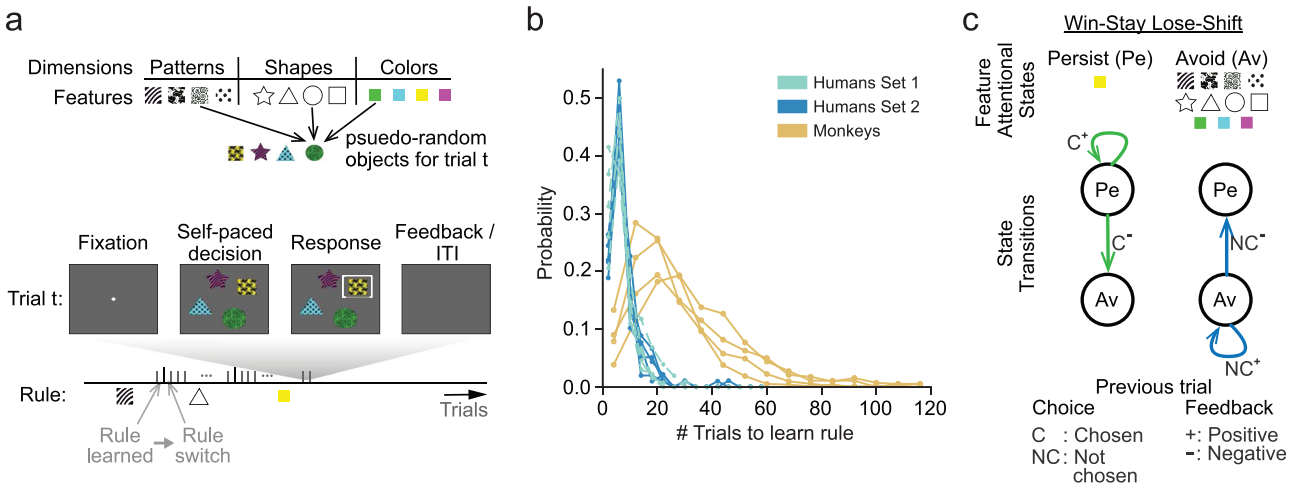

Figure 1. Monkeys rapidly learn rules in a rule-switching task but are slower than humans. ***a***, Rule-switching task structure. Each trial is composed of fixation, decision, response, feedback, and ITI epochs. After fixation, the subject is presented with four objects that are pseudorandomly composed of three features—a pattern, shape, and color. The features composing each object are mutually exclusive with respect to other objects. Each block of continuous trials is governed by a rule (1 of the 12 features). The subject receives positive feedback only for choosing the object with that feature. The identity of the rule is hidden and must be discovered. An uncued rule switch to a random new feature occurs when the subject demonstrates they have learned the current rule. ***b***, Distribution of trials-to-learning criteria in four monkey subjects (brown) and 10 human subjects (5 subjects, Dataset 1, green; 5 subjects, Dataset 2, blue). All subjects rapidly learn the rule, but on average, monkeys are over four times slower than humans. ***c***, Decision process for the WSLS learning strategy in two-armed bandit problems. The decision to choose an arm can be in one of two states: persist when it is chosen and avoid when it is not. The decision to choose an arm stays in the persist state as long as positive feedback is received (win–stay) and switches to the avoid state otherwise. It then stays in the avoid state as long as positive feedback is received and switches to the persist state when negative feedback is received (lose–shift). See Extended Data Figure 1-1 for more details.

many species, particularly in the two-armed bandit problem where the identity of the more rewarding arm must be learned and can change over trials. Here, one of the arms is repeatedly chosen as long as this produces positive feedback (win–stay). When negative feedback is received, the other arm is chosen on the next trial (lose–shift). This strategy can be cast as a decision process comprised of two behavioral states—persist and avoid—where the choice of the currently rewarded arm is in the persist state and transitions to the persist or avoid states subject to positive or negative feedback, respectively, while the choice of the other arm is in the avoid state and transitions to the persist state when that arm was not chosen on the previous trial and negative feedback was received (Fig. 1*c*).

The WSLS strategy is computationally efficient and requires that the subject attend to and maintain only a single arm's identity in working memory. By replacing arm identity with feature identity, the approach is readily adapted to solve problems in our rule-switching task and always finds the rule. However, feedback is equally informative about all three features in the chosen object, not just the attended one. Due to this neglect of information about unattended features, a simulated WSLS agent learns rules much slower than optimal: indeed humans learn more rapidly (Extended Data Fig. 1-1; WSLS agent mean, 13.31 trials; SD, 12.85 trials; Human Dataset 1, 5.98 ± 0.52 trials). This underscores a trade-off between computational and information efficiency in multidimensional environments. Simultaneously maintaining and updating beliefs about multiple features is more information efficient but increases computational complexity and working-memory demands. In contrast, attending to a single feature at a time is computationally simpler but inefficient in its integration of trial outcomes. In the following sections, we address how the two species solve this trade-off.

Dynamic model uncovers hidden states during rule learning

Prior cognitive model comparisons of human behavior in rule-switching tasks provide evidence for rule-learning strategies wherein subjects selectively attend to and learn about individual features or dimensions, rather than feature configurations (i.e., objects; Bishara et al., 2010; Wilson and Niv, 2012; Niv et al., 2015). It is argued that such a mental representation of stimuli in terms of features resolves the curse-of-dimensionality which impairs learning efficiency in high-dimensional environments. For example, it is more efficient to learn the value of 12 features than the dozens of objects they can be combined into. Drawing on these findings, we developed a behavioral model to predict the probability of a subject choosing individual features given their choices and outcomes on previous trials. However, in contrast to earlier work, our model does not postulate a specific internal belief structure and update rule, thus making fewer assumptions regarding the learning algorithm underlying a subject's behavior. Instead, it aims to discover in an unbiased manner how the decision-making process evolves as a function of feedback. Recently, this approach has been successful at revealing previously unobserved behavioral states underlying human, monkey, rodent, and fruit fly decision-making (Ebitz et al., 2018; Calhoun et al., 2019; Roy et al., 2021; Bolkan et al., 2022).

For each feature, we model whether the feature is chosen or not (denoted as c) as a function of past choices and their outcomes (h) via a Bernoulli GLM (Fig. 2*a*; see Materials and Methods). The choice outcome on an earlier trial is represented by a four-dimensional binary vector where the dimensions represent whether positive feedback was received after choosing the feature on the trial (C^+), negative feedback was received after choosing the feature (C^-), positive feedback was received after not choosing the feature on the trial (NC^+), or negative feedback was received after not choosing the feature (NC^-). This allows us to assess separately how the present choice depends on past choice outcomes both when that feature was chosen (direct feedback) and when it was not (indirect feedback). Furthermore, the model permits dynamic changes in how past choices and outcomes are transformed into a present choice via hidden states (s). A feature's associated hidden state also undergoes a transition

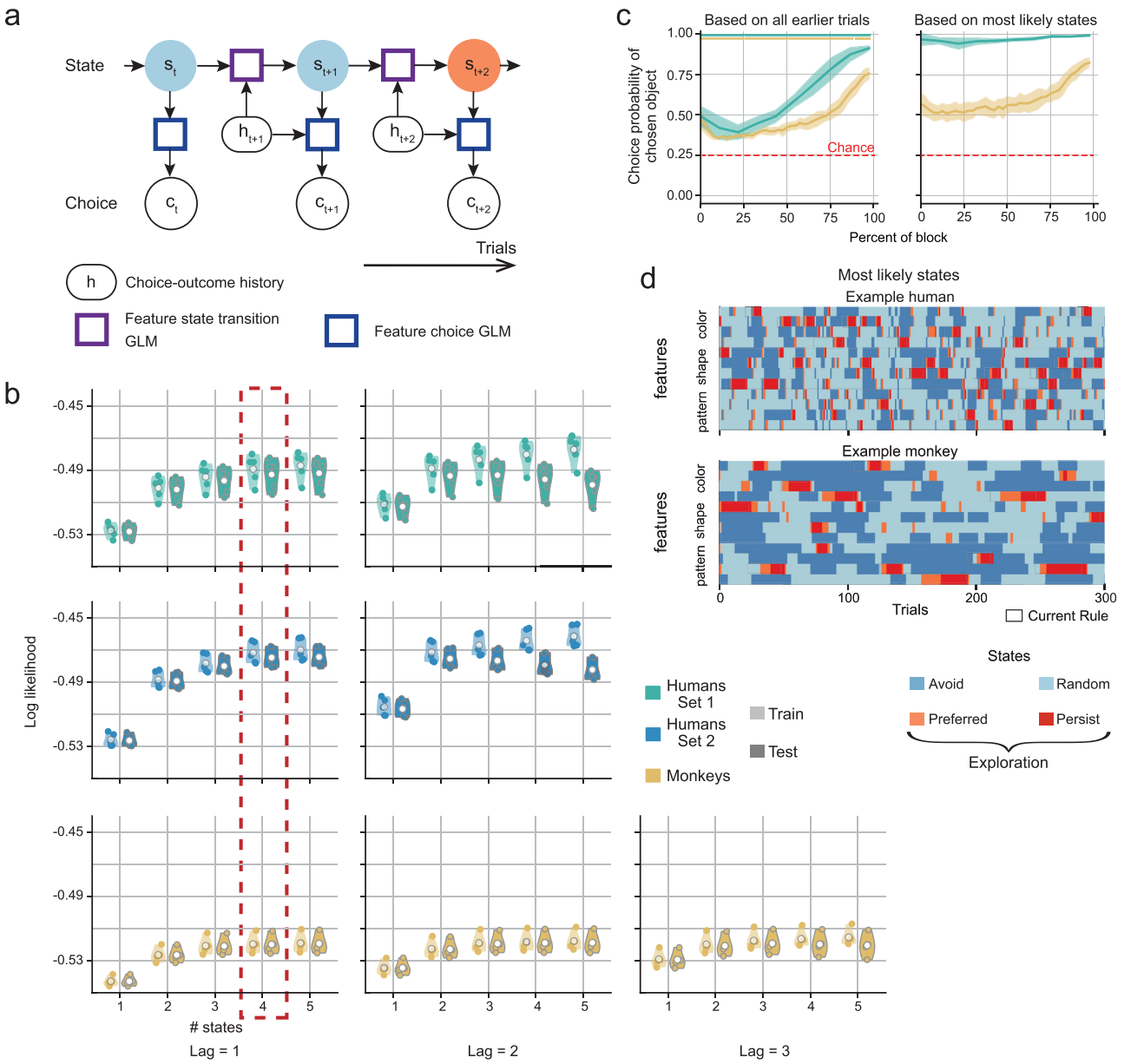

Figure 2. IOHMM-GLM model fits uncover dynamic changes in choice behavior during rule learning. ***a***, IOHMM-GLM model architecture fit to data. The model predicts the choice of a feature *c* at each trial *t* from the choice–outcome trial history *h* via a GLM. Hidden states *s* determine the GLM's parameters. These states can transition at each trial also based on the choice–outcome trial history via a separate state transition GLM. ***b***, Model fit log-likelihoods on training and test datasets for each human (Dataset 1, green; Dataset 2, blue) and monkey (brown) subject in models with varying numbers of states and that use choice outcomes from varying numbers of previous trials (lag) to determine the feature choice and state transition probabilities. Each point represents a single subject's mean over a fivefold cross-validation and over 5 (monkey) or 10 (human) different model initializations. Each subject's best-fit model with four states and lag 1 (dashed red box) was chosen for further analysis. ***c***, Probability of selecting the chosen object produced by a model extension based on feature choice probabilities predicted only from choice outcomes on earlier trials (left) and on feature choice probabilities computed from most-likely state estimates derived from past, present, and future choice outcomes (right). The probability on each trial was binned according to the trial's relative position in the rule block and averaged across blocks. Line and shading represent the mean and SD across subjects for each species. Dots represent block percentiles at which the average object selection probability is significantly above chance (bootstrap test with *t* statistic, $p < 0.05$). ***d***, Most-likely states estimated by the model for 300 trials in an example human (top) and monkey (bottom) subject. The rule on each block is outlined in black. See Extended Data Figures 2-1 and 2-2 for more details.

at the end of each trial depending on past choices and outcomes, which may reflect updates to the feature's value based on past choice outcomes, or a change in the level of attention to the feature or even a shift in strategy (i.e., how a feature's history is weighted in determining its choice). Note that while the model permits these possibilities and others, it does not prescribe the nature and function of the states. Rather, the states and their dynamics emerge upon fitting the model to behavioral data. These hidden state dynamics are modeled as an input-dependent or IOHMM (Bengio and Frasconi, 1994).

The IOHMM-GLM's goodness of fit to behavior depends both on the number of previous trials determining a subject's choice (lag) and on the number of possible hidden states. Accordingly, we fit IOHMM-GLMs to each subject's behavior while systematically varying these two parameters (Fig. 2*b*). Across subjects in both species, model accuracy showed a stronger dependence on the number of states than on the lag. Crucially, accuracy plateaued as the number of states increased and exhibited overfitting at higher lags. In this task, subjects choose objects rather than individual features. Therefore, we extended our model to compute

the probability of choosing each object in a trial, based on the model's predicted probability of choosing individual features on that trial (see Materials and Methods). Fits of this model extension to each subject's behavior based on each of the IOHMM-GLMs in Figure 2*b* revealed a qualitatively similar relationship between model accuracy and the underlying parameters (Extended Data Fig. 2-1*a*). For each subject, the best-fit model comprised of four states and with lag 1 (history from the previous trial only) does not overfit the data while producing prediction accuracies at or very close to the performance plateau. Therefore, we selected these models (Fig. 2*b*, dashed red box; Extended Data Fig. 2-1*a*) for further analysis.

Figure 2*c* (left) shows the choice probability predicted by these four-state lag 1 models for the chosen object at each trial after a rule switch, averaged over rule blocks; averaging across blocks is achieved by normalizing the trial number by the block length. The results show that the model's prediction of the chosen object is significantly above chance (0.25) in both species (monkeys, 0.47 ± 0.02; Human Dataset 1, 0.63 ± 0.02). Also, prediction accuracy improves as the rule is learned over the block's time course. Our primary goal, however, is to find the most accurate explanation for each subject's rule-learning behavior rather than predict their future choices based on past choice outcomes. For this, we consider the most-likely sequence of states across trials inferred by the model for each feature. This is given by the maximum a posteriori probability (MAP) estimate of the sequence of states across all trials in an experimental session. In this formulation, each estimated or inferred state best explains not only the present choice but also past and future choices subject to the model's choice probabilities in the inferred past/future states and its state transition probabilities between the inferred present and past/future states (see Materials and Methods; Fig. 2*d*). The model is generally quite confident in its MAP estimates of the most-likely sequence of states, as measured by the cumulative density of their posterior probabilities (Extended Data Fig. 2-1*b*). Moreover, since the inferred states for each feature are estimated from past, present, and future choices, they yield more accurate estimates of the choice probabilities for chosen objects (Fig. 2*c*, right). For this reason, we rely on the inferred states to identify the rule-learning strategy in each species and to interpret the interspecies differences therein.

To gain insight into the interpretation of our model fits, we similarly analyzed the choices of a simulated WSLS agent (Extended Data Fig. 1-1). By construction, we know that the model's choices only rely on the previous trial. As expected, higher lag models tend to overfit the agent's choices (Extended Data Fig. 2-2*a*). While the agent's true behavior has only two states (Fig. 1*c*), we find that a three-state model provides a better fit. Our model splits the WSLS algorithm's "avoid" state into two states—a random state in which a feature is selected with chance probability and an avoid state in which a feature is selected below chance. This is due to a combination of the task's structure and our model setup. By picking one feature consistently across trials, the agent necessarily avoids other features in the same dimension. However, the agent's choices of features in other dimensions appear random to the model, since an object is composed of one feature from each dimension and objects compositions are generated randomly on each trial. Thus, the appearance of this additional state results from our model's treatment of each feature independent of its relationship to other intradimensional features, a simplifying assumption that allows for tractable fitting. Nevertheless, the model largely recovers the hidden states and state transitions that drive the agent's behavior—it correctly identifies when a feature is associated with the persist state 57.6% of the time and accurately determines the underlying decision process (Extended Data Fig. 2-2*c*,*d*). Collectively, these results show the reliability of this modeling approach to explain rule learning in both species.

Hidden states reflect feature-based attention and reveal qualitatively similar strategies in the two species

Learning is often conceptualized as updates to a decision-making schema based on past decisions and their outcomes (Behrens et al., 2007; Niv et al., 2015). We sought to identify hidden states that capture this decision-making process and to explore what they reveal about the dynamics of human and monkey rule learning in our task. We compute the choice probability of features associated with each state by marginalizing the model's predicted choice probability under each state and history (Extended Data Fig. 3-1*a*) over the choice–outcome histories (Extended Data Fig. 3-1*b*). In both humans and monkeys, a comparison of these choice probabilities revealed that the model determines states based on distinct probabilities of choosing the associated feature, ranging from below chance (avoid) to chance (random) to above chance (preferred) to very high (persist; Fig. 3*a*). That is, the model states correspond to levels of attention paid to each feature. Moreover, this result was consistently observed in the majority of the models fit to the behavior in both species, as well as in a simulated WSLS agent (Extended Data Fig. 2-2). Since features associated with the preferred or persist state are favored during rule learning, we refer to them as being under exploration. We will show that the estimation of the attentional state toward each feature at each trial permits a systematic analysis of when features are selected for or withdrawn from exploration and how the choice–outcome history informs these decisions. This exercise fosters an exposition of the decision-making process that describes the rule-learning strategy in both species, the resulting learning dynamics between rule switch and rule learning, and the identification of the differences in the decision-making process that most prominently explain the learning performance difference between the two species.

Since these analyses rely heavily on the most-likely state estimates, we validated the consistency of these estimates with the model's parameters. First, we compared the feature choice probability per history and state computed directly from the fit parameters (model) and measured based on the state estimate for each feature on each trial (empirical). The two measurements yield consistent results, demonstrating that the estimated most-likely states not only best explain the sequence of choices but also conform with the model's parameters. Next, we similarly compared state transition probabilities per history computed directly from the fit parameters (model) and measured based on the state estimate for each feature on each trial (empirical; Extended Data Fig. 3-2). Here again, we find that the transition probabilities computed from the fit parameters (model) are consistent with empirical measurements of the transition statistics based on the most-likely state estimates. Extended Data Figure 2-2*b* schematizes the decision process in the two species derived from their state transition probabilities. The thickness of an arrow indicates the probability of the respective transition; extremely rare transitions have been pruned. Similar to the WSLS agent, we find that a feature is most often associated with the avoid state, while an intradimensional feature is simultaneously under exploration (Extended Data Fig. 3-1*c*). Since the avoid state likely emerges due to this interdependence between the choices of interdimensional features, which our model forgoes for tractability, we do not treat it as distinct from the random state.

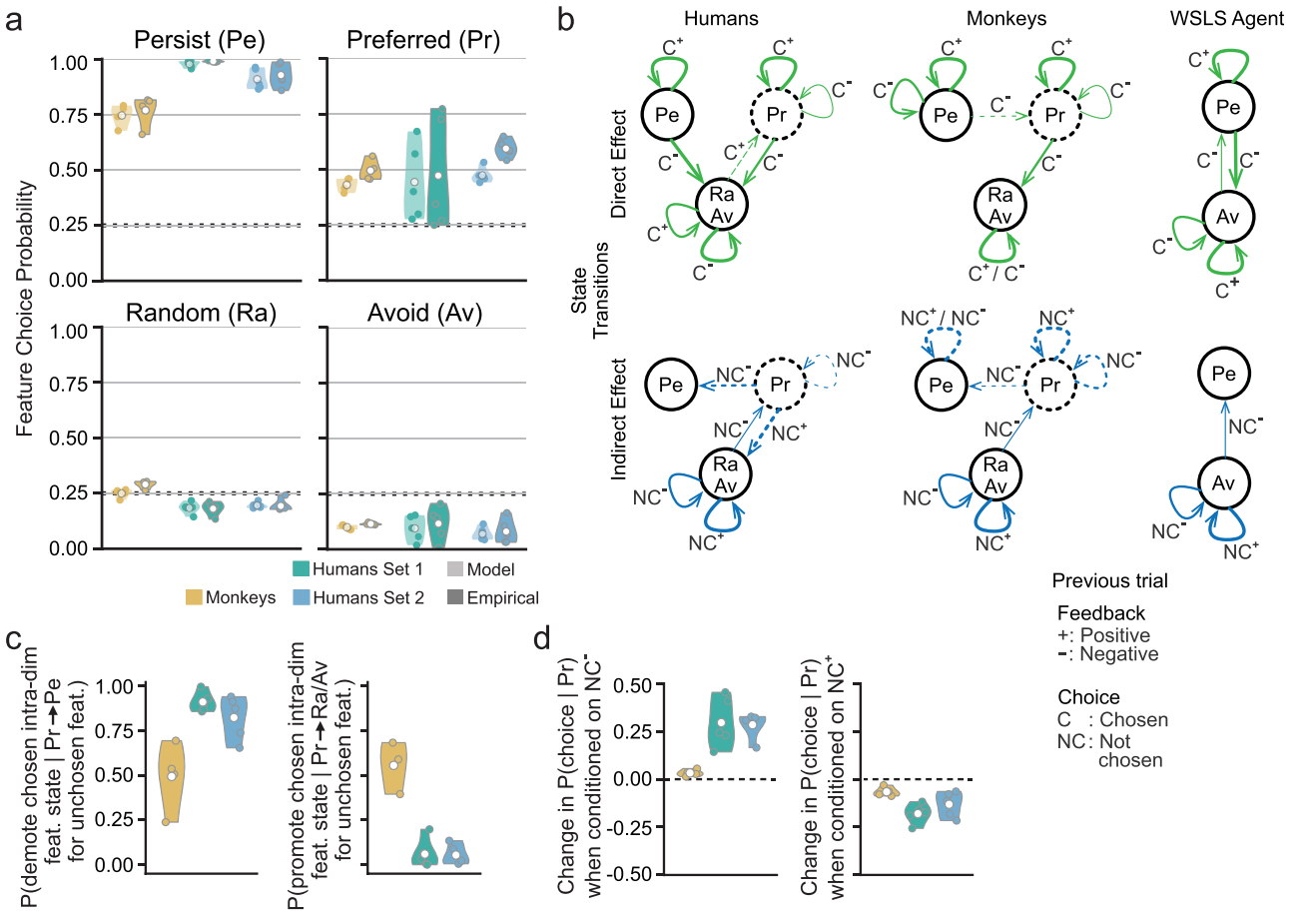

Figure 3. Model describes rule-learning dynamics in terms of changes in feature-attentional states. ***a***, Choice probability of features associated with each state in human (Dataset 1, green; Dataset 2, blue) and monkey (brown) subjects computed directly from model parameters and measured empirically based on most-likely state estimates. Choice probabilities order feature states akin to levels of attention. ***b***, Decision process describing how humans, monkeys, and the WSLS agent start, continue, and stop exploring a feature, derived from their history-dependent state transition probabilities. Process is decomposed based on outcome-dependent transitions when the feature is chosen (direct effect) or not chosen (indirect effect). Arrow thickness indicates probability of the transition. Dashed lines highlight deviations from the WSLS strategy. ***c***, Probability of demoting the state of a chosen feature to a lower probability state when an unchosen intradimensional feature is promoted from the preferred to persist state (left). Probability of promoting a chosen feature to a higher probability state when an unchosen intradimensional feature is demoted from the preferred to the random/avoid state (right). Measurements test to what extent indirect effects of promoting or demoting features in the preferred state result from changing the state, and therefore the choice probability, of a chosen intradimensional feature. Perfect causality would coincide with a probability of 1.0. ***d***, Change in the choice probability of a feature in the preferred state after receiving negative (left) or positive (right) feedback for choosing a different feature. The indirect effect significantly increases the feature choice probability in the former situation and decreases it in the latter. See Extended Data Figures 3-1–3-3 for more details.

We compared the observed behavior in our task with a WSLS strategy taking into account task structure differences between our task and the two-armed bandit task (Fig. 1*c*). The composition of a chosen object by three features forces the choice of features in the avoid state of the WSLS strategy. Thus, a WSLS decision process for our task must define transitions for such features when they are chosen. Moreover, the multidimensional environment of the task offers multiple alternatives for a subject to shift attention to during lose–shift, compared with the two-armed bandit task. An updated WSLS strategy that accounts for these differences is depicted in Extended Data Figure 2-2*b* (right). We can now compare the decision process inferred by the model for the two species (Extended Data Fig. 2-2*b*, left-middle) to this WSLS decision process, revealing salient differences that are delineated by dashed lines. Key among these is the existence of a preferred state where items are not chosen with certainty (or near-certainty) as in the persist state but above chance. The effect of direct feedback (as a result of choosing the feature) on these states and the random/avoid states is similar to those in the WSLS decision process. For example, both species select a feature in the random/avoid state for exploration (by promoting it to the preferred state) seemingly at random after receiving negative feedback for choosing other features. However, an interesting exception is that humans sometimes choose to explore such a feature upon receiving positive feedback for choosing it.

Larger differences emerge with regard to indirect effects of feedback. A feature may not be chosen on a trial when it is associated with the preferred state (feature choice probability in preferred state <1, Fig. 3*a*). However, its state may still transition subject to the feedback received at the end of the trial—an indirect effect. For example, humans and, to a lesser extent, monkeys demote features from the preferred to the random/avoid state upon receiving positive feedback for choosing a different feature. Consequently, their probability of subsequently choosing an unchosen feature that was associated with the preferred state decreases (Fig. 3*d*, right). They also promote features from the preferred to the persist state upon receiving negative feedback for choosing a different feature. Consequently, their probability of subsequently choosing an unchosen feature that was

associated with the preferred state increases (Fig. 3*d*, left). This is striking because it is the only way a feature can transition into the persist state, which appears to be reserved mainly for a feature that the subject determines to be the rule (Fig. 2*d*). Receiving positive feedback for choosing a feature in the preferred state does not definitively confirm that it is the rule, since the rule may be among the other two features in the chosen object. Confidence in the rule's identity may be increased based on the consistency of receiving such direct positive feedback across many trials. Alternatively, it may be done by ruling out other candidates, that is, after receiving negative feedback for choosing an object with a different candidate feature. Consistent with this interpretation, measurements show that when a feature under exploration is not chosen, the object that is chosen often contains a different feature that is also under exploration (Fig. 4*d*).

This approach of promoting a feature to the persist state as an inferred consequence of ruling out an alternative candidate, rather than integrating direct positive feedback across trials in favor of the feature, may be favored by both species due to its computational simplicity—it relies on the outcome of just the previous trial rather than multiple trials and thereby reduces working-memory demands. However, it is possible that these inference-like computations are not deliberate but an inadvertent consequence of demoting or promoting an intradimensional chosen feature. For example, given that the probability of choosing all shapes must sum to 1, when one shape is demoted after its choice produces negative feedback, the probability of choosing another shape that was associated with the preferred state may automatically increase, forcibly promoting it to the persist state. Measurements of the probability of demoting or promoting the chosen feature while promoting or demoting, respectively, an unchosen intradimensional feature in the preferred state are mixed: monkeys do so at chance levels; humans always demote the chosen feature while promoting the unchosen feature but seldom promote the chosen feature while demoting the unchosen feature (Fig. 3*c*). Nevertheless, these indirect-effect transitions directly and significantly alter the subsequent choice probability of the unchosen feature (Fig. 3*d*). In summary, the best-fit models discover feature-based attentional states whose dynamics show marked deviations from a WSLS strategy.

Both species simultaneously evaluate multiple features over several trials during rule learning

The explore–exploit dilemma pits the benefit of continuing to select a recently rewarded option (exploit) against the benefit of selecting a different and potentially more rewarding (but possibly less rewarding) option (explore). While much work has been done to determine how humans and other animals navigate this dilemma (Hills et al., 2015; Gershman, 2018; Wilson et al., 2021), how they deal with it in a multidimensional environment with transiently overlapping options remains unclear. Which of the three features of a chosen and rewarded object should be exploited on the next trial, given that they are unlikely to appear colocated in the same object on the following trial? How should the trade-off between the computational complexity and information efficiency of exploring several features at once be resolved?

The model finds that both species continuously explore one or more features (Fig. 4*a*). In the process, they explore multiple features over the course of a block before ultimately identifying the rule (Fig. 4*b*). Moreover, each feature is often explored for a series of several trials in both species (Fig. 4*c*). But the number of these trials is substantially larger in monkeys, a finding we analyze more closely in the following sections. The model also indicates that both species often explore multiple, but not all, features at a time (Fig. 4*e*). This is consistent with the theory of selective attention (Driver, 2001; Corbetta and Shulman, 2002) wherein objects are selectively attended to (or filtered for higher processing) subject to an internally maintained set of relevant perceptual features

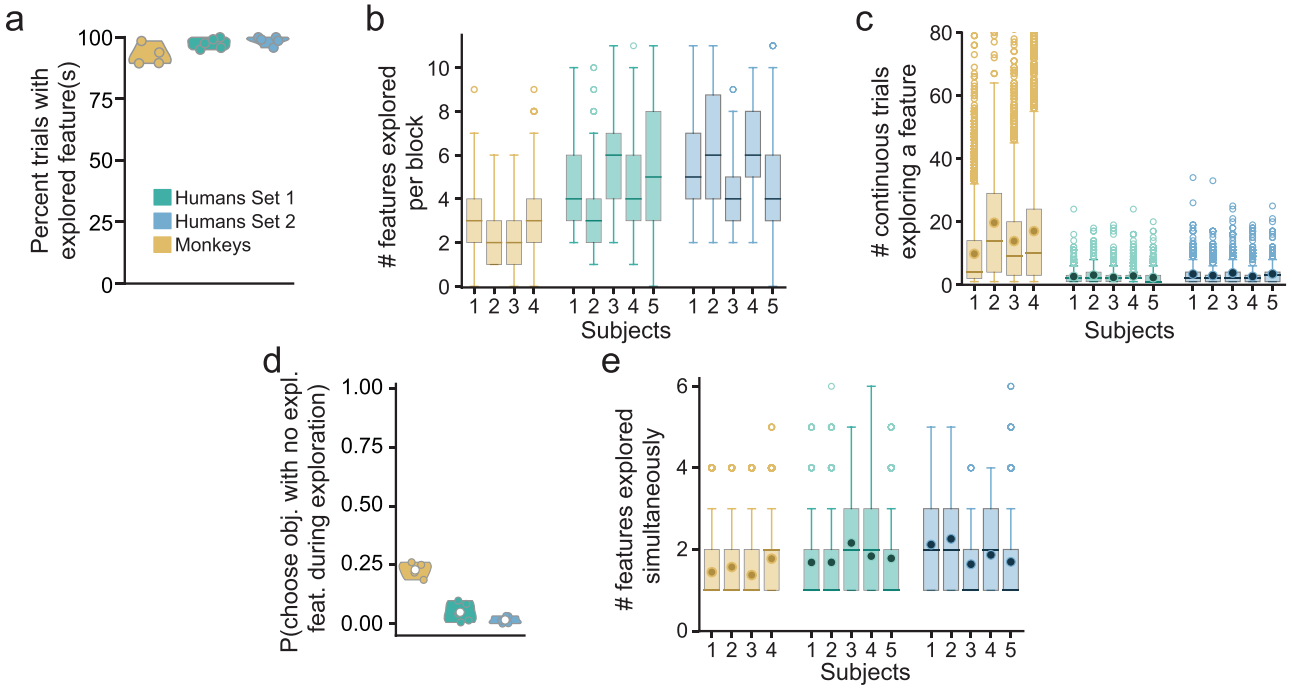

Figure 4. Monkeys and humans explore multiple features for several trials in a row to evaluate them. ***a***, The percentage of all trials where at least one feature is under exploration by humans (Dataset 1, green; Dataset 2, blue) and monkeys (brown). ***b***, Distribution of the number of features explored by each monkey and human subject in a block. ***c***, Distribution of number of continuous trials with a feature in an exploration state. ***d***, Probability of choosing an object with all features in the random or avoid state, while at least one other feature is in the preferred or persist state. ***e***, Distribution of number of features simultaneously explored by each monkey and human subject in trials where at least one feature is under exploration.

(or attentional filters). This also underscores how both species solve the computational complexity–information efficiency trade-off. Since it is computationally challenging to simultaneously attend to and evaluate all 12 features over several trials but inefficient to attend to one feature at a time, both species evaluate a small subset of all features at a time. Indeed, during exploration, it is uncommon for either species to select an object where none of the features is in the preferred or persist states (Fig. 4*d*). However, monkeys do engage in such random exploration much more frequently than humans.

From these results, we conclude that both species exhibit a deliberate form of exploration to address the challenges inherent in the task environment. Features, often more than one at a time, are selected for exploration via promotion to the preferred state after choices of other features produce negative feedback (Fig. 3*b*). They are then continuously explored so long as they produce positive feedback, until alternatives are ruled out. At this point, they are then promoted to the persist state or are ruled out after choosing them produces negative feedback (or, in the case of humans, choosing other features produces positive feedback).

Categorization of feature-attentional states characterizes learning dynamics

Rule learning proceeds through a sequential process that progressively reduces ambiguity regarding the rule's identity until it is ultimately determined. Our model reveals elementary feature-specific computations that individuals in each species apply to maintain and update a small subset of candidate features, of which one may be the rule. To elucidate the over-arching learning dynamics that governs the inferred sequential rule-learning process, we developed a simple approach to categorize individual trials based on the features under exploration and the true rule (Fig. 5*a*). The categories are mutually exclusive and exhaustive—each trial falls into one and only one category. These are as follows:

1. "Perseveration": a continuous series of trials following a rule switch when the feature governing the previous rule is associated with the persist state
2. "Random search": trials when none of the features are under exploration (i.e., associated with the preferred or persist states)
3. "Nonrule exploration": trials when one or more features are under exploration not including the rule feature
4. "Rule-favored exploration": trials when one or more features including the rule are under exploration
5. "Rule preferred": trials when only the rule is associated with the preferred state
6. "Rule exploitation": trials when only the rule is associated with the persist state

We compare the distribution of categories for trials across the course of a rule block between species (Fig. 5*b*). Humans show a progression from perseveration to nonrule exploration, where nonrule features are explored and ruled out, to rule-favored exploration, where the rule feature is simultaneously explored with nonrule features, to rule exploitation, once other candidates are ruled out and the rule is identified. Monkey rule learning is described by similar dynamics except for a much higher incidence of the rule preferred category for most of the block. Also, while nearly all blocks in Human Dataset 1 end in the rule exploitation category, a large proportion of the blocks in the monkey and the second human datasets end in the rule-favored exploration category. A closer analysis of the learning criterion trials, which occur at the end of the block, revealed different reasons for this result in the latter two datasets. The stronger learning criterion (i.e., larger number of trials-to-criterion) in these datasets makes it more likely that the rule and a nonrule feature colocate in the same object on two or more trials in close temporal proximity. Consequently, even though the human subject is exploiting the rule feature in these trials, the model concludes that the nonrule feature is simultaneously under exploration. In contrast, monkeys continue exploring nonrule features even when they do not colocate with the rule feature. This distinction is evidenced by the reward probability, which approaches 1 much sooner in Human Dataset 2 than in the monkey dataset (Extended Data Fig. 5-1*b*). These results show that our categorization approach expresses human and monkey rule-learning dynamics in terms of behaviorally interpretable learning stages, for example, an increase in the reward rate following a rule switch in both species is marked by the onset of rule exploration with the rule-favored exploration category (Fig. 5*c*, bottom).

Examining the number of trials spent in each category determined bottleneck categories that produce the rule-learning performance deficit in monkeys (Fig. 5*c*). Specifically, monkeys spend much longer perseverating on the previous rule, in disambiguating the rule feature from nonrule features (rule-favored exploration) and demonstrating that they have learned the rule (rule preferred or exploitation). The latter two sources of the learning performance deficit in monkeys also explain a majority of the variance in their performance across blocks (Extended Data Fig. 5-1*a*, bottom). In contrast, the number of trials humans spend exploring nonrule features before selecting the rule feature for exploration (nonrule exploration) largely determines the variance in their rule-learning performance.

Random exploration prolongs the expression of learning in monkeys

A key difference between the two species identified via this trial categorization is that monkeys spend many more trials than humans in the rule preferred or exploitation categories. These extra trials spent demonstrating or expressing that the rule has been learned significantly increase both the block length mean and variance (Fig. 5*c*; Extended Data Fig. 5-1*a*). A comparison of monkey learning performance and human performance in Dataset 2 shows that this interspecies difference is not caused by a difference in task parameters or learning criteria. Instead, we hypothesized that the larger mean and variance of the duration of time spent in the rule exploitation category by monkeys compared with that by humans (Fig. 6*a*) may result from their random exploration of other features when a feature is already associated with the persist state (Fig. 3*a*). This behavior is unique to monkeys and is prevalent even during rule exploitation trials (Fig. 6*b*)—after they have identified the rule, monkeys occasionally choose objects that do not include the rule feature.

To test our hypothesis, we simulated agents that select the rule feature with the same probability as monkeys and humans do during the rule exploitation category and asked how many trials it would take these agents to reach a learning criterion. The results revealed that the trial count distributions of the simulated agents were nearly identical to the corresponding subjects (Fig. 6*c*), thus confirming our hypothesis. Similar "random errors" have been observed in humans with focal lateral prefrontal lesions on the WCST (Barcelo and Knight, 2002), where they were attributed to distraction or a failure to maintain the rule in

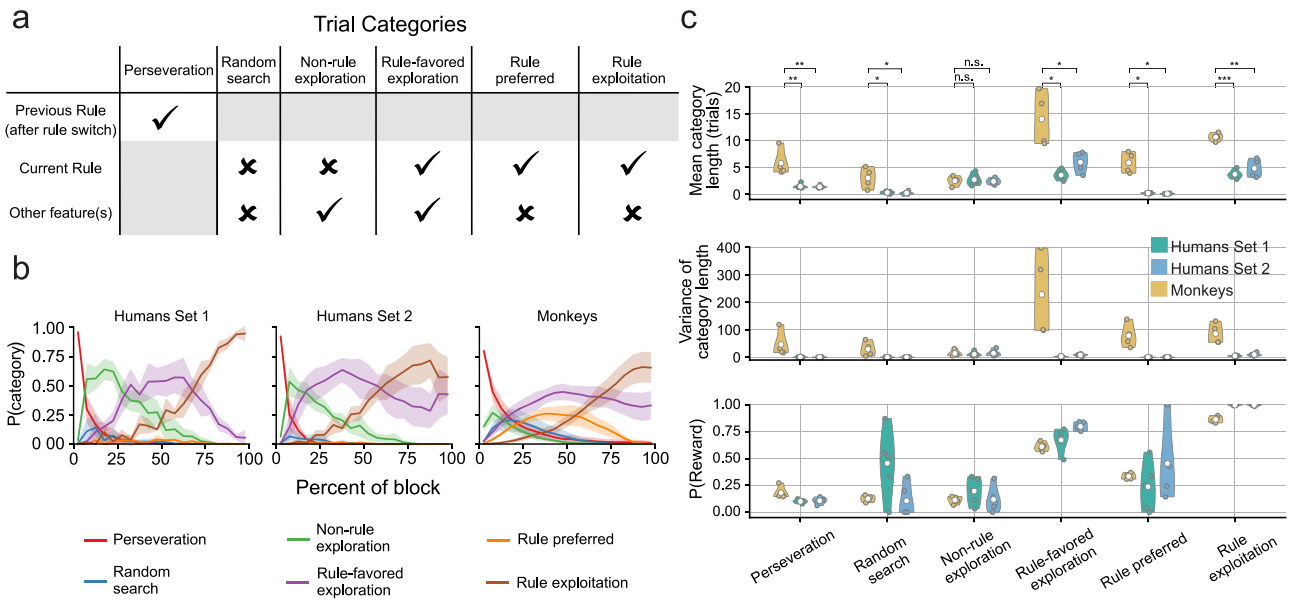

Figure 5. Exploration-based trial categories reveal learning dynamics and identify causes for monkey learning performance deficit. ***a***, Definition of the six trial categories based on whether the features under exploration during the trial include the rule feature. ***b***, Distribution of trial categories at each percentile of rule block. Lines and shaded areas reflect mean values and SEM, respectively, across subjects. ***c***, Trial category summary statistics (top, mean number of trials; middle, variance of number of trials; bottom, reward probability) across rule blocks for human (Dataset 1, green; Dataset 2, blue) and monkey (brown) subjects. Interspecies comparisons of the mean number of trials per category reveal significant differences in the perseveration, random search, rule-favored exploration, rule preferred, and rule exploitation categories (bootstrap test with *t* statistic); n.s., not significant; $*p < 0.1$; $**p < 0.01$; $***p < 0.001$. See Extended Data Figure 5-1 for more details.

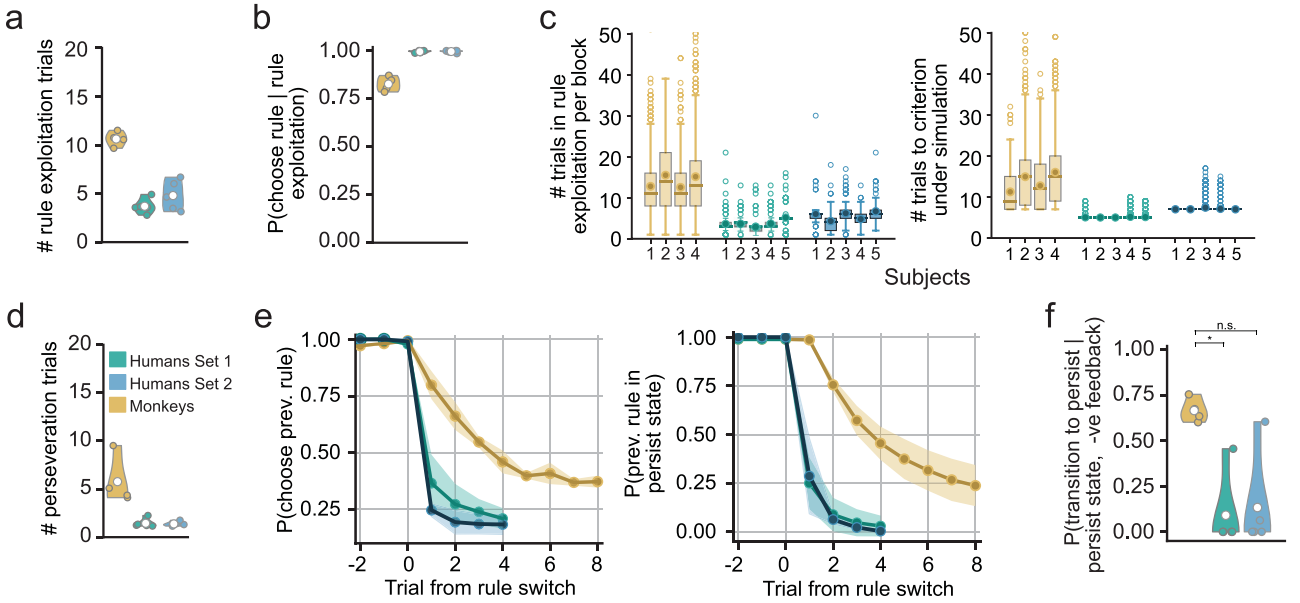

Figure 6. Random exploration and perseverative errors prolong monkey rule learning. ***a***, Mean number of trials spent by human (Dataset 1, green; Dataset 2, blue) and monkey (brown) subjects in the rule exploitation category per rule block. ***b***, Probability of selecting an object with the rule feature across trials in the rule exploitation category. Monkeys occasionally explore other objects compared with humans. ***c***, Distribution of the number of trials spent by human and monkey subjects in the rule exploitation category per rule block (left) and by simulated agents that select the rule feature with probabilities in ***b*** until they reach a learning criterion (right). ***d***, Mean number of trials spent by human and monkey subjects in the perseveration category per rule block. ***e***, The probability of humans and monkeys choosing the previous rule feature at each trial after a rule switch (left) is commensurate with the probability of the previous rule feature being associated with the persist state (right). ***f***, The probability of the previous rule feature transitioning back into the persist state after its selection produces negative feedback is higher in monkeys (bootstrap test with *t* statistic); n.s., not significant; $*p < 0.1$.

working memory. However, it remains unclear whether the monkeys in our experiments were more distractable than their healthy human counterparts or deliberately adopted occasional random exploration as part of their strategy—for example, to prolong a highly rewarded state.

Reduced sensitivity to negative feedback increases perseverative errors in monkeys

Perseverative errors occur when the rewarded feature on the previous block continues to be chosen following the rule switch despite receiving negative feedback for the choice. Such errors

are characteristic of frontal lobe damage and dysfunction in the WCST (Milner, 1963; Gläscher et al., 2019), where they are believed to reflect a cognitive deficit in adapting to changes in task contingencies. Pronounced perseveration error rates in the WCST are also observed in patients with neuropsychiatric (Sullivan et al., 1993; Ozonoff and McEvoy, 1994; Everett et al., 2001) and substance abuse disorders (Sullivan et al., 1993; Bishara et al., 2010). Interestingly, our model's association of the persist state with the previous rule feature during consecutive trials immediately following a rule switch suggests that monkeys persevere on the previous rule for several more trials than humans (Fig. 6*d*). Indeed, direct measurements showed that the probability of choosing the previous rule after a rule switch is consistent with such a state estimate in both species (Fig. 6*e*).

To determine the cause of the elevated perseveration in monkeys, we asked which choice outcome(s) best explained the difference in continued persistence with the previous rule between the two species. Our analysis showed that humans were far more likely to demote the chosen previous rule feature from the persist state in response to negative feedback compared with monkeys (Fig. 6*f*). Monkeys' weaker sensitivity to negative feedback parallels that of humans with substance abuse disorders and prefrontal lesions performing the WCST, who also perseverate more than healthy controls (Bishara et al., 2010; Gläscher et al., 2019).

Reduced negative feedback sensitivity compromises efficient credit assignment and prolongs rule learning in monkeys

The largest contribution to the interspecies difference in rule-learning performance is from trials in the rule-favored exploration category where the rule feature is concurrently explored with one or more nonrule features (Fig. 7*a*,*b*, left). While it is reasonable to explore the rule for several consecutive trials as it produces rewards, what must be explained is why nonrule features are concurrently explored for many more trials by monkeys. Indeed, monkeys continuously explore individual nonrule features for many more consecutive trials during the rule-favored exploration category (Fig. 7*b*, right). This explains the lengthier duration of this category in monkeys, resulting from a higher probability of a nonrule feature transitioning back into an exploration state during rule-favored exploration trials (Fig. 7*c*). Analysis further showed that this interspecies difference in transition probability is explained by a lower sensitivity of monkeys to either form of negative feedback—direct, when the nonrule feature is chosen and negative feedback is received, and indirect, when it is not chosen and positive feedback is received (Fig. 7*d*).

While both species also explore nonrule features during nonrule exploration trials, time spent in this category is relatively short in both humans and monkeys (Fig. 5*c*). So what explains the difference in duration between the two categories in monkeys? Since the nonrule exploration category is followed by rule-favored exploration trials, one possibility is that it is cut short by the onset of exploration of the rule feature as the nonrule feature continues to be concurrently explored for many more trials. However, measurements in monkeys showed that nonrule feature exploration only occasionally spans the two categories (probability = 0.27 ± 0.04). Therefore, the number of trials during which a nonrule feature is explored by monkeys is usually much smaller when it happens in the nonrule exploration category than in the rule-favored exploration category.

This difference in duration is reflected in a higher probability of a nonrule feature transitioning back into an exploration state during rule-favored exploration trials (Fig. 7*e*). Since the

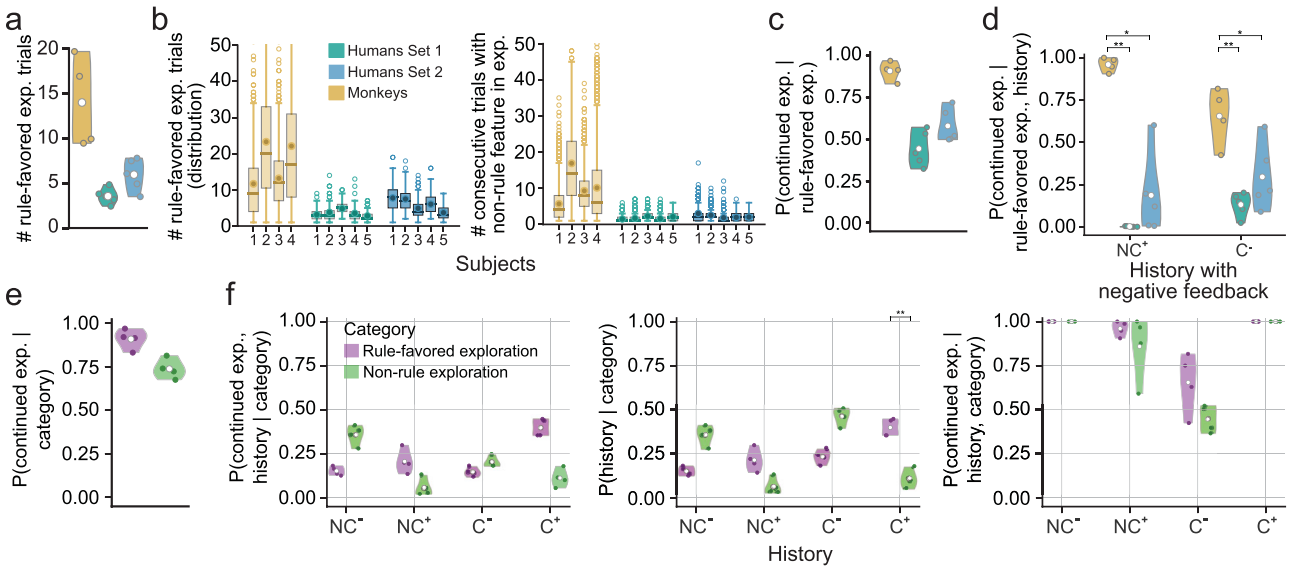

Figure 7. Diminished negative feedback sensitivity prolongs concurrent exploration of rule and nonrule features. ***a***, Mean number of trials spent by human (Dataset 1, green; Dataset 2, blue) and monkey (brown) subjects in the rule-favored exploration category per rule block. ***b***, Distribution across rule blocks of the number of trials spent in the rule-favored exploration category by each subject (left) and of the number of consecutive trials spent by them exploring individual nonrule features during this category (right). ***c***, Probability of a nonrule feature transitioning back into an exploration state during rule-favored exploration trials. ***d***, The probability of a nonrule feature transitioning back into an exploration state upon receiving negative feedback for choosing it (direct negative feedback) or positive feedback for choosing a different feature (indirect negative feedback) during rule-favored exploration trials is higher in monkeys (bootstrap test with *t* statistic). ***e***, Probability of a nonrule feature transitioning back into an exploration state during rule-favored exploration trials and nonrule exploration trials in monkeys. ***f***, Joint probability of a nonrule feature transitioning back into an exploration state and each choice–outcome history occurring during rule-favored exploration trials and nonrule exploration trials in monkeys (left); probability of each choice–outcome history occurring during either category (middle); probability of a nonrule feature transitioning back into an exploration state in response to each choice–outcome history during either category (right). The higher probability of a nonrule feature transitioning back into an exploration state during rule-favored exploration trials compared with that during nonrule exploration trials is explained by a higher incidence of direct positive feedback for choosing the nonrule feature in the former category (bootstrap test with *t* statistic); $*p < 0.1$; $**p < 0.01$. See Extended Data Figure 7-1 for more details.

probability of transitioning back into the explore state is a marginalization of its joint probability with the choice outcome it follows, we asked what choice–outcome history best explains the transition probability difference between the two categories. Measurements showed that receiving positive feedback for choosing the nonrule feature (C^+) is the key differentiator between the joint probabilities for the two categories (Fig. 7*f*, left). This could either be because the transition probability in response to this history is different under the two categories or because the frequency of the history is different under them. We found that while the transition probabilities are similar (Fig. 7*f*, left), monkeys are more likely to have received positive feedback for choosing a nonrule feature under exploration during the rule-favored, exploration category (Fig. 7*f*, middle). When the rule feature is concurrently explored with a nonrule feature (rule favored, exploration), the probability of selecting them both when they colocate in an object is higher. This increases the probability of receiving positive feedback for choosing the nonrule feature, which makes appropriately assigning credit to the rule feature challenging. This underscores the importance of negative feedback sensitivity in demoting nonrule features from exploration states, in the absence of which the duration of concurrent exploration of the rule and nonrule feature(s) is prolonged.

Discussion

Methodological and technological advances in training and recording from animal models allow for the study of increasingly complex behaviors in nonhumans. However, before interpreting their brain activity as a human-like model of neural computation, it is important to ascertain whether their computational algorithms are human-like. Usually, macaque monkeys and humans learn the structure of tasks in different ways, particularly when monkeys must learn via impoverished reward-based feedback, while humans learn via rich verbal instruction plus feedback. This raises the possibility that while they both learn the same tasks, they may enlist different abstractions, cognitive operations, and neural mechanisms (Melloni et al., 2019). Indeed, this critical issue has been considered in interspecies studies of the cognitive processes involved in short-term memory (Wittig et al., 2016), strategic behavior (Brosnan et al., 2011, 2017; Moeller et al., 2023), and task-switching (Caselli and Chelazzi, 2011), with carefully matched experimental approaches. Our study aims to assess whether macaques and humans employ similar mental representations and operations to perform a cognitively complex rule-switching task that relies on several interdependent cognitive processes. Our findings demonstrate that both species employ similar overall strategies to perform the task (Fig. 3*a*,*b*). However, key differences in the decision criteria of these strategies explain monkey performance deficits on the task.

The task that motivated our study, the WCST, was originally developed to test cognitive flexibility (Grant and Berg, 1948; Heaton, 1981). Ensuing research has made it clear that rather than engaging a single cognitive process for task set switching, the WCST relies on a variety of cognitive functions including working memory, attention, decision-making, inhibitory control, and reasoning (Dehaene and Changeux, 1991; Barcelo, 2001; Buchsbaum et al., 2005; Lie et al., 2006; Gläscher et al., 2019). This has inspired systematic studies on WCST performance with two related goals. First, research has focused on accurately characterizing rule-learning strategies and/or the cognitive processes that support their underlying computations (Bishara et al., 2010; Wilson and Niv, 2012; Gläscher et al., 2019). We developed an approach to identify the rule-learning strategy in humans and monkeys based on hidden behavioral states. The best-fit models for both species ascribe these hidden states to varying levels of attention toward individual task-relevant visual features (Fig. 3*a*). This is consistent with the conclusions of earlier studies that humans contend with the "curse-of-dimensionality" in the WCST with selective attention toward individual features during exploration (Bishara et al., 2010; Wilson and Niv, 2012; Gläscher et al., 2019). Our findings clarify these results by showing that in a high-dimensional variant of the task (12 instead of 3 possible rules), both humans and monkeys must further contend with a trade-off between computational complexity and information efficiency while exploring for the rule, and they do so by selectively attending to a few, but not all, features at a time (Fig. 4*e*).

Our approach differs from earlier studies in that it does not postulate a specific learning algorithm (Bishara et al., 2010; Wilson and Niv, 2012). Rather, it discovers the decision process that determines the rule-learning strategy. In doing so, it illustrates important differences between human/monkey rule-learning strategies and the commonly observed WSLS strategy (Fig. 3*b*). For example, a key function of the preferred state, which is not part of the WSLS strategy, is to support the simultaneous exploration of multiple features at a time. This state is also associated with inference-like computations that support a computationally efficient strategy of narrowing down the rule by eliminating other candidates using unambiguous negative feedback.

The second goal, with stronger clinical implications, is the assessment and categorization of error types toward identifying accurate behavioral markers for different types of neuropsychiatric disorders and dysfunction or lesions of different brain regions (Drewe, 1974; Owen et al., 1991; Robbins, 1996; Barcelo and Knight, 2002; Buchsbaum et al., 2005; Nagahama et al., 2005; Lie et al., 2006; Buckley et al., 2009; Bishara et al., 2010; Gläscher et al., 2019). In support of this goal, we have developed a learning-stage categorization method that delineates learning stages by the features under exploration and their relationship to the rule (Fig. 5*a*). Intuitively, this approach tracks how far along a subject is from learning the rule and reflects this in the reward rates across categories (Fig. 5*c*, bottom). Crucially, it allows us to precisely ascribe differences in learning performance between subjects to differences in individual categories (Fig. 5*a*; Extended Data Fig. 5-1*a*).

Consequently, we identify various known error types, but also newer ones that may prove useful in future investigations of behavioral markers for neuropsychiatric disorder and cognitive impairment. Our results distinguish perseverative errors (made during the perseveration category) from nonperseverative ones. Consistent with earlier work in humans with PFC lesions (Barcelo and Knight, 2002), it also subcategorizes the latter into random errors that occur when choices are inconsistent with the hypotheses being tested by the subject and efficient errors that occur when they are. It further identifies two forms of random errors: one occurs during rule search (before the rule preferred or exploitation categories) when subjects occasionally choose none of the features they are currently exploring (Fig. 4*d*); the other occurs after they have found the rule and while they are demonstrating this (Fig. 6*b*). It remains unclear if these random errors are a feature of cognitive flexibility and result from random exploration or are caused by the failure to maintain the attention set in working memory. Indirect evidence in humans has been found in favor of the latter interpretation (Figueroa and Youmans, 2013). If in fact it is a result of higher distractability in monkeys, monkey performance may be

improved by imposing stronger controls on potential environmental distractors (Malmo, 1942). Most errors during rule-favored exploration trials are repeated despite unambiguous (direct or indirect) negative feedback (Fig. 7). These "disambiguation" errors are neither random nor efficient but arise from a deficit in disambiguating the rule feature that is under exploration from a simultaneously explored nonrule feature. This newly identified error type is consistent with the observation that macaque monkeys are more sensitive to negative feedback when the average reward is low and more sensitive to positive feedback when the average reward is high (Wittmann et al., 2020) and bears further exploration in patient populations.

The higher incidence of these error types in monkeys may have more to do with how they learn rather than some fundamental cognitive constraints. Since they cannot receive a rich verbal description of the task's structure as humans do and must learn about it via trial and error, monkeys may misinterpret uncued rule switches as stochasticity in the environment resulting in a maladaptive strategy. To address this confound, we trained a second set of human subjects (Dataset 2) with minimal verbal instructions that excluded details on the task's structure, forcing them to learn the task's structure via trial and error. Yet, these subjects learned the task's structure within a couple of rule blocks, thereafter learning new rules as rapidly as the first human subject set (Dataset 1). Model fits also revealed similar rule-learning strategies in these subjects as in the first set. However, this does not necessarily limit the cause of the interspecies learning differences to differences in cognitive capacity. Whereas the monkeys were motivated by food reward, human motivation was driven by an internally maintained goal of maximizing the total number of correct feedback responses. This may produce widely different motivational states in the two species and thereby engage only partially overlapping learning pathways in their brains. In addition, different response modalities were employed by the two species: monkeys used fixation and saccades to choose the target, whereas humans use key presses. The eye movements during many repeated trials are known to drive risk-seeking attitudes in macaques, while the more effortful and deliberate key presses might promote more conservative, attentive attitudes in humans. Therefore, caution is warranted given the potential bias in cognitive states introduced by these variations in task design for the two species.

Nevertheless, it is intriguing that many of the errors that contribute to monkey rule-learning deficits have also been implicated in the poor performance of humans with cognitive impairment. A higher incidence of perseverative errors in patients with prefrontal cortex pathology was first reported by Milner (1963). Random errors are frequently observed in patients with frontal lobe dysfunction (Barcelo and Knight, 2002). Poor sensitivity to negative feedback is more pervasive in patients with schizophrenia and substance abuse (Bishara et al., 2010; Gläscher et al., 2019). Interestingly, these observations were made in subjects performing the WCST that bears salient differences from our task, which affects the kinds of ambiguities that a subject must contend with and will likely alter their rule-learning strategy as a consequence. First, the WCST relies on a different decision-making computation: rather than select one of four randomly generated objects, WCST subjects must match a randomly generated target object to one of the four fixed reference objects. Second, the matching rule corresponds to a feature dimension rather than an individual feature, for example, under the color rule, subjects are rewarded when they match the target object to the reference object with the same color. While this reduces the number of rules from12 to 3, WCST rules are more abstract.

Ultimately, the validity and generality of these similarities between rule-learning monkeys and cognitively impaired WCST human subjects will need to be resolved through more studies that compare the two species on a wider range of cognitive functions, with broader controls, and using a variety of different tasks. In addition, the basis of these similarities can be resolved through interspecies neural data comparisons. Our work produces several testable neural hypotheses in both species. First, does the neural representation of the current rule persist during and across rule exploitation trials (i.e., after its identity has been learned; Wallis et al., 2001; Mansouri et al., 2006; Bernardi et al., 2020; Minxha et al., 2020)? Second, since the set of explored features must be maintained across several trials, are they represented by neural activity during and across trials? Our model indicates that this attention set is typically small (Fig. 4*e*), and longer bouts of exploring multiple features simultaneously (Fig. 7*b*) require a choice alternation between these features. This drives the need to maintain the explored features in working memory, particularly to support recall of one of them after it is not chosen on one or more previous trials. Third, are the distinct error types (perseverative, random, efficient, disambiguation) differentially represented in the brain? Error coding neurons have been reported in the prefrontal cortex of monkeys performing a WCST analog (Mansouri et al., 2006; Kuwabara et al., 2014). Moreover, perseverative, random, and disambiguation errors signal the need to disengage from the previous rule, address a working-memory error, and remove a nonrule feature from the attention set, respectively. Due to this difference in their function, they may be represented differently, either eliciting stronger responses in different brain regions or eliciting differential responses in the same region (Barcelo, 1999). Fourth, is the strength of these error signals or their modulation of the attention set representation (Mansouri et al., 2006) larger on trials when they serve their function? For example, are perseverative error signals stronger on trials after which perseveration halts compared with when it does not? These analyses can reveal the reason for the interspecies performance difference: What neurocognitive differences explain the relative prevalence of perseverative, random, and disambiguation errors in monkeys compared with those in humans, and are they also observed in humans with cognitive impairment?

There exist several avenues to clarify and improve upon our approach. A key difference between earlier models and ours is our assumption that each feature is associated with discrete states, which our model relates to feature-based attention. In contrast, Bayesian and reinforcement learning approaches posit that subjects reason about features by assigning continuous-valued functions such as belief (Wilson and Niv, 2012) and value (Bishara et al., 2010; Niv et al., 2015) to them, respectively. In future work, we will test whether a model with continuous-valued states provides a better fit to behavior. Our model has also been simplified to keep analysis tractable—it does not explicitly account for interactions between features. This had the unintended consequence of discovering the "phantom" avoid state. Future model improvements will incorporate such interactions explicitly.

In conclusion, we have applied a hypothesis-free state–characterization method to identify and compare the strategies of humans and monkeys on a rule-switching task. The hidden attentional states and state transitions inferred by the model facilitated the determination of the decision process underlying this strategy as well as the various stages of rapid rule learning. The inferred states substantively explain human and monkey choice behavior (Fig. 2*c*). Our overall approach reveals differences in cognitive strategy between the two species and isolates the identity and relative contribution of various error types to the performance

difference between the two species. It shows that random exploration or distraction and poorer sensitivity to negative feedback underlies a higher incidence of these error types in monkeys, leading to their underperformance. The high fidelity demonstrated by the model in inferring hidden attentional and decision states holds promise in advancing the search for more accurate behavioral markers of various types of cognitive dysfunction and in motivating targeted analyses to determine and compare the neural correlates of the various cognitive processes involved in rule learning and cognitive flexibility.

References

Barcelo F (1999) Electrophysiological evidence of two different types of error in the Wisconsin Card Sorting Test. Neuroreport 10:1299–1303.

Barcelo F (2001) Does the Wisconsin Card Sorting Test measure prefrontal function? Span J Psychol 4:79–100.

Barcelo F, Knight RT (2002) Both random and perseverative errors underlie WCST deficits in prefrontal patients. Neuropsychologia 40:349–356.

Barton RA, Venditti C (2013) Human frontal lobes are not relatively large. Proc Natl Acad Sci U S A 110:9001–9006.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214–1221.

Bengio Y, Frasconi P (1994) An input output hmm architecture. Adv Neural Inf Process Syst 7.

Bernardi S, Benna MK, Rigotti M, Munuera J, Fusi S, Salzman CD (2020) The geometry of abstraction in the hippocampus and prefrontal cortex. Cell 183:954–967.

Birch J, Schnell AK, Clayton NS (2020) Dimensions of animal consciousness. Trends Cogn Sci 24:789–801.

Bishara AJ, Kruschke JK, Stout JC, Bechara A, McCabe DP, Busemeyer JR (2010) Sequential learning models for the Wisconsin Card Sort Task: assessing processes in substance dependent individuals. J Math Psychol 54:5–13.

Bolkan SS, et al. (2022) Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. Nat Neurosci 25: 345–357.

Brosnan SF, Parrish A, Beran MJ, Flemming T, Heimbauer L, Talbot CF, Lambeth SP, Schapiro SJ, Wilson BJ (2011) Responses to the assurance game in monkeys, apes, and humans using equivalent procedures. Proc Natl Acad Sci U S A 108:3442–3447.

Brosnan SF, Price SA, Leverett K, Prétôt L, Beran M, Wilson BJ (2017) Human and monkey responses in a symmetric game of conflict with asymmetric equilibria. J Econ Behav Organ 142:293–306.

Buchsbaum BR, Greer S, Chang WL, Berman KF (2005) Meta-analysis of neuroimaging studies of the Wisconsin card-sorting task and component processes. Hum Brain Mapp 25:35–45.

Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PG, Kwok SC, Phillips A, Tanaka K (2009) Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science 325:52–58.

Calhoun AJ, Pillow JW, Murthy M (2019) Unsupervised identification of the internal states that shape natural behavior. Nat Neurosci 22:2040–2049.

Caselli L, Chelazzi L (2011) Does the macaque monkey provide a good model for studying human executive control? A comparative behavioral study of task switching. PLoS One 6:e21489.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215.

Deacon TW (1997) What makes the human brain different? Annu Rev Anthropol 26:337–357.

Dehaene S, Changeux JP (1991) The Wisconsin Card Sorting Test: theoretical analysis and modeling in a neuronal network. Cereb Cortex 1:62–79.

Donahue CJ, Glasser MF, Preuss TM, Rilling JK, Van Essen DC (2018) Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. Proc Natl Acad Sci U S A 115:E5183–E5192.

Drewe E (1974) The effect of type and area of brain lesion on Wisconsin Card Sorting Test performance. Cortex 10:159–170.

Driver J (2001) A selective review of selective attention research from the past century. Br J Psychol 92:53–78.

Ebitz BR, Albarran E, Moore T (2018) Exploration disrupts choice-predictive signals and alters dynamics in prefrontal cortex. Neuron 97:450–461.

Everett J, Lavoie K, Gagnon JF, Gosselin N (2001) Performance of patients with schizophrenia on the Wisconsin Card Sorting Test (WCST). J Psychiatry Neurosci 26:123.

Figueroa IJ, Youmans RJ (2013) Failure to maintain set: a measure of distractibility or cognitive flexibility? Proc Hum Factors Ergon Soc Annu Meet 57:828–832.

Fuster J (2015) *The prefrontal cortex*. Cambridge: Academic Press.

Gabi M, Neves K, Masseron C, Ribeiro PF, Ventura-Antunes L, Torres L, Mota B, Kaas JH, Herculano-Houzel S (2016) No relative expansion of the number of prefrontal neurons in primate and human evolution. Proc Natl Acad Sci U S A 113:9617–9622.

Gershman SJ (2018) Deconstructing the human algorithms for exploration. Cognition 173:34–42.

Gershman SJ, Niv Y (2010) Learning latent structure: carving nature at its joints. Curr Opin Neurobiol 20:251–256.

Gläscher J, Adolphs R, Tranel D (2019) Model-based lesion mapping of cognitive control using the Wisconsin card sorting test. Nat Commun 10:20.

Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR (1997) Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. Arch Gen Psychiatry 54:159–165.

Grant DA, Berg E (1948) A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. J Exp Psychol 38:404.

Heaton RK (1981) *Wisconsin card sorting test manual*. Lutz: Psychological Assessment Resources.

Herculano-Houzel S (2009) The human brain in numbers: a linearly scaled-up primate brain. Front Hum Neurosci 3:31.

Hills TT, Todd PM, Lazer D, Redish AD, Couzin ID, Cognitive Search Research Group (2015) Exploration versus exploitation in space, mind, and society. Trends Cogn Sci 19:46–54.

Kopp B, Lange F, Steinke A (2021) The reliability of the Wisconsin Card Sorting Test in clinical practice. Assessment 28:248–263.

Kuwabara M, Mansouri FA, Buckley MJ, Tanaka K (2014) Cognitive control functions of anterior cingulate cortex in macaque monkeys performing a Wisconsin Card Sorting Test analog. J Neurosci 34:7531–7547.

Lie CH, Specht K, Marshall JC, Fink GR (2006) Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. Neuroimage 30:1038–1049.

Linderman S, Antin B, Zotowski D, Glaser J (2020) SSM: Bayesian learning and inference for state space models (version 0.0.1). Available at: https://github.com/lindermanlab/ssm

Malmo RB (1942) Interference factors in delayed response in monkeys after removal of frontal lobes. J Neurophysiol 5:295–308.

Mansouri FA, Freedman DJ, Buckley MJ (2020) Emergence of abstract rules in the primate brain. Nat Rev Neurosci 21:595–610.

Mansouri FA, Matsumoto K, Tanaka K (2006) Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. J Neurosci 26:2745–2756.

Melloni L, et al. (2019) Computation and its neural implementation in human cognition. In Strüngmann Forum Reports 27:323–346.

Milner B (1963) Effects of different brain lesions on card sorting: the role of the frontal lobes. Arch Neurol 9:90–100.

Minxha J, Adolphs R, Fusi S, Mamelak AN, Rutishauser U (2020) Flexible recruitment of memory-based choice representations by the human medial frontal cortex. Science 368:eaba3313.

Moeller S, Unakafov AM, Fischer J, Gail A, Treue S, Kagan I (2023) Human and macaque pairs employ different coordination strategies in a transparent decision game. Elife 12:e81641.

Nagahama Y, Okina T, Suzuki N, Nabatame H, Matsuda M (2005) The cerebral correlates of different types of perseveration in the Wisconsin Card Sorting Test. J Neurol Neurosurg Psychiatry 76:169–175.

Nelson HE (1976) A modified card sorting test sensitive to frontal lobe defects. Cortex 12:313–324.

Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, Radulescu A, Wilson RC (2015) Reinforcement learning in multidimensional environments relies on attention mechanisms. J Neurosci 35:8145–8157.

Owen AM, Roberts AC, Polkey CE, Sahakian BJ, Robbins TW (1991) Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man. Neuropsychologia 29:993–1006.

Ozonoff S, McEvoy RE (1994) A longitudinal study of executive function and theory of mind development in autism. Dev Psychopathol 6:415–431.

Passingham RE (1972) Non-reversal shifts after selective prefrontal ablations in monkeys (*Macaca mulatta*). Neuropsychologia 10:41–46.
Passingham RE, Smaers JB (2014) Is the prefrontal cortex especially enlarged in the human brain? Allometric relations and remapping factors. Brain Behav Evol 84:156–166.
Paszke A, et al. (2019) PyTorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst 32.
Robbins TW (1996) Dissociating executive functions of the prefrontal cortex. Philos Trans R Soc Lond B Biol Sci 351:1463–1471.
Rogers RD, Andrews T, Grasby P, Brooks D, Robbins T (2000) Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. J Cogn Neurosci 12:142–162.
Roy NA, Bak JH, Laboratory TIB, Akrami A, Brody CD, Pillow JW (2021) Extracting the dynamics of behavior in sensory decision-making experiments. Neuron 109:597–610.
Semendeferi K, Lu A, Schenker N, Damásio H (2002) Humans and great apes share a large frontal cortex. Nat Neurosci 5:272–276.
Semendeferi K, Teffer K, Buxhoeveden DP, Park MS, Bludau S, Amunts K, Travis K, Buckwalter J (2011) Spatial organization of neurons in the frontal pole sets humans apart from great apes. Cereb Cortex 21:1485–1497.
Sullivan EV, Mathalon DH, Zipursky RB, Kersteen-Tucker Z, Knight RT, Pfefferbaum A (1993) Factors of the Wisconsin Card Sorting Test as measures of frontal-lobe function in schizophrenia and in chronic alcoholism. Psychiatry Res 46:175–199.
Viterbi A (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans Inf Theory 13:260–269.
Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. Nature 411:953–956.
Watson TD, Azizian A, Squires NK (2006) Event-related potential correlates of extradimensional and intradimensional set-shifts in a modified Wisconsin Card Sorting Test. Brain Res 1092:138–151.
Wilson RC, Bonawitz E, Costa VD, Ebitz RB (2021) Balancing exploration and exploitation with information and randomization. Curr Opin Behav Sci 38:49–56.
Wilson RC, Niv Y (2012) Inferring relevance in a changing world. Front Hum Neurosci 5:189.
Wittig JH, Morgan B, Masseau E, Richmond BJ (2016) Humans and monkeys use different strategies to solve the same short-term memory tasks. Learn Mem 23:644–647.
Wittmann MK, Fouragnan E, Folloni D, Klein-Flügge MC, Chau BK, Khamassi M, Rushworth MF (2020) Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. Nat Commun 11:3771.