MIT Open Access Articles

Hierarchical reasoning by neural circuits in the frontal cortex

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Sarafyazd, Morteza and Mehrdad Jazayeri. "Hierarchical reasoning by neural circuits in the frontal cortex." Science 364, 6441 (May 2019): eaav8911 © 2019 The Authors

As Published: <http://dx.doi.org/10.1126/science.aav8911>

Publisher: American Association for the Advancement of Science (AAAS)

Persistent URL: <https://hdl.handle.net/1721.1/122955>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike

Hierarchical reasoning by neural circuits in the frontal cortex

Morteza Sarafyazd¹, Mehrdad Jazayeri^{1,*}

¹Department of Brain & Cognitive Sciences, McGovern Institute for Brain Research, MIT

*Corresponding author. Email: mjaz@mit.edu

Abstract

Humans process information hierarchically. In the presence of hierarchies, sources of failures are ambiguous. Humans resolve this ambiguity by assessing their confidence after one or more attempts. To understand the neural basis of this reasoning strategy, we recorded from the dorsomedial frontal cortex (DMFC) and anterior cingulate cortex (ACC) of monkeys in a task in which negative outcomes were either caused by misjudging the stimulus, or by a covert switch between two stimulus-response contingency rules. We found that both areas harbored a representation of evidence supporting a rule switch. Additional perturbation experiments revealed that ACC functioned downstream of DMFC and was directly and specifically involved in inferring covert rule switches. These results reveal the computational principles of hierarchical reasoning as implemented by cortical circuits.

Introduction

A hallmark of cognition is the ability to process information hierarchically. Consider the deliberations of a doctor helping a patient with equivocal symptoms. The doctor has to choose a judicious diagnostic test, interpret the results, prescribe a suitable medicine based on test results, and finally, evaluate the outcome. The hierarchical nature of this decision process makes failures ambiguous. When facing an unfavorable outcome, the doctor may question the dosage, the medicine, the test results, or the suitability of the test. Resolving this ambiguity demands a sophisticated causal inference strategy. Although human capacity to make such inferences is well established (1–7), the link between the key computational principles and the underlying neurobiology is not well understood.

When decisions are organized hierarchically, causal inference about errors demand two critical computations. First, one has to compute a graded expectation of potential outcomes – also known as confidence – based on the quality of evidence. Decades of work has provided strong evidence that humans and animals compute confidence over their choices (4, 8–15) and use it to improve subsequent decisions (3, 16, 17). Second, one must monitor performance at multiple timescales to tease apart proximal versus higher order causes of a failure (e.g., wrong choice of drug versus wrong assumption about the disease). Numerous experiments have found strong error-dependent signals in the dorsomedial frontal cortex (DMFC) and anterior cingulate cortex (ACC) consistent with performance monitoring in a variety of instrumental and conditioning tasks (18–27). An important observation has been that both cortical areas carry performance monitoring signals, and that ACC harbors representations of reward on a longer timescale that could be used to regulate strategic exploratory behavior in non-stationary environments (5, 28–38).

However, the neural substrates and mechanisms that allow humans and animals to compute and integrate confidence about low-level decisions to make strategic adjustments to higher level decisions are not known. To tackle this problem, we designed a hierarchical decision-making task for monkeys in which the rule relating the sensory evidence to behavioral response changed covertly throughout the experiment so that animals had to compare outcomes to expected outcomes across multiple trials to infer rule changes (Fig. 1A). Behavioral results indicated that monkeys, like humans, make such causal inferences based on their confidence. Concurrent neural recordings revealed that both DMFC and ACC carry signals related to performance monitoring with ACC playing a key causal role in making inferences about covert rule changes on longer timescales.

A behavioral task for hierarchical reasoning

In a volatile environment with two alternating response contingency rules (C), animals had to discriminate a sample interval, t_s (Fig. 1B). The interval varied between 530 and 1170 ms and was demarcated by two flashes, one at the fixation point and one in the periphery (Fig. 1C). In rule 1 (C_1), animals had to look toward the second flash (prosaccade) when t_s was shorter than the median interval, 850 ms, and away from it (antisaccade) when t_s was longer than 850 ms (Fig. 1B). In rule 2 (C_2), response contingencies were reversed: prosaccade (“Pro”) for “Long” and an antisaccade (“Anti”) for “Short”. Since generating antisaccades requires inhibition of a prepotent prosaccade response (25, 39, 40), we refer to C_1 and C_2 as “late inhibition” and “early inhibition”, respectively (Fig. 1B). The two rules alternated every 10–28 trials (Fig. 1C). On each trial, animals

had to make two hierarchically organized decisions. First, they had to report their belief about the current rule by making a saccade to one of two colored targets (blue for C_1 and red for C_2). Subsequently, t_s was presented and animals had to make a pro- or antisaccade depending on t_s and the rule. In the main experiment, which we refer to as the “inferred rule experiment”, rules were not cued and rule switches occurred covertly. Therefore animals had to infer the rule based on the outcome of previous trials. Animals also performed a control “instructed rule experiment” in which rule switches were explicitly cued. To receive reward, animals had to correctly report the rule and correctly discriminate t_s (Fig. 1C).

To assess animals’ performance, we measured the proportion of antisaccades ($\text{Pr}(\text{Anti})$) as a function of the sample interval. As a matter of convenience, we will express the interval in each trial relative to the criterion, which we refer to as the discriminant interval, $t_d = t_s - 850$. In the instructed rule experiment, $\text{Pr}(\text{Anti})$ increased lawfully with t_d for the late inhibition rule (Fig. 1D, blue; Table S1), and decreased for the early inhibition rule (Fig. 1D, red; Table S1). In the inferred rule experiment, we analyzed responses both in terms of the subjective rule (i.e., the rule reported by the animal; Fig. 1E) and in terms of the experimentally imposed objective rule (Fig. 1F). Animal’s performance with respect to the subjective rule was not significantly different from that of the instructed rule experiment (Table S1) verifying that animals understood the hierarchical structure of the task and followed the response contingencies according to their belief about the rule. Errors, when measured with respect to the objective rule, increased by a factor of 1.76 ± 0.04 and 1.62 ± 0.05 in Monkey K and I respectively (Wilcoxon rank sum test, $p < 0.001$). This is a straightforward consequence of the incorrect rule reports after covert rule switches.

Next, we focused on animal’s decision about the rule (Fig. 1G,H). In the instructed rule experiment, animals switched their selection of rule immediately after every objective rule switch indicating that animals successfully learned to follow instructions (Fig. 1G, yellow). In the inferred rule experiment, rule switches were covert and thus had to be inferred based on the pattern of feedbacks. A rational observer would make the following considerations: (1) any rewarded trial indicates that the rule was selected correctly and no rule switch is needed; (2) after a negative outcome, there is a chance that a covert rule switch might have occurred, and the chances increase after multiple consecutive negative outcomes; (3) when t_d is far from the median value (i.e., trials in which interval discrimination is relatively easy), the error is more likely to be due to a covert rule switch than incorrect timing. Qualitatively, animals’ pattern of subjective rule switches was consistent with all these predictions. For example, the probability of choosing a new rule was low in the first trial after an objective rule switch and increased monotonically afterwards (Fig. 1G, turquoise). Similarly, the probability of subjective rule switches ($\text{Pr}(\text{Sw})$) was small after rewarded trials (Fig. 1H; Rw: green), increased after one error (Fig. 1H; 1B-Er: filled red), and further increased after two consecutive errors (Fig. 1H; 2B-Er: open red). Moreover, $\text{Pr}(\text{Sw})$ after both one and two errors seemed to increase systematically with t_d (Fig. 1H). To assess these observations quantitatively, we used logistic regression to measure the effect of trial difficulty and number of consecutive errors on $\text{Pr}(\text{Sw})$. The regression coefficient associated with both trial difficulty (indexed by $|t_d|$) and the number of consecutive errors (indexed by $n\text{B-Er}$) were positive (monkey K: $\beta_{|td|} = 5.78 \pm 0.37$, $p < 10^{-16}$, $\beta_{n\text{B-Er}} = 0.91 \pm 0.07$, $p < 10^{-16}$; monkey I: $\beta_{|td|} = 1.73 \pm 0.26$, $p < 10^{-8}$, $\beta_{n\text{B-Er}} = 1.17 \pm 0.09$, $p < 10^{-8}$). These characteristics were present for both saccade directions (Fig. S1), both rules, and both response types (Fig. S2). Therefore, we concluded that animals (1) updated their belief about covert rule switches based on their trial-by-trial confidence in their interval judgments and (2) accumulated evidence across consecutive errors.

These findings demonstrate that monkeys, like humans (3, 4, 16, 41–47), are capable of adopting a sophisticated causal inference strategy in a hierarchical decision task.

A computational model for hierarchical reasoning

We tested various classes of models to infer the relevant latent variables that guide the animals' behavior (Fig. S3,4). In one model, we assumed that animals implemented a probabilistic switching behavior, which is a generalization of the well-known win-stay lose-switch strategy (48). According to this model, the agent switches the rule with fixed probabilities depending on trial outcome without regard to trial difficulty and/or the number of consecutive errors (Fig. S4A). Although variants of this model have successfully captured the behavior of monkeys and rodents in a number of simple tasks (49–52), it failed to explain the behavioral characteristics of monkeys during hierarchical reasoning (Fig. S4D) and was unable to reach the animals' level of performance (Fig. S4E). In another model, we assumed that monkeys learned the hazard rate of environmental switches and delayed their subjective switches accordingly (Fig. S4B). This model also failed to capture animals' switch behavior (Fig. S4D) and level of performance (Fig. S4E).

We also considered other models of hierarchical reasoning that were used to explain how humans decompose tasks into hierarchies (6), track higher-order parameters (i.e., hyperparameters) and environmental states (2, 4, 7, 53), and reduce failures by comparing new, old and counterfactual strategies (1). Although these models have successfully captured human behavior in a range of hierarchical decision-making tasks, they could not straightforwardly be adapted to our experiment as they were not intended to account for failures caused by misjudgments of unreliable stimuli, which is a central component of our task.

To develop a suitable model for our experiment, we first formulated the problem in terms of the behavior of an ideal observer, similar to a recent behavioral study in humans (3). The ideal observer computes the posterior probability of a rule switch by integrating information about (1) the expected accuracy of trial-by-trial interval judgments, and (2) the outcome of preceding trials (see Methods). We then created a simplified confidence-based model (CBM) that computes the evidence for or against a covert rule switch by a single graded latent variable, X_{Σ} , whose value as a function of task difficulty (indexed by t_d) and the number of consecutive errors was inferred from the ideal observer model (see Methods). After updating the value of X_{Σ} , the model sets a binary switch decision, $X_{y/n}$, to 'switch' or 'no switch' depending on whether X_{Σ} is larger or smaller than a threshold, θ (Fig. 1I). To capture both animal's behavior quantitatively, we had to augment the model with a perseveration factor. This factor enabled us to account for the deviation of each animals' behavior from the ideal observer model that computes switch evidence optimally (see Methods). The model captured the characteristic dependence of switch behavior on trial difficulty and number of consecutive errors for both animals (Fig. S4D, Fig. 1H; see Table S1 and S2 for model parameters). Further interrogation of the model using cross validation, parameter identification and *in-silico* lesioning indicated that the key parameters were both necessary and identifiable (Fig. S5). We also verified the predictive validity of the model by confirming that simulations of the model fitted to the subjective psychometric function (Fig. 1E) in the presence of experimentally imposed covert switches was able to capture the objective psychometric function without additional fitting or parameterization (Fig. 1F). Based on the success of this model and the failure of alternatives in capturing the animals' switch behavior (Fig. 1H), we hypothesized that the computational logic of the underlying neural circuitry could be understood in terms X_{Σ} and $X_{y/n}$.

Fig. 1. Hierarchical causal inference, behavioral task, performance and model. (A) Causal inference when the task involves two hierarchically organized decisions. The observer has to infer the stimulus-response contingency rules and judge the sensory evidence. Since both are uncertain, when a decision leads to an error (i.e., negative feedback), the observer has to decide whether to attribute the error to an incorrect decision about the rule or an incorrect judgement about the stimulus. (B) Task contingencies. The sample interval (t_s) varied between 530 and 1170 ms and was designated 'Short' or 'Long' depending on whether it was shorter or longer than 850 ms. For the first rule (C_1), the animal had to make a prosaccade (Pro) when $t_s < 850$ ms and an antisaccade (Anti) when $t_s > 850$ ms. For the second rule (C_2), the response contingencies were reversed. Response contingency rules were volatile and switched throughout the experiment (arrows). (C) Experimental conditions and trial structure. The order of events in every trial were as followed: (i) fixation: animal had to fixate a central spot; (ii) decision about the rule: animal had to report its belief about the underlying rule by making a saccade to one of two targets presented above and below the fixation spot corresponding to two rules (blue: C_1 ; red, C_2); (iii) re-fixation: choice targets were removed cueing the animal to immediately re-fixate; (iv-vi) t_s presentation: two flashes, one around the fixation point (iv), and one at left or right of fixation point (vi), separated by t_s were presented (v); (vii) response: depending on its belief about the rule and the duration of t_s , animal had to either make a prosaccade (Pro) toward the second flash or an antisaccade (Anti) away from it. Pro and Anti response by the animal followed by reward if both rule and interval discrimination were correct. The animal performed two variants of the experiment. In the 'Instructed' rule experiment, a colored cue around the fixation spot indicated the correct rule. This rule cue was only provided during the initial fixation before the rule decision was made. In the 'Inferred' rule experiment, no cue was provided regarding the rule and animal had to infer the rule from the pattern of errors and its own expected accuracy. As shown above the 'rule report' screen in panel b, rule switches occurred in a blocked fashion. In instructed-rule experiment, the rule was cued after each switch and intermittently during the block. In inferred-rule experiment, there was no external information about the rule (gray cue). Top right: Animal could be correct (green checkmark) or incorrect (red cross) about either the rule or the action (Pro/Anri). Reward was provided only if both were correct. (D) The proportion of Anti (Pr(Anti)) as a function of interval ($t_d = t_s - 850$ ms) for the instructed rule experiment for the two monkeys. Lines show model fits (see Methods). (E) Same as (D) for the inferred rule experiment plotted with respect to the rule reported by the animal (i.e., 'subjective'). (F) Same as (E) plotted with respect to the experimentally imposed rule (i.e., 'objective'). (G) Probability of choosing a new rule by the animal (P(new rule)) after an objective rule switch for the instructed and inferred rule experiments. (H) Probability of switching (Pr(Sw)) as a function of interval (t_d) for the instructed and inferred rule experiments. (I) Confidence-based model (CBM) diagram showing the decision process based on accumulated evidence $\hat{X}_\Sigma$.

(yellow) and inferred (turquoise). **(H)** Proportion of subject rule switches ($\text{Pr}(\text{Sw})$) as a function of t_d after reward (Rw: green), 1-back error (1B-Er: filled red circle), and 2-back error (2B-Er: open red circle) after a rewarded trial. Lines show model fits (see Methods). **(I)** Confidence-based switch model. This model updates the value of switch evidence (different levels of red), X_z , based on the outcome of the previous trials (Reward versus Error, denoted 'Rw/Er'), and the animal's belief about t_s ($\hat{t}_s$). After rewarded trials (green drop), X_z is reset to zero. After each error (red cross), X_z is incremented. We modeled X_z as a Gaussian distribution whose mean and standard deviation (μ_{X_z} and σ_{X_z}) as a function of t_s and the number of consecutive errors were set such that the model's behavior approximates the behavior of an ideal observer (see Methods). When X_z breaches a threshold (dashed, θ), the model switches the rule. Modeling X_z as a Gaussian distribution, the probability of switch at any trial can be related to the area under the distribution beyond the threshold (red region under the Gaussian). Panels D-H include error bars (SEM across sessions) but for most points, the bars are not visible as they are smaller than the symbol size.

Electrophysiology

Previous work has established a central role for the dorsomedial frontal cortex (DMFC) and dorsal anterior cingulate cortex (ACC) in monitoring and predicting outcomes (18–25), using outcomes to regulate actions (48, 54–58) and strategic decisions (5, 28–33). Therefore, we recorded neural activity in DMFC comprising supplementary eye field, dorsal supplementary motor area (i.e., excluding the medial bank), and pre-supplementary motor area, and ACC (stereotactic coordinates in Table S3) while animals performed the task. Since our main focus was to understand how the animal used decision outcomes to disambiguate errors, we focused our analyses of neural activity during the inter-trial-interval (ITI) after the trial outcome was revealed.

Previous work has established a central role for DMFC and ACC in monitoring and predicting outcomes (18–25), using outcomes to regulate actions (48, 54–58) and strategic decisions (5, 28–33). Therefore, we recorded neural activity in DMFC and ACC while animals performed the task. Since our main focus was to understand how the animal used decision outcomes to disambiguate errors, we focused our analyses of neural activity during the inter-trial-interval (ITI) after the trial outcome was revealed.

As a first step, we characterized individual neurons in terms of their sensitivity to trial outcome, trial difficulty and the number of consecutive errors (Fig. 2). In both areas a large proportion of neurons responded differently depending on trial outcome (DMFC: Fig. 2A–C; ACC: Fig. 2J–L). Many of the neurons that signaled error trials were differentially modulated depending on trial difficulty (i.e., magnitude of t_d). This difficulty-dependent modulation was present at the level of single neurons (Fig. 2D,E,M,N) and across the population (Fig. 2F,O) as evident from a comparison of firing rates associated with relatively easy ($|t_d| \geq 160$ ms) and difficult trials ($|t_d| < 160$ ms). Moreover, the firing rate of many error-modulated neurons was modulated depending on the number of preceding errors (Fig. 2G–I,P–R). The sensitivity of DMFC and ACC neurons to error, difficulty and trial history corroborates previous findings in a variety of sensorimotor, cognitive and economic decision making tasks (18, 59), and suggest that these areas might serve as the neural substrate for making causal inferences about errors in our hierarchical decision task.

Fig. 2. DMFC and ACC selectivity patterns after feedback. (A-B) Average firing rate of two example DMFC neurons in monkey K (A) and I (B) for rewarded (green) and unrewarded (red) trials relative to the time of feedback (dashed line). (C) Histogram of selectivity to trial outcome (reward versus no reward) across DMFC neurons. Selectivity was computed using ROC analysis based on spike counts within a 600 ms window after feedback (gray in panels A and B). Black corresponds to neurons with significant selectivity (259/624; permutation test, 100 times, $p < 0.05$). (D-E) Average firing rate of the same neurons in A and B sorted with respect to relatively 'easy' (dark: $|t_d| \geq 160$ ms) and 'difficult' (light: $|t_d| < 160$ ms) trials. (F) Histogram of selectivity to trial difficulty ('easy' versus 'difficult') across DMFC neurons using an ROC analysis similar to panel C. Black corresponds to neurons that were significantly modulated by trial difficulty (147/259; permutation test, 100 times, $p < 0.05$). (G-H) Average firing rate of the same neurons shown in A and B after one error (solid: 1B-Er) and two consecutive errors (dashed: 2B-Er) following a rewarded trial. (I) Histogram of selectivity to consecutive errors (1B-Er versus 2B-Er) across DMFC neurons using an ROC analysis similar to panel C. Black corresponds to neurons that were significantly modulated with respect to the number of consecutive errors (139/259; permutation test, 100 times, $p < 0.05$). (J-R) Same as (A-I) for ACC.

Integration of DMFC error-related signals in ACC

Rule inference during ITI depends on two sequential computations, a retrospective computation to evaluate the outcome of the previous trial (i.e., reward versus error), and a subsequent prospective computation to decide whether or not to switch the rule in the upcoming trial. Previous work has found that ACC encodes errors with longer latencies than DMFC (18), and suggested that it may act as a “storage buffer” (60) tracking task-relevant variables across trials (24, 28, 33). Following these suggestions, we analyzed the sensitivity of neurons in DMFC and ACC during the ITI to 1) error versus reward in the previous trial, and 2) switch versus no switch in the next trial. Our prediction was that DMFC would exhibit strong and early error-modulated activity but would be less sensitive to animal’s switch behavior in the next trial. For ACC, on the other hand, we expected neurons to show a longer latency modulation reflecting subsequent rule switches. To test these predictions, we used a simple linear regression model that aimed to explain modulations of spiking activity in each neuron in terms of two indicators, one specifying whether the animal made an error in the previous trial, and one specifying whether the animal switched the rule in the next trial (see Methods).

We found that both DMFC and ACC signals during ITI were modulated depending on both error in the previous trial and the animal’s switch behavior in the next trial (Fig. 3A,B). DMFC selectivity (i.e., regression slope, see Methods) averaged across neurons that were significantly modulated with respect to error (permutation test for regression slope for each neuron, $p < 0.01$) exhibited strong sensitivity early during ITI. The peak sensitivity occurred at a latency of 145 ms and was highly significant (peak sensitivity = 2.48, Wilcoxon rank sum test, $p < 0.001$). ACC was also strongly modulated with respect to trial outcome early during ITI (peak latency = 187 ms, peak sensitivity = 2.21, $p < 0.001$). Additionally, ACC exhibited a strong and late modulation that forecasted animal’s switch behavior in the subsequent trial (peak latency = 415 ms, peak sensitivity = 2.1, $p < 0.001$). DMFC was also modulated by the animal’s switch behavior but these modulations were significantly weaker than ACC (selectivity ratio of DMFC to ACC = 0.74, $p < 0.001$) and lacked the distinctive temporal structure evident in ACC that was indicative of a switch-related computation. Consistent with our hypothesis, these results suggest that DMFC signals errors, and ACC receives error signals and helps the animal decide whether and when to switch (Fig. S6 shows the same results for the two monkeys separately).

To validate this hypothesis, we performed an experiment to test whether the late responses in ACC were sensitive to early signals in DMFC. In a random 50% of error trials, we used electrical microstimulation to alter DMFC activity within a window of 50-150 ms in ITI while recording the ensuing spiking activity in ACC. Our choice of the time window of microstimulation was informed by our analysis of the temporal profile of sensitivity of DMFC to error versus reward (Fig. 3A), and the stimulation current was weak (see Methods) to avoid changes in animal’s overt behavior (e.g., evoking saccades). Consistent with our hypothesis, microstimulation of DMFC changed the post-stimulation spiking activity of single neurons in ACC (Fig. 3C). To summarize this effect across the population of ACC neurons, we used a modulation index to quantify the relative increase in firing rate after microstimulation. This analysis revealed that ACC neurons were strongly and significantly modulated as a result of DMFC microstimulation (Fig. 3D). Importantly, we verified that DMFC microstimulation had no significant effect on behavior ensuring that modulation of ACC activity was not due to an indirect change in switch probability (Fig. S7).

To ensure that the communication channel from DMFC to ACC played a task-dependent functional role and was not due to some non-specific microstimulation-induced cortico-cortical interactions, we additionally examined the effect of DMFC microstimulation on ACC in the instructed rule experiment using an identical protocol (same experimental sessions, same time window, same current levels, etc). The key difference in the instructed rule experiment compared to the inferred rule experiment was that rule switches were instructed and animals did not have to rely on errors in the previous trial to decide whether to switch. In other words, errors did not inform rule switches and therefore, there was no need to engage the communication channel from DMFC to ACC to integrate early error-related signals. Accordingly, we found that the same microstimulation protocol had no significant effect on ACC activity in the instructed rule experiment (Fig. 3D, Fig. S8 for the individual animals). These results provide evidence that ACC integrates error-related signals in DMFC selectively when there is need to evaluate errors with respect to a hierarchy of potential causes and decide whether to switch.

Fig. 3. Retrospective and prospective computations in DMFC and ACC during ITI. (A) Time course of DMFC selectivity to error in the previous trial and to rule switches in the next trial. The blue trace shows the time course of selectivity to error versus reward averaged across neurons with significant error-modulated activity. The brown trace shows the time course of selectivity to switch versus no switch averaged across DMFC neurons with significant switch-predictive activity. (B) Same as panel (A) for ACC neurons. (C) Modulation of spiking activity in example ACC neurons due to weak electrical microstimulation of DMFC. DMFC microstimulation early during ITI (shaded region: 50-150 ms) was followed by an increase in the firing rate of ACC neurons (red) compared to no stimulation (black). (D) Modulation of ACC activity due to DMFC microstimulation during the inferred and instructed rule experiments. For each ACC neuron, we computed a microstimulation modulation index quantifying the relative increase in firing rate as a result of microstimulation. Denoting the firing rate in the stimulated and non-stimulated trials by r_1 and r_2 (measured by a sliding window of 150 ms) respectively, the modulation index was computed as $(r_1 - r_2) / (r_1 + r_2)$. The averaged modulation index was then computed across neurons before and after the stimulation window separately for the inferred (green) and instructed (yellow) rule experiments. Stimulation sensitivity index was significantly increased within 150-300 ms relative to feedback in the inferred rule experiment but not for the instructed rule experiment (Wilcoxon rank sum test; $p < 0.01$).

ACC represents cumulative switch evidence and drives switching behavior

Both the task-selective effect of DMFC microstimulation on ACC (Fig. 3C,D) and the strong switch-predictive signals in ACC (Fig. 3B) motivated a hypothesis that ACC may be involved in computing the decision to switch based on evidence in support of a possible covert switch. For the hypothesis to hold, ACC must harbor both a graded representation of the evidence supporting a switch and a binary signal predicting animal's future switches. To test this hypothesis, we examined ACC signals in the context of our model of behavior. The model formalized animal's subjective switches in terms of a binary variable, $X_{y/n}$, whose state (switch or no switch) was set based on the value of cumulative switch evidence relative to a threshold (Fig. 1I). The cumulative switch evidence, in turn, was captured by a single graded latent variable, $\hat{X}_\Sigma$, whose value absorbed the dependence of behavior on both trial difficulty (e.g., larger for easier trials) and the number of consecutive errors.

We performed a regression analysis on individual ACC neurons to investigate whether $\hat{X}_\Sigma$ and $X_{y/n}$ were encoded by distinct subpopulations of neurons or were mixed across the population. Consistent with numerous recent studies in various higher cortical areas (35, 61–65), this analysis indicated that $\hat{X}_\Sigma$ and $X_{y/n}$ were mixed across the population (Fig. S9). Therefore, we used a recently-developed targeted dimensionality reduction technique (63) that allowed us to tease apart signals encoding $\hat{X}_\Sigma$ and $X_{y/n}$ independently across the population (see Methods).

We found a strong representation of $\hat{X}_\Sigma$ in ACC population activity (Fig. 4A, top). We quantified the sensitivity of ACC to $\hat{X}_\Sigma$ by measuring the slope of the regression line relating $\hat{X}_\Sigma$ to ACC population activity associated with $\hat{X}_\Sigma$ based on firing rates 200-400 ms after feedback (Fig. 4A, bottom). The regression slope was significantly positive for both animals independent of animal's binary choice about rule switch (monkey K: [0.809*, CI: 0.533,1.085] when $X_{y/n}=1$ and [1.098*, CI: 0.844,1.351] when $X_{y/n}=0$; monkey I: [0.532*, CI: 0.475,0.589] when $X_{y/n}=1$ and [0.884*, CI: 0.729,1.038] when $X_{y/n}=0$). This observation suggests that ACC either computed or received inputs that encode $\hat{X}_\Sigma$.

ACC also had a clear representation of $X_{y/n}$ as evidenced by the convergence of the population activity associated with $X_{y/n}$ on to one of two distinct states depending on whether the animal would switch in the next trial or not (Fig. 4B, top). We used linear regression to quantify the degree to which ACC responses 200-400 ms after feedback were modulated by $X_{y/n}$ (Fig. 4B, bottom). The baseline for the regression was significantly larger when the animal switched in the next trial (monkey K: [2.408*, CI: 1.977,2.840]; monkey I: [1.325*, CI: 0.993,1.657]) compared to when it did not (monkey K: [0.707*, CI: 0.514,0.900]; monkey I: [0.788*, CI: 0.627,0.950]).

Note that all the analyses relating ACC activity to $\hat{X}_\Sigma$ and $X_{y/n}$ were performed on cross-validated data and were not the result of an overfitted regression model. Moreover, the encoding of $\hat{X}_\Sigma$ and $X_{y/n}$ in ACC was robust to various nuisance parameters (Fig. S10) including saccade direction and response type (Pro versus

Anti). These results provided evidence in support of the hypothesis that ACC may be responsible for converting the cumulative switch evidence, $\hat{X}_{\Sigma}$, to the binary decision to switch, $X_{y/n}$ (Fig. S11 and S12 show the results of the same analysis for DMFC and ACC, respectively, separately for the two animals).

As a final test for our hypothesis, we used electrical microstimulation to perturb ACC during the window of 200-400 ms in which the sensitivity of the population activity to $\hat{X}_{\Sigma}$ and $X_{y/n}$ was evident (Fig. 4A,B). We reasoned that if animals relied on ACC to decide when to switch, perturbation of signals in ACC in this window would interfere with this computation and would result in a measurable difference in the observed switch probability. Consistent with this prediction, we found that switch probabilities increased significantly in the random subset of error trials after which ACC was electrically stimulated (Fig. 4C, Fig. S13). This experiment provided evidence that ACC played a functional role in making causal inferences about covert rule switches. Although the effect of microstimulation was strong and significant in both animals, in one animal the stimulation enhanced the sensitivity of the animal to preceding errors as evidenced by an increase in slope (Fig. 4C, bottom left), whereas in the other animal, the effect was an overall increase the switch probability evident as a change in baseline (Fig. 4C, bottom left).

To ensure that this effect was indeed related to ACC supporting the animal's attempt to infer covert rule switches, we additionally tested the effect of ACC microstimulation in the instructed rule experiment. Since switches in this control experiment were instructed and not based on cumulative switch evidence, we predicted that ACC perturbation should have minimal effect on the animal's switch probability. Results of this experiment supported our prediction. ACC microstimulation in the instructed rule task had nearly no effect on switch probability. We note that it is possible that animals made inferences about the rule in both tasks, but this inference was overridden in the instructed task by the reliable visual cue. The positive effect of microstimulation in the inferred rule task coupled with the negative effect in the instructed task further substantiated a central role for ACC in making causal inferences about errors in a hierarchical setting.

Fig. 4. Representation of switch evidence and causal manipulation of switch behavior in ACC. (A) Encoding of the graded switch evidence (X_z) by population activity in ACC. Using targeted dimensionality reduction, we identified the pattern of population activity in ACC that encoded X_z , inferred from animal's behavior. We then derived a sensitivity index (d') to quantify the distance between activity associated with different levels of X_z (Low to High shown in three colors) and activity associated with rewarded trials (i.e., $X_z=0$). The sensitivity is shown separately for trials that led to a switch (filled symbols) and those that did not (open symbols). The top panel shows sensitivity as a function of time, and the bottom panel shows the response tuning to X_z within a time window of 200-400 ms after feedback (gray region). Results for both panels correspond to cross-validated data (see Methods). (B) Encoding of the binary switch variable ($X_{y/n}$) by population activity in ACC. Applying the same analysis technique as in (A) revealed that ACC activity patterns that were sensitive to $X_{y/n}$ were organized in a binary fashion (i.e., switch versus no switch) regardless of the value of X_z (cross-validated; see Methods). (C) Top: The effect of ACC electrical microstimulation within the 200-400 ms relative to feedback on switch probability (Pr(Sw)) in the inferred and instructed task. Open circles represent individual sessions. Electrical stimulation was applied on ~50% of error trials (red: Stim) and the effect of Pr(Sw) was compared to the non-stimulated trials (black: No-Stim). In both animals, ACC microstimulation led to an increased Pr(Sw) in the inferred task (Wilcoxon rank sum test, *: $p<0.05$, **: $p<0.01$). In contrast, switches in the instructed task did not increase due to ACC microstimulation (n.s.; $p>0.05$). For the instructed task, Pr(Sw) is only shown for trials where the subjective switch was not cued. Bottom: The effect of ACC stimulation in the inferred rule experiment on Pr(Sw) as a function of X_z of 1B-Er trials (Wilcoxon rank sum test, *: $p<0.05$, **: $p<0.01$).

Discussion

Our results provide an understanding of the computational principles and neurobiological underpinnings of adaptive decision making when the agent/brain has to reason about errors in the presence of multiple hierarchically organized sources of uncertainty. In terms of behavior, we were able to establish that macaque monkeys are capable of implementing a causal inference strategy with a level of sophistication that is comparable to humans (3). Both monkeys treated each negative outcome as evidence for a covert rule switch, but did so rationally by taking into account the fact that (1) errors after easy trials (i.e., higher expectation of reward) were more likely due to a rule switch than a timing error, and (2) repeated errors provided stronger evidence for a rule switch. In terms of neurobiology, we found that the animals' behavior was supported by distributed and hierarchically organized neural circuits in the frontal cortex (45, 66). Specifically, we found that DMFC and ACC, two areas that have been implicated in error monitoring (18, 26, 67–73) and the control of adaptive behavior (5, 28–36), carried signals relevant for causal inference. ACC however, seemed to function downstream of DMFC and played a direct role in integrating evidence and computing whether and when to attribute failures to a covert rule switch.

Both DMFC and ACC were modulated retrospectively by the outcome of the low-level decisions, which is consistent with previous findings suggesting a role for these areas in performance monitoring (18–21, 23, 25). In other words, both areas had a representation of switch evidence. However, the hierarchical nature of our task allowed us to uncover a distinct and longer latency form of modulation in ACC that was predictive of the animal's subsequent switch behavior. This result coupled with an experiment involving simultaneous DMFC perturbation and ACC recording enabled us to reveal the sequential nature of computations in DMFC and ACC, and provided a coherent explanation of their distinct functional link to behavior, their different activation latencies, and their circuit-level interactions. In particular, our results suggest that the long latency signals in ACC result from an integration of early error-related modulations in DMFC, much like what is seen in well-established first-order perceptual decision making tasks between sensory and association/premotor areas (74–81). An important outstanding question that our study motivates is the problem of maintaining a desired behavioral strategy between switches, which may involve other brain areas such as the dorsolateral prefrontal cortex (85–91).

ACC is modulated by reward history (23, 92) and reward expectations (24) and is linked to reward-dependent control of adaptive behavior in many situations including movement selection (58, 93), decision making under risk (33), foraging (28), and regulation of exploration versus exploitation (94). Our work extends the role of ACC in adaptive control to the general class of cognitive tasks that demand hierarchical reasoning about errors. As we found in our work, and others in other behavioral settings (3, 16, 95), reasoning about errors can be captured by a model comprised of two key computational variables: 1) a latent switch evidence variable that accumulates confidence-dependent error information, and 2) a latent threshold that controls a binary decision to switch depending on the strength of the cumulative switch evidence. We found that ACC harbors distinct activity patterns associated with both the cumulative switch evidence and the binary decision to switch that were mixed at the level of single neurons but dissociated across the population. Moreover, we found that perturbation of ACC interfered with the animal's rationality during causal reasoning about errors. Together, these results suggest that ACC uses graded evidence derived from errors in low-level processes in a decision hierarchy to select between longer-term behavioral strategies associated with higher levels of the hierarchy.

Materials and Methods

All experimental procedures conformed to the guidelines of the National Institutes of Health and were approved by the Committee of Animal Care at the Massachusetts Institute of Technology. Experiments involved two awake, behaving monkeys (species: *M. mulatta*; ID: K and I; weight: 7.5 and 10.5 kg; age: 4 and 5 yrs old, respectively). Animals were head-restrained and seated comfortably in a dark and quiet room, and viewed stimuli on a Acer H236HL LCD monitor (23 inch; refresh rate: 60 Hz, resolution: 1920x1080). All reported stimulus presentation times have to be rounded to a multiple of the frame duration (16.67 ms). Eye movements were registered by an infrared camera and sampled at 1kHz (Eyelink 1000, SR Research Ltd, Ontario, Canada). The MWorks software package (<http://mworks-project.org>) was used to present stimuli and to register eye position. A photodiode was used for registering the timing of events during stimulus presentation.

Neurophysiology recordings were made by 24- or 36-channel laminar probes with 100- μ m interelectrode spacing (V-probe, Plexon Inc.) through a bio-compatible cranial implant whose position was determined based on stereotaxic coordinates and structural MRI scan of the two animals. Extracellular signals were band-pass filtered (300 Hz to 6 kHz) and digitized (sampling rate: 30 kHz, resolution: 16-bit) using a digital Intan headstage (Intan Technologies; <http://intantech.com/>), and collected the data using the OpenEphys system (<http://www.open-ephys.org/>). Single or multi-unit action potential waveforms were detected and sorted offline using MKsort (<https://github.com/ripple-neuro/mksort>). Analysis of both behavioral and spiking data was performed using custom MATLAB code (Mathworks, MA).

Electrophysiology. We recorded neural activity in (1) the dorsomedial frontal cortex (DMFC) comprising supplementary eye field, dorsal supplementary motor area (i.e., excluding the medial bank), and pre-supplementary motor area, and (2) the anterior cingulate cortex (ACC). Recording sites and number of sessions/trials are reported in Table S3.

ACC electrical microstimulation. On a random 50% of error trials, we stimulated ACC in a 200-400 ms time window after the saccade reached to target. We generated biphasic current pulses (amplitude: 40-60 μ m peak-to-peak; pulse duration: 300 μ s, frequency: 200 Hz, total stimulation duration: 200 ms) using a commercial stimulator (CereStim 96, Blackrock Microsystems) and injected current using a custom bipolar system with two tungsten microelectrodes (10-50 k Ω , Microprobes), one serving as a current source and another as a dedicated return. To cover the region of interest in ACC, we placed the current source just below the border of gray-white matter in ACC margin and the return electrode, ~3 mm above and ~2 mm lateral to the source electrode. The number of stimulation sessions/trials are reported in Table S1. To have an appropriate control for these stimulation experiments, both the inferred and instructed experiments were tested on every stimulation session.

Simultaneous electrical microstimulation of DMFC and recording in ACC. On a random 50% of error trials, we stimulated DMFC in a 50-150 ms time window after the saccade reached to target, and recorded the effect of stimulation on ACC afterwards. We injected current (amplitude: 20-30 μ m peak-to-peak; pulse duration: 300 μ s, frequency: 200 Hz, total stimulation duration: 100 ms) through a tungsten electrode placed ~2-mm deep just below the border of gray-white matter in DMFC and used a guide tube piercing the dura as

the return, while recording simultaneously with a separate V-probe. The stimulation current was weak to make sure the behavior would not change. This was particularly important because stimulation of DMFC could evoke saccades (e.g., around supplementary eye fields), and such explicit change in behavior would complicate the interpretation of any observed changes in ACC activity. The number of stimulation sessions/trials are reported in Table S3. To have an appropriate control for these stimulation experiments, both the inferred and instructed experiments were tested on every stimulation session.

REFERENCES AND NOTES

1. M. Donoso, A. G. E. Collins, E. Koechlin, Human cognition. Foundations of human reasoning in the prefrontal cortex. *Science*. **344**, 1481–1486 (2014).
2. C. M. Glaze, A. L. S. Filipowicz, J. W. Kable, V. Balasubramanian, J. I. Gold, A bias–variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*. **2**, 213–224 (2018).
3. B. A. Purcell, R. Kiani, Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4531–40 (2016).
4. A. Zylberberg, D. M. Wolpert, M. N. Shadlen, Counterfactual Reasoning Underlies the Learning of Priors in Decision Making. *Neuron*. **99**, 1083–1097.e6 (2018).
5. N. Kolling, J. Scholl, A. Chekroud, H. A. Trier, M. F. S. Rushworth, Prospection, Perseverance, and Insight in Sequential Behavior. *Neuron*. **99**, 1069–1082.e7 (2018).
6. Q. J. M. Huys *et al.*, Interplay of approximate planning strategies. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3098–3103 (2015).
7. C. Summerfield, T. E. Behrens, E. Koechlin, Perceptual classification in a rapidly changing environment. *Neuron*. **71**, 725–736 (2011).
8. R. Kiani, M. N. Shadlen, Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*. **324**, 759–764 (2009).
9. A. Kepecs, N. Uchida, H. A. Zariwala, Z. F. Mainen, Neural correlates, computation and behavioural impact of decision confidence. *Nature*. **455**, 227–231 (2008).
10. S. M. Fleming, H. C. Lau, How to measure metacognition. *Front. Hum. Neurosci.* **8**, 443 (2014).
11. P. G. Middlebrooks, M. A. Sommer, Neuronal correlates of metacognition in primate frontal cortex. *Neuron*. **75**, 517–530 (2012).
12. T. J. Pleskac, J. R. Busemeyer, Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol. Rev.* **117**, 864–901 (2010).
13. R. Ratcliff, J. J. Starns, Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychol. Rev.* **120**, 697–719 (2013).
14. J. Drugowitsch, R. Moreno-Bote, A. Pouget, Relation between belief and performance in perceptual decision making. *PLoS One*. **9**, e96511 (2014).
15. Y. Komura, A. Nikkuni, N. Hirashima, T. Uetake, A. Miyamoto, Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat. Neurosci.* **16**, 749–755 (2013).
16. R. van den Berg, A. Zylberberg, R. Kiani, M. N. Shadlen, D. M. Wolpert, Confidence Is the Bridge between Multi-stage Decisions. *Curr. Biol.* **26**, 3157–3168 (2016).
17. A. Lak *et al.*, Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron*. **84**, 190–201 (2014).
18. S. Ito, V. Stuphorn, J. W. Brown, J. D. Schall, Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science*. **302**, 120–122 (2003).
19. J. D. Schall, V. Stuphorn, J. W. Brown, Monitoring and control of action by the frontal lobes. *Neuron*.

- 36, 309–322 (2002).
20. J. W. Brown, T. S. Braver, Learned predictions of error likelihood in the anterior cingulate cortex. *Science*. **307**, 1118–1121 (2005).
 21. C. Amiez, J.-P. Joseph, E. Procyk, Anterior cingulate error-related activity is modulated by predicted reward. *Eur. J. Neurosci.* **21**, 3447–3452 (2005).
 22. A. W. MacDonald 3rd, J. D. Cohen, V. A. Stenger, C. S. Carter, Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*. **288**, 1835–1838 (2000).
 23. H. Seo, D. Lee, Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27**, 8366–8377 (2007).
 24. M. Shidara, B. J. Richmond, Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science*. **296**, 1709–1711 (2002).
 25. M. Schlag-Rey, N. Amador, H. Sanchez, J. Schlag, Antisaccade performance predicted by neuronal activity in the supplementary eye field. *Nature*. **390**, 398–401 (1997).
 26. F. Bonini *et al.*, Action monitoring and medial frontal cortex: leading role of supplementary motor area. *Science*. **343**, 888–891 (2014).
 27. S. W. Kennerley, T. E. J. Behrens, J. D. Wallis, Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat. Neurosci.* **14**, 1581–1589 (2011).
 28. B. Y. Hayden, J. M. Pearson, M. L. Platt, Neuronal basis of sequential foraging decisions in a patchy environment. *Nat. Neurosci.* **14**, 933–939 (2011).
 29. M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter, J. D. Cohen, Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624–652 (2001).
 30. J. G. Kerns *et al.*, Anterior cingulate conflict monitoring and adjustments in control. *Science*. **303**, 1023–1026 (2004).
 31. A. Shenhav, M. M. Botvinick, J. D. Cohen, The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*. **79**, 217–240 (2013).
 32. K. Johnston, H. M. Levin, M. J. Koval, S. Everling, Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron*. **53**, 453–462 (2007).
 33. S. W. Kennerley, M. E. Walton, T. E. J. Behrens, M. J. Buckley, M. F. S. Rushworth, Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* **9**, 940–947 (2006).
 34. J. Wang, D. Narain, E. A. Hosseini, M. Jazayeri, Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* **21**, 102–110 (2018).
 35. E. D. Remington, D. Narain, E. A. Hosseini, M. Jazayeri, Flexible Sensorimotor Computations through Rapid Reconfiguration of Cortical Dynamics. *Neuron*. **98**, 1005–1019.e5 (2018).
 36. X. Lu, M. Matsuzawa, O. Hikosaka, A neural correlate of oculomotor sequences in supplementary eye field. *Neuron*. **34**, 317–325 (2002).
 37. J. D. Murray *et al.*, A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661–1663 (2014).

38. T. E. J. Behrens, M. W. Woolrich, M. E. Walton, M. F. S. Rushworth, Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
39. S. Everling, M. C. Dorris, D. P. Munoz, Reflex suppression in the anti-saccade task is dependent on prestimulus neural processes. *J. Neurophysiol.* **80**, 1584–1589 (1998).
40. N. Amador, M. Schlag-Rey, J. Schlag, Primate antisaccades. I. Behavioral characteristics. *J. Neurophysiol.* **80**, 1775–1786 (1998).
41. M. M. Botvinick, Y. Niv, A. C. Barto, Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition.* **113**, 262–280 (2009).
42. D. Badre, Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends Cogn. Sci.* **12**, 193–200 (2008).
43. M. M. Botvinick, Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* **12**, 201–208 (2008).
44. D. W. Schneider, G. D. Logan, Hierarchical control of cognitive processes: switching tasks in sequences. *J. Exp. Psychol. Gen.* **135**, 623–640 (2006).
45. E. Koechlin, C. Ody, F. Kouneiher, The architecture of cognitive control in the human prefrontal cortex. *Science.* **302**, 1181–1185 (2003).
46. T. D. Kim, M. Kabir, J. I. Gold, Coupled Decision Processes Update and Maintain Saccadic Priors in a Dynamic Environment. *J. Neurosci.* **37**, 3632–3645 (2017).
47. A. Shenhav, J. D. Cohen, M. M. Botvinick, Dorsal anterior cingulate cortex and the value of control. *Nat. Neurosci.* **19**, 1286 (2016).
48. C. H. Donahue, H. Seo, D. Lee, Cortical signals for rewarded actions and strategic exploration. *Neuron.* **80**, 223–234 (2013).
49. K. Iigaya, M. S. Fonseca, M. Murakami, Z. F. Mainen, P. Dayan, An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nat. Commun.* **9**, 2477 (2018).
50. D. J. Barraclough, M. L. Conroy, D. Lee, Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci.* **7**, 404–410 (2004).
51. C. H. Donahue, D. Lee, Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nat. Neurosci.* **18**, 295–301 (2015).
52. B. Massi, C. H. Donahue, D. Lee, Volatility Facilitates Value Updating in the Prefrontal Cortex. *Neuron.* **99**, 598–608.e4 (2018).
53. M. R. Nassar, R. C. Wilson, B. Heasly, J. I. Gold, An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci.* **30**, 12366–12378 (2010).
54. K. A. Hadland, M. F. S. Rushworth, D. Gaffan, R. E. Passingham, The anterior cingulate and reward-guided selection of actions. *J. Neurophysiol.* **89**, 1161–1164 (2003).
55. W. H. Alexander, J. W. Brown, Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* **14**, 1338 (2011).
56. B. Y. Hayden, M. L. Platt, Neurons in anterior cingulate cortex multiplex information about reward

- and action. *J. Neurosci.* **30**, 3339–3346 (2010).
57. N. S. Narayanan, M. Laubach, Top-down control of motor cortex ensembles by dorsomedial prefrontal cortex. *Neuron*. **52**, 921–931 (2006).
 58. K. Shima, J. Tanji, Role for cingulate motor area cells in voluntary movement selection based on reward. *Science*. **282**, 1335–1338 (1998).
 59. V. Stuphorn, T. L. Taylor, J. D. Schall, Performance monitoring by the supplementary eye field. *Nature*. **408**, 857–860 (2000).
 60. S. R. Heilbronner, B. Y. Hayden, Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annu. Rev. Neurosci.* **39**, 149–170 (2016).
 61. M. Rigotti *et al.*, The importance of mixed selectivity in complex cognitive tasks. *Nature*. **497**, 585–590 (2013).
 62. M. L. R. Meister, J. A. Hennig, A. C. Huk, Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. *J. Neurosci.* **33**, 2254–2267 (2013).
 63. V. Mante, D. Sussillo, K. V. Shenoy, W. T. Newsome, Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. **503**, 78–84 (2013).
 64. M. T. Kaufman, M. M. Churchland, S. I. Ryu, K. V. Shenoy, Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* **17**, 440–448 (2014).
 65. D. Raposo, M. T. Kaufman, A. K. Churchland, A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* **17**, 1784–1792 (2014).
 66. L. T. Hunt, B. Y. Hayden, A distributed, hierarchical and recurrent framework for reward-based choice. *Nat. Rev. Neurosci.* **18**, 172–182 (2017).
 67. C. B. Holroyd, M. G. H. Coles, The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* **109**, 679–709 (2002).
 68. V. Stuphorn, J. D. Schall, Executive control of countermanding saccades by the supplementary eye field. *Nat. Neurosci.* **9**, 925–931 (2006).
 69. N. S. Narayanan, J. F. Cavanagh, M. J. Frank, M. Laubach, Common medial frontal mechanisms of adaptive control in humans and rodents. *Nat. Neurosci.* **16**, 1888–1895 (2013).
 70. H. Gemba, K. Sasaki, V. B. Brooks, “Error” potentials in limbic cortex (anterior cingulate area 24) of monkeys during motor learning. *Neurosci. Lett.* **70**, 223–227 (1986).
 71. H. Niki, M. Watanabe, Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Res.* **171**, 213–224 (1979).
 72. C. Shen *et al.*, Anterior Cingulate Cortex Cells Identify Process-Specific Errors of Attentional Control Prior to Transient Prefrontal-Cingulate Inhibition. *Cereb. Cortex*. **25**, 2213–2228 (2015).
 73. M. Ullsperger, D. Y. von Cramon, Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *Neuroimage*. **14**, 1387–1401 (2001).
 74. M. N. Shadlen, W. T. Newsome, Motion perception: seeing and deciding. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 628–633 (1996).

75. J. D. Roitman, M. N. Shadlen, Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* **22**, 9475–9489 (2002).
76. A. Hernandez *et al.*, Decoding a Perceptual Decision Process across Cortex. *Neuron.* **66**, 300–314 (2010).
77. A. K. Churchland, R. Kiani, M. N. Shadlen, Decision-making with multiple alternatives. *Nat. Neurosci.* **11**, 693–702 (2008).
78. G. D. Horwitz, W. T. Newsome, Separate signals for target selection and movement specification in the superior colliculus. *Science.* **284**, 1158–1161 (1999).
79. T. D. Hanks *et al.*, Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature.* **520**, 220–223 (2015).
80. P. Znamenskiy, A. M. Zador, Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature.* **497**, 482–485 (2013).
81. C.-T. Law, J. I. Gold, Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.* **11**, 505–513 (2008).
82. E. Seidemann, E. Zohary, W. T. Newsome, Temporal gating of neural signals during performance of a visual discrimination task. *Nature.* **394**, 72–75 (1998).
83. R. J. Krauzlis, L. P. Lovejoy, A. Zénon, Superior colliculus and visual spatial attention. *Annu. Rev. Neurosci.* **36**, 165–182 (2013).
84. J. I. Gold, M. N. Shadlen, The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J. Neurosci.* **23**, 632–651 (2003).
85. M. K. Mian *et al.*, Encoding of rules by neurons in the human dorsolateral prefrontal cortex. *Cereb. Cortex.* **24**, 807–816 (2014).
86. N. P. Rougier, D. C. Noelle, T. S. Braver, J. D. Cohen, R. C. O'Reilly, Prefrontal cortex and flexible cognitive control: rules without symbols. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7338–7343 (2005).
87. M. F. S. Rushworth, K. A. Hadland, D. Gaffan, R. E. Passingham, The effect of cingulate cortex lesions on task switching and working memory. *J. Cogn. Neurosci.* **15**, 338–353 (2003).
88. E. K. Miller, J. D. Cohen, An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
89. R. E. Passingham, *The Frontal Lobes and Voluntary Action*. Oxford Psychology Series (1993).
90. W. F. Asaad, G. Rainer, E. K. Miller, Task-specific neural activity in the primate prefrontal cortex. *J. Neurophysiol.* **84**, 451–459 (2000).
91. N. Zarr, J. W. Brown, Hierarchical error representation in medial prefrontal cortex. *Neuroimage.* **124**, 238–247 (2016).
92. C. Amiez, J. P. Joseph, E. Procyk, Reward encoding in the monkey anterior cingulate cortex. *Cereb. Cortex.* **16**, 1040–1055 (2006).
93. Z. M. Williams, G. Bush, S. L. Rauch, G. R. Cosgrove, E. N. Eskandar, Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat. Neurosci.* **7**, 1370–1375 (2004).

94. R. Quilodran, M. Rothé, E. Procyk, Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron*. **57**, 314–325 (2008).
95. C. M. Glaze, J. W. Kable, J. I. Gold, Normative evidence accumulation in unpredictable environments. *Elife*. **4** (2015), doi:10.7554/eLife.08825.
96. C. Malapani, S. Fairhurst, Scalar Timing in Animals and Humans. *Learn. Motiv.* **33**, 156–176 (2002).
97. J. Gibbon, Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* **84**, 279 (1977).

ACKNOWLEDGMENTS

We wish to thank Rossana Chung and Tiffany Vincenza Parks for help with animal care, and Robert Desimone, Sabbi Lall, Seth W. Egger, Hansem Sohn, Nicolas Meirhaeghe, Jing Wang and Evan Remington for helpful comments on the manuscript. **Funding:** M.J. was supported by the Sloan Foundation, the Klingenstein Foundation, the Simons Foundation, the McKnight Foundation, and the McGovern Institute. M.S. was supported by the Jack Hilibrand Fellowship, the Henry E. Singleton fellowship and a fellowship from the Friends of the McGovern Institute. **Author Contributions:** M.S. and M.J. conceived the study and designed the experiments. M.S. trained the animals, collected the data, and analyzed the data. M.J. supervised the project. Both authors were involved in writing the manuscript. **Competing Interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary material.

SUPPLEMENTARY MATERIALS

Materials and Methods

Figs. S1 to S14

Tables S1 to S3

References (96-97)

Supplementary Materials for
Hierarchical reasoning by neural circuits in the frontal cortex

Morteza Sarafyazd, Mehrdad Jazayeri

correspondence to: mjaz@mit.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S14
Tables S1 to S3

Materials and Methods

Inferred rule experiment

A trial was initiated when the animal fixated a gray central fixation spot (diameter = 0.5 deg). After a random delay (400-800 ms, uniform hazard), a blue and a red visual target (diameter = 1 deg) were presented above and below the fixation point (distance = 7 deg). The animal had to report its belief about the rule by making a saccade to one of the targets within 1000 ms of their appearance (blue for C_1 and red for C_2). After the rule report epoch, the animal had to shift its gaze back to the fixation point to initiate the interval discrimination epoch (trial was aborted if re-fixation took longer than 500 ms). The trial proceeded regardless of whether the rule was reported correctly. After another random delay (400-900 ms, uniform hazard), two 50-ms flashes (diameter = 0.75 deg) were presented sequentially, a flash at the center of the screen followed by a flash to the left or right of the fixation point (distance = 9 deg). These two flashes demarcated the sample interval, t_s , which was drawn randomly from a 9-point discrete uniform distribution ranging between 530 and 1170 ms. In less than 700 ms after the second flash, the animal had to make a prosaccade toward the second flash or an antisaccade away from it to report whether its estimate of t_s was shorter or longer than a criterion of 850 ms, which corresponds to the mean of t_s distribution. We limited the reaction time to 700 ms to motivate the monkey to control the preparatory prepotent response inhibition dynamically during the interval measurement epoch. At the end of each trial, reward was provided (with a green visual feedback at the target location) if both the rule report in the first epoch and the interval discrimination in the second epoch were correct. No visual feedback or reward was given in error trials. When t_s was equal to the criterion (850 ms) in which case there was no correct answer for the interval discrimination, the animal got reward on a random 50% of the trials if it reported the rule correctly. The trial was aborted if the animal broke fixation before the second flash.

Instructed rule experiment

This task was identical to the rule inference experiment task with the only difference that the rule was cued explicitly and therefore the animal did not need to infer the rule based on history of errors. To specify the rule, we presented a red or blue disc around the fixation spot during the first 3-5 trials after every rule switch and 50-80% of trials until the next rule switch. The color of the disc served as an explicit cue for the current rule. To make sure the structure of the experiments was identical, in the instructed rule experiment, similar to the inferred rule experiment, animals had to first report the rule. On trials that the disc was not presented, the animal could infer it from memory of the previous trial since all rule switches were explicitly cued. The colored disc around the fixation spot disappeared immediately after the animal made a saccade to report the rule so that the interval discrimination epoch of the two tasks were identical.

Model for behavior

We developed a model with two hierarchically organized stages to capture behavior. The model is illustrated schematically in Fig. S3 and S4. The first stage captures the psychometric function describing the probability of Anti ($\text{Pr}(\text{Anti})$) within a trial as a function of rule and t_d in that trial (Fig. S3). The second stage captures the probability of

switch ($\text{Pr}(\text{Sw})$) across trials as a function of feedback, t_d and rule in the preceding trials (Fig. S4). 1 and 2 refer to the parameters associated with the first and second stages, respectively. We use simple letter names for either experimentally controlled variables (e.g., t_s) or observable behavioral variables (e.g., action: A), and letter names with superscript (^) for behaviorally unobservable variables (e.g., internal belief about rule, $\hat{C}$, or internal estimate of discriminant interval, $\hat{t}_d$) and intended actions (e.g., intended pro- or antisaccade, $\hat{A}$).

Modeling the first level of the decision hierarchy

We assumed that the observer makes a noisy measure of t_s ($\hat{t}_s$). Consistent with scalar variability of timing (96, 97), we modeled the noise by a zero-mean Gaussian whose standard deviation scales with t_s with the constant of proportionality w_m , as follows:

$$p(\hat{t}_s|t_s, w_m) \sim \mathcal{N}(t_s, w_m t_s)$$

The observer compares $\hat{t}_s$ to an internal criterion, $\hat{t}_c$ ($\hat{t}_d = \hat{t}_s - \hat{t}_c$), to categorize the interval as “Short” or “Long”. We treated $\hat{t}_c$ as a parameter that was fixed throughout each behavioral session but was allowed to change across sessions. Moreover, we included a lapse rate, Γ_t to accommodate trials in which the animal categorized t_s randomly (50% “Short” and 50% “Long”). With this formulation, the probability of an intended antisaccade can be written as follows:

$$p(\hat{A} = \text{Anti}|\hat{C}, t_s) = \frac{\Gamma_t}{2} + (1 - \Gamma_t) \int_{\hat{t}_s} p(\hat{A} = \text{Anti}|\hat{C}, \hat{t}_s, \hat{t}_c) p(\hat{t}_s|t_s, w_m) d\hat{t}_s$$

where the first term in the integrant is determined by task contingencies as follows:

$$p(\hat{A} = \text{Anti}|\hat{C}, \hat{t}_s, \hat{t}_c) = \begin{cases} 1, & \text{if } \hat{C} = C_1 \text{ and } \hat{t}_s > \hat{t}_c \\ 0, & \text{if } \hat{C} = C_1 \text{ and } \hat{t}_s < \hat{t}_c \\ 0, & \text{if } \hat{C} = C_2 \text{ and } \hat{t}_s > \hat{t}_c \\ 1, & \text{if } \hat{C} = C_2 \text{ and } \hat{t}_s < \hat{t}_c \end{cases}$$

We assumed that the observer uses $\hat{A}$ to generate a Pro or Anti action (A). In decision making studies, typically there is no need to distinguish between intended and executed actions. In our work, this distinction is warranted because animals occasionally fail to execute an intended antisaccade (39) when the inhibition to suppress the competing prepotent prosaccade is not sufficiently strong (25, 39, 40). To account for this asymmetry, the model was augmented with a parameter, Γ_{Anti} , to account for the small percentage of trials in which an intended antisaccade resulted in an erroneous prosaccade. The full model thus used 4 parameters ($\Psi_1 = \{w_m, \hat{t}_c, \Gamma_t, \Gamma_{\text{Anti}}\}$) to capture the probability of antisaccades as a function of t_s and C :

$$p(A = \text{Anti}|\hat{C}, t_s) = (1 - \Gamma_{\text{Anti}}) \left[\frac{\Gamma_t}{2} + (1 - \Gamma_t) \int_{\hat{t}_s} p(\hat{A} = \text{Anti}|\hat{C}, \hat{t}_s, \hat{t}_c) p(\hat{t}_s|t_s, w_m) d\hat{t}_s \right]$$

Finally, we assumed that actions across trials were conditionally independent and used the sum of the logarithm of individual conditional probabilities to derive the maximum likelihood estimate (MLE) of the model parameters:

$$\Psi_1(\text{MLE}) = \arg \max_{\Psi} \left[\sum_{i=1}^{N_{\text{Anti}}} \log p(A^i = \text{Anti}|\hat{C}^i, \hat{t}_s^i, \Psi) + \sum_{i=1}^{N_{\text{Pro}}} \log(1 - p(A^i = \text{Anti}|\hat{C}^i, \hat{t}_s^i, \Psi)) \right]$$

where indices i and j count over N_{Anti} antisaccade and N_{Pro} prosaccade trials, respectively. We fit this model to behavioral data in individual sessions of both the inferred rule (subjective rule) and instructed rule experiments.

Modeling the second level of the decision hierarchy

As a first step, we developed an ideal observer model (IOM) whose switch behavior on trial k is governed by posterior odds of rule change compared to no rule change (3), denoted Q_{IOM}^k . If we index the last rewarded trials by $k_0 - 1$, Q_{IOM}^k has to be computed based on the consecutive error trials $k_0, \dots, k$ (denote $[k_0:k]$) following the last rewarded trials:

$$Q_{IOM}^k = \frac{Pr(C^{k+1} \neq \hat{C}^k | t_s^{[k_0:k]}, \hat{C}^{[k_0:k]})}{Pr(C^{k+1} = \hat{C}^k | t_s^{[k_0:k]}, \hat{C}^{[k_0:k]})}$$

The denominator is the probability of no rule change despite negative feedback across all $[k_0:k]$ trials. For every trial i , the probability of being incorrect without a rule change is the product of the probability that no objective rule change occurred multiplied by the probability that the interval was discriminated incorrectly. If we denote the probability of correctly discriminating t_s in trial i by y^i , and the probability of objective covert rule change in trial i by its hazard function, denoted λ^i , the probability of an incorrect discrimination in the absence of rule change becomes $(1 - \lambda^i)(1 - y^i)$. The denominator therefore is the product of this probability across all $[k_0:k]$ trials. Using a similar logic, the numerator is the sum of multiple probabilities including the probability that the rule change occurred on trial k_0 , that it occurred on trial $k_0 + 1$ but not on trial k_0 , and the all the other terms until the last terms, which corresponds to a rule change on trial k with no rule change in the preceding $[k_0:k-1]$ trials. Accordingly, Q_{IOM}^k can be written as follows:

$$Q_{IOM}^k = \frac{\lambda^{k_0} + \sum_{i=k_0+1}^k \lambda^i \prod_{j=k_0}^{i-1} (1 - \lambda^j)(1 - y^j)}{\prod_{i=k_0}^k (1 - \lambda^i)(1 - y^i)}$$

Since y^i is a measure of expected accuracy (i.e., the probability of being correct), we quantified it based on our model of the subjective psychometric function after marginalizing over t_s (which the ideal observer does not know). For λ^i , to avoid over-parameterizing the model, we did not attempt to model the animal's subjective hazard

function and instead, assumed a constant hazard rate consistent with our experimentally imposed block transition scheme. With y^i and λ^i specified, we can fully derive Q_{IOM}^k with no additional free parameters.

Due to its history-dependent nature, direct use of IOM to characterize animal's behavior is not feasible (i.e., would need far more data that we can collect). As an approximation, we developed a simpler confidence-based model (CBM). The behavior of CBM is governed by a graded latent variable representing trial-by-trial evidence for a covert switch, which we denote by X_Σ . In this model, X_Σ is incremented after each error with a magnitude that depends on the previous trial's expected accuracy, and leads to a switch when X_Σ breaches a threshold (θ). To match CBM as closely as possible to IOM, we assumed that X_Σ on each trial has a Gaussian distribution. We set the mean and standard deviation (denoted by μ_{X_Σ} and σ_{X_Σ}) of X_Σ such that Q_{CBM}^k (probability that X_Σ is above θ over the probability it is below θ) was equal to Q_{IOM}^k for 1B-Er. Moreover, we assumed that switch evidence is subject to diffusive noise across trials; i.e., that variance grows linearly with the number of error trials (i.e., $\sigma_{X_\Sigma}^2$ for N errors = $N \sigma_{X_\Sigma}^2$ for 1 error). We also considered a nonlinear relationship between variance and number of trials (Fig. S5D-F) but the simple diffusive (linear) model was able to capture animals' switch behavior more accurately.

$$p(X_\Sigma) \sim \mathcal{N}(\mu_{X_\Sigma}, \sigma_{X_\Sigma})$$

$$Q_{CBM}^k = \frac{P(X_\Sigma \geq \theta)}{P(X_\Sigma < \theta)} = \frac{\int_{\theta}^{\infty} \mathcal{N}(\mu_{X_\Sigma}, \sigma_{X_\Sigma}) dX_\Sigma}{1 - \int_{\theta}^{\infty} \mathcal{N}(\mu_{X_\Sigma}, \sigma_{X_\Sigma}) dX_\Sigma} = Q_{IOM}^k$$

Rearranging this equation, we can write:

$$\int_{-\infty}^{\theta} \mathcal{N}(\mu_{X_\Sigma}, \sigma_{X_\Sigma}) dX_\Sigma = 1 - \frac{Q_{IOM}^k}{1 + Q_{IOM}^k}$$

which can be solved analytically and numerically to derive μ_{X_Σ} in terms of Q_{IOM}^k and σ_{X_Σ} . This formulation enabled CBM to closely approximate IOM. To fit CBM to animal's behavior, we made two simplifying assumptions: $X_\Sigma = 0$ after a rewarded trial, and $\theta = 1$. These are inconsequential because their effect on behavior can be fully subsumed by μ_{X_Σ} and σ_{X_Σ} .

To account for the deviation of each animal's behavior from the ideal observer model, we augmented μ_{X_Σ} by a less-than-one multiplicative 'perseverance factor', α . This factor reduces the magnitude of increments in X_Σ and causes an overall drop in the probability of switch. The full model thus has three parameters, μ_{X_Σ} , σ_{X_Σ} , and α . However because of the relationship between μ_{X_Σ} , σ_{X_Σ} , and the known Q_{IOM}^k , μ_{X_Σ} is not a free parameter. Therefore, our model for rule inference had only two free parameters ($\Psi_2 = \{\sigma_{X_\Sigma}, \alpha\}$). In this formulation, the probability of switch ($X_{y/n}=1$) can be simplified as follows:

$$p(X_{y/n} = \text{Sw} | \Psi_2) = \int_1^{\inf} \mathcal{N}(\alpha \mu_{X_\Sigma}, \sigma_{X_\Sigma}) dX_\Sigma$$

Finally, we assumed that switch decisions were conditionally independent across trials and used the sum of the logarithm of individual conditional probabilities to derive the maximum likelihood estimate (MLE) of the model parameters. The model fits were based on data in the 1B-Er and 2B-Er trials.

$$\Psi_2(\text{MLE}) = \arg \max_{\Psi} \left[\sum_{i=1}^{N_{\text{Sw}}} \log p(X_{y/n} = \text{Sw} | \Psi) + \sum_{j=1}^{N_{\text{No Sw}}} \log(1 - p(X_{y/n} = \text{Sw} | \Psi)) \right]$$

We performed several additional analyses to validate our assumptions about CBM and to verify that it was a suitable model: (1) *In-silico* model lesioning: we assessed the necessity of σ_{X_Σ} and α by testing the capacity of the model to capture the observed switch behavior when either of those parameters were fixed. Results in Fig. S5G-L show two examples, one in which σ_{X_Σ} was set to 0.5, and α was fitted, and another when α was set to 1 and σ_{X_Σ} was fitted. However, our conclusions were general and applied to other fixed values of σ_{X_Σ} and α so long as the fixed values were not identical to those recovered by the full model. (2) Model identification: we applied the model fitting procedure to simulated data generated by CBM to ensure that parameters could be recovered. Results are shown in Fig. S5C. (3) 2-fold cross-validation: to ensure that the success of the model was not due to overfitting, we repeatedly fit the model to a random half of each dataset and test the model on the remaining half. Results are shown in Fig. S5B. (4) Predictive validity: we simulated the model fit to the subjective psychometric function in the presence of experimentally imposed covert rule switches, and asked whether it could predict the observed objective psychometric function without additional parameterization and/or fitting. Results are shown in Fig. 1F.

Model free analysis of switch behavior

We also evaluated whether switch probability depended on trial difficulty (indexed by t_d) and the number of consecutive trials using a simple logistic regression:

$$P(X_{y/n}^{i+1} = \text{Sw} | t_d^i, n\text{B-Er}) = \frac{1}{1 + e^{-(\beta_0 + \beta_{|t_d|} |t_d^i| + \beta_{n\text{B-Er}} [n\text{B-Er}])}} \quad \text{Equation 1}$$

where $P(X_{y/n}^{i+1})$ represents the probability of switch in trial $i+1$, t_d^i , the discriminant interval in trial i , and $n\text{B-Er}$, the number of consecutive error before trial $i+1$. The parameters of the logistic model (β_0 , $\beta_{|t_d|}$, $\beta_{n\text{B-Er}}$) capture the baseline, dependence on $|t_d^i|$, and dependence on $n\text{B-Er}$ of $P(X_{y/n}^{i+1})$, respectively. We used MLE to fit the model to behavior, and used fits to $\beta_{|t_d|}$ and $\beta_{n\text{B-Er}}$ to assess the statistical significance of the effect of $|t_d^i|$ and $n\text{B-Er}$ on switch probability, respectively.

Analysis of single neurons

We estimated firing rates of each neuron per condition by computing average spike counts in 25 ms time bins followed by convolution with a Gaussian boxcar filter ($\sigma=45$

ms). For each neuron, we compared average firing rates during the inter-trial-interval (ITI) after rewarded and unrewarded trials (Fig. 2). We also compared responses for both rewarded and unrewarded trials further broken down according to whether the preceding trial was relatively “easy” or “difficult. We defined “easy” and “difficult” as trial in which $|t_d| \geq 160$ ms and $|t_d| < 160$ ms, respectively. Collapsing the full range of t_d to a small number of categories was necessary to overcome the limited number of error trials per sessions, but the results were qualitatively the same regardless of the criterion for defining easy versus difficult trials (we also did all the analyses for $|t_d| \geq 80$ ms and $|t_d| < 80$ ms). We also compared average firing rates as a function of number of preceding consecutive errors either across all values of t_d or after further dividing trials of the last error according to trial difficulty ($|t_d| \geq 160$ ms versus $|t_d| < 160$ ms).

Retrospective and prospective information in DMFC and ACC

We used a time-dependent linear regression model to examine the time course of the representations of trial outcome (Rw/Er) and switch behavior ($X_{y/n}$) across individual neurons in DMFC and ACC. We z-scored the firing rates of each neuron across all trials and times ($r(t)$), and formed a regression model that aimed to explain the firing rates as a function of trial outcome, $X_{y/n}$, and other experimental and behavioral variables as follows:

$$r(t) = \beta_{outcome} I_{Rw/Er} + \beta_{Sw} I_{y/n} + \beta_{Target} I_{left/right} + \beta_A I_{Pro/Anti} + \beta_{td} t_d + \beta_{Rule} I_{C1/C2} + \beta_0$$

Equation 2

In this regression, $I_{Rw/Er}$ was set to 1 (Er) or 0 (Rw), $I_{y/n}$ was set to 0 (no switch) or 1 (switch), $I_{left/right}$, to -1 or 1 depending on the saccade direction, $I_{Pro/Anti}$, to -1 or 1 depending on the animal’s action, $I_{C1/C2}$ to -1 or 1 depending on rule, and t_d was the experimentally imposed variable. Solving this equation yielded $\beta_{outcome}$ and β_{Sw} as a function of time for each neurons. Since our interest was to evaluate the relative strength of sensitivity across neurons and areas, we converted $\beta_{outcome}$ to a selectivity index by 1) taking the absolute value of each $\beta_{outcome}$ and 2) normalizing all $\beta_{outcome}$ at each time point by the average $\beta_{outcome}$ across all neurons, areas, and time during ITI. We did the same for β_{Sw} .

$$SI_{outcome} = \frac{|\beta_{outcome}|}{\langle |\beta_{outcome}| \rangle_{\text{neurons, areas}}}$$

$$SI_{Sw} = \frac{|\beta_{Sw}|}{\langle |\beta_{Sw}| \rangle_{\text{neurons, areas}}}$$

Finally, we averaged the selectivity index associated with $\beta_{outcome}$ and β_{Sw} across significantly sensitive neurons in each area and compared the time course of these two selectivity indices across the two areas (Fig. 3).

Sensitivity of population activity in DMFC and ACC to X_Σ and $X_{y/n}$

We employed a targeted dimensionality reduction (63) to assess the encoding of information about X_Σ and $X_{y/n}$ across the population in DMFC and ACC. This approach

has been carefully described before (63). Here, we briefly describe the procedure. First, we z-scored firing rates of each neuron across all trials and times, and formed a regression model that aimed to explain the entire vector of firing rates across the population (Z) at each point in time as a function of various experimental and behavioral variables as follows:

$$Z = \beta_{outcome} I_{Rw/Er} + \beta_{X\Sigma} X_{\Sigma} + \beta_{Sw} I_{y/n} + \beta_{Target} I_{left/right} + \beta_A I_{Pro/Anti} + \beta_{td} t_d + \beta_{Rule} I_{C1/C2} + \beta_0$$

Equation 3

The regressors are the same as described for Equation 2, with the addition of X_{Σ} which was inferred on a trial-by-trial basis from the fit of the confidence-based model to behavior. Second, we solved the regression model to derive the regression coefficients. Third, we denoised the coefficients by finding their projection in the subspace spanned by the first 10 principal components (PCs) of data that explained ~70% (or more) of the total variance in both areas and animals. To compute the PCs, we applied PCA to data resorted to a matrix of dimensions $[N_{units}, N_C \times T]$ (N_{units} : number of neurons, N_C : number of conditions, T : number of time bins). Fourth, for each regressor, we found the time at which the coefficient has the largest norm (i.e., the encoding of the corresponding variable of interest was strongest). Fifth, we orthogonalized the coefficients. To do so, we applied QR-decomposition to a matrix whose columns were the coefficients, and used the columns of the matrix Q as the new orthogonalized coefficients.

Next, we used the orthogonalized coefficients, $\beta_{X\Sigma}$ and β_{Sw} , to examine how DMFC and ACC encoded X_{Σ} and $X_{y/n}$ at the population level. We first describe the procedure of $\beta_{X\Sigma}$: 1) we found the projection of data onto $\beta_{X\Sigma}$, which we refer to as $Z_{X\Sigma}$; 2) we divided $Z_{X\Sigma}$ into two groups, one containing all the trials that did not lead to a switch and another for all the trials that did lead to a switch; 3) we divided trials in each group to 6 overlapping bins depending on the value of X_{Σ} (i.e., low X_{Σ} , intermediate X_{Σ} , and high X_{Σ}). The bin and overlapping window were chosen so that the number of trials in each bin were nearly equal. 4) we used d' as a distance metric to quantify the separation between the distributions of $Z_{X\Sigma}$ in each bin to the distribution of $Z_{X\Sigma}$ associated with rewarded trials (i.e., when $X_{\Sigma}=0$). This distance metric was used to assess the sensitivity of DMFC and ACC to X_{Σ} . To examine the statistics, the population analysis method was cross-validated by resampling dataset into equally-sized training and testing sets (bootstrapping method, 100-times). The procedure for β_{Sw} was identical except that we used the projections of data onto β_{Sw} (instead of $\beta_{X\Sigma}$), which we refer to as Z_{Sw} . Analysis of Z_{Sw} with respect to X_{Σ} enables us to assess the extent to which DMFC and/or ACC encoded $X_{y/n}$ as a binary variable (i.e., independent from X_{Σ}). In addition, as a control analysis, we conditioned the same subsets of trials on anti/pro-saccades, and separately on left/right saccades, and the projections into $\beta_{X\Sigma}$ were measured to quantify the action dependency of X_{Σ} neural state (Fig. S10).

Fig. S1. Switch behavior of the two monkeys as a function of saccade location. (A-D) Results are shown with the same ordering and format as in Fig. 1E-H. Behavior is largely independent of the direction of saccade. There were no significant difference between corresponding points on each plot (point-by-point Wilcoxon rank sum test; $p > 0.05$).

Fig. S2. Switch behavior of the two monkeys as a function of rule (C_1 versus C_2) and action (Pro versus Anti). Results are shown with the same ordering and format as in Fig. 1G-H. Behavior is largely independent of the direction of saccade. There were no significant difference between corresponding points on each plot (point-by-point Wilcoxon rank sum test, $p > 0.05$).

Time interval discrimination (Action within a trial)

Fig. S3. Description of the observer model of time interval discrimination. The observer model makes a noisy measurement of the sample interval, t_s , denoted $\hat{t}_s$. The measurement noise has a zero-mean Gaussian distribution and its standard deviation (σ) scales with t_s with constant of proportionality, w_m ($\sigma = w_m t_s$). The model determines the intended action, $\hat{A}$ based on rule $\hat{C}$ ($\hat{C} = C_1$ or $\hat{C} = C_2$) and $\hat{t}_s$ relative to an internal criterion ($\hat{t}_c$). When $\hat{C} = C_1$ and $\hat{t}_s > \hat{t}_c$, $\hat{A}$ is an antisaccade (Anti), and when $\hat{C} = C_1$ and $\hat{t}_s < \hat{t}_c$, $\hat{A}$ is a prosaccade (Pro). For $\hat{C} = C_2$, these response contingencies are reversed. The model also allows for a percentage of ‘lapse’ trials (Γ_t), in which $\hat{A}$ is randomly assigned to Pro or Anti (shown as $\hat{t}_s?$). The model’s action, A usually follows from $\hat{A}$, but it also allows for a percentage of additional lapses parametrized by Γ_{Anti} in which $\hat{A} = \text{Pro}$ despite $\hat{A} = \text{Anti}$ due to a failure to inhibit the prepotent prosaccade response. In sum, the model has 4 parameters ($\Psi_1 = \{w_m, \hat{t}_c, \Gamma_t, \Gamma_{\text{Anti}}\}$) to capture the probability of antisaccades as a function of t_s and $\hat{C}$. The two psychometric functions on the right are examples of the model’s behavior for the two rules shown in terms of proportion of antisaccades ($\text{Pr}(\text{Anti})$) as a function of the discriminant interval, t_d .

Fig. S4. Alternative models for switching behavior. (A) Probabilistic switch model. This model implements a variant of the win-stay lose-switch strategy in which the animal switches the rule with different fixed probabilities (p and q) depending on trial outcome (Rw: rewarded trial, Er: error trial). (B) Delayed switch model. This model switches after an error only after a certain number of trial since the last successful switch. The model emulates an observer that exploits knowledge about the block-structure in the experiment. (C) Confidence-based switch model. This model computes a graded variable, $\hat{X}_\Sigma$, whose value is determined by the outcome (Rw/Er) and difficulty (indexed by $\hat{t}_s$) of preceding trials. $\hat{X}_\Sigma$ is reset to zero after each trial with positive outcome, and is incremented by a sample of Gaussian distribution after error trials. The mean and standard deviation of the distribution ($\mu_{X\Sigma}$ and $\sigma_{X\Sigma}$) are set so that the model's behavior approximates the behavior of an ideal observer (see Methods). When the updated $\hat{X}_\Sigma$ breaches a threshold, θ , the model switches its belief about the rule ($\hat{C}$). The model also uses a perseveration factor (α) to account for the deviation of each animal's behavior from optimality (not shown in C but see Methods). Since $\mu_{X\Sigma}$ was determined by the ideal observer, the behavior of this stage of the model was fully explained by two parameters ($\Psi_2 = \{\sigma_{X\Sigma}, \alpha\}$). (D) Switch probability ($\text{Pr}(\text{sw})$) based on simulations of three models. Each model was constructed such that it matched as closely as possible to each animal's subjective psychometric function. The data from the probabilistic switch model was generated assuming no switch after rewarded trials and 50% switch after error trials irrespective of confidence. The data for the delay switch model was generated by assuming that no switch for the first 10 trials (consistent with the experimentally imposed hazard) after each successful switch, and 50% switch for each error trial afterwards. The data for the confidence-based switch model are the same as those shown in the main manuscript. For each model, the results are shown for 1-back error (1B-Er; solid) and 2-back error (2B-Er; dashed) conditions. (E) Percentage error as a function of the discriminant interval, t_d , for each of the three models with the same specifications as in D. The Confidence-based switch model has the lowest probability of error in comparison to the the other two models.

Fig. S5. Cross-validation, identification and lesion studies of the confidence-based belief update model. (A) Full CBM model with two parameters α and σ_{X_Σ} . (B) Validation of the full CBM model. Fit of CBM (line) to switch behavior of two monkeys (symbols) in the same format as Fig. 1H. The model fits were cross-validated and repeated 10 times for random train and validation set. (C) Model identifiability. Fits of CBM to data generated from CBM with random choices of α and σ_{X_Σ} . Inset: Difference between fitted and actual value across simulations. There is not significant difference between the fit and true values ($p > 0.05$). (D) Altered CBM in which the standard deviation (instead of variance) of noise increases with consecutive trials. The altered component (the one that differs from the full CBM) is highlighted by the shaded box. (E) Same as panel B for alternate model in D. (F) Comparison of the full and alternate CBM by using RMSE across 10 cross-validated fits to behavioral data (error bar: SEM). (G) Altered CBM in which α is set to 1 (no perseverance) while σ_{X_Σ} is fitted to behavior. (H,I) Same as E,F for the altered CBM in G. (J-L) Same as panels (G-I) for an altered CBM in which σ_{X_Σ} is set to 0.5 (instead of being inferred from the ideal observer) while α is fit.

Fig. S6. Retrospective and prospective computations in DMFC and ACC during ITI, separately for the two animals. Results are shown in the same format as in Fig. 3A,B, separately for the two animals.

Fig. S7. No effect of DMFC microstimulation on switch probability. The ordinate shows average switch probability ($\text{Pr}(\text{Sw})$) across all sample intervals after 1-back error trials separately for trials when DMFC was and was not stimulated. There was no significant change in $\text{Pr}(\text{Sw})$ in the inferred or instructed experiments as a result of DMFC microstimulation (n.s.: $p > 0.05$). For the instructed experiment, $\text{Pr}(\text{Sw})$ is only shown for trials where the subjective switch was not cued.

Fig. S8. Modulation of ACC activity due to DMFC microstimulation shown for the two animals separately. Results are shown in the same format as Fig. 3D. We examine the statistical significance of the stimulation modulation index in the inferred and instructed experiments independently. In each case, we compared the distribution of firing rates in the stimulated trials to a null distribution created by random labeling of trials as stimulated and unstimulated. We used a shuffling procedure to avoid making arbitrary assumptions about the sampling distributions. In the inferred experiment, DMFC microstimulation significantly increased the stimulation modulation index in ACC within 150-300 ms relative to feedback (Monkey K: mean=0.091, $p<0.001$; Monkey I: mean=0.102, $p<0.01$). In contrast, there was no significant change in ACC activity in the same window after DMFC microstimulation in the instructed experiment (Wilcoxon rank sum test, $p>0.05$). Since the results for the instructed experiment were not significant, we did not perform a direct statistical comparison of the inferred and instructed experiments. Although the results for both animals were statistically significant, we wish to highlight the difference between the two animals as this difference may provide stronger support for our hypothesis. The animal for which the effect of stimulation in the instructed task was stronger (“monkey I”), was less flexible at alternating between the inferred and instructed tasks. This is evident from a careful examination of the behavior. In monkey I, who occasionally did not switch following an instructed switch (compare yellow traces in Fig. 1G), DMFC stimulation influenced ACC during the instructed experiment. In contrast, in monkey K who followed the rule instruction nearly perfectly (Fig. 1G), the DMFC-stimulation had nearly no effect on ACC in the instructed experiment. In other words, the stimulation effects, when considered in relation to each animal’s behavior, seem to suggest that when rule inference via feedback is at play, DMFC stimulation has a more prominent effect on ACC, which is consistent with our hypothesis.

Fig. S9. Mixed representation of $\hat{X}_{\Sigma}$ and $X_{y/n}$ at the level of individual neurons in ACC. The selectivity of each neuron to switch evidence and switch choice was measured using linear regression (see Methods). Colored dots represent neurons with significant selectivity (permutation test, 100 times, $p < 0.01$) to $\hat{X}_{\Sigma}$ (red) and $X_{y/n}$ (blue) or both (green).

Fig. S10. Sensitivity of ACC to switch evidence was robust to nuisance parameters. (A) After fitting the regression model to ACC (Fig. 4), we separated trials based on saccade direction (left versus right) and response type (prosaccade versus antisaccade) to characterize the extent to which the encoding of switch evidence (X_{Σ}) was dependent on these nuisance variables. Results indicated that representation of X_{Σ} in ACC was present for both saccade directions and both response types. All regression slopes were significantly larger than zero ($p < 0.01$). (B) Same as panel (A) for the switch choice selectivity ($X_{y/n}$). Results indicated that representation of switch choice in ACC was present for both saccade directions and both response types ($p < 0.01$).

Fig. S11. Representations of switch evidence and switch choice in DMFC. Results are shown in the same format as in Fig. 4A,B, separately for the two animals. For the bottom panel, the slope of the regression line and the corresponding confidence intervals are reported. Regression slopes that were significantly different from zero are in bold.

Fig. S12. Representations of switch evidence and switch choice in ACC. Results are shown in the same format as in Fig. 4A,B, separately for the two animals. For the bottom panel, the slope of the regression line and the corresponding confidence intervals are reported (slope [confidence interval]). Regression slopes that were significantly different from zero are in bold. Note that in ACC, selectivity to switch is largely independent of X_z (bottom row, second and fourth panels from left), and that neural responses strongly predict animals' switch behavior in the next trial (Table below shows the baseline of the regression line as a function of $X_{y/n}$ for the two animals). These results support our hypothesis that the decision to switch (i.e., threshold crossing) is likely made in ACC.

	$X_{y/n}$ (Sw)	Sensitivity to $X_{y/n}$ b_0 [CI]
Monkey K	Yes No	2.408 , [1.977, 2.840] 0.707 , [0.514, 0.900]
Monkey I	Yes No	1.325 , [0.993, 1.657] 0.788 , [0.627, 0.950]

* CI: 95% Confidence interval. Significant estimates ($p < 0.01$) are in bold.

Fig. S13. The effect of ACC microstimulation on switch probability of the two animals separately. The effect of ACC microstimulation (red) compared to stimulation (black) is shown as a function of t_d in the preceding trial for monkey K (A) and monkey I (B). Stimulation was performed after a randomly selected 50% of 1-back error trials. The lines represent the fit of the confidence-based model (see Methods) to each animal for each condition.

Fig. S14. Coronal section containing ACC from the structural MRI scan of the two animals. Magenta shows the recording site in ACC (AG: anterior to the genu of arcuate sulcus).

Parameter	Inferred rule experiment		Instructed rule experiment	
	Monkey K	Monkey I	Monkey K	Monkey I
w_m	0.0581 ± 0.0042	0.1024 ± 0.0107	0.0596 ± 0.0043	0.1083 ± 0.0093
$\hat{t}_c$ (ms)	850.6 ± 129.7 (C_1) 846.8 ± 129.1 (C_2)	853.3 ± 164.2 (C_1) 846.0 ± 162.8 (C_2)	848.4 ± 129.4 (C_1) 856.0 ± 130.5 (C_2)	856.6 ± 164.8 (C_1) 838.7 ± 161.4 (C_2)
Γ_{Anti}	0.0084 ± 0.0013	0.0108 ± 0.0021	0.0080 ± 0.0012	0.0095 ± 0.0018
Γ_t	0.0122 ± 0.0019	0.0150 ± 0.0029	0.0096 ± 0.0015	0.0240 ± 0.0046

Table S1. Model parameters for the psychometric functions ($M \pm \text{SEM}$). Since the responses were asymmetric (prosaccade versus antisaccade), we fit the $\hat{t}_c$ for the two rules (C_1 and C_2) separately. However, results indicated that there was not significant difference in $\hat{t}_c$ between the two rules ($p > 0.05$). There was no significant difference between fitted model parameters in inferred and instructed rule experiments ($p > 0.05$).

Parameters	Monkey K	Monkey I
α	$0.72^{\pm 0.03}$	$0.44^{\pm 0.07}$
$\sigma_{X\Sigma}$	$1.56^{\pm 0.30}$	$1.09^{\pm 0.11}$

Tabel. S2. Model parameters for switch probability (M $\pm$ SEM)

	Monkey K		Monkey I	
Area	DMFC	ACC	DMFC	ACC
Stereotactic coordinates	ML 2-6mm AG 0-6 mm	ML 2-5mm AG 0-8 mm	ML 2-6mm AG 0-7 mm	ML 2-5mm AG 0-8 mm
Recording sessions/trials	18/32148	16/27463	14/18008	17/24672
Stimulation sessions/trials	4/5738	4/6362	6/8685	8/10062

Table S3. DMFC/ACC electrophysiology. ML: mediolateral relative to midline; AG: Anterior-posterior relative to the genu of arcuate sulcus. DMFC stimulation pertains to experiments in which we simultaneously recorded from ACC. Fig. S14 shows a coronal section of the two animal's frontal cortex containing a sample recording site in ACC.