COGNITIVE NEUROSCIENCE

Dorsal raphe nucleus controls motivation-state transitions in monkeys

Luke Priestley[1]*, Mark Chiew[2]†‡, Mo Shahdloo[1], Ali Mahmoodi[1], Xinghao Cheng[3], Robin Cleveland[3], Matthew Rushworth[1]§, Nima Khalighinejad[1]§

The dorsal raphe nucleus (DRN) is an important source of serotonin in the brain, but fundamental aspects of its function remain elusive. Here, we present a combination of minimally invasive recording and disruption studies to show that DRN brings about changes in motivation states. We use recently developed methods for identifying temporal patterns in behavior to show that monkeys change their motivation depending on the availability of rewards in the environment. Distinctive patterns of DRN activity occur when monkeys transition between a high-motivation state occupied when rewards are abundant, to a low-motivation state engendered by reward scarcity. Disrupting DRN diminishes sensitivity to the reward environment and perturbs transitions in motivational states.

INTRODUCTION

Animals need rewards for their survival, and they need to obtain them as efficiently as possible. In many naturalistic scenarios, this means tracking general features of the surrounding environment: The foraging behavior of many species, for example, involves comparing the opportunity an animal is currently confronted with against the general richness and stochasticity of opportunities it has encountered in the recent past, which guide its expectations for the future (1, 2). These are rational considerations for animals given the biological constraints that encumber them. Finding and pursuing rewards consumes precious metabolic resources that must later be replenished, and so it is critical that animals organize their reward-seeking activities in ways that exploit their external milieu—to press their advantage when things are good and conserve their energy when things are bad.

How does the brain reconcile motivation for rewards with the environment in this way? Here, we argue for a critical role for the dorsal raphe nucleus (DRN)—a phylogenetically ancient part of the brainstem that is distinguished by its serotonergic innervation of the mammalian forebrain. Although fundamental aspects of DRN's function remain elusive, two major themes are discernible: (i) that DRN controls changes in an animal's behavior and (ii) that DRN responds to reward-related features of an animal's milieu, like the value, valence, and uncertainty of recent outcomes (3–10).

We build on this previous work in arguing that DRN controls transitions between motivational states that reconcile an animal's behavior with the distribution of rewards in the environment. We use a novel behavioral paradigm to demonstrate that rhesus monkeys are more motivated to pursue rewards that occur in rich environments with many high-value opportunities. We implement recent innovations in quantitative behavioral modeling to show that this pattern is explicable by state-like changes in an animal's motivation that match the current environment's reward distribution. We take advantage of the whole-brain perspective afforded by functional magnetic resonance imaging (fMRI) to show that brain activity in DRN—but no other neuromodulatory nucleus—covaries with transitions in motivation-states, specifically when monkeys transition between a high-motivation state occupied when rewards are abundant, to a low-motivation state engendered by reward scarcity. Such comparisons are critical for identifying the different functions of different neuromodulatory nuclei (11). Interpreting the causal significance of fMRI activation patterns is possible when the impact of disrupting the activity can be examined (12–14). Accordingly, we modulate neural activity with minimally invasive transcranial ultrasound stimulation (TUS) to show that DRN, but not a functionally related neuromodulatory nucleus—ventral tegmental area (VTA)—is causally involved in motivational-state transitions (15–17). In doing so, we provide the first demonstration that minimally invasive modulation of DRN is possible and a new perspective on DRN's behavioral function.

RESULTS

Animal behavior is modulated by the environment

Four rhesus monkeys (*Macaca mulatta*) performed a simple decision-making task involving sequential encounters with reward opportunities that varied in reward magnitude and reward probability (Fig. 1A). Upon each encounter, the animals could either pursue the opportunity and incur a short temporal cost or let the opportunity pass and proceed to the next encounter. Each session comprised four blocks of 40 to 50 trials. The mean reaction time (RT) after the go cue was 0.76 s (SD = 0.07 s). The mean premature response rate was 3% (SD = 1.8%) of trials per session. The animals performed one session per day at the same time of the day.

We systematically controlled the reward probability values for reward opportunities within each block to engender different reward environments. In other words, the distribution of a given reward-probability value varied between blocks. Blocks varied on two dimensions: (i) richness, defined by the mean reward probability of opportunities—i.e., the value of an average opportunity in a block, and (ii) stochasticity, defined as trial-to-trial variability in reward probability, and implemented by changing the width of reward-probability distributions (Fig. 1, B and C; see Materials and Methods

[1]Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, Tinsley Building, Oxford, UK. [2]Wellcome Centre for Integrative Neuroimaging, Oxford Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, University of Oxford, John Radcliffe Hospital, Oxford, UK. [3]Oxford institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Old Road Campus Research Building, Oxford, UK.
*Corresponding author. Email: luke.priestley@psy.ox.ac.uk
†Present address: Physical Sciences, Sunnybrook Research Institute, Toronto, Canada.
‡Present address: Medical Biophysics, University of Toronto, Toronto, Canada.
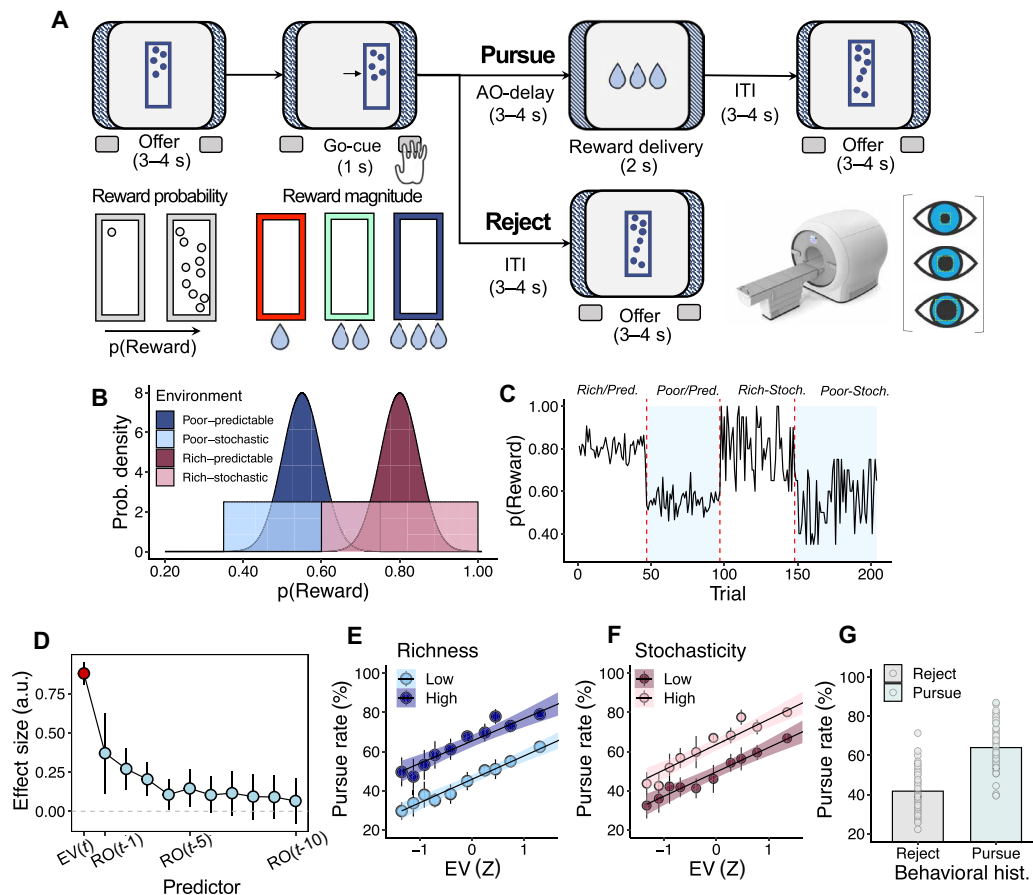§These authors contributed equally to this work.

Fig. 1. Animals are more likely to pursue reward opportunities in rich environments. (**A**) The behavioral task involved sequential encounters with reward opportunities that varied in reward magnitude (stimulus color) and reward probability (dots per stimulus). Animals performed the task while undergoing functional MRI and pupillometry. (**B**) The probability density of reward probability in different reward environments. Rich blocks [purple; mean(reward probability) = 0.80] had higher mean reward probability than poor blocks [blue; mean(reward probability) = 0.55]. Stochastic blocks [light colors; SD(reward probability) = 0.13] were uniformly distributed, and predictable blocks were normally distributed [dark colors; SD(reward-probability) = 0.05]. Block-type was cued with a visual stimulus bordering the screen (see Materials and Methods). (**C**) A depiction of changes in reward probability as a function of reward environment in an example session. Each session comprised four environments covering each permutation of richness and stochasticity. (**D**) Effect size of reward outcomes (RO) on past trials on pursue-versus-reject decision on the current trial. For comparison, the effect size of the expected value (EV) of the current trial is shown with a red point. Points and whiskers indicate effect size (i.e., regression weight) ± 95% confidence interval (CI) for each predictor. Effects for outcomes past $t - 5$ are nonsignificant at 0.05 level. (**E**) Animals were more likely to pursue opportunities as the richness of the environment increased. *x* axis indicates expected value of the reward opportunity on each trial. *y* axis indicates rate of reward pursuit. Color scale indicates mean-split according to richness of the environment, i.e., high is trials where richness$_t$ > μ$_{richness}$ and low is richness$_t$ ≤ μ$_{richness}$. Dots and whiskers indicate means (±SEM) of the pursue rate in deciles expected value over each animal in each session. (**F**) Animals were more likely to pursue opportunities as the stochasticity of the environment increased. *x* axis, *y* axis, dots, whiskers, and color scale follow conventions of (E). (**G**) Animals were likely to repeat pursue-versus-reject decisions over consecutive trials. Dots indicate mean level of responding per animal per session. a.u., arbitrary units.

for details). We refer to the reward probability associated with a specific reward opportunity as reward probability. We refer to blocks in which reward probabilities were higher, on average, as rich blocks, and to blocks in which they were lower on average as poor blocks. Similarly, we refer to blocks in which there was high variance in reward-probability values as stochastic blocks, and to blocks in which there was low variance as predictable blocks.

We first confirmed that animals were more likely to pursue reward opportunities as a function of increases in an opportunity's reward probability (fig. S1, A-i) and reward magnitude (fig. S1, A-ii). Then, we examined how animals changed their behavior in response to the environment. We reasoned that an animal's understanding of the environment would depend on its reward history. To determine

the time horizon of sensitivity to past rewards, we used a mixed-effects logistic regression with individual terms for reward outcomes on trials $t - 1$ to $t - 10$ (where t is the current trial; GLM1.1, see Materials and Methods). This suggested that behavior was modulated by rewards as distant as five trials into the past, but no further (Fig. 1D). We therefore operationalized the richness of the environment as the average reward accumulated in the previous five trials and the stochasticity of the environment as the SD of the reward rate over the same five trial period. Changing the time window used to operationalize environments, or operationalizing the environment using the blocks specified in the experimental design rather than reward history, did not affect the inferences drawn from any following analyses (see figs. S1, B and C, and S2).

We quantified the influence that the richness and stochasticity of the environment had on behavior with a series of mixed-effect binomial general linear models (GLM) that accounted for animal-specific variation in behavior. This revealed that animals were more likely to pursue opportunities as a function of increases in the richness and stochasticity of the environment (GLM1.2, see Materials and Methods; $\beta_{\text{environment-richness}} = 0.46$, SE = 0.08, $P < 0.001$; Fig. 1E; see also fig. S1, B to D; $\beta_{\text{environment-stochasticity}} = 0.08$, SE = 0.03, $P = 0.020$; Fig. 1F; see also fig. S1E). These effects were independent of trial-by-trial reward value—For example, opportunities with the same expectation value were more likely to be pursued in rich relative to poor environments, meaning that the effect was not due to the correlation between rich environments and high-value reward opportunities on a given trial (Fig. 1, E and F; see also fig. S1, B, D, and E).

The fact that the richness of the environment predicted whether animals would pursue reward opportunities indicated a subtle but important feature of behavior. Animals could only obtain rewards by pursuing reward opportunities, meaning that rich environments effectively corresponded to periods where many rewards had been pursued in the recent past. Insofar as receiving rewards in the past encouraged animals to pursue opportunities in the present, the richness of the environment effect therefore indicated that pursue-versus-reject decisions were autocorrelated over consecutive trials. We tested this hypothesis by predicting the pursue-versus-reject decision taken on each trial using the corresponding decision on the previous trial (behavioral history, hence). This confirmed that animals tended to repeat pursue-versus-reject decisions over consecutive trials, although each trial featured a separate reward opportunity with distinct magnitude and probability parameters (GLM1.3; $\beta_{\text{behavioral-history}} = 0.64$, SE = 0.24, $P < 0.001$; Fig. 1G).

A hidden Markov model identifies motivation states in behavior

We reasoned that autocorrelations in animal behavior might reflect changes in motivation states—that is, changes in an animal's intrinsic propensity to pursue rewards that are independent of the value of the opportunity currently at stake and that might persist over many trials (Fig. 2, A and B). It has recently been shown that latent time-varying dispositions in behavior can be uncovered using a general linear model hidden Markov model (GLM-HMM) approach (18, 19). We therefore tested the motivation-state hypothesis using a GLM-HMM.

GLM-HMMs are an extension of the classical GLM in which the parameter values of a GLM are allowed to change over time-varying latent states. Transitions between latent states are governed by Markovian dynamics and are thus called "HMM states." In simple terms, this means that if animals behave in fundamentally different ways—for example, if they adopt different motivation states—over the course of the task, the GLM-HMM can detect these states and specify how the relationship between the task and behavior changes in each state via the weights in the GLM part of the model. We formulated a GLM-HMM where the pursue-reject decision on each trial was predicted with a binomial GLM that included the following
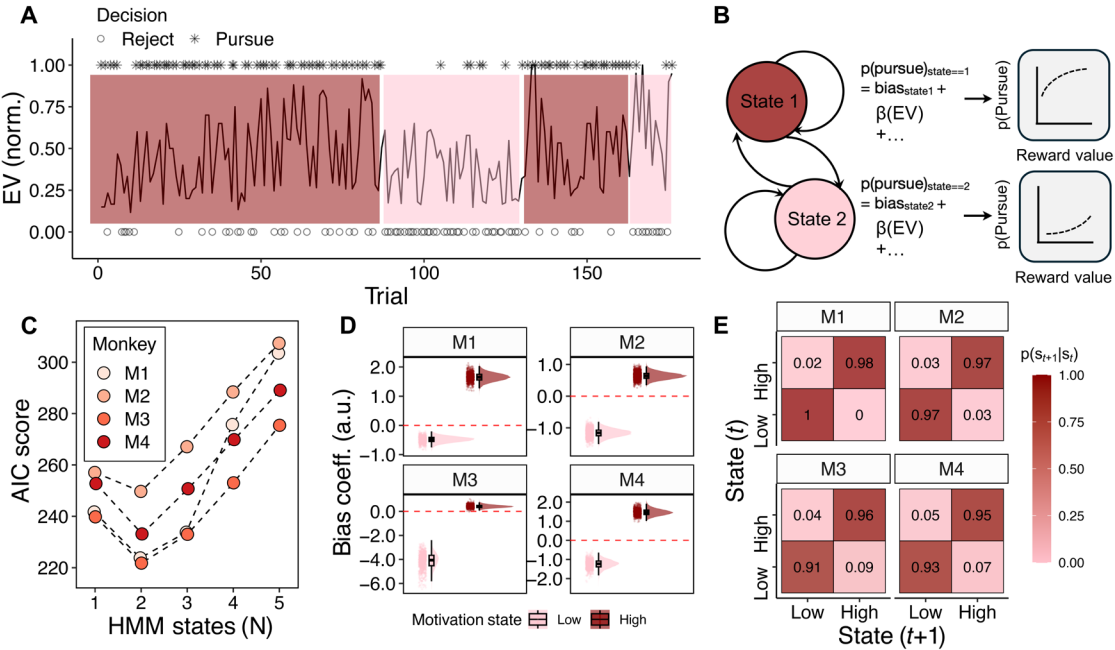

Fig. 2. A general linear model–hidden Markov model for animal behavior. (**A**) Pursue-versus-reject decisions (stars at the top indicate reward pursuits; dots at the bottom indicate reward rejections) as a function of expectation value (*y* axis) over time (*x* axis) in an example session. The animals showed prolonged biases to pursue (red) or reject (pink) opportunities over many consecutive trials. (**B**) We captured these patterns using a GLM-HMM that featured state-dependent bias parameters. These bias parameters produced state-dependent decision functions that reflected changes in the way animals made pursue-versus-reject decisions over time. (**C**) A two-state GLM–HMM improved fit relative to other models according to AIC—a penalized form of log-likelihood (see fig. S3 for full cross-validation results). Points show mean session-wise AIC for each animal. (**D**) The posterior probabilities of state-specific bias parameter values in each animal (M1–M4) suggested that HMM-states were associated with meaningful changes in an animal's bias to pursue-versus-reject rewards. (**E**) Transition matrices indicated that in each animal, HMM-states were strongly autocorrelated over consecutive trials (M1–M4).

terms: (i) a bias or "intercept" term, (ii) a predictor for the expectation value of the reward opportunity available on each trial, and (iii) predictors for external environment cues that were presented throughout the task (Fig. 2B and fig. S3; see Materials and Methods for details). Our hypothesis was that animals' experienced changes in their intrinsic predisposition to pursue-versus-reject reward opportunities, over and above any influence of extrinsic factors such as the expectation value of the current opportunity. We therefore let only the bias parameter change between HMM states and held the weights on other predictors constant (but see figs. S5 and S6 for alternative models in which additional parameters changed between HMM states). Incorporating predictors for task features in the GLM part of the model ensured that the behavioral changes identified by the GLM-HMM are specifically due to shifts in the animal's intrinsic propensity to pursue rewards, after taking into account external factors like the reward value on the current trial.

We used a previously reported model-selection procedure, which involved testing models with different numbers of HMM states using fivefold cross-validation (see fig. S3 for details) (18). We validated the procedure using a series of parameter recovery tests (see fig. S8). After confirming successful parameter recovery, we implemented it on behavioral data. This indicated that cross-validated log-likelihoods for two-state GLM-HMMs were higher than one-state GLM-HMMs for all animals (Fig. 2C; see fig. S3 for full cross-validation results) but that invoking additional states beyond this number was not warranted for any animal. We therefore used two-state GLM-HMMs for all further analyses.

Each animal-specific two-state GLM-HMM featured clearly distinct state-specific bias parameters (Fig. 2D) and profoundly autocorrelated HMM-state transition matrices, which were consistent with state-like fluctuations in motivation to pursue rewards (Fig. 2E). To further examine the validity of the model, we simulated data from a fitted two-state GLM-HMM and compared the output with: (i) the observed behavioral data recorded from animals and (ii) data simulated from a conventional binomial GLM with no HMM states. The two-state GLM-HMM data reproduced key features of animal behavior including trial-to-trial autocorrelation in pursue-versus-reject decisions, (Fig. 3, A and B) and the richness of the environment and behavioral history effects described earlier (Fig. 3, C and D). The binomial GLM with no HMM states was unable to reproduce this

pattern. This demonstrates that the key behavioral effects described above—the tendency of animals to pursue more rewards in rich environments and to repeat their pursue-versus-reject decisions over time—can be explained parsimoniously by a two-state GLM-HMM in which there are time-varying changes in the bias term, which determines an animal's predisposition to pursue or reject opportunities. Further analyses demonstrated that HMM states were not reducible to satiety, fatigue, or time on task (fig. S4).

We tested the whether the statistical patterns identified by the GLM-HMM corresponded to biological motivation states with a series of follow-up analyses. To do this, we first decoded the maximum a posteriori HMM state on each trial using a well-established procedure called the Viterbi algorithm. We validated the Viterbi algorithm on simulated data, which demonstrated that it identified the correct HMM state on 90% of trials and identified HMM-state transitions within $\pm$ three trials of a true transition in 85% of cases (Fig. 4, A to C). We then examined how decoded HMM states were related to behavior. As expected, the animals were markedly more likely to pursue reward opportunities in the putative high-motivation state compared to the low-motivation state, over and above any influence of the reward value available (GLM2.1; $\beta_{motivation-level} = 2.66$, SE = 0.56, $P < 0.001$; Fig. 4D). Behavior during transitions between HMM states was characterized by abrupt changes in pursue rates, suggesting that the GLM-HMM captured state-like differences in behavior with a high degree of temporal precision (GLM2.2; $\beta_{before-versus-after-transition(high-to-low)} = -1.48$, SE = 0.08, $P < 0.001$; $\beta_{before-versus-after-transition(low-to-high)} = 1.80$, SE = 0.30, $P < 0.001$; Fig. 4G). Last, we established the model's convergent validity by comparing HMM states to: (i) trial-by-trial pupil size, a well-validated indicator of physiological arousal (20), and (ii) trial-by-trial RTs for pursue decisions, a simple way of quantifying vigor. Consistent with the link between HMM states and motivation, putative high-motivation states featured faster RTs for pursue decisions (GLM2.3; $\beta_{motivation-level} = -0.20$, SE = 0.03, $P < 0.001$; Fig. 4E) and increases in pupil diameter during decision-making (GLM2.4; $\beta_{motivation-level} = 0.20$, SE = 0.02, $P < 0.001$; Fig. 4F). The GLM-HMM, thus, provided quantitative evidence for discrete, persistent, and biologically meaningful internal motivation-states in monkey behavior.

Last, we tested the hypothesis that motivation states were related to the external environment: Was it the case, for example, that animals
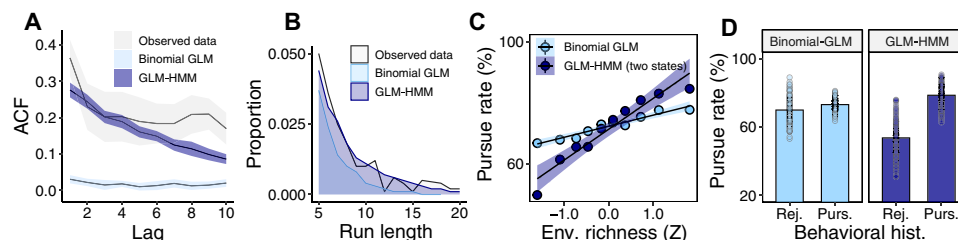
Fig. 3. A two-state GLM-HMM recapitulates qualitative patterns of animal behavior. (**A**) Data simulated from a two-state GLM-HMM reproduced the autocorrelations in decision-making that characterized animal behavior. A binomial GLM did not reproduce this pattern. y axis shows autocorrelation function (ACF) between the decision on trial t and decisions on $(t − 1):(t − 10)$ for recorded behavior, the two-state GLM-HMM and a binomial GLM for an example animal. Lines and shaded areas show the means and SEM, respectively, of the ACF at each individual lag. (**B**) A complementary perspective on autocorrelated decisions is decision run lengths—the number of times that a decision is repeated over consecutive trials [e.g., rejecting three consecutive opportunities is a run length of 3; see also (18)]. The GLM-HMM recapitulates observed patterns of run lengths, especially for run lengths spanning five or more consecutive trials. A binomial GLM is unable to reproduce this pattern. Data from the same example animal and simulations as previous panels. (**C** and **D**) Implementing the GLMs (GLM1.1 to 1.3) used to analyze behavior on simulated data from a two-state GLM-HMM recapitulated both richness of the environment (C) and behavioral history effects (D). The same effects did not occur in data simulated from one-binomial GLMs. See Fig. 1G and fig. S1D for comparison. In (C), dots indicate mean pursue rate in deciles of richness of the environment across simulations, and shaded area indicates SEM around line of best fit. In (D), dots indicate mean pursue rates in individual simulations.
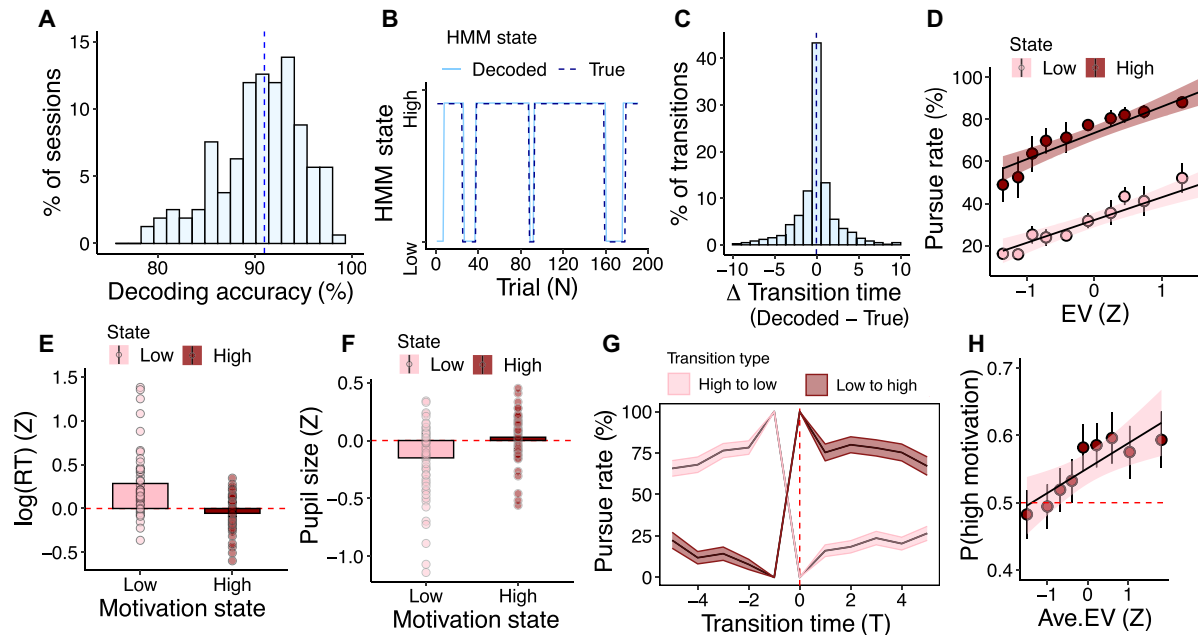
Fig. 4. HMM states correspond to distinct, persistent, and biologically meaningful motivation states during behavior. We decoded the maximum a posteriori HMM state on each trial using the Viterbi algorithm and compared it to behavior. (**A**) We established Viterbi algorithm's accuracy using simulated datasets where the true HMM state was known. Decoding was successful on 90.68% ($\pm$0.72%) of trials. Histogram shows distribution of session-level decoding accuracy (%) across simulated datasets. (**B**) Comparison of true and Viterbi-decoded states in example simulated session. (**C**) We assessed the accuracy of state-transition decoding by quantifying the distance between decoded and true transition trials. Most transition trials were exactly identified (mean distance = −0.03) and 84% of decoded transitions occurred within $\pm$3 trial window of a true transition. Histogram shows proportion of decoded-versus-true transition distances. (**D**) Animals were more likely to pursue reward opportunities in the high-motivation state compared to the low-motivation state, over and above any influence of an opportunity's expected value. Dots and whiskers indicate means $\pm$ SEM pursue rates in deciles of expected value. (**E**) RTs were faster when animals occupied the high-motivation state. Dots indicate mean RT for each animal in each session. (**F**) Pupil size during the decision phase (see Materials and Methods) of the task was greater when animals occupied the high-motivation state relative to the low-motivation state. Dots indicate mean pupil size in each animal in each session. (**G**) Pursue rates changed abruptly at motivation-state transitions. This suggests that the GLM-HMM identified state-like rather than gradual changes in behavior. Color shades indicate transition direction. Lines and shaded areas indicate means ($\pm$SEM) of pursue rate across all animals in all sessions (**H**) Animals are more likely to occupy the high-motivation state increases as the expectation value of recent reward opportunities (Ave. EV) increases. Dots and whiskers indicate means ($\pm$SEM) of rate of motivation-state occupancy in deciles average expectation value.

matched their motivation state with the overarching distribution of rewards in the task milieu? We quantified the availability of rewards by calculating, for each trial, the average expectation value of reward opportunities in the preceding five trials. We did not use the richness of the environment metric introduced earlier because this was calculated on the basis of an animal's reward history and was therefore contingent on its recent behavior—Relating an animal's motivation state to its recent behavior, thus, would be a logically circular analysis. In contrast, the expectation value of reward opportunities was experimentally controlled via the reward distributions characterizing the task (Fig. 1B) and was therefore not affected by an animal's decisions. This showed that animals were more likely to occupy high-motivation states as the expected value of recent opportunities increased (GLM2.5; $\beta_{Ave.EV} = 0.19$, SE = 0.08, $P = 0.022$; Fig. 4H; see also fig. S7). The GLM-HMM therefore demonstrated not just that animals experienced state-like changes in motivation for rewards but that these states were reconciled with the distribution of rewards in the world around them.

Brain activity in DRN reflects an animal's reward environment and changes in its motivation state

Animals performed the task under fMRI. Our analysis of fMRI recordings focused on a priori regions of interest (ROIs; see figs. S9 and S10) comprising the ascending neuromodulatory systems (ANS)—an

assemblage of phylogenetically ancient nuclei that includes the serotonergic DRN, in addition to the VTA and substantia nigra (SN), the cholinergic nucleus basalis (NB), and the noradrenergic locus coeruleus (LC). We also examined habenula (Hb)—an epithalamic nucleus with diverse subcortical connections that interact with the ANS in a reciprocal fashion (*21, 22*). Broadening our analysis beyond DRN provided a comparative perspective on its function, which is valuable given that reward functions are sometimes jointly attributed to different subcortical nuclei—The richness of the environment, for example, has been linked to both VTA and DRN-related signals (*9, 10, 23, 24*), while various forms of reward uncertainty have been ascribed to DRN, LC, and the cholinergic basal forebrain alike (*11, 25–28*). fMRI's wholistic perspective enabled us to address this by comparing signals in DRN and other ANS nuclei (*12*). This was accomplished with a novel suite of fMRI acquisition and preprocessing methods that optimized blood oxygen level–dependent (BOLD) signal from subcortical regions and minimized artefacts and noise sources for midbrain and brainstem regions (see the section "Acquisition, reconstruction, and preprocessing of MRI data" in Materials and Methods).

We first examined brain activity during the pursue-versus-reject decision made on each trial (GLM3.2; Materials and Methods). We focused on activity representing the richness and stochasticity of the reward environment at this time. Note that the richness of the

environment was correlated with an animal's behavioral history (see the section "Animal behavior is modulated by the environment"). We therefore performed our analysis separately for trials in which the previous encounter was (i) pursued and (ii) rejected to ensure that any patterns of brain activity that we identified were specific to the richness of the environment, and not related to behavioral history. We note, however, that the key aspects of DRN's relationship with the richness of the environment obtain regardless of whether the data are analyzed separately in this way (Fig. 5A).

DRN activity was negatively correlated with the richness of the environment after rejection of opportunities [one-sample two-tailed t test; GLM3.2; $t_{DRN;rejected}(58) = -2.80$, $P = 0.034$; $t_{DRN;pursued}(58) = -0.37$, $P = 0.713$ after Bonferroni correction for multiple comparisons, as are all subsequent t tests; Fig. 5A, see legend for further discussion of timing of BOLD signal]. In contrast, VTA activity was prominent following opportunity pursuits [GLM3.2; $t_{VTA;pursued}(58) = 3.30$, $P = 0.009$; $t_{VTA;rejected}(58) = 1.60$, $P = 0.350$; Fig. 5A], and Hb exhibited aspects of both DRN and VTA patterns [GLM3.2; $t_{Hb;pursued}(58) = 3.32$, $P = 0.009$; $t_{Hb;rejected}(58) = -3.37$, $P = 0.008$; $t_{Hb;rejected-versus-pursued}(58) = 4.95$, $P < 0.001$; Fig. 5A]. A two-way analysis of variance (ANOVA) followed by pairwise comparisons confirmed that the effects in DRN and VTA after rejections and pursuits,

respectively, were different from one another but not from the corresponding Hb effect [$F_{ROI}(2, 348) = 8.106$, $P < 0.001$, $F_{Behav.-history}(1, 348) = 22.305$, $P < 0.001$; $F_{ROI-by-Behav.-history}(2, 348) = 4.521$, $P = 0.012$; $t_{VTA versus DRN;pursued}(58) = 2.501$, $P = 0.015$; $t_{DRN versus VTA;rejected}(58) = -3.403$, $P < 0.001$; $t_{VTA versus Hb;pursued}(58) = 0.650$, $P = 0.518$; $t_{DRN versus Hb;rejected}(58) = -0.214$, $P = 0.831$]. We did not find an effect of environment richness as a function of behavioral history in any other ROI (all Ps > 0.05). Despite previous evidence linking DRN with reward uncertainty, we found no relationship between the environment's stochasticity and brain activity in DRN [GLM3.1; $t_{DRN}(58) = 1.10$, $P = 0.824$] and no evidence that stochasticity was represented in other ROIs (GLM3.1; see fig. S11).

After determining that DRN, VTA, and Hb were the key regions tracking the environment, we tested whether they were linked to the motivation states identified with the GLM-HMM. We first quantified the relationship between brain activity and motivation-state level (high versus low). This showed that high motivation states were signaled by increased activity in VTA and Hb but not in DRN [GLM3.4; $t_{Hb;motivation state}(58) = 2.34$, $P = 0.045$; $t_{VTA;motivation state}(58) = 2.82$, $P = 0.019$; $t_{DRN;motivation state}(58) = 1.75$, $P = 0.084$; Fig. 5B]. We next examined activity related to motivation-state transitions (change versus no change) by identifying transition trials on which
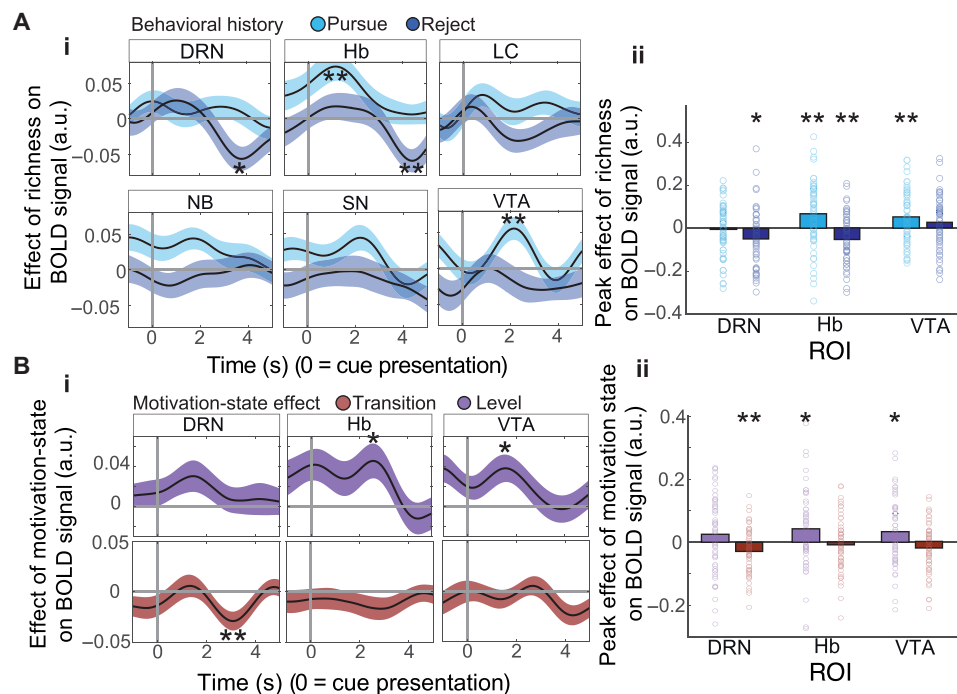
Fig. 5. Brain activity in DRN represents the richness of an animal's environment and transitions between motivation-states. (**A-i**) Time course of the effect that the richness of the environment (β-richness) has on BOLD signal in subcortical ROIs. Time (x axis) depicted relative to the onset of reward opportunities when pursue-versus-reject decisions are made (i.e., time = 0 corresponds to decision time). BOLD signal in macaque monkeys peaks approximately 2 to 4 s after neural activity. The timing of peak β weights in DRN, VTA, and Hb is therefore consistent with brain activity that occurs at the time of decision-making. We did not find an effect of environment richness as a function of behavioral history in any other ROI (all Ps > 0.05). The panel shows the analysis separately as a function of behavioral history (GLM3.2; see Results for explanation). (**A-ii**) Distribution of peak effect sizes of richness of the environment (β-richness) on BOLD signal. Dots indicate peak effect sizes for individual sessions. Separating the data in this way revealed notable patterns in the representation of the environment across regions. DRN represented the richness of the environment with negative sign after the previous opportunity was rejected, VTA represented the richness of the environment with positive sign when animals pursued the previous opportunity, and Hb exhibited distinct positive and negative signals on trials after pursuits and rejections, respectively. (**B-i**) Time course of the effect of motivation-state level (high- versus low-motivation state, purple) and motivation-state transitions (red) on BOLD activity in ROIs that represented the reward environment. (**B-ii**) Distribution of peak effect sizes for motivation-state level (purple) and motivation-state transitions (red) on brain activity. Only DRN represents transitions between motivation-states. Dots indicate peak effect sizes for individual sessions. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

the maximum a posteriori motivation state was different from the preceding trial. For each transition trial, we defined a corresponding transition period comprising a symmetric seven-trial window centered on the transition trial (i.e., the transition trial ± three trials). We did this for two reasons: (i) The biological process of transitioning between motivation states is likely to unfold over time instead of occurring instantaneously, meaning that brain activity related to such transitions will occur over a series of trials. (ii) Validation of the GLM-HMM on simulated data showed that state transitions could be accurately decoded within a ± three-trial window (see Fig. 4C). On average, these transition periods comprised 35% of the total trials in a given session. Brain activity in these periods showed a notable dissociation whereby DRN—but neither Hb nor VTA—signaled transitions between motivation states [GLM3.4; $t_{DRN;state\ transition}(58) = -3.211$, $P = 0.006$; $t_{Hb;state\ transition}(58) = -1.43$, $P = 0.160$; $t_{VTA;state\ transition}(58) = -1.80$, $P = 0.152$; Fig. 5B; see fig. S14 (B and C) for data from individual animals].

The motivation-state transition effect was localized to DRN: Analyzing BOLD signal from adjacent anatomical features, such as the medial raphe nucleus (MRN) and fourth ventricle, showed that there were no corresponding patterns of activity in these areas, and the motivation-state effect in DRN was, indeed, greater than the null effects in MRN and the fourth ventricle [Fig. 6A; $t_{DRN}(58) = -3.21$,

$P = 0.006$, $t_{MRN}(58) = 1.01$, $P = 0.638$, $t_{Vent}(58) = -0.007$, $P = 0.994$, $t_{DRN-versus-MRN}(58) = -3.13$, $P = 0.003$; $t_{DRN-versus-Vent}(58) = -2.21$; $P = 0.031$]. We then probed the temporal specificity of the DRN activity pattern by varying the position of the transition periods examined in the fMRI analysis relative to the transition trials identified by the GLM-HMM. Activity changes in DRN were most prominent when the analyzed period was closely aligned with the transition trial decoded by the GLM-HMM, suggesting that they were indeed specific to motivation-state transition events (Fig. 6C; see fig. S14A for comparison between analysis periods that include and exclude the transition trial). Last, we parsed motivation-state transitions according to direction (i.e., high to low and low to high). This indicated that DRN activity covaried with high-to-low transitions but not low-to-high transitions [GLM3.4 $t_{DRN;high-to-low\ transitions}(58) = -2.51$, $P = 0.015$; $t_{DRN;low-to-high\ transitions}(58) = -0.91$, $P = 0.365$; Fig. 6B]. Together, these analyses suggested that DRN—and DRN specifically—implemented negative changes in intrinsic motivation for rewards.

Noninvasive disruption of DRN perturbs motivation-state transitions

fMRI recordings linked DRN to behavior in two complementary ways; (i) DRN coded the richness of the environment, which modulated an animal's pursue-versus-reject decisions toward specific
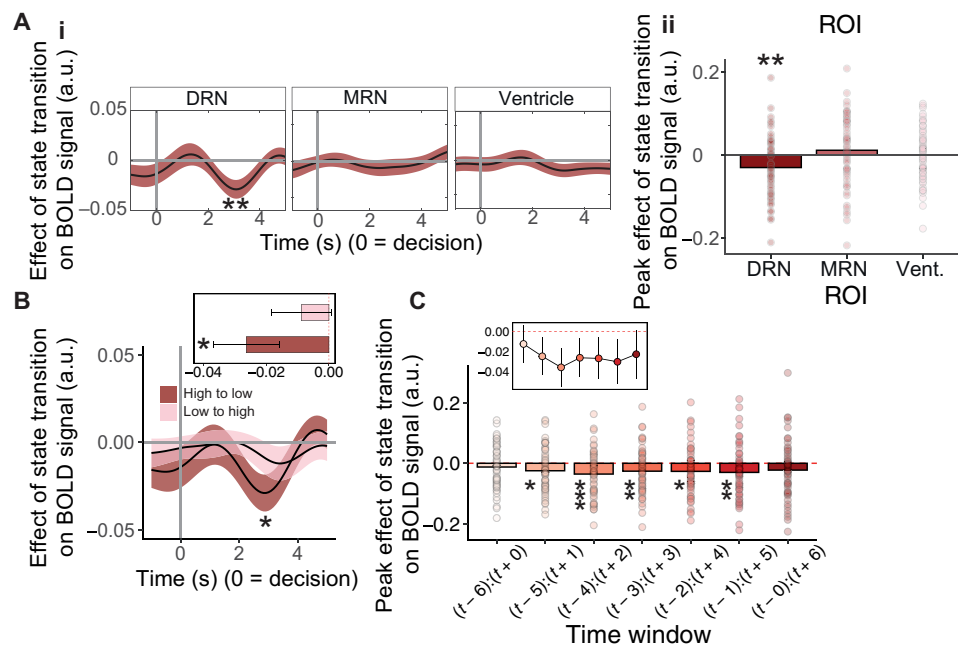
Fig. 6. Motivation-state activity in DRN is spatially and temporally specific. (Ai-ii) Motivation-state transitions affected BOLD signal in DRN but not neighboring features like MRN and the 4th ventricle, confirming that the pattern was DRN specific. (**B**) The time course of the effect that low-to-high (pink) and high-to-low (red) motivation-state transitions have on DRN BOLD activity. The effect is strongest for high-to-low state transitions. Inset panel shows distribution of peak effect sizes. (**C**) We initially tested the effect of motivation-state transitions on brain activity in a symmetric seven-trial window centered on the transition trial (i.e., the transition trial ±3 trials). This indicated that transitions were represented in the BOLD activity of DRN (Fig. 5B). We confirmed that this effect was not an artefact of our initial window selection by iterating the analysis over different seven-trial windows comprising transitions events. This showed that the effect was robust to window position [$t_{(t-5):(t+1)}(58) = -2.63$, $P = 0.011$; $t_{(t-4):(t+2)}(58) = -3.69$, $P < 0.001$; $t_{(t-3):(t+3)}(58) = -2.70$, $P = 0.009$; $t_{(t-2):(t+4)}(58) = -2.60$, $P = 0.012$; $t_{(t-1):(t+5)}(58) = -2.69$, $P = 0.009$; $t_{(t-0):(t+6)}(58) = -1.88$, $P = 0.065$; GLM3.6]. Note that the windows featuring the least overlap with decoded motivation transitions show nonsignificant effects [$t_{(t-6):(t+0)}(58) = -1.34$, $P = 0.180$; $t_{(t-0):(t+6)}(58) = -1.88$, $P = 0.065$]. Inset (C) shows mean ± 95% CI of transition effect across time windows. x axis indicates position of transition period relative to decoded transition time (where $t = 0$ represents decoded transition trial). In time course graphs (A-i and B), lines and shadings show the mean and SE of the β weights across the sessions, respectively. Effect-size graphs (A-ii and C) show size of the peak regression weight in the time course selected using an unbiased leave-one-out procedure (see Materials and Methods). Dots indicate peak regression weight from individual sessions. Error bars in (B) inset show SD of session-wise regression weights. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

reward opportunities, and (ii) DRN coded transitions between motivation states, and therefore changes in reward pursuit over multitrial timescales. We tested DRN's causal contribution in these respects with a second experiment using TUS (Fig. 7A)—a minimally invasive and reversible technique that disrupts brain activity via kinetic interactions between focused ultrasound waves and the neuron and astrocyte membrane that, in turn, induce changes in adjacent synapses (29, 30). As a result, short TUS trains of the type we used are known to produce short-term changes in neural activity by inducing N-methyl-D-aspartate–dependent plasticity (31–35) leading to changes in activity in spatially circumscribed gray brain regions that are targeted. As a result, a region targeted with TUS alters its responsiveness to activity in interconnected areas, while nonstimulated areas show no such change (17).

We used an offline TUS protocol that modulates neural activity for over an hour after sonication (17, 36, 37). We compared DRN-TUS to three control conditions: a sham condition (sham-TUS), an active cortical control (superior temporal sulcus; STS-TUS), and an additional active subcortical control in VTA. We chose VTA as a subcortical control region because it was implicated in motivation states during fMRI recordings in distinct ways to DRN, thereby allowing us to test the specificity of DRN's function. We chose STS as our active cortical control condition because it is a region closest to the transducer position when TUS is targeted to DRN and through which the ultrasound beam might pass when targeting the DRN. It is therefore ideally located to rule out any potential side effect of sound beams on brain structures due to some process of reflection or refraction from skull.
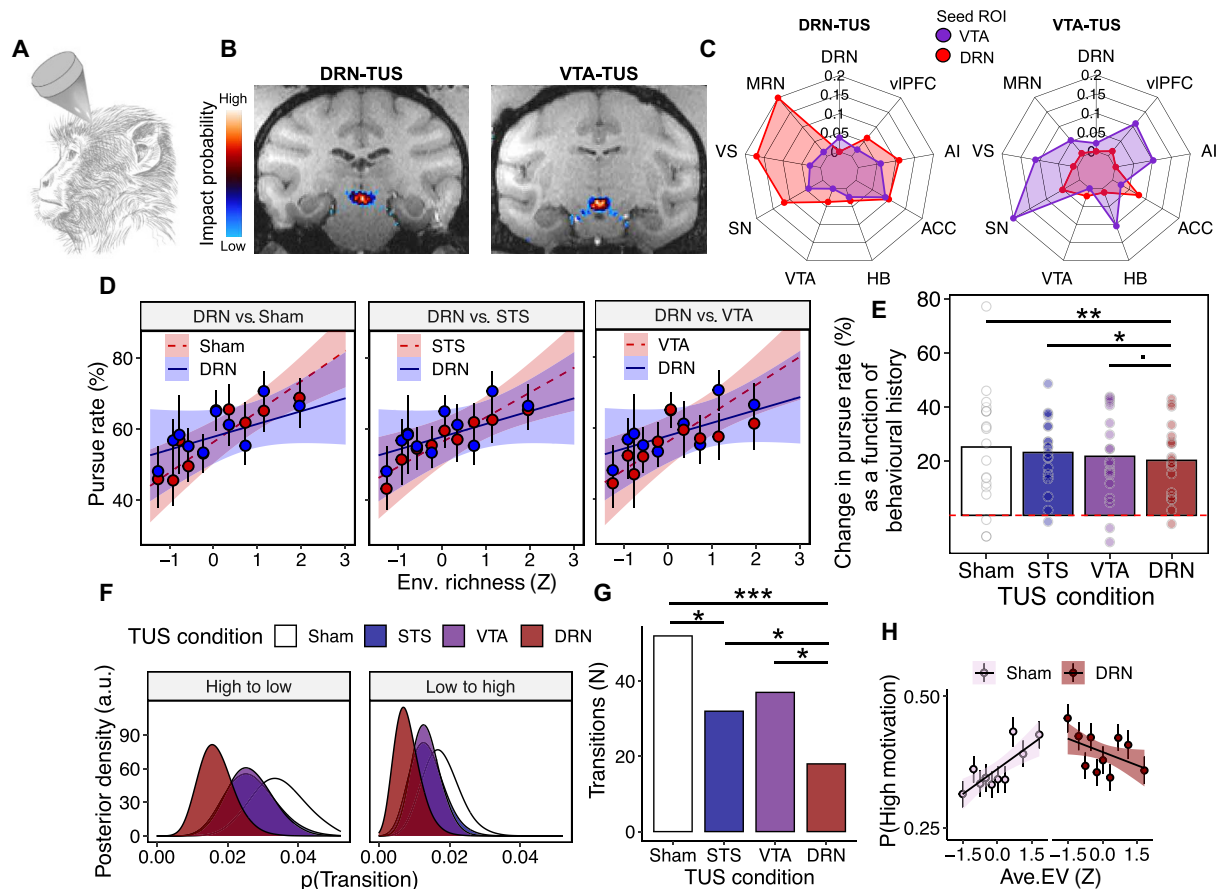

Fig. 7. DRN is causally involved in the relationship between the environment and behavior. (A) We investigated the causal contribution of DRN to behavior with TUS. (B) Simulation of ultrasound propagation under the TUS protocol suggested target-specific sonication after DRN-TUS and VTA-TUS. Color scale indicates the probability of neuromodulation after bilateral DRN/VTA TUS (see fig. S15). The low impact probability corresponds to 0.5 MPa. High impact probabilities at each acoustic focus are shown in fig. S15. (C) DRN-TUS disrupted DRN (red) connectivity with key brain regions while leaving VTA (blue) connectivity intact. The opposite pattern occurred after VTA-TUS. Radial axis shows absolute difference in connectivity between seed and target ROIs pre-versus-post TUS (see Materials and Methods). Targets included vlPFC, AI, ACC, Hb, SN, VERSUS, MRN, and DRN. (D) DRN-TUS attenuated the richness of the environment's effect on behavior relative to sham-TUS, STS-TUS, and VTA-TUS. Points and whiskers indicate means (±SEM) pursue rate across deciles of environmental richness. (E) DRN-TUS attenuated behavioral history's effect on current behavior relative to sham-TUS and STS-TUS, and marginally compared to VTA-TUS. y axis indicates difference in pursue rate based on behavioral history (pursue-rate$_{behav.history==pursue}$ − pursue-rate$_{behav.history==reject}$). Positive values indicate that animals were more likely to pursue a reward if they also pursued the previous one. Points indicate differences in mean pursue rates per session. (F) Posterior probability of state transitions in GLM-HMMs fitted to each TUS condition. The central tendency of DRN-TUS posteriors is lower than STS-TUS, VTA-TUS, and sham-TUS posteriors. (G) Fewer motivation-state transitions occurred in DRN-TUS sessions versus all other TUS conditions. Data shown as sum of all motivation-state transitions in each TUS condition. (H) DRN-TUS diminished the relationship between the average value of recent opportunities and motivation-state level. Dots and whiskers indicate means (±SEM) level of high-motivation state occupancy in deciles of average expected value (Ave. EV). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

We first performed simulations of acoustic wave propagation under the TUS protocol, which established that selective perturbation of DRN and VTA was feasible despite their diminutive size and anatomical location (Fig. 7B and fig. S15). We then empirically tested the TUS protocol's anatomical specificity by measuring changes in functional connectivity between DRN, VTA and a series of key brain areas before-versus-after DRN-TUS and—as a control—before-versus-after VTA TUS. This showed notable and selective changes in region-specific functional connectivity: DRN-TUS disrupted DRN's ordinary patterns of coactivation with key interconnected regions but left VTA connectivity unchanged (Fig. 7C; see also fig. S18). Conversely, after VTA-TUS, VTA connectivity changed but DRN connectivity did not (Fig. 7C). In combination with the sonication simulations, this indicated that TUS induced safe, efficacious, and highly localized disruption of DRN and VTA activity, analogous to previous investigations of TUS in cortical brain regions (17, 36, 37).

We tested the same monkeys from the fMRI experiment on five sessions of the behavioral task per TUS condition in counterbalanced and pseudorandomized order (see Materials and Methods). We first investigated how TUS modulated the influence of (i) richness of the environment and (ii) behavioral history, which were the key factors driving the pursue-versus-reject decisions made on each trial. DRN-TUS diminished the effect of richness of the environment relative to all other conditions (GLM4.1; $\beta_{\text{DRN-TUS versus sham-TUS*richness}} = -0.18$, SE $= 0.05$, $P < 0.001$; $\beta_{\text{DRN-TUS versus STS-TUS*richness}} = -0.12$, SE $= 0.06$, $P = 0.025$; $\beta_{\text{DRN-TUS versus VTA-TUS*richness}} = -0.12$, SE $= 0.06$, $P = 0.035$; Fig. 7D). Similarly, DRN-TUS clearly reduced the effect of behavioral history relative to sham-TUS and STS-TUS and produced a marginal difference relative to VTA-TUS (GLM4.2; $\beta_{\text{DRN-TUS versus sham-TUS*behave-history}} = -0.42$, SE $= 0.11$, $P < 0.001$; $\beta_{\text{DRN-TUS versus STS-TUS*behave-history}} = -0.24$, SE $= 0.11$, $P = 0.027$; $\beta_{\text{DRN-TUS versus VTA-TUS*behave-history}} = -0.20$, SE $= 0.11$, $P = 0.057$; Fig. 7E). There were no differences between active control conditions (STS-TUS and VTA-TUS) and sham-TUS (GLM4.1; Fig. 7, D and E; $\beta_{\text{VTA-TUS versus sham-TUS*richness}} = -0.06$, SE $= 0.06$, $P = 0.282$; $\beta_{\text{STS-TUS versus sham-TUS*richness}} = -0.06$, SE $= 0.05$, $P = 0.306$; GLM4.2; Fig. 7E; $\beta_{\text{VTA-TUS versus sham-TUS*behavioral-history}} = -0.20$, SE $= 0.10$, $P = 0.060$; $\beta_{\text{STS-TUS versus sham-TUS*behavioral-history}} = -0.18$, SE $= 0.11$, $P = 0.090$). DRN, therefore, was specifically and causally involved in the impact of an animal's recent reward environment and behavior on its present decisions to pursue reward.

Next, we applied the GLM-HMM to behavioral data from the TUS experiment (see Materials and Methods). Our analysis focused on motivational-state transitions, which were a distinguishing feature of DRN activity in fMRI recordings. These recordings suggested that DRN was specifically involved in high-to-low motivational-state transitions, but we first performed the analysis in a direction-neutral way given the interdependence of time-series observations: Because the animals typically began the task in a high-motivation state [$P(\text{init-state} = \text{high}) = 0.70$], reducing the likelihood of high-to-low transitions would necessarily reduce the frequency of low-to-high transitions in a fixed-length time series, even if the underlying low-to-high transition probability was unchanged. The analysis showed a notable pattern whereby DRN-TUS reduced the likelihood of transitions relative to all other TUS conditions (GLM4.3; $\beta_{\text{DRN-TUS versus sham-TUS}} = -1.08$, SE $= 0.27$, $P < 0.001$; $\beta_{\text{DRN-TUS versus VTA-TUS}} = -0.71$, SE $= 0.29$, $P = 0.013$; $\beta_{\text{DRN-TUS versus STS-TUS}} = -0.57$, SE $= 0.29$, $P = 0.051$; Fig. 7, F and G). There was no effect of VTA-TUS relative to sham-TUS

(GLM4.3; $\beta_{\text{VTA-TUS versus sham-TUS}} = -0.37$, SE $= 0.22$, $P = 0.090$). There was a moderate difference between STS-TUS and sham-TUS, but this was outweighed by the fact that DRN-TUS reduced the likelihood of transitions even relative to STS-TUS (GLM4.3; $\beta_{\text{STS-TUS versus sham-TUS}} = -0.51$, SE $= 0.23$, $P = 0.022$; $\beta_{\text{DRN-TUS versus STS-TUS}} = -0.57$, SE $= 0.29$, $P = 0.051$). DRN-TUS was, moreover, the only condition that reduced high-to-low transitions specifically (GLM4.4; $\beta_{\text{DRN-TUS versus sham-TUS}} = -0.77$, SE $= 0.32$, $P = 0.016$; $\beta_{\text{VTA-TUS versus sham-TUS}} = -0.33$, SE $= 0.28$, $P = 0.238$; $\beta_{\text{STS-TUS versus sham-TUS}} = -0.43$, SE $= 0.29$, $P = 0.139$). Last, DRN-TUS caused animals to spend more time in high-motivation states relative to sham-TUS (GLM4.5; $\beta_{\text{DRN-TUS versus sham-TUS}} = 0.28$, SE $= 0.07$, $P < 0.001$), consistent with the idea that DRN-TUS prevented animals from transitioning to low-motivation states. Note that motivation-state transitions are rare events, regardless of the TUS condition. However, our analysis clearly shows that DRN disruption significantly reduces or even eliminates these events, compared to other conditions (see fig. S16 for data from individual animals).

Last, we sought to specify how DRN controlled transitions between motivation states. To do so, we returned to an earlier analysis demonstrating that motivation states were matched to the availability of rewards: In brief, animals were more likely to occupy high-motivation states when there were many high-value rewards available, analogous to the way that animals pursued more opportunities in rich environments (Fig. 4H). Given DRN's critical involvement in the latter phenomenon, we asked whether its control of motivational-state transitions was mediated by the availability of rewards. Consistent with this view, DRN-TUS diminished the relationship between the availability of rewards and motivation states relative to all other TUS conditions (GLM4.5; Fig. 7H; $\beta_{\text{DRN-TUS versus sham-TUS*Ave.EV}} = -0.28$, SE $= 0.07$, $P < 0.001$; $\beta_{\text{DRN-TUS versus STS-TUS*Ave.EV}} = -0.29$, SE $= 0.07$, $P < 0.001$; $\beta_{\text{DRN-TUS versus VTA-TUS*Ave.EV}} = -0.30$, SE $= 0.07$, $P < 0.001$). There were no effects of VTA-TUS or STS-TUS relative to sham-TUS ($\beta_{\text{VTA-TUS versus sham-TUS*Ave.EV}} = -0.03$, SE $= 0.06$, $P = 0.610$; $\beta_{\text{STS-TUS versus, sham-TUS*Ave.EV}} = -0.05$, SE $= 0.07$, $P = 0.499$). Intriguingly, DRN-TUS's influence was more prominent for the link between low-value reward environments and low-motivation states, which dovetailed with fMRI signals implicating DRN specifically in high-to-low transitions (GLM4.5; for Ave. EV $< \mu_{\text{Ave.EV}}$ $\beta_{\text{DRN-TUS*Ave.EV}} = -0.52$, SE $= 0.19$, $P = 0.006$; for Ave. EV $\geq \mu_{\text{Ave.EV}}$ $\beta_{\text{DRN-TUS*Ave.EV}} = -0.13$, SE $= 0.15$, $P = 0.403$). Together, these results suggest that DRN has a fundamental role in ensuring that an animal's motivation state is appropriate to the distribution of rewards in the environment.

A cortico-subcortical circuit reconciles behavior with the environment

In a final analysis, we examined interactions between DRN and other brain regions that might relate to its behavioral function. To do so, we first reverted to the analysis of brain activity and behavior that did not rely on the GLM-HMM framework (Fig. 5A). Here, DRN, VTA, and Hb represented an animal's environment with complementary patterns of activity, which suggested that they might form a circuit for reconciling decisions with the surrounding environment. To test the circuit hypothesis, we expanded our purview to functionally related cortical ROIs in supplementary motor area (SMA), anterior cingulate cortex (ACC), and anterior insula (AI) based on previous work implicating these regions in behavioral change (9, 36, 38–40). Only AI activity signaled the richness of the environment with the

distinctive contingency on behavioral history seen in subcortical ROIs, and we therefore retained AI as a cortical ROI for connectivity analysis (see fig. S13).

Next, we asked which ROIs coded the pursue-versus-reject decision taken on each trial—in other words, which ROIs might be the output of the putative decision-making circuit. This demonstrated that BOLD activity time-locked to decision-making in both Hb and AI represented pursue-versus-reject decisions [GLM3.3; $t_{pursue;Hb}(58)$ = 3.24, P = 0.011; $t_{pursue;AI}(58)$ = 2.78, P = 0.028; fig. S12]. We then conducted psychophysiological interaction (PPI) analyses to probe changes in pairwise connectivity between regions as a function of the environment (41). PPIs mirrored the patterns observed earlier: Connectivity between DRN and AI increased as a function of richness of the environment after rejections [GLM3.6; $t_{PPI(DRN-AI}$ $_{by\ richness);rejected}(58)$ = 2.84, P = 0.006; Fig. 8B] and increased between VTA and AI after pursuits [GLM 3.6; $t_{PPI(AI-VTA\ by\ richness);pursued}(58)$ = 2.14, P = 0.036; Fig. 8D]. Adopting the GLM-HMM analysis approach showed a corresponding pattern of results (GLM3.7; Figs. 8B and 6A-iii). Connectivity between Hb and AI was not modulated

by the richness of the environment but instead by the pursue/reject decision taken on each trial, consistent with a role in translating motivation to action (42) [GLM3.8 $t_{PPI(AI-Hb\ by\ action)}(58)$ = 2.25, P = 0.028; Fig. 8C].

DISCUSSION

We provide converging evidence that DRN controls changes in an animal's motivation for rewards. We observed distinctive patterns of DRN activity corresponding to the richness of an animal's environment and transitions between statistically delineated motivation states. We followed this with minimally invasive DRN disruption, which diminished the environment's effect on decision-making and reduced the frequency of motivation-state transitions. This suggests that DRN is causally involved in transitions between a high-motivation state occupied when rewards are abundant and a low-motivation state engendered by reward scarcity.

We used a GLM-HMM approach to describe features of animal behavior that are difficult to capture with a standard GLM. In
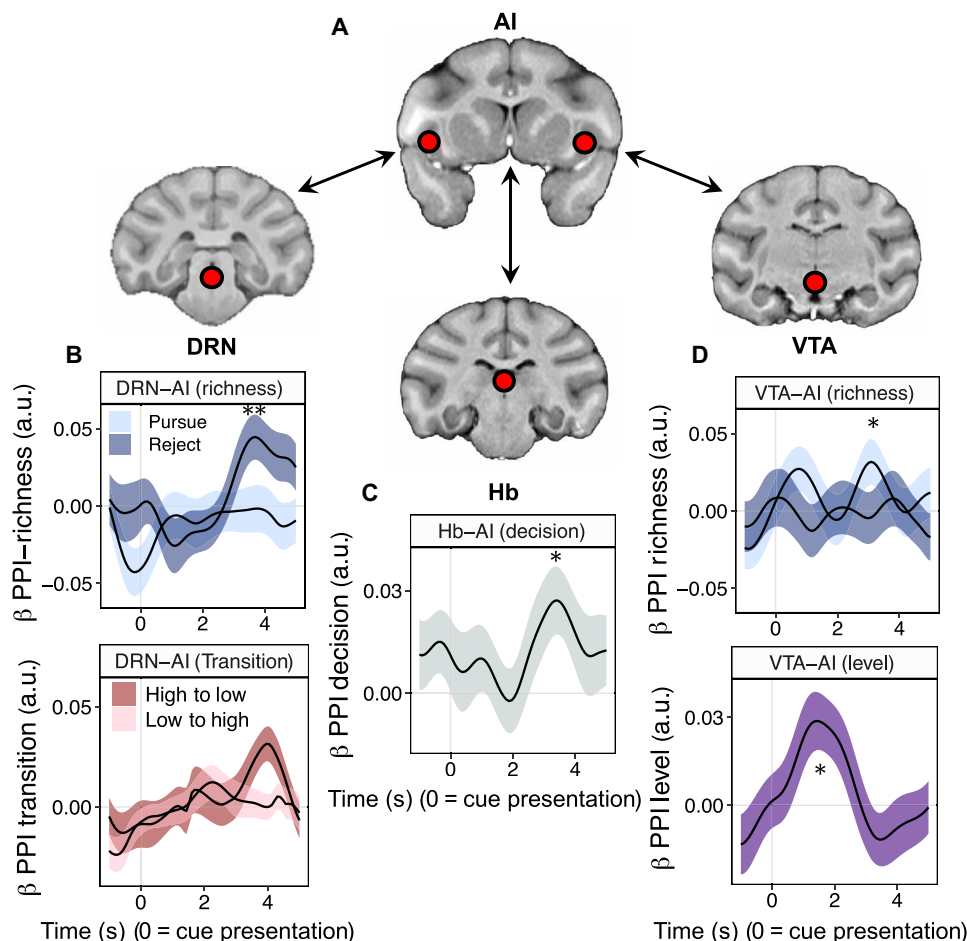
Fig. 8. A brain circuit for reconciling an animal's motivation with its environment. (**A**) PPIs indicated the following patterns of cortico-subcortical connectivity. (**B**) Connectivity between DRN and AI increased as a function of richness of the environment in the aftermath of rejections (GLM3.6; top), and as a function of high-to-low motivation-state transitions [bottom; GLM3.7; $t_{PPI(high-to-low)}$ = 3.64, P < 0.001], consistent with DRN's fMRI signals (**C**) connectivity between AI and Hb increased as a function of an animal's ultimate decision about whether to pursue a reward opportunity (GLM3.8). (**D**) In contrast to DRN, connectivity between VTA and AI increased as a function of richness of the environment in the aftermath of pursuits (GLM3.6; top) and during high-motivation states (GLM3.7; $t_{motivation-level}$ = 2.47, P = 0.016). *P < 0.05, **P < 0.01, ***P < 0.001.

particular, we tested the hypothesis that animals adopted different motivation states using a GLM-HMM in which only the bias parameter changed between HMM states. This model allowed us to identify periods where animals were more likely—all else equal—to pursue rewards, and periods where animals were less likely—all else equal—to pursue rewards, and to identify points of transition between these periods. The number of trials in our dataset did not allow us to implement more complicated versions of the GLM-HMM where, for example, the transition matrix is modeled as a function of covariates instead of as fixed parameters (43). A GLM-HMM with a covariate-dependent transition matrix would be an interesting subject for future research. Similarly, we acknowledge that other descriptive and/or mechanistic models would likely reveal further interesting features of behavior (44–46). Here, however, we focused on the specific hypothesis that animals would experience shifts in the baseline likelihood of pursuing rewards.

Our findings dovetail with two prevailing views on DRN function: The first has emphasized DRN's role in behavioral changes, including between exploration and exploitation or patience and impulsivity (5, 47, 48). For example, the firing rate of DRN serotonin neurons increase during waiting for delayed reward (49). The second has linked DRN activity to statistical aspects of rewards, like the general value and/or uncertainty of the options characterizing an environment (6–10, 50). We observed similar brain-behavior relationships here, and the co-occurrence of both phenomena suggests that they are fundamentally related. In support of this view, we found that

DRN disruption impaired animals from matching their motivation state with the availability of rewards, and specifically from entering low-motivation states during low-value environments, which was an important motif in their normal behavior. In tandem with previous studies, this indicates that DRN is critical for reconciling an animal's behavior with the reward statistics of the surrounding world (Fig. 9). An alternative interpretation of the DRN disruption effect is that fewer transitions are decoded because the two motivation states were more similar to one another. Although the motivation states identified during DRN TUS sessions were moderately more similar to one another than the motivation states identified in other conditions (see fig. S17), it is important to emphasize that the animals clearly exhibited two discrete motivation states after DRN TUS and that fewer transitions between states occurred in the DRN TUS condition (Fig. 7, F and G). Furthermore, the motivation-state transition interpretation is consistent with the correlation between transition events and DRN activity observed in fMRI recordings (Figs. 5B and 6). We found no evidence that DRN activity was related to motivation-state level (high versus low; Fig. 5B), which one would expect if DRN's role was to maintain or distinguish the motivation-states themselves.

Several previous studies have identified DRN activity in tasks where there is fundamental uncertainty about the reward available on each trial (3, 51, 52). In contrast, the properties of specific reward opportunities were explicitly signaled in our paradigm, and the key manipulation was the distribution of reward opportunities over
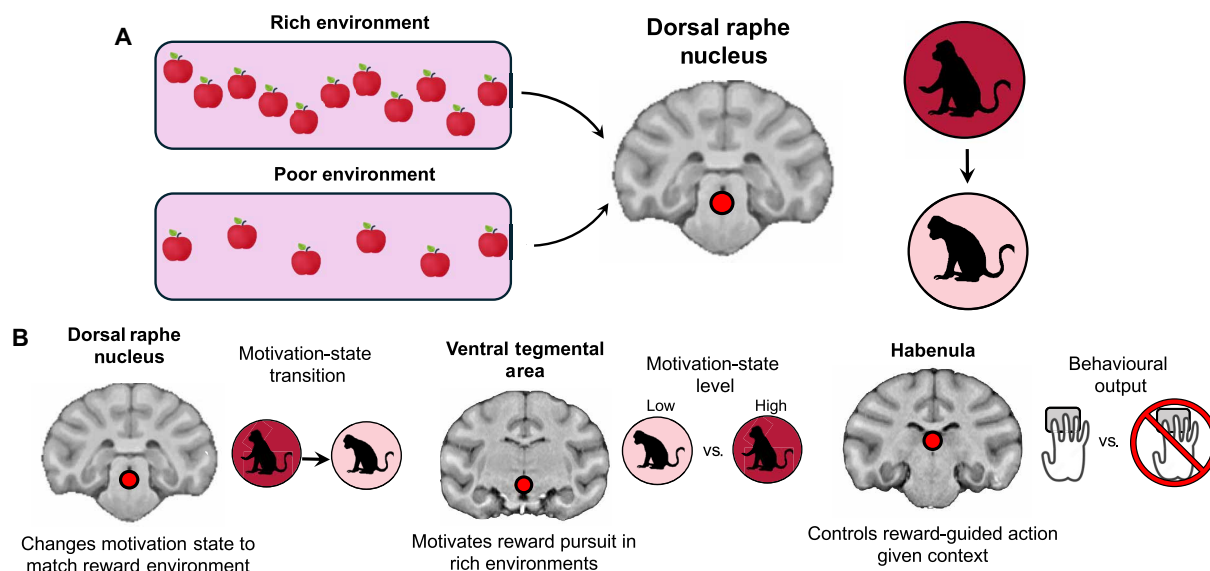
Fig. 9. DRN controls motivation-state transitions as a function of the environment. (A) We provide converging evidence from recoding and disruption techniques that DRN controls transitions between high-motivation states occupied when rewards are abundant, to low-motivation states engendered by reward scarcity. We started by showing that the purse-versus-reject decision made on each trial was influenced by an animal's recent history of reward (richness of the environment) and previous pursue-versus-reject behavior (behavioral history) on (Fig. 1, D to G). These effects indicated that behavior was autocorrelated, consistent with the idea that animals occupied an underlying hidden motivation state that could not be explained solely by the external environment (Fig. 2A). To test this hypothesis, we used a GLM-HMM framework. We identified state-like fluctuations in motivation to pursue rewards (Fig. 2, C to E) that were discrete, persistent, and biologically meaningful (Fig. 4). Using an optimized fMRI approach, we then identified three nuclei (DRN, HB, and VTA) that tracked animals' external environment and GLM-HMM–derived motivation states. HB and VTA encoded the level of motivation state (whether animal was in a high or low motivation-state; Fig. 5B). DRN specifically encoded transitions between high-to-low motivation states (Fig. 5B). Last, we complemented the fMRI results with a causal intervention. We showed that disrupting DRN perturbed the previously observed effects of the external environment on behavior (Fig. 7, D and E) and, importantly, reduced the likelihood of transitions in motivation-states (Fig. 7, F to H). (B) We thereby distinguish DRN's function from interconnected subcortical nuclei, which are also implicated in motivation. In combination with previous studies, we suggest a complementary scheme in which DRN (left) controls transitions between high-to-low motivation states, VTA (middle), implements and maintains the changes in reward sensitivity that characterize motivation states, and habenula (right) integrates an animal's internal and external context to control its behavioral output.

time. This might explain the absence of reward-uncertainty signals in DRN fMRI recordings. It might similarly explain VTA's involvement in the task. Although it is perhaps unexpected that VTA did not feature more prominently given its well-documented links to reward and motivation, the task did not require reinforcement learning, which is a key function of dopaminergic nuclei (53). Instead, we observed patterns of VTA activity corresponding to (i) the richness of the environment after pursuits and (ii) trial-by-trial motivation-state level, which suggest a cognate role in driving behavior in proximity to rewards (Fig. 9) (23).

How might DRN coordinate with other regions of the primate brain to control decisions? Unlike other subcortical ROIs, we found that activity in Hb covaried with both an animal's pursue/reject decisions and the individual factors that shaped them. This is consistent with an emerging theory that Hb integrates motivationally salient information from its diverse range of afferent connections to control motor output via the basal ganglia (Fig. 9) (42, 54–56). Similar patterns of activity occurred in AI, which is perhaps the computational source of representations concerning rewards obtained over multiple time points (9, 42, 57). Although it is difficult to specify directions of influence from fMRI recordings, we observed patterns of functional connectivity consistent with communication between AI-DRN, AI-VTA, and AI-Hb—a scheme that matches well-documented patterns of anatomical connection (58, 59). Cortico-subcortical interactions of this kind echo a common neural motif whereby information synthesized in the cortex is transmitted to subcortical nuclei, which have wide-ranging efferent connections capable of orchestrating brain-wide activity (42, 55).

The fact that animals were ceteris paribus more likely to pursue rewards in high-value environments controverts prominent formulations of optimal reward-seeking (1), which argue that animals should be less selective during periods of reward scarcity on account of their homeostatic requirements. Nevertheless, the pattern observed ensures that animals maximize their reward intake in the manner that these theories envisage: Given that animals usually lack perfect knowledge about the richness or sparseness of future opportunities, a reasonable strategy is to concentrate their efforts on rich environments where they have the most to gain (9, 10, 23). In support of this interpretation, the GLM-HMM analysis identified persistent motivation states in monkey behavior that were positively correlated with the availability of rewards.

Last, many neurons in DRN are serotonergic and that DRN is the principal source of serotonin to the mammalian forebrain (60, 61). Although the techniques we implemented are not serotonin specific, our results are consistent with several proposals about its function. We found (i) a positive correlation between DRN activity and reward scarcity, evoking classic theories that link serotonin to punishments and reward omissions (24, 62–64), and (ii) a causal relationship between DRN and changes in behavior, akin to serotonin's role in choice perseveration (4, 5). The key finding in our experiment relative to previous work is that DRN controlled patterns of behavior that unfolded over multiple time points and not just specific reinforcement events or decisions. In a similar vein, it is possible that serotonin is important not for moment-to-moment decision-making per se but for changing the overarching strategy of an animal's behavior and perhaps especially for changes that involve adverse feedback from the environment. Testing this hypothesis is a promising avenue for future experiments.

MATERIALS AND METHODS

Experimental model and subject details
The experiments were performed with four male rhesus macaques (*M. mulatta*). The sample size corresponded to those used in previous studies in which it had been possible to identify significant and reliable whole-brain fMRI recording and TUS across multiple testing sessions but with the minimum number of animals. The animals were 10 to 12 years of age and weighed 14.30 to 17.09 kg. They lived in group housing with a 12-hour light-dark cycle and were afforded access to water for 12 to 16 hours on testing days and free access on nontesting days. All procedures were conducted under a license issued by the UK Home Office in accordance with the Animal (Scientific Procedures) Act 1986 and the European Union guidelines (EU Directive 2010/63/EU).

Behavioral training
All monkeys were fitted with MRI compatible cranial implants that facilitated head fixation during testing and training. Training was conducted in MRI compatible chairs designed to accommodate monkeys in a sphinx position and took place in custom-built environments that replicated an MRI scanner. The animals were trained on simplified versions of the task in which they did not need to wait go-cue and action-outcome delays and were not exposed to controlled changes in their reward environment. Go-cue and action-outcome delays were then introduced and carefully increased over the course of training until they were sufficient for an fMRI experiment (see below). Changes in the reward environment were introduced after the animals were accustomed to the final delay timings. Testing for both the fMRI and TUS experiments began when the proportion of pursued trials per session was stationary over consecutive days.

Behavioral task
Animals performed a simple decision-making task involving sequential encounters with reward opportunities that appeared on a computer screen. Reward opportunities were presented in visual form via colored boxes (size of the box: 8 cm by 26 cm) that were filled with dots. The color of the stimulus indicated the number of juice drops that the opportunity was worth (red = 1; green = 2; blue = 3). The number of dots comprising the stimulus indicated the probability that reward would be delivered if the opportunity was pursued. Reward probabilities ranged between (0.05, 1), and dots indicated 0.05 linear increments of reward probability. The stimuli were designed so that dots gradually filled the box from the top downward.

Opportunities first appeared in the center of the screen before displacing either right or left (Fig. 1A). The displacement of the opportunity stimulus functioned as a go cue, which indicated that the opportunity was available for pursuit. Monkeys could pursue opportunities by manually responding to an infrared sensor corresponding to the opportunity's on-screen location—For example, if the opportunity displaced right, they needed to touch a sensor with their right hand, and vice versa if it displaced left. The go cue was designed to temporally dissociate decisions about pursuing opportunities from the motoric processes, which realized decisions in behavior. Similarly, the randomized right/left displacement of the opportunity prevented monkeys from motor planning during the decision phase. The durations separating opportunity onsets and go cues were drawn from *go-cue* ~ uniform(3, 4) and were optimized

for a canonical macaque hemodynamic response function (HRF) of approximately 4 s. Monkeys had 1 s to pursue opportunities after the go cue. After pursuing opportunities, the monkeys needed to wait action-outcome delays spanning *A-O-delay* ~ uniform(3, 4) before receiving potential juice rewards accompanied by visual reward feedback. The next trial then began after an inter-trial interval (ITI) spanning *ITI* ~ N(1, 4). If the monkeys did not pursue an opportunity, they bypassed the action-outcome and reward-delivery durations and proceeded to the next trial. If the monkeys pursued opportunities before the go cue—i.e., prematurely—they needed to wait for the remaining duration of the opportunity presentation, in addition to go-cue, action-outcome, and reward-feedback durations. The opportunity that elicited the premature response was then repeated on the next trial.

Each session of the task comprised four blocks of 40 to 50 trials. Blocks were used to engender different reward environments by systematically controlling the reward-probability aspect of opportunities. Environments varied on two dimensions: (i) Richness was defined by the mean of the reward-probability distribution, where $\mu_{rich} = 0.80$ and $\mu_{poor} = 0.55$, and (ii) stochasticity was defined by the width of reward-probability distributions, where predictable environments were generated from Gaussian distributions with $\sigma_{predictable} = 0.05$ and unpredictable environments from random-uniform distributions with ranges of 0.40 such that $\sigma_{unpredictable} = 0.13$. Blocks were characterized along both richness and stochasticity dimensions, which yielded four distinct environment types—rich-predictable, rich-stochastic, poor-predictable, and poor-stochastic. Each session featured one of each environment-type, and the order of environments was counterbalanced with respect to the richness dimension to avoid long periods of low-value offers (e.g., during consecutive poor environments), which were difficult for monkeys to perform. Environment types were indicated by visual cues that bordered the screen and indicated what kind of block the animals were currently in. The visual stimuli were abstract black and white patterns (see the diagonal dashed lines bordering the screen in the diagram of the task in Fig. 1A).

The task was implemented in MATLAB v2019 by MathWorks using Psychophysics Toolbox v3 (*65*) and presented on MRI-compatible screens (23in BOLD screen; Cambridge Research Systems) approximately 30 cm away from the subjects. Juice rewards consisted in a solution of water, blackcurrant cordial, and banana, and each juice drop was 1 ml. Pupillometric data were obtained during fMRI sessions via an MR-compatible infrared EYElink 1000 eye-tracker device by SR Research Ltd. recording pupil diameter and gaze direction along *x* and *y* planes at a 250-Hz sampling rate.

Preprocessing and analysis of pupillometric data were performed in R and entailed the following steps [adapted from (*66*)]. Samples reflecting eye blinks were identified with a detection algorithm native to the EYElink device. Artefacts were defined as consecutive samples with discrepancies of >50 a.u. All samples within symmetric 25 sample (i.e., 0.1 s) windows of artefacts or eye blinks were scrubbed and linearly interpolated, and interpolated time series were then low-pass filtered with a 4-Hz cutoff. These processes were performed for pupil size, *x*-gaze direction, and *y*-gaze direction time series, respectively, which were *z*-scored. We then regressed variance due to *x*-gaze direction and *y*-gaze direction from pupil size via linear regression and performed subsequent analysis using the residuals.

To incorporate pupillometry in our behavioral analysis, we extracted pupil size in specific epochs of the task, with a focus on periods just before and after the presentation of reward opportunities in which animals were making decisions. In particular, we quantified—for each trial—mean pupil size 0.5 to 0 s before (ITI phase) and 0.5 to 1.0 s after (decision phase) stimulus presentation. We did not select the first 0.5 s to avoid pupillometric responses reflecting sudden luminance changes, as when a visual stimulus first appears. When comparing the pupil size between motivation states (GLM2.4; Fig. 4F), we focused on pupil size during the decision phase. On all other occasions, when using it as a confound regressor to control for arousal, we focused on pupil size during the ITI phase. We chose this phase to make sure that it is not confounded by the brightness of the stimulus in front of the animal. Pupil size data were also used to reject periods of behavior, in which animals had their eyes shut (5.1% of the trials on average).

Animals were given four blocks of 40 to 50 trials. They performed one session per day, at the same time of the day.

Behavioral analysis

Our initial analysis of behavior characterized the influences on binomial pursuit/reject decisions. We accomplished this with mixed-effect binomial GLMs with subject identity as a random variable—This means that all GLMs account for inter-subject variability for the effect of interest. We began by investigating the time horizon of an animal's sensitivity to past rewards with the following GLM (GLM1.1)

$$
\begin{aligned}
\text{logit}(\text{decision}_t) = {} & \beta_0 + \beta_1 \text{reward-magnitude}_t \\
& + \beta_2 \text{reward-probability}_t \\
& + \beta_3 \left(\text{reward-magnitude}_t{}^* \text{reward-probability}_t\right) \\
& + \beta_4 \text{reward-outcome}_{t-1} + \beta_5 \text{reward-outcome}_{t-2} \\
& + \beta_6 \text{reward-outcome}_{t-3} \\
& + \beta_7 \text{reward-outcome}_{t-4} \\
& + \beta_8 \text{reward-outcome}_{t-5} + \beta_9 \text{reward-outcome}_{t-6} \\
& + \beta_{10} \text{reward-outcome}_{t-7} + \beta_{11} \text{reward-outcome}_{t-8} \\
& + \beta_{12} \text{reward-outcome}_{t-9} + \beta_{13} \text{reward-outcome}_{t-10} + \mu_0 + \varepsilon
\end{aligned}
$$

where β_{0-13} are fixed-effects (β_0 is the intercept), μ_0 is the by-subject random intercept, and ε is an error term. Decision is animals' pursue (coded as 1) versus reject (coded as 0) decisions. Reward-magnitude is the reward magnitude of the current opportunity, which varies between one and three drops of juice. Reward-probability is the reward probability of the current opportunity, which varies between 0.35 and 1. Reward-outcome is the actual received reward and varies between zero (no reward), one, two, or three drops of juice. The small number of subjects precluded fitting GLM1.1 with the full random-effects structure—that is, affording all predictors random effects for each subject. To assess the significance of past reward outcomes while also taking inter-subject variability into account, we therefore fit an iterative series of GLMs in which one of the previous outcome predictors was given a random slope (e.g., in the first GLM, reward-outcome$_{t-1}$ was given a random slope). We assessed the statistical significance of past

outcome regressors using the random effect in the relevant GLM. Figure 1D shows the outcome of this analysis. Given the results from GLM1.1, we operationalized the richness of an animal's reward environment as a moving average with a retrospective five-trial window, and the stochasticity of its environment as the SD of that reward rate as follows (richness and stochasticity of environment)

$$richness_t = \text{mean}\left(\text{reward-received}_{t-5} : \text{reward-received}_{t-1}\right)$$

$$stochasticity_t = \text{sd}\left(richness_{t-5} : richness_{t-1}\right)$$

We implemented the following model to assess the factors influencing pursuit/reject decisions (GLM1.2)

$$\begin{aligned}
\text{logit}\left(\text{decision}_t\right) = {} & \beta_o + \beta_1 \text{reward-magnitude}_t \\
& + \beta_2 \text{reward-probability}_t + \beta_3 \text{environment-richness}_t \\
& + \beta_4 \text{environment-stochasticity}_t \\
& + \beta_5 \left(\text{reward-magnitude}_t {}^* \text{reward-probability}_t\right) \\
& + \beta_6 \text{trial-number}_t + \mu_0 + \varepsilon
\end{aligned}$$

where β_{0-6} are fixed-effects, μ_0 is the by-subject random intercept, and ε is an error term. The small number of subjects precluded fitting GLM1.2 with the full random-effects structure. We therefore fit an iterative series of models in which the key predictors of interest—environment-richness and environment-stochasticity—were individually afforded random effects by animal. All mixed-effects GLMs were performed in R, and models were fit via maximum likelihood estimation as implemented in the lme4 toolbox. We then fit the following GLMs in using the same procedure (GLM1.3)

$$\begin{aligned}
\text{logit}\left(\text{decision}_t\right) = {} & \beta_o + \beta_1 \text{reward-magnitude}_t \\
& + \beta_2 \text{reward-probability}_t + \beta_3 \text{behavioral-history}_t \\
& + \beta_4 \text{environment-richness}_t + \beta_5 \text{environment-stochasticity}_t \\
& + \beta_6 \left(\text{reward-magnitude}_t {}^* \text{reward-probability}_t\right) \\
& + \beta_7 \text{trial-number}_t + \mu_0 + \varepsilon
\end{aligned}$$

where β_{0-6} are fixed-effects, μ_0 is the by-subject random intercept and ε is an error term.

Behavioral modeling

We used a GLM-HMM to probe time-varying patterns in behavior that might reflect internal motivation states. A GLM-HMM is an extension of the GLM that consists of two parts. The HMM part of the model assumes that time series events arise from so-called hidden states, which produce observations according to state-dependent probability distributions (18, 19). These states bear Markov relations to one another, such that the state at time t is determined strictly by the state at time $t - 1$, and the respective probabilities of transition between states. A simple HMM of animal behavior in this context, for example, might comprise states in which responses arise from binomial distributions with state-dependent probability parameters that reflect changes in an animal's motivation (i.e., probability of pursuing) across time.

The GLM part of a GLM-HMM parameterizes state-dependent probability distributions according to predictors (18). The weights

afforded to each predictor can vary from state to state, which enables the model to capture time-varying differences in an animal's decision-making process. We implemented a GLM-HMM in which pursuit/reject decisions were parameterized by bias terms, in addition to predictors for the expectation value of the specific opportunity animals were faced with on the current trials (cued to the animals by the color of the stimulus and the number of dots comprising the stimulus) and binary environment cues indicating the experimentally manipulated richness and stochasticity of the reward environment (cued to the animals by abstract black and white patterns bordering the screen). We were interested in internal states of motivation—that is, changes in an animal's intrinsic propensity to pursue rewards. We therefore let only the bias parameter(s) change between states and constrained the weight on expectation value and environment-cue parameters to be the same across states. Formally

$$p\left(y_t = pursue \mid z_t = k, \mathbf{x}_t, \mathbf{w}_k\right) = \frac{1}{1 + e^{-\mathbf{w}_k \cdot \mathbf{x}_t}}$$

where $y \in \{pursue, reject\}^{\text{T}}$ is the animal's decisions across T trials, z_t is the HMM state at trial t and $z_t \in \{1, ..., K\}$ such that there are K states, $x_t \in \mathbb{R}^{\text{M}}$ is a matrix of M predictors in the GLM part of the model at time t, and $\mathbf{w}_k \in \mathbb{R}^{\text{M}}$ represents M GLM weights over predictors that are specific to state k. Transitions between the states unfold according to a transition matrix $A \in \mathbb{R}^{K \times K}$

$$p\left(z_{t+1} = k \mid z_j, A\right) = A_{j,k}$$

where transition matrix is stationary. The joint probability of behaviors and HMM states is given by

$$p\left(\mathbf{y}, \mathbf{z} \mid \{\mathbf{x}\}_{t=1}^T, \{\mathbf{w}\}_{k=1}^K, A, \tau\right) = p\left(z_1\right) \prod_{t=2}^{T} p\left(z_{t+1} \mid z_t\right) p\left(y_t, \mathbf{x}_t\right)$$

where $\tau \in \mathbb{R}^K$ is the initial state distribution such that $\tau_k = p\left(z_1 = k\right)$. The likelihood function was fit to behavioral data via Markov chain Monte Carlo optimization implemented in STAN (67). For fMRI sessions, GLM-HMMs were fit to each individual animal's behavior. For TUS sessions, GLM-HMMs were fit for each stimulation condition.

We probed the explanatory benefit of the GLM–HMM framework by fitting a series of models with $K \in \{1, 2, 3, 4, 5\}$ states. This involved a five-fold cross-validation protocol in which models were iteratively fit to 80% of sessions and tested on the remaining 20% of sessions (18). This process was performed within each animal. The best fitting model was determined on the basis of the log-likelihood of held-out sessions (see fig. S3). In Fig. 2, we further report session-wise Akaike Information Criterion (AIC) scores—a well-established and easily interpretable metric of model fit that penalizes model complexity and is a demanding test for GLM-HMMs with $K \geq 2$ because additional states increase complexity exponentially.

We further validated the GLM-HMM framework using simulated data. To begin, we tested whether the fitting procedure accurately recovered the parameters used to generate simulated datasets by taking a fitted binomial GLM (i.e., $K = 1$) and two-state GLM-HMM for an example animal and generating 10 datasets comprising 16 sessions from each model (i.e., the typical size of the behavioral data recorded from each animal). We then implemented the fitting procedure on simulated datasets and tested (i) whether the correct number of HMM states was recovered (one state versus two states), and (ii) whether the correct parameters were identified in the two-state case

(see fig. S8). We additionally examined simulated data for qualitative patterns that were reminiscent animal behavior (Fig. 3). After model selection, we decoded the maximum a posteriori sequence of HMM states in each session via the Viterbi algorithm (*67*)

$$\mathbf{z}_s^* = argmax_{\mathbf{z}_{1:T}} p(\mathbf{z}_{s,1:T} \mid \mathbf{y}_{s,1:T})$$

We used simulated data from a two-state GLM-HMM to validate the Viterbi decoding algorithm (Fig. 4, A to C). Decoding the HMM states allowed us to quantify their impact on behavior, which we did using the following mixed-effects GLMs with subject identity as the grouping variable (GLM2.1)

$$\begin{aligned}\text{logit}(\text{decision}_t) &= \beta_o + \beta_1 \text{reward-magnitude}_t \\ &+ \beta_2 \text{reward-probability}_t + \beta_3 \text{motivation-level}_t \\ &+ \beta_4 \text{ reward-magnitude}_t{}^* \text{reward-probability}_t \\ &+ \beta_5 \text{trial-number}_t + \mu_0 + \mu_1 \text{motivation-level}_t + \varepsilon\end{aligned}$$

where β_{0-5} are fixed-effects, μ_{0-1} are by-subject random effects for intercepts and motivation level, respectively, and ε is an error term. Motivation-level $\in \{0,1\}$ is a binary variable reflecting the motivation state (low versus high) occupied on trial t. We identified motivation-state transitions by searching for cases where decoded motivation states were different across consecutive trials (i.e., motivation-level$_t \neq$ motivation-level$_{t-1}$). We compared behavior before-versus-after motivation-state transitions with the following mixed-effects GLM (GLM2.2):

$$\begin{aligned}\text{logit}(\text{decision}_t) &= \beta_o + \beta_1 \text{reward-magnitude}_t \\ &+ \beta_2 \text{reward-probability}_t \\ &+ \beta_3 \text{before-versus-after-transition}_t \\ &+ \beta_4 \text{ reward-magnitude}_t{}^* \text{reward-probability}_t \\ &+ \beta_5 \text{trial-number}_t \ \mu_0 + \mu_1 \text{motivation-level}_t + \varepsilon\end{aligned}$$

where β_{0-5} are fixed-effects, μ_{0-1} are by-subject random effects for intercepts and time from transition, respectively, and ε is an error term. Before-versus-after transition was a binary variable that grouped trials into two categories: (i) "before-transition" trials, defined as the five trials preceding a motivation-state transition, and (ii) "after-transition" trials, defined as the five trials in the immediate aftermath of a transition. We performed GLM2.2 separately for low-to-high and high-to-low motivation-state transitions. We then characterized the GLM-HMM's convergent validity by comparing decoded states with variables that should change with motivation: (i) reaction-time, and (ii) pupil-size. To do this, we implemented the following mixed-effects GLMs (GLM2.3 and GLM2.4)

$$\begin{aligned}\text{log}(\text{Reaction-time}_t) &= \beta_o + \beta_1 \text{reward-magnitude} \\ &+ \beta_2 \text{reward-probability} + \beta_3 \text{HMM-state} \\ &+ \beta_4 \text{trial-number} + \mu_1 \text{HMM-state} + \varepsilon\end{aligned}$$

$$\begin{aligned}\text{Pupil-size}_t &= \beta_o + \beta_1 \text{reward-magnitude} + \beta_2 \text{reward-probability} \\ &+ \beta_3 \text{HMM-state} + \beta_4 \text{trial-number} + \mu_1 \text{HMM-state} + \varepsilon\end{aligned}$$

where β_{0-4} are fixed effects and μ_1 are random effects within each subject. We did not implement random intercepts (i.e., μ_0) for either model because pupil size and reaction time were z-scored within

subjects during preprocessing. This meant that the central tendency of within-subject distributions for reaction time and pupil size was the same value (i.e., 0), and it was not necessary to add terms capturing inter-subject differences.

We used changes in decoded HMM states to examine the temporal dynamics of motivation states. We were particularly interested in links between motivation states and aspects of an animal's external environment, like the recent history of reward opportunities it had been presented with. We quantified the latter by calculating a moving average of the expectation value of reward opportunities received in the previous five trials.

$$Ave.EV_t = \text{mean}(\text{EV}_{t-5}: \text{EV}_{-1})$$

We used availability of rewards (Ave. EV) to predict motivation states via the following mixed-effects binomial GLM (GLM2.5)

$$\begin{aligned}\text{logit}(\text{motivation-state}_t) &\in \{0, 1\} = \beta_o \\ &+ \beta_1 \text{EV} + \beta_2 \text{Ave. EV} + \beta_3 \text{trial-number} + \mu_0 + \mu_1 \text{Ave. EV} + \varepsilon\end{aligned}$$

where β_{0-3} are fixed effects and μ_{0-1} are random effects within each subject. We used the EV—the expected value of the current offer on a given trial—to control for the effects of the current offer on an animal's motivation state. We used EV rather than the separate reward-magnitude and reward-probability dimensions of offers because Ave. EV and reward probability were meaningfully correlated ($r = 0.49$). This correlation was an inherent feature of the task because reward environments were engendered by manipulating the distribution of reward probabilities over time. This created autocorrelations in reward probabilities over time and, therefore, correlations between reward-probability and retrospective summary statistics of task events like the Ave. EV. Scaling reward probabilities by reward magnitude as in the expectation-value formula dampened this correlation ($r = 0.23$), meaning that EV and Ave. EV could be part of the same regression without collinearity.

Acquisition, reconstruction, and preprocessing of MRI data

During collection of fMRI data, the monkeys were head-fixed in a sphinx posture within an MRI compatible chair by Rogue Research. MR images were acquired with a horizontal bore clinical 3T scanner with a 15-channel nonhuman primate–specific receive coil by RAPID Biomedical. Structural images were acquired during a previous experiment (*36*). Functional images were acquired via the CMMR multiband gradient-echo T2* echo planar imaging (EPI) sequence designed specifically to achieve high signal-to-noise in subcortical structures (*68*, *69*). This was characterized by 1.25-mm isotropic voxels with a repetition time (TR) of 1282 ms, echo time (TE) of 25.40 ms, multiband acceleration factor MB = 2, in-plane acceleration factor R = 2, and flip angle of 63°. The functional scans covered the whole brain (field of view, 120 mm). There were 42 slices at coronal orientation with foot-to-head phase encoding direction and 1.25-mm thickness.

Offline reconstruction of the raw functional data was performed following the dynamic off-resonance correction method developed by Shahdloo and colleagues (*70*). In summary, standard Nyquist ghost correction and dynamic zeroth-order B0 correction were applied first. Then, the EPI reference navigator data acquired at every time point were compared to navigator data from single-band references to estimate first-order dynamic off-resonance perturbations

arising from the awake animal's body movements. Last, the off-resonance estimates were used to correct the raw data before reconstruction.

Preprocessing of MR images was performed with a combination of FMRIB's software library, Advanced Normalization Tools, the Human Connectome Project Workbench, and Magnetic Resonance Comparative Anatomy Toolbox (71). Although monkeys were head-fixed during MRI acquisition, incidental limb and body movements caused time-varying distortions in the B_0 magnetic field and therefore nonlinear motion artefacts along the phase encoding direction. To account for this, a low-noise EPI volume was identified for each session and then implemented as a reference to which other volumes were nonlinearly registered slice by slice along the phase-encode direction. Aligned and distortion-corrected EPIs were then registered nonlinearly first to monkey-specific high-resolution images and then to a group template in CARET f99 macaque space. Further details of the group template construction are described elsewhere (37). Last, the functional images were temporally filtered (high-pass temporal filtering, 3-dB cutoff of 100 s) and spatially smoothed (Gaussian spatial smoothing, full-width half maximum of 2.5 mm).

Three measures were used to detect artefacts in the data: (i) For each slice in each volume the linear transform (in the y plane) from that slice to the corresponding slice in the mean reference image; (ii) the normalized correlation between that slice and the corresponding slice in the mean reference image; (iii) for each volume, the correlation between that volume (mean-filtered across z slices) and the mean reference image after correction. Volumes were removed when they exceeded 2.5 SDs above the median of each measure. The threshold was chosen to keep the number of censored volumes less than 10% of the total volumes. We also added 13 principal components analysis components describing, for each volume, the warping from that volume to the mean reference image when correcting motion artefacts (i.e., they capture signal variability associated with motion induced distortion artefacts), as parametric regressors of noninterest that were not convolved in our GLMs.

fMRI data analysis

We focused our analysis of fMRI data on circumscribed a priori ROIs comprising DRN, VTA, SN, NB, LC, and Hb in the subcortex, and ACC, AI, and SMA in the cortex. Subcortical and SMA ROIs consisted in anatomical masks that were drawn on a group structural template in CARET F99 macaque monkey space and then warped to individual structural and functional spaces by nonlinear transformation. These masks were constructed separately by two different assessors based on the Rhesus Monkey Brain Atlas (72) and then evaluated on for convergence across assessors. ACC and AI ROIs were defined as 3-mm spheres centered on the peak of functionally relevant activation contrasts obtained in previous studies (9, 36).

We extracted the filtered time series of BOLD signal from each ROI. The extracted signals were then averaged, normalized, and up-sampled by a factor of 15 (9, 36, 37). The upsampled data were then epoched to 6-s time windows spanning 1 s before to 5 s after the appearance of the reward-opportunity stimulus on each trial. We then examined the relationship between behavior and brain activity with ordinary least squares GLMs performed at each time point in each epoch.

Inferential statistics for time-course GLMs were performed using a leave-one-out cross-validation procedure designed to estimate the peak regression coefficient in each session without selection bias (9, 36, 37). For each session s (N = 59), we determined the time point t at which the largest absolute-value regression coefficient occurred in the remaining N − 1 (i.e., 58) sessions. We restricted our search for t to a 4-s window from 1 s after to 5 s after the decision-making time—that is, a 4-s window centered on the mean macaque HRF. We then calculated the regression coefficient in session s at time t. We repeated this iteratively for each session, which yielded a series of 59 regression coefficients. We performed significance testing on regression coefficients with one-sample two-tailed t tests. To control for multiple comparisons, we applied the Bonferroni-Holm correction for any analysis performed on more than one ROI. For each GLM, we repeated this for each regressor in each ROI. All time course analysis was conducted in MATLAB using custom analysis scripts.

We implemented the following series of GLMs (GLM3.1, GLM3.2, GLM3.3, and GLM3.4)

$$\text{BOLD}_{\text{ROI}} = \beta_o + \beta_1 \text{reward-magnitude}$$
$$+ \beta_2 \text{reward-probability} + \beta_3 \text{environment-richness}$$
$$+ \beta_4 \text{environment-stochasticity} + \beta_5 \text{behavioral-history}$$
$$+ \beta_6 \text{environment-richness}^* \text{behavioral-history}$$
$$+ \beta_7 \text{environment-stochasticity}^* \text{behavioral-history}$$
$$+ \beta_8 \text{pupil-size} + \beta_9 \text{trial-number}$$

where BOLD_{ROI} indicates a t-by-s matrix containing time-series data for a given ROI (where t is trial, and s is time sample). Pupil size and trial number were added as confound regressors to control for the effects of arousal and fatigue, respectively.

$$\text{BOLD}_{\text{ROI}} = \beta_o + \beta_1 \text{reward-magnitude}$$
$$+ \beta_2 \text{reward-probability} + \beta_3 \text{environment-richness}$$
$$+ \beta_4 \text{environment-stochasticity} + \beta_5 \text{pupil-size} + \beta_6 \text{trial-number}$$

GLM3.2 was performed separately for subsets of trials in which animals pursued and rejected the previous opportunity, respectively.

$$\text{BOLD}_{\text{ROI}} = \beta_o + \beta_1 \text{decision} + \beta_2 \text{pupil-size} + \beta_3 \text{trial-number}$$

where decision is the pursue-versus-reject decision made on a given trial.

$$\text{BOLD}_{\text{ROI}} = \beta_o + \beta_1 \text{motivation-transition-period}$$
$$+ \beta_2 \text{motivation-state-level} + \beta_3 \text{pupil-size}$$
$$+ \beta_4 \text{trial-number}$$

where transition-period is a dummy-coded variable covering symmetric 7-trial windows (t − 3:t + 3) around decoded motivation-state transitions and motivation-state-level is the decoded motivation-state (high versus low) for a given trial. GLM3.4 was subsequently performed with 7-trial transition-periods aligned to from (t − 6:t + 0) to (t − 0:t + 6) with respect to decoded transitions to compare the timing of transition-related signals (see Fig. 6). It was also performed separately for low-to-high and high-to-low transition events (see Fig. 6).

We tested whether DRN coded the value of recently available rewards (Ave. EV) with the following GLM (GLM3.5)

$$\text{BOLD}_{\text{DRN}} = \beta_o + \beta_1 \text{EV} + \beta_2 \text{Ave. EV}$$
$$+ \beta_3 \text{pupil-size} + \beta_4 \text{trial-number}$$

GLM3.5 was performed on data from DRN alone, as it was designed to test a specific hypothesis arising from the combined neural and behavioral data. We used EV to control for the effects of the currently available offer instead of separate reward-magnitude and reward-probability regressors as in previous GLMs of BOLD signal. This is due to collinearity between Ave. EV and reward-probability regressors (see GLM2.5 for further explanation).

We probed changes in connectivity as a function of richness of the environment using the following PPI-GLM (GLM3.6)

$$\text{BOLD}_{\text{ROI}} = \beta_o + \beta_1\text{BOLD}_{\text{seed}} + \beta_2\text{PPI}$$
$$+ \beta_3\text{environment-richness} + \beta_4\text{reward-magnitude}$$
$$+ \beta_5\text{reward-probability} + \beta_6\text{environment-stochasticity}$$
$$+ \beta_7\text{pupil-size} + \beta_8\text{trial-number}$$

where $\text{BOLD}_{\text{seed}}$ is a t-by-s matrix containing time-series data for seed regions in PPI analysis, and PPI is the interaction between $\text{BOLD}_{\text{seed}}$ and environment-richness regressors. GLM3.6 was performed separately for subsets of trials in which an animals pursued and rejected the previous opportunity, respectively. Analogously, we tested differences in functional connectivity with respect to (i) motivation-state level, and (ii) motivation-state transitions with the following PPI (GLM3.7)

$$\text{BOLD}_{\text{ROI}} = \beta_o + \beta_1\text{BOLD}_{\text{seed}}$$
$$+ \beta_2\text{PPI} + \beta_3\text{motivation-transition-period}$$
$$+ \beta_4\text{motivation-state-level} + \beta_5\text{pupil-size}$$
$$+ \beta_6\text{trial-number}$$

where $\text{BOLD}_{\text{seed}}$ is a t-by-s matrix containing time-series data for seed regions in PPI analysis. The analysis was performed twice—once where PPI was the interaction between $\text{BOLD}_{\text{seed}}$ and motivation-transition-period regressors, and once where PPI was the interaction between $\text{BOLD}_{\text{seed}}$ and motivation-state-level regressors. Last, we tested differences in functional connectivity with respect to pursue/reject decisions with the following PPI (GLM3.8)

$$\text{BOLD}_{\text{ROI}} = \beta_o + \beta_1\text{BOLD}_{\text{seed}} + \beta_2\text{PPI}$$
$$+ \beta_3\text{decision} + \beta_4\text{pupil-size} + \beta_5\text{trial-number}$$

where $\text{BOLD}_{\text{seed}}$ is a t-by-s matrix containing time-series data for seed regions in PPI analysis, and PPI is the interaction between $\text{BOLD}_{\text{seed}}$ and decision regressors.

Transcranial ultrasound stimulation
TUS was performed using a four-element annular array transducer (NeuroFUS CTX-500, 64-mm active diameter, Brainbox Ltd., Cardiff, UK) combined with a programmable amplifier (Sonic Concept Inc.'s Transducer Power Output System, TPO-105, Brainbox Ltd., Cardiff, UK). The transducer was paired with a transparent coupling cone filled with degassed water and sealed with a latex membrane. The water was degassed for 4 to 5 hours before each stimulation session and was replaced after each session. The resonance frequency of the ultrasonic wave was set to 500 kHz. The stimulation protocol was based on previously established protocols in macaques (73). We used the following protocol: duty cycle, 30%; pulse length, 30 ms; pulse repetition interval, 100 ms; and total stimulation duration, 30 s.

The pressure field from the transducer was measured in a water tank with a 75-µm diameter polyvinylidene difluoride needle hydrophone (Precision Acoustics, Dorset UK), which had been calibrated at 500 kHz by the National Physical Laboratory (Teddington, UK). The free-field spatial-peak pulse-average intensity (Isppa) at 60-mm focal depth was 120 W/cm^2, which was consistent with the output given by the transducer manufacturer.

At the beginning of each stimulation session, the animal's skull was shaved and a conductive gel (SignaGel Electrode; Parker Laboratories Inc.) was applied to the skin. The water-filled coupling cone and the gel was used to ensure ultrasonic coupling between the transducer and the animal's head. Next, the ultrasound transducer/coupling cone was placed on the skull and a Brainsight Neuronavigation System (Rogue Research, Montreal, CA) was used to position the transducer so that the focal spot would be centered on the targeted brain region. There were four stimulation conditions: (i) DRN (target of interest); (ii) VTA (subcortical active control condition); (iii) STS (cortical active control condition); (iv) sham (passive control condition). DRN and VTA targets were approximately 50 mm, and the STS 35 mm, from the surface of the transducer (the exact focal distance depended on the subject). All targets were sonicated bilaterally for 60 s in total, with 30 s of stimulation applied to a target from each hemisphere. Sonication of the midline targets (DRN and VTA) from one hemisphere was immediately followed by sonication of the same target from the contralateral hemisphere [cross-beam stimulation; (36)]. Sonication of the STS in one hemisphere was immediately followed by sonication of a homologous target in the contralateral hemisphere. Hemispheres were sonicated in a pseudo-random order. After stimulation, the monkeys were immediately moved to a testing room for behavioral data collection. The sham condition completely matched a typical stimulation session (setting, stimulation procedure, neuro-navigation, targeting, transducer preparation, and timing of its bilateral application to the shaved skin on the head of the animal) except that sonication was not triggered. During the sham session, the montage was pseudo-randomly positioned to target DRN, VTA, or STS. Each stimulation condition was repeated five times, on separate days, and the order of the stimulation sessions was pseudo-randomized for each animal. The stimulation was always performed at the same time of the day, and there was always a 24-hour gap between each session, regardless of it being a real or sham stimulation session.

Acoustic modeling
We simulated the propagation of acoustic waves produced by the TUS protocol as described by Yaakub and colleagues (74). In brief, we used k-wave—a k-space pseudospectral solver (75)—and kArray tools to obtain estimates for the pressure amplitude, peak intensity, and spatial distribution of TUS at steady state. First, we simulated the acoustic wave propagation in water (free field) to characterize ultrasound beam for a target intensity of 120 W/cm^2 at 50-mm focal depth (focal depth of VTA and DRN from transducer). Next, we performed the simulation for DRN and VTA targets in the skull. The skull was estimated from pseudo-CT images obtained from each monkey using a Black Bone MRI sequence (76). The skull was obtained by thresholding the pseudo-CT images at 1400 to 2100 Hounsfield Units (HU). A linear relationship between the pseudo-CT Hounsfield Units and the sound speed, as well as the density, and absorption coefficient was assumed as described elsewhere (77, 78). The simulation grid size was set to the size of the T1-weighted MRI with a grid

spacing of 0.5 mm, which results in approximately 6 points per wavelength in water and tissue and up to 12.4 points per wavelength in bone.

Resting-state imaging data acquisition, preprocessing, and analysis

We further validated the TUS protocol by examining its effect on resting state coactivation patterns between VTA/DRN and key interconnected cortical and subcortical regions. Awake resting-state fMRI (rs-fMRI) data were acquired for all four monkeys (the same animals as in experiments 1 and 2) pre-versus-post DRN-TUS and VTA-TUS, respectively. Preprocessing and analysis of rs-fMRI data have been described elsewhere (17, 73).

We characterized the effects of TUS on the coactivation of DRN/VTA and ROIs by comparing rs-fMRI data collected before DRN/VTA TUS with the rs-fMRI data collected immediately after DRN/VTA TUS. Prestimulation and poststimulation rs-fMRI were collected on the same day. The impact of DRN/VTA TUS on coactivation patterns was quantified with seed-based connectivity analyses, which involved calculating a series of pairwise linear correlations in BOLD activity between a seed region (DRN or VTA) and the remaining ROIs. The resulting prestimulation connectivity fingerprints for DRN-TUS and VTA-TUS were then contrasted with poststimulation connectivity fingerprints before DRN-TUS and VTA-TUS (see Fig. 7C).

TUS data analysis

We characterized the behavioral effect of TUS by examining its two-way interactions with key predictors of decision-making in a series of binomial GLMs. In these GLMs, the effect for TUS condition was constructed to compare each individual active stimulation condition (STS, VTA, and DRN) to sham-TUS as a reference. On some occasions, we compared DRN-TUS to one of the other active stimulation conditions—for example, DRN-TUS versus STS-TUS. We did this by performing GLMs on subsets of data that included only the TUS conditions of interest. These analyses are specifically noted in the main text. We tested the following (GLM4.1, GLM4.2, GLM4.3, GLM4.4, and GLM4.5) to examine how TUS modulates the richness of the environment and behavioral history effects, respectively.

$$\text{logit}\left(\text{decision}_t\right) = \beta_o + \beta_1\text{reward-magnitude}$$
$$+ \beta_2\text{reward-probability} + \beta_3\text{environment-richness}$$
$$+ \beta_4\text{environment-stochasticity} + \beta_5\text{ TUS-condition}$$
$$+ \beta_6\left(\text{reward-magnitude}^*\text{reward-probability}\right)$$
$$+ \beta_7\left(\text{environment-richness}^*\text{TUS-condition}\right)$$
$$+ \beta_8\left(\text{trial-number}\right) + \mu_0 + \varepsilon$$

where β_{0-8} are fixed effects, μ_0 is the by-subject random intercept, and ε is an error term.

$$\text{logit}\left(\text{decision}_t\right) = \beta_o + \beta_1\text{reward-magnitude}$$
$$+ \beta_2\text{reward-probability} + \beta_3\text{behavioral-history}$$
$$+ \beta_4\text{TUS-condition} + \beta_5\left(\text{reward-magnitude}^*\text{reward-probability}\right)$$
$$+ \beta_6\left(\text{behavioral-history}^*\text{TUS-condition}\right) + \beta_7\left(\text{trial-number}\right) + \mu_0 + \varepsilon$$

where β_{0-7} are fixed effects, μ_0 is the by-subject random intercept, and ε is an error term. The amount of data in each TUS condition for each subject prevented us from fitting models with more complex random-effects structures for these models. We examined the influence of TUS on the frequency of transitions between motivation states with the following GLM

$$\text{logit}\left(\text{state-transition}_t\right) = \beta_o + \beta_1\text{reward-magnitude}$$
$$+ \beta_2\text{reward-probability} + \beta_3\text{TUS-condition}$$
$$+ \beta_4\text{reward-magnitude}^*\text{reward-probability}$$
$$+ \beta_5\text{trial-number} + \mu_0 + \varepsilon$$

where, state-transition $\in \{0,1\}$, β_{0-5} are fixed-effects, μ_0 is the by-subject random intercept, and ε is an error term. We performed a similar GLM to examine the influence of TUS on high-to-low transitions specifically

$$\text{logit}\left(\text{state-decrease}_t\right) = \beta_o + \beta_1\text{reward-magnitude}$$
$$+ \beta_2\text{reward-probability} + \beta_3\text{TUS-condition}$$
$$+ \beta_4\text{reward-magnitude}^*\text{reward-probability}$$
$$+ \beta_5\text{trial-number} + \mu_0 + \varepsilon$$

where, state-decrease $\in \{0,1\}$, β_{0-5} are fixed-effects, μ_0 is the by-subject random intercept, and ε is an error term. We analyzed TUS's influence relationship between motivation states and the availability of rewards (Ave. EV) with the following GLM

$$\text{logit}\left(\text{motivation-state}_t\right) = \beta_o + \beta_1\text{EV}$$
$$+ \beta_2\text{Ave. EV} + \beta_3\text{TUS-condition}$$
$$+ \beta_4\text{TUS-condition}^*\text{Ave. EV}$$
$$+ \beta_5\text{trial-number} + \mu_0 + \varepsilon$$

where, motivation-state $\in \{0,1\}$ and the high-motivation state = 1 (i.e., positive coefficients correspond to increases in the likelihood of occupying the high-motivation state), β_{0-5} are fixed-effects, μ_0 is the by-subject random intercept, and ε is an error term. We performed GLM4.5 separately on behavioral data mean-split according to Ave. EV to determine whether DRN-TUS principally affected the influence between low-value environments and low-motivation states.

Supplementary Materials

REFERENCES AND NOTES

1. J. R. Krebs, J. T. Erichsen, M. I. Webber, E. L. Charnov, Optimal prey selection in the great tit (Parus major). *Anim. Behav.* **25**, 30–38 (1977).
2. D. W. Stephens, J. R. Krebs *Foraging Theory* (Princeton Univ. Press, 1986).
3. K. W. Miyazaki, K. Miyazaki, K. F. Tanaka, A. Yamanaka, A. Takahashi, S. Tabuchi, K. Doya, Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. *Curr. Biol.* **24**, 2033–2040 (2014).
4. H. E. M. den Ouden, N. D. Daw, G. Fernandez, J. A. Elshout, M. Rijpkema, M. Hoogman, B. Franke, R. Cools, Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* **80**, 1090–1100 (2013).
5. H. F. Clarke, J. W. Dalley, H. S. Crofts, T. W. Robbins, A. C. Roberts, Cognitive inflexibility after prefrontal serotonin depletion. *Science* **304**, 878–880 (2004).
6. E. S. Bromberg-Martin, O. Hikosaka, K. Nakamura, Coding of task reward value in the dorsal raphe nucleus. *J. Neurosci. Off. J. Soc. Neurosci.* **30**, 6262–6272 (2010).

7. K. Nakamura, M. Matsumoto, O. Hikosaka, Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus. *J. Neurosci. Off. J. Soc. Neurosci.* **28**, 5331–5343 (2008).

8. K. Hayashi, K. Nakao, K. Nakamura, Appetitive and aversive information coding in the primate dorsal raphé nucleus. *J. Neurosci.* **35**, 6195–6208 (2015).

9. M. K. Wittmann, E. Fouragnan, D. Folloni, M. C. Klein-Flügge, B. K. H. Chau, M. Khamassi, M. F. S. Rushworth, Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. *Nat. Commun.* **11**, 3771 (2020).

10. N. Khalighinejad, S. Manohar, M. Husain, M. F. S. Rushworth, Complementary roles of serotonergic and cholinergic systems in decisions about when to act. *Curr. Biol. CB* **32**, 1150–1162.e7 (2022).

11. L. Marshall, C. Mathys, D. Ruge, A. O. de Berker, P. Dayan, K. E. Stephan, S. Bestmann, Pharmacological fingerprints of contextual uncertainty. *PLoS Biol.* **14**, e1002575 (2016).

12. K. Miyamoto, T. Osada, R. Setsuie, M. Takeda, K. Tamura, Y. Adachi, Y. Miyashita, Causal neural network of metamemory for retrospection in primates. *Science* **355**, 188–193 (2017).

13. S. Bestmann, A. O. de Berker, J. Bonaiuto, Understanding the behavioural consequences of noninvasive brain stimulation. *Trends Cogn. Sci.* **19**, 13–20 (2015).

14. E. Dayan, N. Censor, E. R. Buch, M. Sandrini, L. G. Cohen, Noninvasive brain stimulation: From physiology to network dynamics and back. *Nat. Neurosci.* **16**, 838–844 (2013).

15. J. Kubanek, J. Brown, P. Ye, K. B. Pauly, T. Moore, W. Newsome, Remote, brain region–specific control of choice behavior with ultrasonic waves. *Sci. Adv.* **6**, eaaz4193 (2020).

16. T. Deffieux, Y. Younan, N. Wattiez, M. Tanter, P. Pouget, J.-F. Aubry, Low-intensity focused ultrasound modulates monkey visuomotor behavior. *Curr. Biol. CB* **23**, 2430–2433 (2013).

17. D. Folloni, L. Verhagen, R. B. Mars, E. Fouragnan, C. Constans, J.-F. Aubry, M. F. S. Rushworth, J. Sallet, Manipulation of subcortical and deep cortical activity in the primate brain using transcranial focused ultrasound stimulation. *Neuron* **101**, 1109–1116. e5 (2019).

18. Z. C. Ashwood, N. A. Roy, I. R. Stone, A. E. Urai, A. K. Churchland, A. Pouget, J. W. Pillow, Mice alternate between discrete strategies during perceptual decision-making. *Nat. Neurosci.* **25**, 201–212 (2022).

19. W. Zucchini, *Hidden Markov Models for Time Series: An Introduction Using R [Electronic Resource]. Monographs on Statistics and Applied Probability* (Series), (CRC Press/Taylor & Francis Group, 2009).

20. S. Mathôt, Pupillometry: Psychology, physiology, and function. *J. Cogn.* **1**, 16 (2018).

21. O. Hikosaka, The habenula: From stress evasion to value-based decision-making. *Nat. Rev. Neurosci.* **11**, 503–513 (2010).

22. M. Matsumoto, O. Hikosaka, Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* **447**, 1111–1115 (2007).

23. Y. Niv, N. D. Daw, D. Joel, P. Dayan, Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology (Berl)* **191**, 507–520 (2007).

24. N. D. Daw, S. Kakade, P. Dayan, Opponent interactions between serotonin and dopamine. *Neural Netw. Off. J. Int. Neural Netw. Soc.* **15**, 603–616 (2002).

25. A. J. Yu, P. Dayan, Uncertainty, Neuromodulation, and Attention. *Neuron* **46**, 681–692 (2005).

26. I. E. Monosov, D. A. Leopold, O. Hikosaka, Neurons in the primate medial basal forebrain signal combined information about reward uncertainty, value, and punishment anticipation. *J. Neurosci.* **35**, 7443–7459 (2015).

27. N. M. Ledbetter, C. D. Chen, I. E. Monosov, Multiple mechanisms for processing reward uncertainty in the primate basal forebrain. *J. Neurosci.* **36**, 7852–7864 (2016).

28. K. Zhang, C. D. Chen, I. E. Monosov, Novelty, salience, and surprise timing are signaled by neurons in the basal forebrain. *Curr. Biol.* **29**, 134–142.e3 (2019).

29. S. Yoo, D. R. Mittelstein, R. C. Hurt, J. Lacroix, M. G. Shapiro, Focused ultrasound excites cortical neurons via mechanosensitive calcium accumulation and ion channel amplification. *Nat. Commun.* **13**, 493 (2022).

30. D. G. Blackmore, D. Razansky, J. Götz, Ultrasound as a versatile tool for short- and long-term improvement and monitoring of brain function. *Neuron* **111**, 1174–1190 (2023).

31. Y. Shamli Oghli, T. Grippe, T. Arora, T. Hoque, G. Darmani, R. Chen, Mechanisms of theta burst transcranial ultrasound induced plasticity in the human motor cortex. *Brain Stimulat.* **16**, 1135–1143 (2023).

32. N. Samuel, K. Zeng, I. E. Harmsen, M. Y. R. Ding, G. Darmani, C. Sarica, B. Santyr, A. Vetkas, A. Pancholi, A. Fomenko, V. Milano, K. Yamamoto, U. Saha, R. Wennberg, N. C. Rowland, R. Chen, A. M. Lozano, Multi-modal investigation of transcranial ultrasound-induced neuroplasticity of the human motor cortex. *Brain Stimulat.* **15**, 1337–1347 (2022).

33. G. Darmani, T. O. Bergmann, K. Butts Pauly, C. F. Caskey, L. de Lecea, A. Fomenko, E. Fouragnan, W. Legon, K. R. Murphy, T. Nandi, M. A. Phipps, G. Pinton, H. Ramezanpour, J. Sallet, S. N. Yaakub, S. S. Yoo, R. Chen, Non-invasive transcranial ultrasound stimulation for neuromodulation. *Clin. Neurophysiol.* **135**, 51–73 (2022).

34. K. R. Murphy, J. S. Farrell, J. L. Gomez, Q. G. Stedman, N. Li, S. A. Leung, C. H. Good, Z. Qiu, K. Firouzi, K. Butts Pauly, B. P. T. Khuri-Yakub, M. Michaelides, I. Soltesz, L. de Lecea, A tool for monitoring cell type–specific focused ultrasound neuromodulation and control of chronic epilepsy. *Proc. Natl. Acad. Sci. U.S.A>* **119**, e2206828119 (2022).

35. T. Di Ianni, K. P. Morrison, B. Yu, K. R. Murphy, L. de Lecea, R. D. Airan, High-throughput ultrasound neuromodulation in awake and freely behaving rats. *Brain Stimulat.* **16**, 1743–1752 (2023).

36. N. Khalighinejad, A. Bongioanni, L. Verhagen, D. Folloni, D. Attali, J.-F. Aubry, J. Sallet, M. F. S. Rushworth, A basal forebrain-cingulate circuit in macaques decides it is time to act. *Neuron* **105**, 370–384.e8 (2020).

37. A. Bongioanni, D. Folloni, L. Verhagen, J. Sallet, M. C. Klein-Flügge, M. F. S. Rushworth, Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature* **591**, 270–274 (2021).

38. D. G. R. Tervo, E. Kuleshova, M. Manakov, M. Proskurin, M. Karlsson, A. Lustig, R. Behnam, A. Y. Karpova, The anterior cingulate cortex directs exploration of alternative strategies. *Neuron* **109**, 1876–1887.e6 (2021).

39. R. Quilodran, M. Rothé, E. Procyk, Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* **57**, 314–325 (2008).

40. F. M. Stoll, V. Fontanier, E. Procyk, Specific frontal neural dynamics contribute to decisions to check. *Nat. Commun.* **7**, 11990 (2016).

41. K. J. Friston, C. Buechel, G. R. Fink, J. Morris, E. Rolls, R. J. Dolan, Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* **6**, 218–229 (1997).

42. N. Khalighinejad, N. Garrett, L. Priestley, P. Lockwood, M. F. S. Rushworth, A habenula-insular circuit encodes the willingness to act. *Nat. Commun.* **12**, 6329 (2021).

43. Z. Mohammadi, Z. C. Ashwood, T. I. B. Laboratory, J. W. Pillow, Identifying the factors governing internal state switches during nonstationary sensory decision-making. bioRxiv [Preprint] (2024); 10.1101/2024.02.02.578482.

44. J. Reisinger, S. Basu, R. Bayardo, Hierarchical mixtures of glms for combining multiple ground truths, (2011).

45. M. Ito, K. Doya, Parallel representation of value-based and finite state-based strategies in the ventral and dorsal striatum. *PLoS Comput. Biol.* **11**, e1004540 (2015).

46. L. Hannah, D. Blei, W. Powell, Dirichlet Process Mixtures of Generalized Linear Models, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (JMLR Workshop and Conference Proceedings, 2010), pp. 313–320; https://proceedings.mlr.press/v9/hannah10a.html.

47. H. F. Clarke, S. C. Walker, H. S. Crofts, J. W. Dalley, T. W. Robbins, A. C. Roberts, Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. *J. Neurosci. Off. J. Soc. Neurosci.* **25**, 532–538 (2005).

48. E. Lottem, D. Banerjee, P. Vertechi, D. Sarra, M. oude Lohuis, Z. F. Mainen, Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nat. Commun.* **9**, 1000 (2018).

49. K. Miyazaki, K. W. Miyazaki, K. Doya, Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. *J. Neurosci. Off. J. Soc. Neurosci.* **31**, 469–479 (2011).

50. C. D. Grossman, B. A. Bari, J. Y. Cohen, Serotonin neurons modulate learning rate through uncertainty. *Curr. Biol.* **32**, 586–599.e7 (2022).

51. K. Miyazaki, K. W. Miyazaki, A. Yamanaka, T. Tokuda, K. F. Tanaka, K. Doya, Reward probability and timing uncertainty alter the effect of dorsal raphe serotonin neurons on patience. *Nat. Commun.* **9**, 2048 (2018).

52. Y.-Y. Feng, E. S. Bromberg-Martin, I. E. Monosov, Dorsal raphe neurons integrate the values of reward amount, delay, and uncertainty in multi-attribute decision-making. *Cell Rep.* **43**, 114341 (2024).

53. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).

54. O. Hikosaka, S. R. Sesack, L. Lecourtier, P. D. Shepard, Habenula: Crossroad between the basal ganglia and the limbic system. *J. Neurosci.* **28**, 11825–11829 (2008).

55. N. Khalighinejad, L. Priestley, S. Jbabdi, M. F. S. Rushworth, Human decisions about when to act originate within a basal forebrain–nigral circuit. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11799–11810 (2020).

56. E. S. Bromberg-Martin, Y.-Y. Feng, T. Ogasawara, J. K. White, K. Zhang, I. E. Monosov, A neural mechanism for conserved value computations integrating information and rewards. bioRxiv [Preprint] (2022); https://doi.org/10.1101/2022.08.14.503903.

57. D. Folloni, E. Fouragnan, M. K. Wittmann, L. Roumazeilles, L. Tankelevitch, L. Verhagen, D. Attali, J.-F. Aubry, J. Sallet, M. F. S. Rushworth, Ultrasound modulation of macaque prefrontal cortex selectively alters credit assignment–related activity and behavior. *Sci. Adv.* **7**, eabg7700 (2021).

58. L. Yetnikoff, A. Y. Cheng, H. N. Lavezzi, K. P. Parsley, D. S. Zahm, Sources of input to the rostromedial tegmental nucleus, ventral tegmental area, and lateral habenula compared: A study in rat. *J. Comp. Neurol.* **523**, 2426–2456 (2015).

59. P. L. A. Gabbott, T. A. Warner, P. R. L. Jays, P. Salway, S. J. Busby, Prefrontal cortex in the rat: Projections to subcortical autonomic, motor, and limbic centers. *J. Comp. Neurol.* **492**, 145–177 (2005).

60. W. Fu, V. Le Maître, V. Fabre, J.-F. Bernard, Z.-Q. David Xu, T. Hökfelt, Chemical neuroanatomy of the dorsal raphe nucleus and adjacent structures of the mouse brain. *J. Comp. Neurol.* **518**, 3464–3494 (2010).

61. B. L. Jacobs, E. C. Azmitia, Structure and function of the brain serotonin system. *Physiol. Rev.* **72**, 165–229 (1992).

62. M. J. Crockett, L. Clark, T. W. Robbins, Reconciling the role of serotonin in behavioral inhibition and aversion: Acute tryptophan depletion abolishes punishment-induced inhibition in humans. *J. Neurosci.* **29**, 11993–11999 (2009).

63. P. Dayan, Q. J. M. Huys, Serotonin in affective control. *Annu. Rev. Neurosci.* **32**, 95–126 (2009).

64. R. Cools, A. C. Roberts, T. W. Robbins, Serotoninergic regulation of emotional and behavioral control processes. *Trends Cogn. Sci.* **12**, 31–40 (2008).

65. M. Kleiner, D. Brainard, D. Pelli, A. Ingling, R. Murray, C. Broussard, What's new in psychtoolbox-3. *Perception* **36**, 1–16 (2007).

66. N. Tardiff, J. D. Medaglia, D. S. Bassett, S. L. Thompson-Schill, The modulation of brain network integration and arousal during exploration. *Neuroimage* **240**, 118369 (2021).

67. Stan Development Team, Stan Modeling Language Users Guide and Reference Manual (2022); stan-dev.github.io.

68. S. Moeller, E. Yacoub, C. A. Olman, E. Auerbach, J. Strupp, N. Harel, K. Uğurbil, Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn. Reson. Med.* **63**, 1144–1153 (2010).

69. K. Setsompop, B. A. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, L. L. Wald, Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Med.* **67**, 1210–1224 (2012).

70. M. Shahdloo, U. Schüffelgen, D. Papp, K. L. Miller, M. Chiew, Model-based dynamic off-resonance correction for improved accelerated fMRI in awake behaving nonhuman primates. *Magn. Reson. Med.* **87**, 2922–2932 (2022).

71. M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, S. M. Smith, FSL. *Neuroimage* **62**, 782–790 (2012).

72. G. Paxinos, *The Rhesus Monkey Brain in Stereotaxic Coordinates* (Academic, San Diego, 2000).

73. L. Verhagen, C. Gallea, D. Folloni, C. Constans, D. E. Jensen, H. Ahnine, L. Roumazeilles, M. Santin, B. Ahmed, S. Lehericy, M. C. Klein-Flügge, K. Krug, R. B. Mars, M. F. Rushworth, P. Pouget, J.-F. Aubry, J. Sallet, Offline impact of transcranial focused ultrasound on cortical activation in primates. *eLife* **8**, e40541 (2019).

74. S. N. Yaakub, T. A. White, J. Roberts, E. Martin, L. Verhagen, C. J. Stagg, S. Hall, E. F. Fouragnan, Transcranial focused ultrasound-mediated neurochemical and functional connectivity changes in deep cortical regions in humans. *Nat. Commun.* **14**, 5318 (2023).

75. B. T. Cox, S. Kara, S. R. Arridge, P. C. Beard, k-Space propagation models for acoustically heterogeneous media: Application to biomedical photoacoustics. *J. Acoust. Soc. Am.* **121**, 3453–3464 (2007).

76. K. A. Eley, S. R. Watt-Smith, F. Sheerin, S. J. Golding, "Black Bone" MRI: A potential alternative to CT with three-dimensional reconstruction of the craniofacial skeleton in the diagnosis of craniosynostosis. *Eur. Radiol.* **24**, 2417–2426 (2014).

77. L. Marsac, D. Chauvet, R. La Greca, A.-L. Boch, K. Chaumoitre, M. Tanter, J.-F. Aubry, Ex vivo optimisation of a heterogeneous speed of sound model of the human skull for non-invasive transcranial focused ultrasound at 1 MHz. *Int. J. Hyperth. Off. J. Eur. Soc. Hyperthermic Oncol. North Am. Hyperth. Group* **33**, 635–645 (2017).

78. J. K. Mueller, L. Ai, P. Bansal, W. Legon, Numerical evaluation of the skull for human neuromodulation with transcranial focused ultrasound. *J. Neural Eng.* **14**, 066012 (2017).