

# Population coding of strategic variables during foraging in freely moving macaques

Received: 24 July 2021

Accepted: 9 January 2024

Published online: 5 March 2024

 Check for updates


Neda Shahidi<sup>1,2,3</sup>, Melissa Franch<sup>1</sup>, Arun Parajuli<sup>1</sup>, Paul Schrater<sup>4,5</sup>, Anthony Wright<sup>1</sup>, Xaq Pitkow <sup>6,7,8,9,10,12</sup>  & Valentin Dragoi <sup>1,7,11,12</sup> 

Until now, it has been difficult to examine the neural bases of foraging in naturalistic environments because previous approaches have relied on restrained animals performing trial-based foraging tasks. Here we allowed unrestrained monkeys to freely interact with concurrent reward options while we wirelessly recorded population activity in the dorsolateral prefrontal cortex. The animals decided when and where to forage based on whether their prediction of reward was fulfilled or violated. This prediction was not solely based on a history of reward delivery, but also on the understanding that waiting longer improves the chance of reward. The task variables were continuously represented in a subspace of the high-dimensional population activity, and this compressed representation predicted the animal's subsequent choices better than the true task variables and as well as the raw neural activity. Our results indicate that monkeys' foraging strategies are based on a cortical model of reward dynamics as animals freely explore their environment.

While foraging, animals must continuously make decisions about where to search for food and when to move between possible food sources. To survive in environments with sparse resources, animals forage more effectively if they can predict future outcomes before they execute costly actions such as relocation<sup>1–3</sup>. Two major limitations of past neuroscience studies of foraging have impeded our understanding of this natural behavior. First, trial-based tasks are unable to expose the continuous decision-making process during food search and selection, and second, restraining body movements may substantially distort prediction of outcomes of dynamic food sources, as the perception of time is tightly linked to freedom of movement<sup>4,5</sup> and cortical dynamics<sup>6</sup>.

Trial-based tasks revealed that animals use reward history to detect changes in reward rates<sup>7,8</sup>. We often quantify how an animal adapts to these reward rates by tracking how often it chooses each available

option. For example, classical foraging theories revolve around the idea of the matching law<sup>9</sup>: animals dedicate time or effort to an option in proportion to its value according to its reward history. However, a neglected aspect of adaptive behavior is that the animals adjust their response rate, meaning that they choose 'when' to forage in addition to 'where' to forage. Choosing the response rate systematically is particularly efficient when the time of choice predicts the chance of receiving a reward, that is, in nature as well as in many foraging studies<sup>7–13</sup>. For a restrained animal engaged in a trial-based foraging task, 'when' to choose is distorted by the trial structure, while 'where' to choose is distorted by a confined spatial distribution of reward. Additionally, examining foraging in trial-based tasks makes it difficult to examine the neural bases of the continuous decisions the animal would make freely about when and where to engage with the task.

<sup>1</sup>Department of Neurobiology and Anatomy, McGovern Medical School, University of Texas, Houston, Houston, TX, USA. <sup>2</sup>Georg-Elias-Müller-Institute for Psychology, Georg August-Universität, Göttingen, Germany. <sup>3</sup>Cognitive Neuroscience Laboratory, German Primate Center, Göttingen, Germany. <sup>4</sup>Department of Computer Science, University of Minnesota, Minneapolis, MN, USA. <sup>5</sup>Department of Psychology, University of Minnesota, Minneapolis, MN, USA. <sup>6</sup>Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA. <sup>7</sup>Department of Electrical and Computer Engineering, Rice University, Houston, TX, USA. <sup>8</sup>Center for Neuroscience and Artificial Intelligence, Baylor College of Medicine, Houston, TX, USA. <sup>9</sup>Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA. <sup>10</sup>Department of Machine Learning, Carnegie Mellon University, Pittsburgh, PA, USA. <sup>11</sup>Present address: Neuroengineering Initiative, Rice University, Houston, TX, USA. <sup>12</sup>These authors contributed equally: Xaq Pitkow, Valentin Dragoi.  e-mail: [xaq@cmu.edu](mailto:xaq@cmu.edu); [vdragoi@rice.edu](mailto:vdragoi@rice.edu)

The second major limitation of past studies is that experimental restraints used when recording neural activity can distort animal behavior<sup>14–16</sup>. The consequences of physical restraints may be especially dramatic on food-seeking behavior because animals use head and body movements to gather information from their environment for foraging<sup>17,18</sup>. Furthermore, the cortical activity differs when the animals aim for targets that are far from their immediate reach<sup>19</sup>. The restraints may also affect when the animals choose to act, particularly for trial-free experiments for which timing is crucial.

In this Article, to circumvent these limitations, we designed a trial-free task where animals can forage freely, without bodily restraints. We examined how dynamic task variables influence rewards and neural activity, and subsequently, how this activity influences foraging behavior. Animals were allowed to continuously interact with the task and explore a wide range of reward expectancies by choosing when and where to act. We found that animals adjust their foraging on the basis of deviations from theoretical reward predictions, reflecting subjective reward expectations and leading us to two hypotheses about brain computations. First, the subjective estimates of reward predictors should be decodable from the brain. Second, animals should choose the time and the place of foraging attempts according to these neural estimates. We tested these hypotheses by recording from neurons in the dorsolateral prefrontal cortex (dlPFC), an area where neural activities encode reward<sup>13,20,21</sup> and are related to memory<sup>22</sup> and action preparation<sup>23,24</sup>. Here, we show that single neurons had mixed selectivity<sup>25</sup> to experimental variables and, conversely, those variables were distributed across many neurons. Finally, this distributed representation accurately predicts where and when an animal will forage next.

## Results

Monkeys ( $n = 2$ ) were exposed to two concurrent reward sources on a variable interval (VI) schedule<sup>10</sup>. We made it costly for the animal to switch between reward sources by placing them 120 cm apart (Fig. 1a, left). Animals freely interacted with the task equipment, and we did not impose a trial structure or a narrow response window (Methods). A multi-electrode Utah array was chronically implanted in the dlPFC (Extended Data Fig. 1), and measured spiking activity was collected using a lightweight, energy-efficient wireless device (Fig. 1a, right and Fig. 1b)<sup>26</sup>. The experimental setup was designed for the effective transmission of a low-power electromagnetic signal (Methods)<sup>27,28</sup>.

Rewards on both sides (box 1 and box 2) became available at exponentially distributed random times after the animal obtained a previous reward. The reward availability was hidden from the monkey. Once becoming available, each reward remained available until the animal pressed a button, at which time the reward was delivered (Fig. 1c). The distribution of waiting times before a reward became available could have different mean times or ‘schedules’ for each side (that is, constant hazard rates; Fig. 1c). Schedules were chosen from 10, 15, 25 or 30 s and were constant for a block of rewards. Multiple schedules allowed us to diversify the response dynamics of the animals<sup>10</sup>. Each experimental session contained two to four blocks with 34 or 66 rewards in each block. Given the constant hazard rate and the fact that rewards never disappeared once available, the probability of reward availability increased exponentially toward 1 with the time elapsed since the last press (the waiting time), with a time constant given by the reward schedule (Methods, Fig. 1d and Extended Data Fig. 2). Since the monkey chose when to respond, its decisions influenced the probability of reward availability (Fig. 1c). An ideal observer that did not know the schedule or availability should track the time and reward histories, so we hypothesize that animals attempt to maximize their reward by tracking these quantities, referred to as the reward predictors, to determine when and where to respond.

We examined whether the firing rates of neurons in dlPFC represent the reward predictors as they are continuously evolving in time (monkey G: 11 sessions and monkey T: 19 sessions;  $n = 1,323$  single and

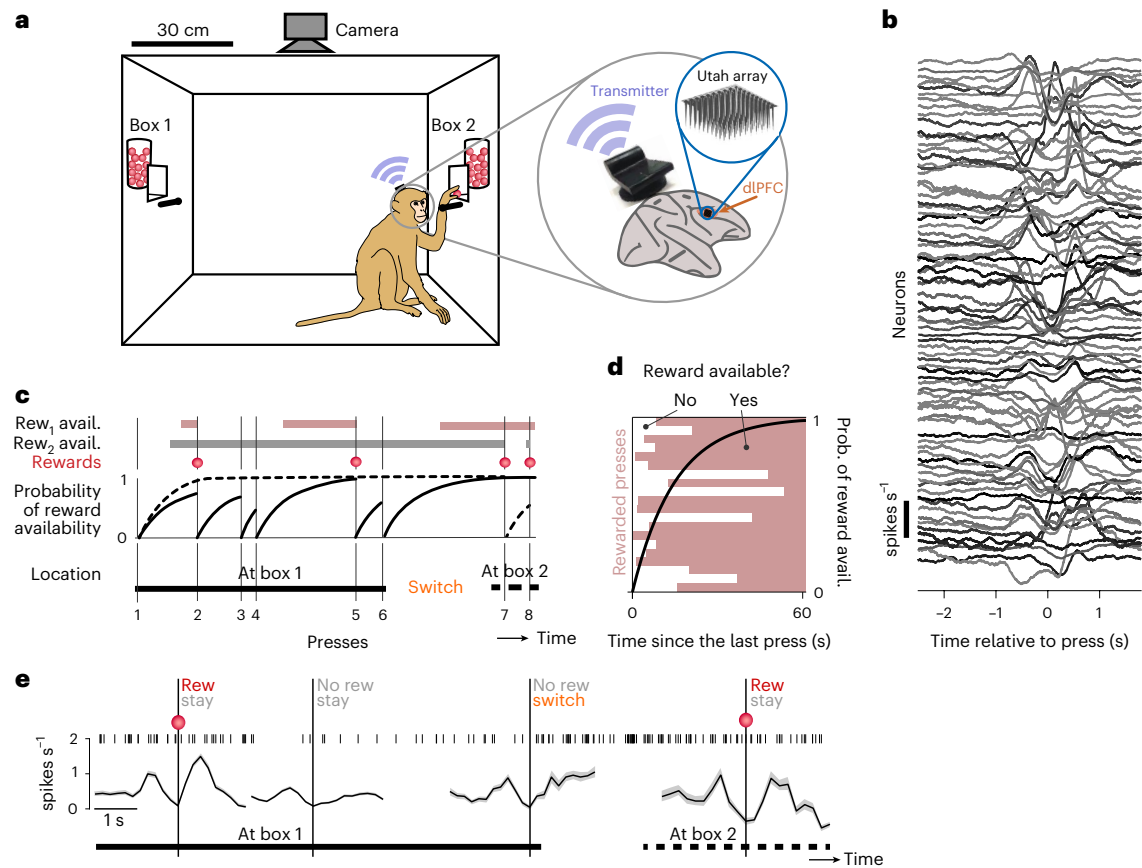
multi-units). Additionally, we extracted the neurons’ press-locked events, that is, firing rates a few seconds before and after each press (Fig. 1b). The continuous-time neural activity allowed us to understand how continuous representations of task variables in dlPFC leads to the animal’s choice to press. The press-locked neural activity explained how the state of these representation, before a press, combined with the new information, which is the reward outcome, predict where and when the animals press next. The continuous spike raster and the press-locked firing rate of a sample neuron (Fig. 1e) are shown for four consecutive box presses with different reward/choice outcomes: an unrewarded press followed by a switch to the other box, an unrewarded press when the animal stayed at the same box and two rewarded presses when the animal stayed at the same box. The fourth outcome, switching to the other box after a rewarded press, accounted for only 2% of the presses, so we do not show it in this example.

Here we explain how we identified reward predictors, variables that the animal can either observe or control and that they potentially use to estimate the chances of rewards. Consequently, we determined whether these variables empirically predicted the next reward outcome in our experiment. As the stochastic rewards do not always match the prediction, we examined the consequences of prediction violations on animals’ choice of box and time of press. Next, we used canonical correlation analysis (CCA) to identify the neural representation of these variables in the population of recorded neurons in dlPFC. Finally, we tested whether these representations predicted the animal’s choices in advance.

## Predictors of the next reward

According to the marginal value theorem of foraging theory<sup>1</sup>, an animal could optimize its reward while minimizing travel costs by estimating the box schedules, tracking the temporal evolution of the probability of reward availability and using them to choose when and where to search for reward. Although the probability of the reward availability is the best predictor of the randomly generated reward, it was completely unobservable to the animals in our experiment. However, other predictive variables were observable or controllable by the animals, such as the waiting time between the presses or the reward ratio, defined as the proportion of the current option’s recently delivered reward compared to all recently delivered rewards from either box. The recent history was defined by applying a causal half-Gaussian filter to the binary sequence of delivered (1) or denied (0) rewards<sup>7,8</sup>. The waiting time, together with the scheduled reward rate, determines the probability of reward availability (Methods and Extended Data Fig. 2a). The reward ratio, when tracked on a timescale relevant to the volatility of the environment<sup>7</sup>, is a proxy for the scheduled reward ratio, defined as the ratio of the scheduled reward rate on the current box and the sum of the scheduled reward rates of two boxes.

As the scheduled reward ratio changes without warning from block to block, we maximized the correlation of the scheduled reward ratio with the animal’s observed reward ratio by tuning the width of the causal half-Gaussian filter mentioned above (Extended Data Fig. 2b). We assessed how well each variable predicted the reward by correlating the rewarded fraction of presses with that variable before each press. Specifically, we pooled 8,862 behavioral presses from 30 sessions of two monkeys, binned them according to each hidden or observable/controllable variable so that there were 50 presses in each bin, calculated the fraction of rewarded presses within each bin (Fig. 2a), and computed the Pearson correlation between the binned variable and rewarded fraction of presses. Naturally, the probability of reward availability was highly correlated ( $r = 0.93$ ; Fig. 2a) with the rewarded fraction of presses. The scheduled reward rate was correlated with the fraction of rewarded presses as well ( $r = 0.43$ ; Fig. 2a). This correlation is weaker than the correlation of the waiting time with the fraction of rewarded presses ( $r = 0.92$ ; Fig. 2a) because the probability of reward availability is determined by both waiting time and the scheduled



**Fig. 1 | Foraging in freely moving monkeys while population activity in the prefrontal cortex is recorded wirelessly.** **a**, Left: schematic of experimental setup with two reward boxes, two buttons and an overhead camera. Right: the location of the Utah array in the dlPFC (area 46) and wireless transmitter. **b**, Press-averaged firing rates of 80 single and multi-units recorded simultaneously. **c**, An illustration of task dynamics with eight hypothetical presses (vertical lines) in the concurrent variable-interval foraging task. In this illustration, the monkey responds six times on box 1, then switches to box 2 and responds twice. Therefore, press 6 is considered a press with a switch choice. The first two rows show the independent telegraph processes determining the reward (rew.) availability at boxes 1 and 2. In the example shown, press numbers 2, 5, 7 and 8 were rewarded (third row, red). The time dependence of the probability of reward availability is shown in the fourth

row (see **d** for a different representation). **d**, An alternative illustration to clarify the relationship between the probability (prob.) of reward availability (avail.) and the waiting time. The shaded area shows when a reward is available after each of the 20 presses on box 1 (some of them shown in **c**). The black trace associated with the y-axis on the right shows the probability of reward availability (Methods and Extended Data Fig. 2a), which at each time is in fact the proportion of pink bars, out of 20 trials with pink bars. **e**, The spike train of one example neuron on the timescale of four consecutive presses showing a variety of events (top row). Event-locked average firing rates of the same neuron are shown in the bottom row for conditions with reward/no reward and stay/switch choices. For ease of visualization, this example used a neuron with a relatively low firing rate compared to others in the population (compared with **b**).

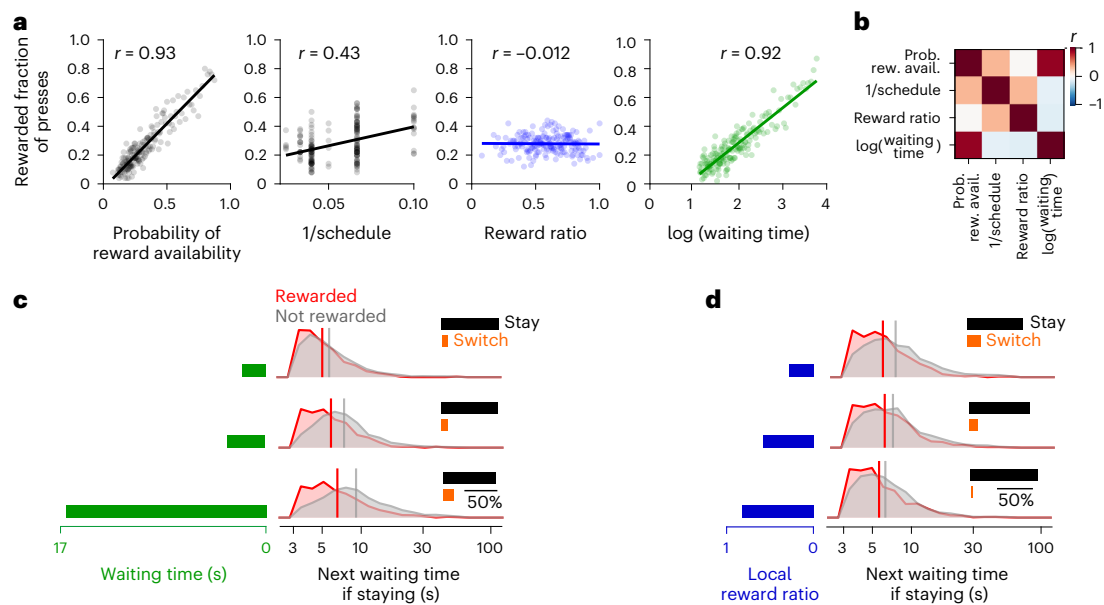
reward rate, and the animals choose a wide range of the waiting times, diluting the prediction of the scheduled reward rate alone.

Although the waiting time was highly predictive of the next reward, the reward ratio potentially plays an important role in animal's subjective reward expectation<sup>8</sup>. The reward ratio was not correlated with the fraction of rewarded presses (Fig. 2a,  $r = -0.012$ ). However, it was positively correlated with the scheduled reward rate on the side that the animal pressed ( $r = 0.32$ ), meaning that it might be used by the animals as an observable estimation of the hidden reward rates. Moreover, it was only weakly correlated with the log of waiting time ( $r = -0.140.14$ ), meaning that it may be considered by the animals as a source of information, independent from the waiting time. We refer to the waiting time and the reward ratio as the reward predictors because they may be used by the animals to predict the reward, and therefore may play a role in determining the animals' reward expectation (for the analysis of other observable reward predictors, see Extended Data Fig. 2c–e).

### Do reward predictors determine 'when' and 'where' to press?

Although the subjective reward expectation is not directly measurable, we might infer changes in the animals' reward expectation

from the animals' next choice, after a reward is delivered or denied. For example, an animal may realize that waiting longer increases its chances of receiving a reward, so we expect that an unrewarded press after a long wait might lead it to wait even longer between presses at the current box. Alternatively, the animal may realize that the waiting time for getting a reward at the current box is too long. Therefore, it may switch to the other box anticipating a better reward rate. We thus hypothesized that the animals' decision on where and when to press depends upon the reward predictors, as the basis of animals' reward expectation. We evaluated this hypothesis by analyzing the effect of such reward predictors on the probability distribution of the next waiting time and the probability of switching. These events were grouped depending on whether presses were rewarded and occurred after a short (3–5 s), medium (5–8 s) or long (8–60 s) wait (Fig. 2c and Extended Data Fig. 3, separated for monkeys). An unrewarded press increased the next wait by 10% (area under the receiver operating characteristic curve (AUC) of  $0.53 \pm 0.03$ ) after a short wait, by 28% (AUC of  $0.59 \pm 0.02$ ) after a medium wait and by 42% (AUC of  $0.59 \pm 0.02$ ) after a long wait, each compared with the corresponding average waiting times for rewarded presses. Moreover, the probability of switching to the other box increased with the duration of



**Fig. 2 | Reward predictors, together with the reward outcome, determine the choices and the next waiting time.** **a**, The predictability of the next reward from experimental and behavioral variables: 8,862 presses from 30 sessions were pooled together and binned into 50 press bins according to each experimental variable. The rewarded fraction of presses was calculated in each bin, then the Pearson correlation coefficient was calculated across bins between the average of the experimental variable and the rewarded fraction of presses. **b**, A correlation

matrix of the task variables in **a**. **c**, Histograms of the next waiting time for rewarded and unrewarded presses that were made after a short, medium or long wait, determined by equal intervals in the percentile of the presses. Inset: an increase in the probability of switching when not rewarded after a short, medium or long wait. The probability of switching after being rewarded was less than 2% and therefore excluded from this analysis. **d**, The same as **c**, but for reward ratio instead of waiting time.

unrewarded waits (9.5%, 10.2% and 16.5% more switches after a short, medium and long waiting time; Fig. 2c, insets). These choice differences (to continue pressing the button for the same box or switch to the other box) and the next waiting time when choosing to press on the same box, demonstrate that animals base their expectation of reward on their waiting time and adjust their behavior by waiting longer before the next press or switching to the other box when this expectation is not met. While previous studies point to melioration, that is, following the current flow of reward delivery<sup>9</sup>, we provide evidence of more temporally structured computations: the animals predict the chance of the next reward as they choose how long to wait before making the next press and adjust the waiting time when their expectation is not met. A key to this finding was a trial-free task, allowing animals to experience a wide range of waiting times and spontaneously discovering that longer intervals yielded a higher chance of receiving a reward.

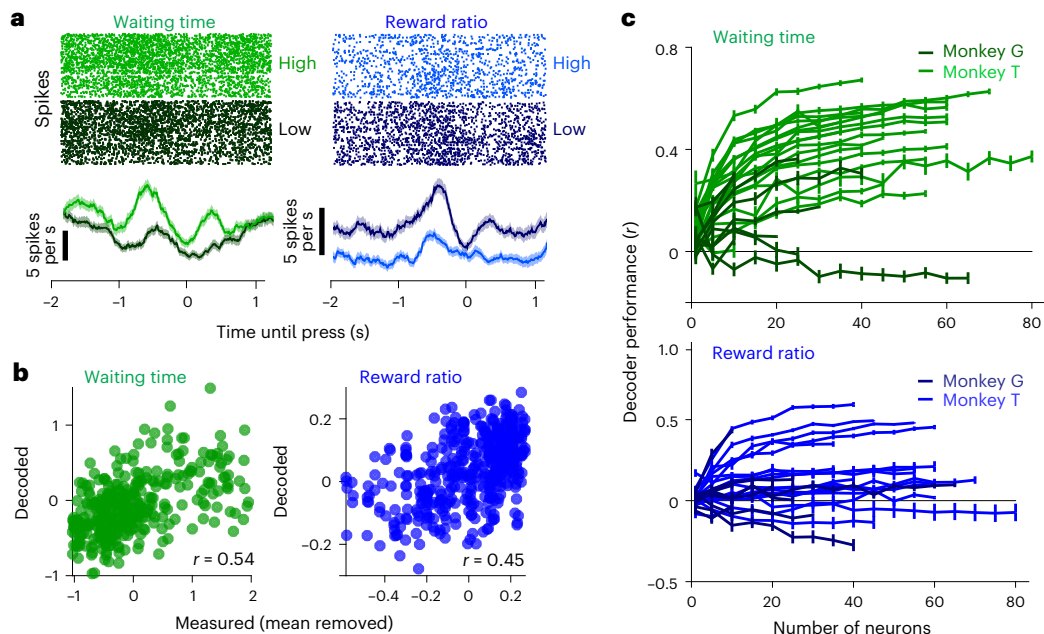
The animals might also develop expectations about the quality of the current box from the reward ratio. Again, we can infer these expectations indirectly through changes in the next waiting time and choices. After unrewarded presses, animals waited longer and switched more, with the smallest changes for biggest reward ratios (Fig. 2d; 23%, 18% and 15% longer unrewarded waits and 19%, 12% and 3% switches after a low, medium and high reward ratio, respectively). This suggests that animals require stronger evidence to override a better reward history.

Altogether, this provides evidence that an animal's policy on when and where to press depends on whether the box delivers a reward, as expected after a long waiting time or a high reward ratio. We inferred that animals update their expectation when those expectations are violated by the lack of an expected reward. This policy is a case of 'learning a guess from a guess'<sup>29</sup>, which is useful in the absence of sensory evidence directly cueing the probability or availability of reward. To provide further evidence that the waiting time and reward ratio underlie animals' reward expectation, we examined their encoding in the recorded neural population.

### Task-relevant activity in dlPFC

Before a motor action, the activity of neurons in the dlPFC is correlated with the value of a visually cued expected reward<sup>20</sup> or the probability of reward, estimated by the recent history of reward delivery<sup>13</sup>. Therefore, we hypothesized that the activity of dlPFC neurons, before each press, encodes the reward expectation for that press, for the range of the reward predictors variables observed or generated by each animal. For example, the neuron in Fig. 3a, left, activates more before a press following a long wait (top 20% of waiting times in that session) compared with a short waiting time (bottom 20%; Wilcoxon rank-sum test,  $P < 10^{-3}$ ). Similarly, the neuron in Fig. 3a, right, activates more when the reward ratio before a press is in the bottom 20% compared with when it was in the top 20% (Wilcoxon rank-sum test,  $P < 10^{-3}$ ).

As task-irrelevant variables such as locomotion, limb and eye movement and pupil size before or after presses may influence dlPFC activity<sup>30</sup>, we performed control experiments to quantify the correlation between task-irrelevant variables and neural activity. First, our control experiments in which animals moved to receive reward from the same boxes as in Fig. 1a revealed that eye movements have only a minor influence on neuronal activity while animals interacted with the box, although they have a stronger influence during locomotion ( $r = 0.16$ ,  $t$ -test,  $P < 10^{-3}$ , for eye velocity, and  $r = 0.13$ ,  $t$ -test,  $P < 10^{-3}$ , for fixation rate<sup>27</sup>). We thus decorrelated the neural activity from the locomotion by projecting neural activity onto the subspace orthogonal to locomotion (Methods) such that the remaining neural activity was uncorrelated with locomotion (Extended Data Fig. 4). Second, one animal performed the same task as presented here, while its arm movements, pupil diameter and eye velocity were monitored using the same eye tracking method as in ref. 27. We found  $\leq 9\%$  of neurons in dlPFC with significant ( $P < 0.01$ ) correlation with the arm movement (Extended Data Fig. 5) in 1 s time intervals starting 2 s before and ending 2 s after presses. Pupil diameter was correlated with  $\leq 10\%$  of neurons. However, after we decorrelated the neural activity from the locomotion, the percentage of neurons with a significant correlation with the pupil diameter dropped to  $\leq 7\%$ . Similarly, the percentage of neurons with a significant



**Fig. 3 | Neuronal populations encode variables of the reward dynamics.** **a**, Sample neurons for which the pre-press firing rate covaries with either waiting time (left) or reward ratio (right). The firing rate was calculated for a 200 ms sliding window starting 2 s before and ending 1 s after presses. Firing rates were averaged across presses with low (<20th percentile, gray) and high (>80th percentile, colored) of either the waiting time or the reward rate. Data are presented as mean values  $\pm$  s.e.m. **b**, Decoded and measured waiting time (left) and reward ratio (right) for two sample sessions. Forty-five neurons in the session

on the left and 60 neurons in the session on the right were used. The shown value of Pearson correlation coefficient is the average value across the cross-validated sets in each session. **c**, Decoding the waiting time (top) or the reward ratio (bottom) for 30 sessions as a function of the number of neurons used as predictors. The predictor neurons were chosen randomly from the population. Data are presented as mean values  $\pm$  s.e.m. across 20 randomly selected subset of units. Sessions of monkey G are shown with a darker color, and sessions of monkey T are shown with a lighter color.

correlation with the eye velocity dropped from  $\leq 9\%$  to  $\leq 4\%$ . As decorrelating the neural activity from locomotion also decreases the correlation between the neural activity and other task-irrelevant variables, we focused our analysis for the rest of this study on the neural activity that was decorrelated from the locomotion.

### Decoding reward predictors from the neural population

Since the waiting time influences both future behavior and the reward probability when the button is pressed, we examined how the neural activity encodes waiting time just before a button press. We measured the spike counts in a 1 s interval (that is, a ‘pre-press’ interval from  $-1.1$  to  $-0.1$  s) for each neuron ( $n = 1,323$  single and multi-units). This time interval was selected since the arm movement starts approximately 0.5 s before the press is recorded, and the modulation of neural activity typically starts around 0.5 s before that movement<sup>31</sup>. The pre-press firing rate of the neuron in Fig. 3a, left, was correlated with the waiting time (Spearman correlation coefficient of 0.24;  $t$ -test,  $P < 10^{-3}$ ; Fig. 3b, left). For the entire population of cells, around 35% of neurons exhibited a significant Spearman correlation ( $t$ -test,  $P < 0.01$ ; 31% positively correlated and 4% negatively correlated; monkey G: 27%, and monkey T: 37%).

To further examine how information about the waiting time is distributed across neurons, we decoded the waiting time from population activity before each press using the spike counts of randomly subsampled sets of neurons (for a description of the regression-based decoder analysis, see Methods). Our decoder analysis revealed that even random neural subpopulations encode the waiting time (Fig. 3c; Wilcoxon rank-sum test with false discovery rate, with multiple comparison correction (WRFDR),  $P \leq 0.01$ ).

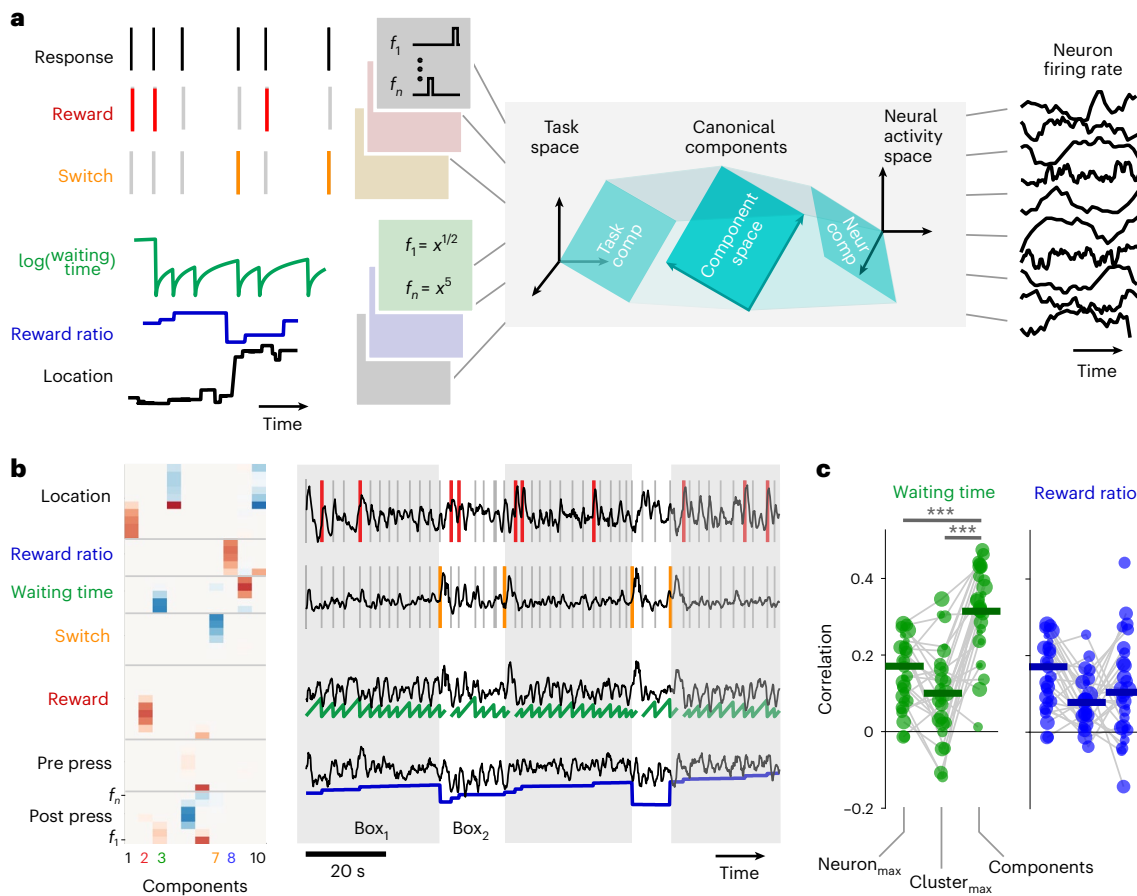
Furthermore, consistent with previous reports<sup>8,13,21,32</sup>, we found that dIPFC neurons encode the reward ratio. Over the entire population, there was a significant correlation between the pre-press firing rate and reward ratio ( $t$ -test,  $P < 0.01$ ) for 23% of the neurons (9% positively

correlated and 14% negatively correlated; monkey G: 12%, and monkey T: 26%). Decoder performance for the reward ratio was higher than chance (WRFDR,  $P < 0.01$ ) when we used a subpopulation of one or more neurons as the predictors. Taken together, these results indicate that both reward predictors are encoded in the pre-press neural activity at the individual neuron and population levels. This finding provides further evidence that the animals’ reward expectation is founded on the chosen reward predictors.

### Identifying continuous task variables in a latent space

Unlike waiting time, the reward ratio jumps discretely at press times. We aimed to gauge the waiting time’s explanatory power for the continuously evolving neuron activity. We attempted to fit the variability in a neuron population using a weighted sum of task-related variables and basis functions<sup>33–35</sup>. Some of these variables were event based (presses, reward delivery and choice to stay or switch location), while others evolved continuously (waiting time, reward ratio and location within the cage). For event-based task variables, each event raster was filtered with a 200 ms boxcar and then shifted to a variety of offsets<sup>36</sup> (Fig. 3e). For continuously evolving task variables, we used monomial basis functions with powers of 0.5, 1, 2, 3 and 5 (Fig. 4a). Neural activity was smoothed with a 1 s sliding window. To concentrate our analysis on times when animals were engaged in the task, we excluded time bins preceding or following any presses by more than 5 s.

To identify the latent representation of task variables in the neural space, we used CCA to find components that are shared between the task and the neural spaces. CCA finds these canonical components by applying singular value decomposition to the cross-correlation matrix between two spaces<sup>37</sup>. To favor interpretable latent components such that each component is associated with a small subset of the task variables, we imposed a sparsification penalty (least absolute shrinkage and selection operator with fullness constant of 0.3 on the weights of



**Fig. 4 | Canonical components of the neural population represent task variables in continuous time. a**, An illustration of the CCA for finding a reduced-dimensional space in the task space and the corresponding subspace in the neural activity space. The canonical components define the maximally correlated subspaces between the task variables and the neural (neur) activity. The 51-dimensional task space was made from 6 task variables by passing each variable through a set of basis functions (Methods). In brief, the basis functions were pulse-shaped temporal delay filters for press, reward and choice events. For continuously evolving task variables (the waiting time, the reward ratio and two-dimensional location), the basis functions were a set of instantaneous power functions. Overall, 51 predictors were made using these six task variables. The neural space was made using all simultaneously recorded neurons. **b**, Left: the weight of the contribution of each task variable in the first ten canonical components, sorted in the descending order of the correlation between the

projection of each component in the task and neural spaces. The indices of the components representing waiting time, reward ratio, reward and choice are color coded for easier association. Right: neural representation of four task variables: reward, choice, waiting time and reward ratio for the same sample session on the left. The component that was associated with each of these four task variables was identified as the component for which the absolute value of the weights was highest, compared with the weights for the other task variables. **c**, Cross-validated Pearson correlation coefficient between the reward predictors, waiting time (left) and reward ratio (right;  $*P < 0.005$ ), with either individual neurons or clusters of five or more neurons (Extended Data Fig. 6) that are maximally correlated with each reward predictor, compared with the correlation coefficient between the reward predictors and the canonical components (WRFDR; left:  $P < 10^{-3}$  for clusters and 0.003 for neurons; right:  $P > 0.1$ ). Each data point is associated with one session ( $n = 30$ ).

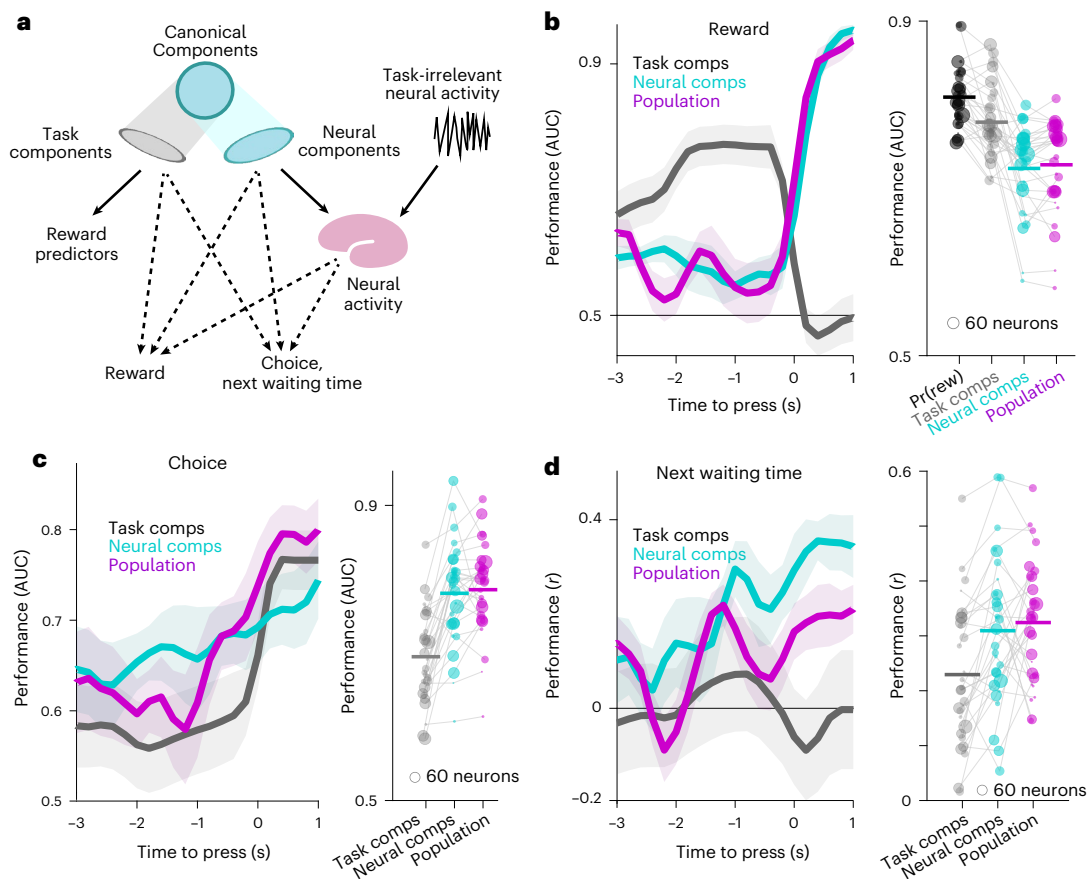
the task variables<sup>37</sup>). This regularization helps reduce overfitting the model. We calculated ten components for each training set (Fig. 4b), and then identified neural components making the greatest contributions to rewards, choices, waiting time and reward ratio. Interestingly, the waiting time neural component ramps up between the consecutive presses (Fig. 4b, third row), suggesting that the latent representation of the waiting time might be used by the brain to generate the next press, in a similar fashion to the evidence accumulation models proposed in decision making<sup>38</sup>. The reward ratio component followed the difference between the reward ratio of the boxes (Fig. 4b, fourth row). The reward and choice components showed sharp post-press elevated activity (Fig. 4b, first and second rows).

We asked whether fitting a model to reconstruct the activity of individually recorded neurons<sup>34</sup> or sites<sup>33</sup>, then clustering the neurons based on the similarity between the reconstructed activity (Extended Data Fig. 6) yields a better representation than the latent variables that we found using the CCA. We calculated the Pearson correlation coefficient between reward predictors and their associated canonical

components analysis and compared them with the correlation between the reward predictors with the neuronal clusters or individual neurons in each session that was maximally correlated with each reward predictor. The average correlation coefficient between the waiting time and the neural components was higher than that with the individual neurons or neural clusters of  $>5$  neurons (WRFDR,  $P < 0.005$ ; Fig. 4c, left). The correlation between the reward ratio and the neural components was the same as that with the individual neurons or clusters (WRFDR,  $P > 0.1$ ; Fig. 4c, right). This indicates that the latent neural components provide better correlates of reward predictors relative to individual neurons or the average activity of groups of neurons that were clustered together based on their task-relevant activity. Furthermore, the latent neural components were uncontaminated by movement-related confounds (Extended Data Fig. 7).

### Predicting reward, choice and the next waiting time

Since the animal cannot know the true hidden reward dynamics, its choices can only be driven by its subjective beliefs about these variables,



**Fig. 5 | Neural population analysis.** **a**, Prediction of rewards, choice or next waiting time from task components, neural components and the entire simultaneously recorded population. The task components are defined as the projection of the canonical component in the ten-dimensional space of waiting time and reward rate (each passed through five basis functions). The neural components are defined as the projection of the canonical component into the neural population space. The prediction was trained and tested for each 200 ms time bin, starting 3 s before and ending 1 s after the presses. **b–d**, The prediction

results for rewards (**b**), choice (**c**) or next waiting time (**d**) for example sessions (left) or summarized across all sessions for the peak within the 2 s time interval before the presses (right). Left: the prediction performances using the post-press components or population activity are shown for comparison. Data are presented as mean values  $\pm$  s.e.m. Right: the peak was calculated as the average of five time bins with the highest prediction performances. The prediction performance for the pre-press time bins provide evidence that the post-press choices or the rewards were in fact predictable, even before the presses were made.

rather than the objective truth from the experiment. For instance, if the monkey overestimates reward probability (perhaps due to misjudging waiting time or scheduled reward rate), he is more likely to switch boxes after an unrewarded push. We predicted switching based on neuronal components corresponding to task variables, interpreting them as current estimates of the animal's subjective beliefs. We decoded the pre-press neural activity by projecting the population activity onto the subspace formed by the first ten canonical components for the reward predictors. This projection accounts for latent representation of reward predictors that could potentially influence the choices or the next waiting time or predict the eventual reward outcome.

We attempted to predict rewards, choices and the next waiting times from three distinct types of predictors: (1) the pre-press reward predictors (canonical components in the reward predictors' space), (2) neural representations of the reward predictors (canonical components in the neural space) and (3) the entire simultaneously recorded neural population (Fig. 5a). For a fair comparison between the components and the entire neural population, we did not sparsify the weights of task variables in canonical components. To predict the reward, we trained binomial logistic regressions on the same data used to find the canonical components, then tested on the held-out data. To assess the prediction performance, we calculated the AUC showing the discriminability of the predictors' output for the rewarded presses from the unrewarded presses. The same method was used for the choice to stay

or switch. To predict the next waiting time, we used generalized linear models instead of logistic regression and evaluated the performance by calculating the Pearson correlation coefficient between the real and the predicted values. All predictors were trained and tested for each 200 ms time bin, starting 3 s before each press and ending 1 s after.

In the example session shown in Fig. 5b, left, the prediction of the reward outcomes using the task components improved as the analysis windows approached the time of the press. The reward outcomes are determined by the actual experimental task variables, and indeed we confirmed that the true pre-press task components (the projection of the canonical component in the space of reward predictors) predict actual rewards better than either their neural representations (the projection of the canonical components on the neural population space) or the entire neural population (Fig. 5b, right, and for monkey-separated results, see Extended Data Fig. 8).

In contrast, the choices and the next waiting time should follow the animal's subjective estimation of the reward dynamic variables. Fascinatingly, the neural activity before a press predicted the subsequent choice (Fig. 5c) and waiting time (Fig. 5d). As the animal's movement to switch to the other box or press the button again occurred after the current press (Extended Data Fig. 9), the prediction of either of these actions by the neural components precedes the execution of the predicted actions. Moreover, the head, arm and eye movements within the pre-press time window were not significantly different between

presses after which the animal stayed and the presses after which the animal switched to the other box (Extended Data Fig. 5c;  $P > 0.12$  for all the comparisons, Wilcoxon rank-sum test). Therefore, we provide further evidence that the animals construct an expectation of reward before a press, based on their subjective understanding of the temporal structure of the task. Subsequently, animals decide when and where to press next based on the expected reward and the actually observed reward. Interestingly, the ten-dimensional neural representation of the pre-press task components predicted the choice and the next waiting time as well as the entire neural population (Fig. 5c,d), indicating that these few canonical neural components successfully capture the relevant signals within the larger neural population space.

It might seem obvious that neural features should be better predictors of when and where to press than experimental variables, after all, the animal's brain is making its choice and not the experimental equipment. However, it is not evident a priori that the relevant neural representations would be found within our recorded dIPFC population, nor whether we record enough neurons to capture enough of the animal's choice-relevant information. Furthermore, even if the dIPFC does contain the choice-relevant signals, it is not obvious that the neural components for our specific hypothesized reward predictors would be the right ones to predict the choices. It is thus noteworthy that these neurally decoded reward predictors predict choices significantly better than the task variables from which they are derived, and equally well as the full neural population. Evidently, our analysis identifies a neural subspace containing correlates of latent variables that are relevant for subsequent choices. This subspace also tends to avoid neural dimensions that contain choice-irrelevant variability, since if present, these variations could contribute to overfitting and would only hinder our ability to predict choice. We conclude that we are capturing neural correlates of the animals' subjective beliefs about the latent reward dynamics that inform their choices.

## Discussion

We used a trial-free, unrestrained-animal approach to demonstrate that freely moving monkeys base their foraging strategy on an internal prediction of reward. This prediction is not based solely on the recent history of reward but relies on an internal estimation of the time they have been waiting since the last time they made a choice, which determines the probability of reward availability. Indeed, we found that neural populations in the prefrontal cortex contain information about reward predictors. Complementary to previous research in restrained animals<sup>8,13</sup>, we revealed that neural signals not only encode reward information, but also significantly predict animal's choices after each press during foraging. These findings challenge and extend long-standing theories of reward-seeking behavior<sup>9</sup> that suggest that animals follow the choice with the maximum recent rate of reward, without constructing a reward model to predict future behavior, according to the matching law<sup>9</sup>.

We argue that matching, while ubiquitous, does not entail a single computational strategy. For our foraging task, matching behavior is consistent with substantially different strategies. One strategy can be simulated using an agent that switches to the other box after the number of unrewarded presses exceeds a noisy threshold (Extended Data Fig. 10). This strategy corresponds to a basic 'win-stay/lose-switch' rule. We implemented this strategy by sampling the threshold from a Gaussian distribution with the same mean and variance as the loss count distribution at times when the animal switches sides. Although this agent is blind to both the average reward rate and the probability of the next reward, it still follows the generalized matching law (Extended Data Fig. 10). The slight undermatching that we observed resembles the behavior of various species in previous studies<sup>7,9</sup>.

We examined a more complex strategy that tracks reward probability and uses foraging theory to make choices by involving three variables: (1) the time since the preceding press and (2) the variable-interval

schedule—which together determine the probability of reward—and (3) the relative cost of switching locations, which affects the threshold for when to switch. We simulated an agent that follows such a strategy by making choices based on the correct probability of reward availability on both boxes. The agent switches to the other side when the probability of reward availability on the other box exceeds that of the current box by a fixed switching cost, and otherwise waits for the probability of reward availability to increase everywhere (Extended Data Fig. 10), in accordance with the marginal value theorem of foraging theory<sup>1</sup>. Unlike the first agent, this agent has complete information about the task. Nonetheless, we again observed nearly matching behavior, now with slight overmatching (Extended Data Fig. 10). These two simulations show that the generalized matching law may arise when following a strategy that is either blind to timing or fully informed. This implies that matching behavior is not, by itself, informative about the underlying strategy or animals' ability to grasp the hidden rule of the task.

Surprisingly, we found that the targeted representation within the high-dimensional space of neural population activity predicts choice better than the behavioral dynamics and as well as the entire population of recorded neural activity. This is an important confirmation of how targeted dimensionality reduction can reveal neural computations better than behavior or unprocessed neural activity. This type of analysis is essential in natural experiments where task variables are correlated.

One limitation of our findings is the extent to which our results can be generalized across other types of reward dynamics. The reward dynamics in our task are stochastic and time based, and they resemble the depletion of food resources found in nature. Follow-up studies are needed to determine whether our findings apply to other reward schemes, such as non-Markovian, more clock-like dynamics or those based on press rate<sup>39</sup>, whereby the reward becomes available after a variable number of presses rather than a variable time interval.

By allowing animals to move freely, our study represents a necessary move toward studying neural correlates of natural cognition in a free-roaming setting. This paradigm shift has been suggested decades ago<sup>40</sup>, but is only feasible now due to advances in low-power, high-throughput electrophysiological devices and large-scale computing<sup>41</sup>. Freely moving experimental paradigms probably increase the engagement of natural decision-making processes in the animal's brain, and possibly reduce the distortions in population dynamics that may be associated with unnatural head-fixed tasks. The free-roaming setting also enabled us to implement a natural switching cost between two reward options by simply allowing the monkey to walk between them. This is commonly implemented for restrained animals as a timeout period immediately after switching decisions. The subjective value and its neural representation potentially differ for a foraging task in which the animal explores by performing an effortful action such as relocation, compared with a task in which explorative actions do not cost effort, but instead cost time. Overall, a shift toward more natural behavior will be inevitable for understanding neural mechanisms of cognition<sup>41–46</sup>.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-024-01575-w>.

## References

1. Charnov, E. L. Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* **9**, 129–136 (1976).
2. Smith, J. M. Evolution and the theory of games. *Am. Sci.* **64**, 41–45 (1976).
3. Kilpatrick, Z. P., Davidson, J. D. & El Hady, A. Normative theory of patch foraging decisions. Preprint at *bioRxiv* (2020).



4. Wiener, M., Zhou, W., Bader, F. & Joiner, W. M. Movement improves the quality of temporal perception and decision-making. *eNeuro* **6**, 1–17 (2019).
5. De Kock, R., Zhou, W., Joiner, W. M. & Wiener, M. Slowing the body slows down time perception. *eLife* **10**, 1–23 (2021).
6. Balasubramaniam, R. et al. Neural encoding and representation of time for sensorimotor control and learning. *J. Neurosci.* **41**, 866–872 (2021).
7. Iigaya, K. et al. Deviation from the matching law reflects an optimal strategy involving learning over multiple timescales. *Nat. Commun.* **10**, 1466 (2019).
8. Sugrue, L. P., Corrado, G. S. & Newsome, W. T. Matching behavior and the representation of value in the parietal. *Cortex* **30A**, 457–461 (2004).
9. Herrnstein, R. J., Rachlin, H. & Laibson, D. I. *The Matching Law: Papers in Psychology and Economics* (ed. Herrnstein, R.) (Harvard Univ. Press, 1997).
10. Herrnstein, R. J. Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* **4**, 267–272 (1961).
11. Aldiss, M. & Davison, M. Sensitivity of time allocation to concurrent-schedule reinforcement. *J. Exp. Anal. Behav.* **44**, 79–88 (1985).
12. Heyman, G. M. A markov model description of changeover probabilities on concurrent variable-interval schedules I. *J. Exp. Anal. Behav.* **31**, 41–51 (1979).
13. Grabenhorst, F., Tsutsui, K. I., Kobayashi, S. & Schultz, W. Primate prefrontal neurons signal economic risk derived from the statistics of recent reward experience. *eLife* **8**, 1–30 (2019).
14. Freedman, E. G. Coordination of the eyes and head during visual orienting upward. *Exp. Brain Res.* **190**, 369–387 (2008).
15. Tollin, D. J., Populin, L. C., Moore, J. M., Ruhland, J. L. & Yin, T. C. T. Sound-localization performance in the cat: the effect of restraining the head. *J. Neurophysiol.* **93**, 1223–1234 (2005).
16. Miller, C. T. et al. Natural behavior is the language of the brain. *Curr. Biol.* **32**, R482–R493 (2022).
17. Bracis, C., Gurarie, E., Van Moorter, B. & Goodwin, R. A. Memory effects on movement behavior in animal foraging. *PLoS ONE* **10**, 1–21 (2015).
18. Knight, K. Head movements give away foraging behaviour. *J. Exp. Biol.* **214**, i–ii (2011).
19. Berger, M., Agha, N. S. & Gail, A. Wireless recording from unrestrained monkeys reveals motor goal encoding beyond immediate reach in frontoparietal cortex. *eLife* **9**, 1–29 (2020).
20. Leon, M. I. & Shadlen, M. N. Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* **24**, 415–425 (1999).
21. Tsutsui, K. I., Grabenhorst, F., Kobayashi, S. & Schultz, W. A dynamic code for economic object valuation in prefrontal cortex neurons. *Nat. Commun.* **7**, 12554 (2016).
22. Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154–5167 (1996).
23. Fuster, J. M. Executive frontal functions. *Exp. Brain Res.* **133**, 66–70 (2000).
24. Voloh, B. et al. Hierarchical action encoding in prefrontal cortex of freely moving macaques. *Cell Rep.* **42**, 113091 (2023).
25. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
26. Yin, M. et al. An externally head-mounted wireless neural recording device for laboratory animal research and possible human clinical use. In *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* <https://doi.org/10.1109/EMBC.2013.6610199> (2013).
27. Milton, R., Shahidi, N. & Dragoi, V. Dynamic states of population activity in prefrontal cortical networks of freely-moving macaque. *Nat. Commun.* <https://doi.org/10.1038/s41467-020-15803-x> (2020).
28. Shahidi, N., Schrater, P., Wright, T., Pitkow, X. & Dragoi, V. Population coding of strategic variables during foraging in freely-moving macaques. Preprint at *bioRxiv* <https://doi.org/10.1101/811992> (2019).
29. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* 2nd edn (The MIT Press, 2018).
30. Tremblay, S., Testard, C., Ditullio, R. W., Inchauspé, J. & Petrides, M. Neural cognitive signals during spontaneous movements in the macaque. *Nat. Neurosci.* <https://doi.org/10.1038/s41593-022-01220-4> (2022).
31. Chandrasekaran, C., Peixoto, D., Newsome, W. T. & Shenoy, K. V. Laminar differences in decision-related neural activity in dorsal premotor cortex. *Nat. Commun.* **8**, 614 (2017).
32. Elliott, R., Friston, K. J. & Dolan, R. J. Dissociable neural responses in human reward systems. *J. Neurosci.* **20**, 6159–6165 (2000).
33. Engelhard, B. et al. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–513 (2019).
34. Pillow, J. W. et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995–999 (2008).
35. Huk, A., Bonnen, K. & He, B. J. Beyond trial-based paradigms: continuous behavior, ongoing neural activity, and natural stimuli. *J. Neurosci.* **38**, 7551–7558 (2018).
36. Shahidi, N. et al. Filter-based models of suppression in retinal ganglion cells: comparison and generalization across species and stimuli. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.12.01.518577> (2022).
37. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–534 (2009).
38. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Ann. Rev. Neurosci.* <https://doi.org/10.1146/annurev.neuro.29.051605.113038> (2007).
39. Pierce, D. W. & Cheney, C. D. *Schedules of Reinforcement* (B. F. Skinner Foundation, 2015).
40. Hernández-Peón, R., Scherrer, H. & Jouvét, M. Modification of electric activity in cochlear nucleus during ‘attention’ in unanesthetized cats. *Science* **123**, 331–332 (1956).
41. Stevenson, I. H. & Kording, K. P. How advances in neural recording affect data analysis. *Nat. Neurosci.* **14**, 139–142 (2011).
42. Pitkow, X. & Angelaki, D. E. Inference in the brain: statistics flowing in redundant population codes. *Neuron* **94**, 943–953 (2017).
43. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., Malvar, M. A. & Poeppel, D. Neuroscience needs behavior: correcting a reductionist bias. *Neuron* **93**, 480–490 (2017).
44. Wu, Z., Kwon, M., Daptardar, S., Schrater, P. & Pitkow, X. Rational thoughts in neural codes. *Proc. Natl. Acad. Sci. USA* **117**, 29311–29320 (2020).
45. Calapai, A. et al. A cage-based training, cognitive testing and enrichment system optimized for rhesus macaques in neuroscience research. *Behav. Res. Methods* **49**, 35–45 (2017).
46. Fagot, J., Gullstrand, J., Kemp, C., Defilles, C. & Mekaouche, M. Effects of freely accessible computerized test systems on the spontaneous behaviors and stress level of Guinea baboons (*Papio papio*). *Am. J. Primatol.* **76**, 56–64 (2014).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise

in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

## Methods

All experiments were performed under protocols approved by the University of Texas at Houston Animal Care and Use Committee and the Institutional Animal Care and Use Committee for the University of Texas Health Science Center at Houston. Two adult male rhesus monkeys (*Macaca mulatta*; monkey G: 15 kg, 9 years old; monkey T: 12 kg, 9 years old) were used in the experiments. An additional adult male rhesus monkey (*Macaca mulatta*; monkey M, 10 kg, 11 years old) was used for the control experiment, tracking the eye and limb movements.

### Behavioral training and testing

The experimental setup was a custom-made cage (120 cm × 60 cm × 90 cm) that was placed in a dedicated room, free from distractions. After habituating each monkey for at least 4 days per week for over 4 weeks, we trained animals to press the button on each box to receive a reward. Over the course of 4–6 months, we gradually increased the mean time in the VI schedule to let the monkeys grasp the concept of probabilistic reward delivery. Once we started using VI 10 (corresponding to an average reward rate of <0.1 rewards s<sup>-1</sup>), monkeys started to spontaneously switch back and forth between the two boxes. If the monkeys disengaged from the task or showed signs of stress, we decreased the VI schedule (increased the reward rate) and kept it constant for 1 or 2 days. If the monkey showed a strong bias toward one reward source, we used unbalanced schedules to encourage the monkeys to explore the less preferred box.

After training, we tested monkeys using a range of balanced and unbalanced reward schedules. For balanced schedules we used VI reward scheduling with the average interreward interval of 20 s or 30 s, that is, VI 20 or VI 30, on both boxes. For unbalanced schedules, we used VI 20 versus VI 40, VI 15 versus VI 25, or VI 10 versus VI 30. The unbalanced schedules may reverse once, twice or three times during a session, for example, after a reversal the box with VI 20 becomes VI 40 and the box with VI 40 becomes VI 20. Each session lasts until the monkey receives 100 or 200 rewards, ranging from 1 to 7 h including a 1 h break after 100 rewards in sessions with 200 rewards. If monkeys were not engaged with the task for more than 2 min, we sometimes interrupted them to encourage them to engage with the task. For the analysis, we exclude all presses that occurred for more than 60 s. For the press-locked analysis, we also excluded presses that were made less than 2 s after the previous press to avoid mixing in the press-locked neural activity.

### Tracking whole-body, limb and eye movements

To determine the physical location and locomotion of the monkey, an overhead wide-angle camera was permanently installed in the experimental cage and the video was recorded at an average rate of six frames per second. Each frame was postprocessed in six steps using custom-made MATLAB code. First, the background image was extracted by averaging all frames in the same experimental session, then it was subtracted from each frame. The background-subtracted image was then passed through a manually determined threshold to identify the dark areas. The same image frame was also processed using standard edge detection algorithms. The thresholded and edge detected images were then multiplied together, and the result was convolved with a spatial filter, which was a circle with the estimated angular diameter of the monkey. The peak of this filtered image was marked as the location of the monkey. We used this heuristic because the illumination of the experimental room and the configuration of objects was constant. We expect novel techniques for motion and posture detection using deep neural network<sup>47,48</sup> to yield similar results. Locomotion (speed) was calculated as the magnitude of the vector difference between monkey locations in consecutive frames divided by their time difference.

To compute the head, arm and whole-body movements of monkey M, we trained DeepLabCut (version 2.0) with 200 frames annotated

with the center of the head, snout, each ear, shoulders, elbows and paws, then tracked these body markers in videos recorded in three sessions. These frames came from the same overhead camera described above but were recorded at 30 Hz. ‘Head movements’ included labels from the center of the head, snout and each ear. ‘Limb or arm’ movement was computed from shoulder, elbow and paw labels. ‘Torso or whole-body’ movement was calculated from the animal’s upper and mid back labels. To calculate the average speed of each label during frames of interest, we calculated the Euclidean distance between the label coordinates across consecutive frames divided by the time between frames. Subsequently, we quantified the overall movement of each body area (head, arm or torso) by computing the mean speed, and averaging the speeds of the corresponding body part labels that were recorded at 30 Hz.

To track the eyes, we used a commercially available eye tracker (ISCAN). To train animals to wear the device without damaging it, its three-dimensional (3D) geometry was modeled (Sketchup Pro), and dummies were 3D printed and fitted with eye mirrors. To properly position the eye tracker and dummies relative to the eye, custom adapters were designed and 3D printed to attach directly to the animal’s head post and serve as an anchor point for the eye tracker. These adapters were designed to interface with the head post, without touching the animal directly, to minimize discomfort and reduce the likelihood of the device being tampered with. These dummy eye trackers were worn by animals for several mock recording sessions to adjust them to wearing the device. Once the animals grew accustomed to wearing the dummy, the real device was used. We used two-dimensional coordinates of the pupil to compute eye velocity from Euclidean distance of these values across consecutive eye camera frames.

### Determining reward availability and calculating probability of reward availability

In each time bin of size  $dt = 10$  ms, reward became available at a given box if a sample from a Bernoulli distribution was 1. The probability of this event was  $dt/VI$ . When the reward became available, it stayed available until collected by the animal. This makes the probability of reward availability a function of the scheduled VI as well as the time since the preceding press:

$$P_{\text{rew}} = 1 - (1 - dt/VI)^{t/dt},$$

where  $t$  is the time since the preceding press (Extended Data Fig. 1).

### Chronic implantation of the Utah array

A titanium head post (Christ Instruments) was implanted, followed by a recovery period (>6 weeks). After acclimatization with the experimental setup, each animal was surgically implanted with a 96-channel Utah array (BlackRock Microsystems) in the dlPFC (area 46; anterior of the Arcuate sulcus and dorsal of the principal sulcus (Extended Data Fig. 1)). The stereotaxic location of dlPFC was determined using magnetic resonance images and brain atlases before the surgical procedure. The array was implanted using the pneumatic inserter (BlackRock Microsystems). The pedestal was implanted on the caudal skull using either bone cement or bone screws and dental acrylic. Two reference wires were passed through the craniotomy under and above the dura mater. After the implant, the electrical contacts on the pedestal were protected always using a plastic cap except during the experiment. Following array implantation, animals had at least a 2 week recovery period before we recorded from the array.

### Recording and preprocessing of neural activity

To record the activity of neurons while minimizing the interference with the behavioral task, we used a lightweight, battery-powered device (Cereplex-W, BlackRock Microsystems) that communicates wirelessly with a central amplifier and digital processor (Cerebus neural signal

processor, Blackrock Microsystems). First, the monkey was head fixed, the protective cap of the array's pedestal was removed, the contacts were cleaned using alcohol and the wireless transmitter was screwed to the pedestal. The neural activity was recorded in the head fixed position for 10 min to ensure the quality of the signal before releasing the monkey in the experimental cage. The cage was surrounded by eight antennas. In the recorded signal, spikes were detected online (Cerebus neural signal processor, Blackrock Microsystems) using a manually selected upper threshold on the amplitude of the recorded signal in each channel or an upper and a lower threshold that were  $\pm 6.25$  times the standard deviation of the raw signal. To minimize the recording noise, we optimized the electrical grounding by keeping the connection of the pedestal to the bone clean and tight. The on-site digitization in the wireless device also showed lower noise than common wired head stages. The remaining noise from the movements and muscle activities of the monkeys was removed offline using the automatic algorithms in offline sorting (Plexon Inc.). Briefly, this was done by removing the outliers (outlier threshold, 4–5 standard deviations) in a 3D space that was formed by the first three principal components of the spike waveforms. Then, the principal components were used to sort single units using the expectation-maximization algorithm (offline sorter version 4.0). Each single and multi-unit signal was evaluated using several criteria: consistent spike waveforms, modulation of activity with 1 s of the button pushes and exponentially decaying inter-spike interval histogram with no inter-spike interval shorter than the refractory period (1 ms). The analyses used all spiking units with consistent waveform shapes (single units) as well as spiking units with mixed waveform shapes but clear pre- or post-press modulation of firing rates (multi-units).

### Removing task-irrelevant components from neural activity

For each neuron  $k$ , we remove movement-related temporal components of the press  $r_{kt}$ , by subtracting its projection onto the subspace spanned by the task-irrelevant variables:  $r_k^\perp = r_k - \Pi r_k$ , where  $\Pi$  is the projection matrix  $\Pi = L(LL^\top)^{-1}L^\top$  and  $L$  is the  $T \times 1$  vector describing the time series of locomotion, calculated as the magnitude of the changes in the two-dimensional location.

### Regression-based and binary decoder analysis

To decode a binary variable, such as the reward or the choice to stay or switch, we used logistic regression. To evaluate this model, we used the AUC to determine the separability of the probability distributions of the held-out samples belonging to either of the classes (reward versus no reward and stay versus switch). To decode continuous-value variables such as waiting time or the reward ratio, we used a linear regression model<sup>49</sup>. To evaluate this model, we calculated the Pearson correlation coefficient between the measured and predicted values. To train and cross-validate these decoders, we divided the presses in each session to 4–18 blocks, holding out one block at a time for testing and using the rest of the blocks for training. To divide the presses into blocks, we found the gaps in press times that were larger than 30 s, then placed all presses between consecutive gaps in one block.

### Selecting task-relevant variables and useful basis functions for continuous-time analyses

Our continuous-time analyses used a set of basis functions applied to the time series of experimental task variables. To use event-based variables in continuous-time predictions, we filtered the variables with different boxcar-shaped delay filters. We used different numbers of these pulse basis functions for different variables: seven basis functions for the pre-press time interval, post-press time and post-choice; and ten basis functions for the post-reward time to include the entire range of reward collection time, starting with the food release sound cue at the press time and ending when the food pellet was consumed (spanning 2 s; Fig. 4a). For task variables  $x_t$  that were already defined continuously over time, we applied a set of nonlinear power functions,  $x_t^a$ , with

powers  $a \in \{\frac{1}{2}, 1, 2, 3, 5\}$ . This 51-dimensional feature vector was used to predict components of neural activity.

### CCA

Canonical components were calculated using singular value decomposition of the cross-covariance matrix between the task variables and the neural activity, specifically the pre-press firing rates of simultaneously recorded neurons. We regularized this linear model using an  $\ell_1$  penalty<sup>37</sup> to calculate canonical components. The cross-validation procedure was the same as for the decoders.

### Statistical analysis

We used the two-sided Wilcoxon signed-rank test except where indicated. We chose this test rather than parametric tests, such as the  $t$ -test, for its greater statistical power (lower type I and type II errors) when data are not normally distributed. When multiple groups of data were tested, we used the false discovery rate multiple comparisons<sup>50</sup> correction whose implementation is a standard function in MATLAB. We used WRFDR abbreviation to indicate using two-sided signed-rank test with false discovery rate multiple comparison correction. No statistical methods were used to predetermine sample sizes. However, the size of our dataset and the number of the experimental sessions are similar to those reported previously<sup>27</sup>. Data collection and analysis were not performed blind to the conditions of the experiments.

### Use of generative artificial intelligence

While preparing the last version of this work, the authors used chatGPT to shorten parts of the text to meet the journal's word count limit. After using this tool, the authors reviewed and edited the content as needed. The authors take full responsibility for the content of the publication.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The preprocessed data used for this study is available at <https://doi.org/10.6084/m9.figshare.24762996.v1>. The raw data will be available upon request.

### Code availability

The custom code used in this study is available at <https://doi.org/10.6084/m9.figshare.24762996.v1>.

### References

- Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
- Bala, P. C. et al. Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nat. Commun.* **11**, 4560 (2020).
- Bishop C. M. *Pattern Recognition and Machine Learning* (Springer, 2006).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

### Acknowledgements

This study was supported in part by NIH 5U01NS094368 to V.D., P.S. and X.P., and a McNair Foundation grant to X.P. N.S. was supported in part by DFG CRC1528.

### Author contributions

N.S., A.W. and V.D. designed the setup and the experiments. N.S. and A.W. collected the data. N.S., X.P. and P.S. performed the analyses.

N.S., X.P., P.S. and V.D. wrote the paper. M.F. collected the data from monkey M used in Extended Data Figs. 5 and 7. A.P. and M.F. analyzed the data from monkey M to quantify the effect of eye, head and limb movements on neural activity and choices (Extended Data Figs. 5 and 7).

### Competing interests

The authors declare no competing financial or personal interests that could influence the reported work.

### Additional information

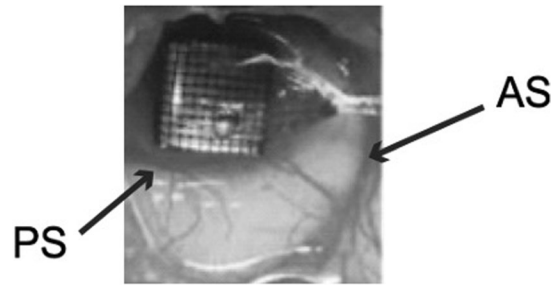
**Extended data** is available for this paper at <https://doi.org/10.1038/s41593-024-01575-w>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41593-024-01575-w>.

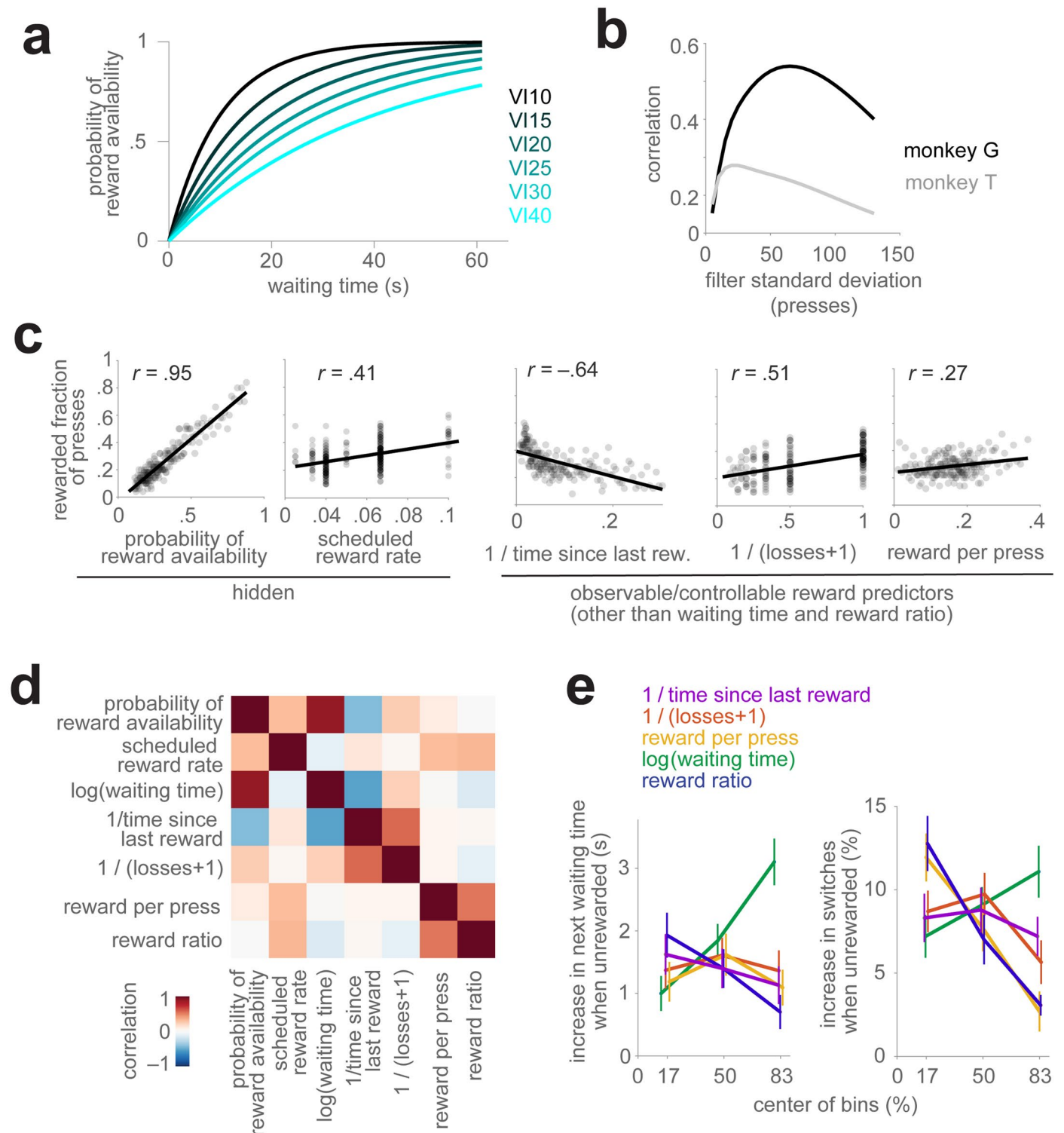
**Correspondence and requests for materials** should be addressed to Xaq Pitkow or Valentin Dragoi.

**Peer review information** *Nature Neuroscience* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | The MEA was implanted in the dorsolateral prefrontal cortex.** The location of a 96-channel Utah array in dlPFC (area 46) on the left hemisphere of monkey G. The arcuate sulcus (AS) and principal sulcus (PS) are marked.

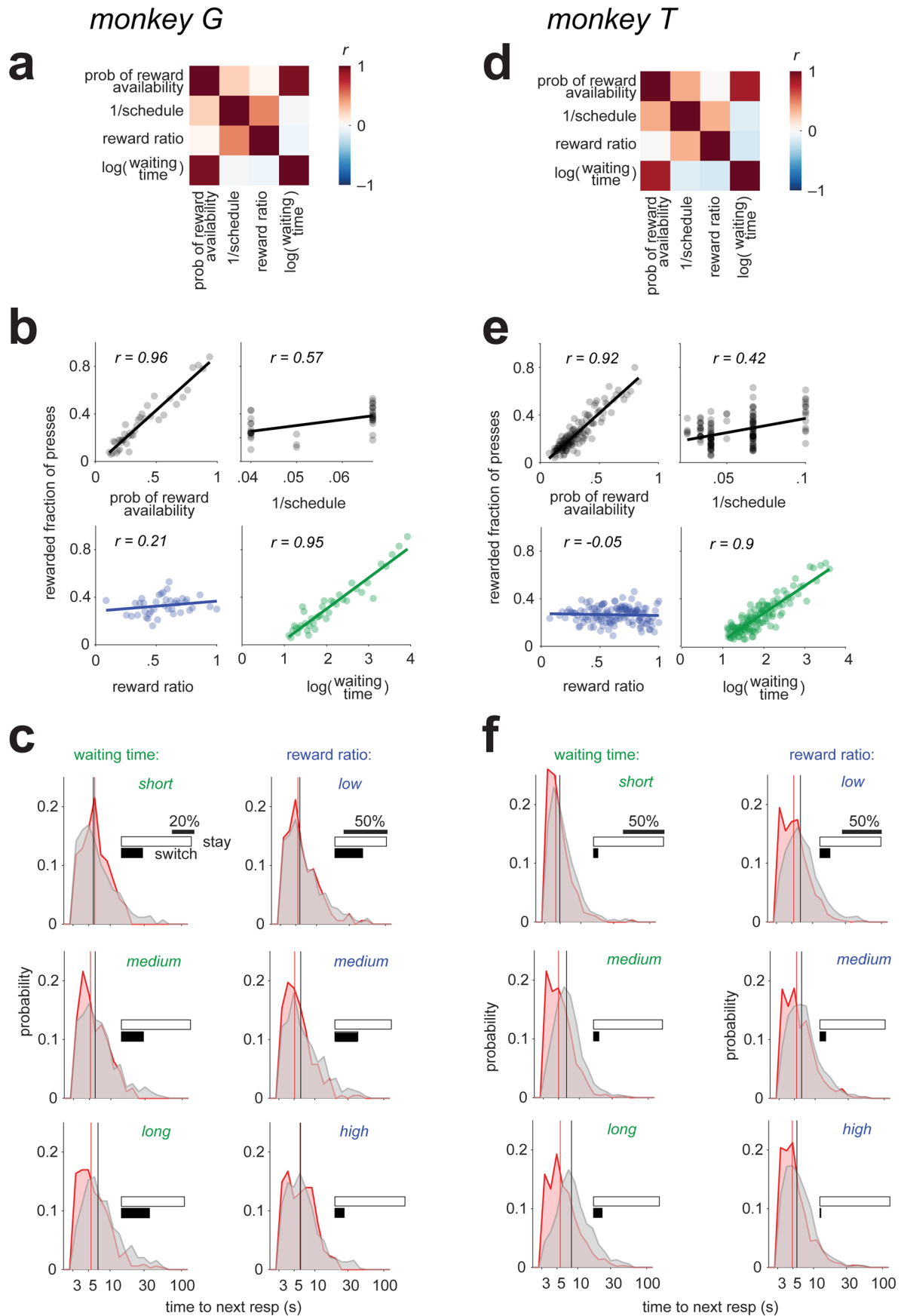


Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | A comprehensive set of behavioral and experimental variables was examined before selecting the reward predictors. (a)** The probability of reward availability as a function of the scheduled reward rate and the time since the preceding response on the same box. **(b)** The Pearson correlation coefficient between the scheduled and the (observable) reward ratios is calculated using a recent sequence of reward outcomes defined in the main text. The recency was imposed by choosing the standard deviation of a causal half-Gaussian filter (x-axis). For each monkey, the reward ratio was calculated using the standard deviation for achieving the maximum correlation with the scheduled reward ratio. **(c-e)** Other observable/controllable variables that the animal might have used to predict rewards. **(c)** Similar to Fig. 2a, but for the inverse of the time since the last reward, the inverse of the number of unrewarded presses (losses), and the reward per press (the binary sequence of rewarded (1) and unrewarded (0) presses, filtered using the same causal half-Gaussian

filter that was used for calculation of the reward ratio). **(d)** Correlation matrix between observable and unobservable reward predictors. Based on this matrix, the waiting time was chosen as one reward predictor because it was maximally correlated with the hidden probability of reward availability, and the reward ratio was selected as another reward predictor because it was maximally correlated with the scheduled reward rate and minimally correlated with the waiting time. Other predictors were omitted because they were correlated with either of these two variables ( $|r| > 0.21$ ). **(e)** Change of the next action ( $n = 8862$ ) as a function of each reward predictor candidate that was discretized into three bins with an equal number of presses in each bin. The x-axis shows the center of the bins. The y-axis shows the excess percentage of switches or waiting time when unrewarded compared to when rewarded. Two chosen reward predictors, waiting time and reward ratio, linearly predict the change of the following action. Data are presented as mean values  $\pm$  SEM.

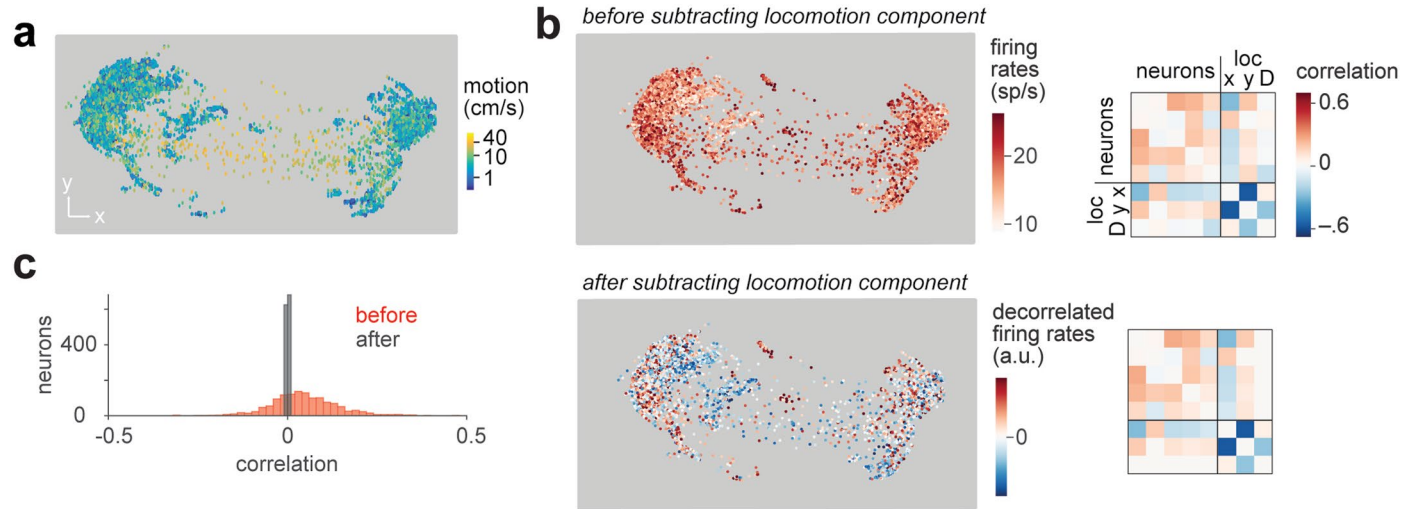




Extended Data Fig. 3 | See next page for caption.

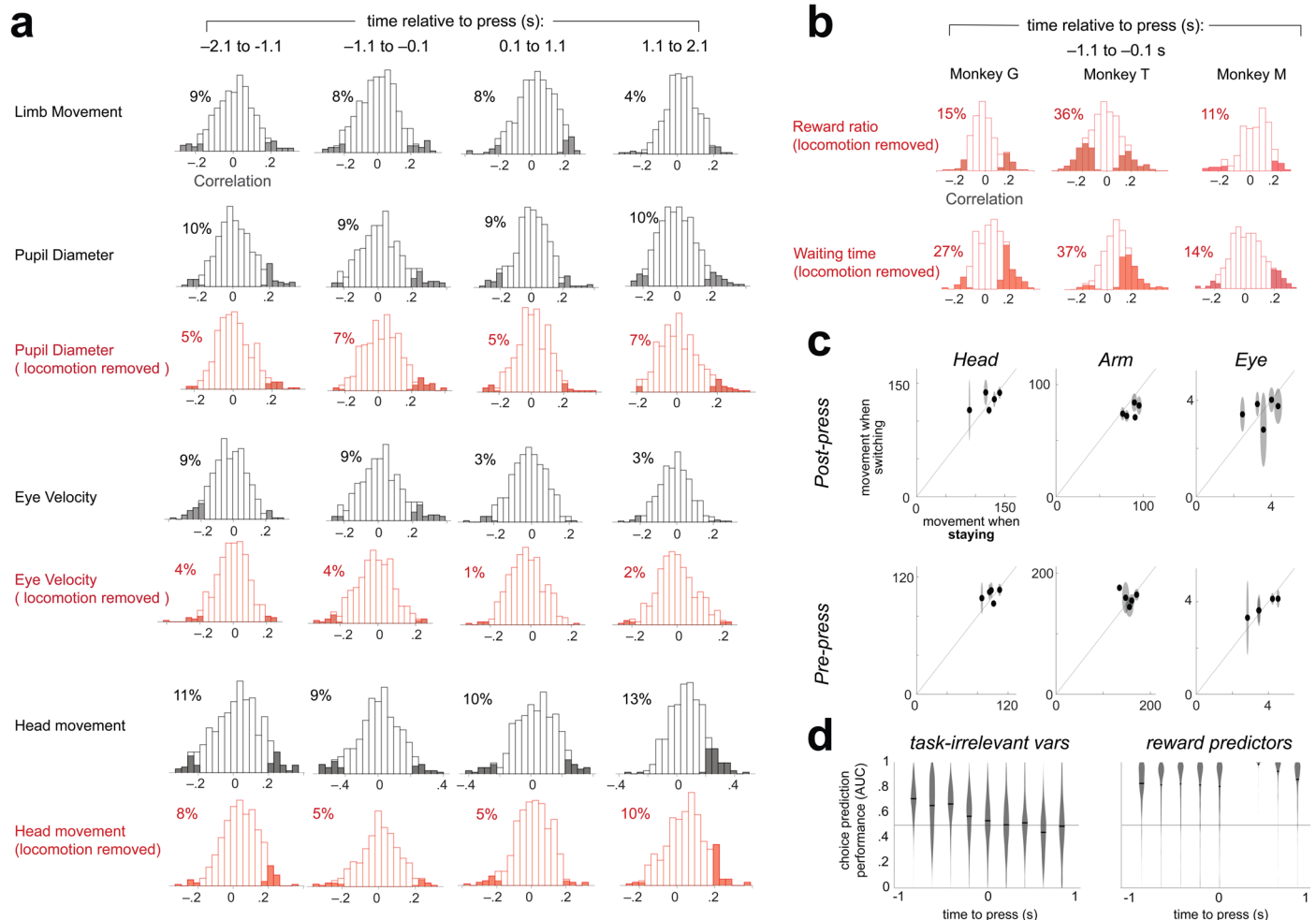
**Extended Data Fig. 3 | Reward predictors, together with the reward outcome, determine the choices and the next waiting time.** Same as Fig. 2, but separated for monkey G and T. **(a,d)** Correlation matrices of the task variables. **(b,e)** Predictability of the next reward from experimental and behavioral variables. The rewarded fraction of presses was calculated in each 50 press bin, then the Pearson correlation coefficient was calculated across bins between the average of the experimental variable and the rewarded

fraction of presses. **(c,f)** Left: histograms of the next waiting time for rewarded and unrewarded presses that were made after a short, medium or long wait, determined by equal intervals in the percentile of the presses. Inset: an increase in the probability of switching when not rewarded after a short, medium or long wait. The probability of switching after being rewarded was less than 2% and therefore excluded from this analysis. Right: the same as left, but for reward ratio instead of waiting time.



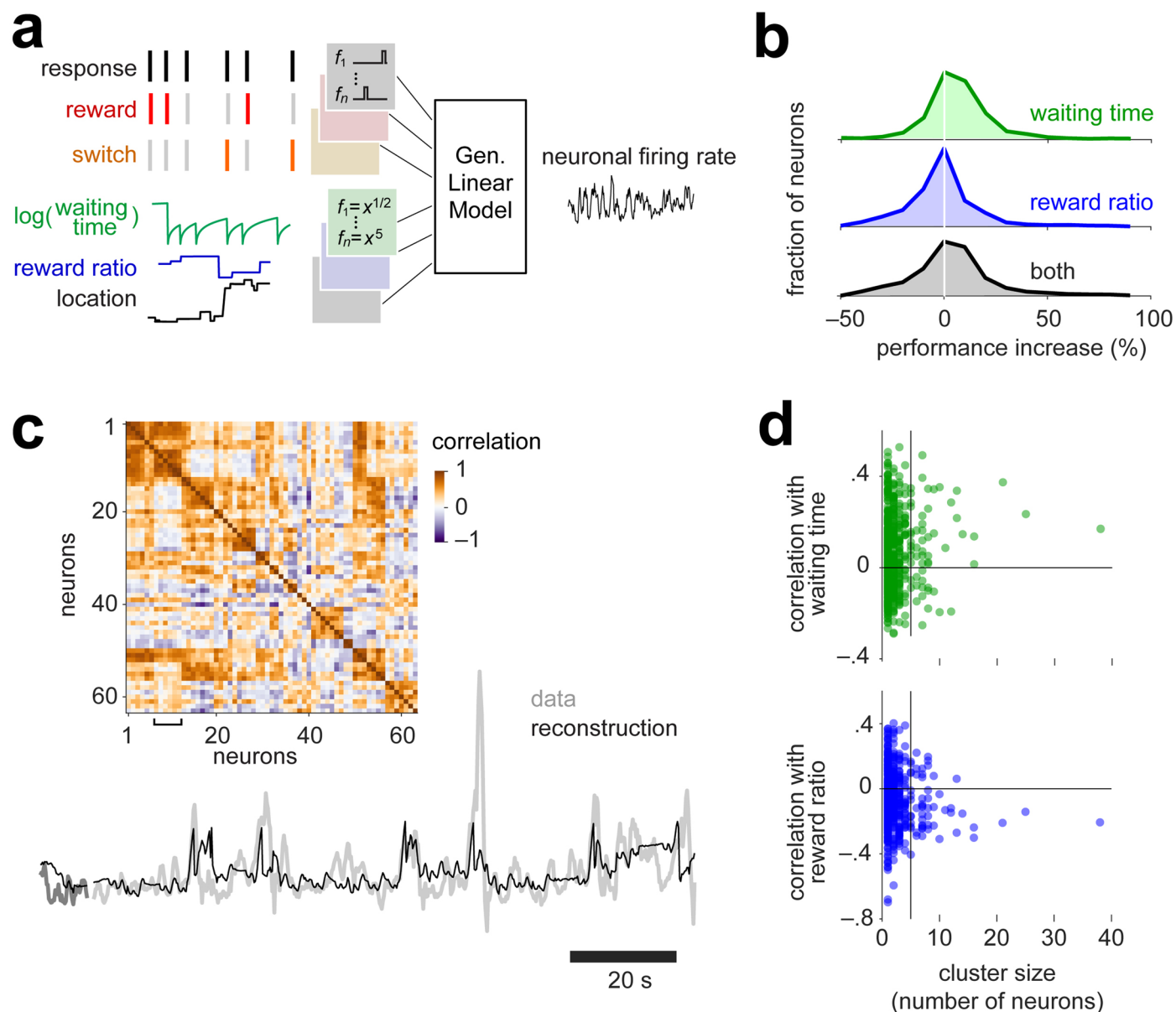
**Extended Data Fig. 4 | The neural activities were orthogonalized to monkeys' locomotion before Figs. 3–5 analyses. (a)** Monkey locomotion in a sample session for the press-locked time bins (< 3 s before or > 1 s after a press). Each dot represents the animal's position in space, sampled every 200 ms. **(b) Left:** Population-averaged firing rates for each time bin in (a). *Top* and *bottom* show

firing rates before and after subtracting the vector projection of locomotion. **Right:** Matrix of correlation coefficients between pre-press firing rate of each neuron and between the monkey's locations Loc X and Loc Y and locomotion Loc D. For clarity, only an arbitrary subset of neurons is shown. **(c)** Histogram of correlation coefficients computed as in (b), but for all recorded neurons.



**Extended Data Fig. 5 | Effect of eye and limb movements on the neural activities and choices. (a)** Histograms of correlation coefficients between the limb and eye movements and individual neuron activities. Bars are shaded for statistically significant correlations, and the inset indicates the percentage of significantly correlated neurons. A deep artificial neural network was trained using DeepLabCut (Mathis et al., 2018) to localize monkey M's shoulder, elbow, and paw of the forelimb contralateral to the recording site (right limb) in each overhead video frame of the foraging animal. We could compute average limb movement in any desired time interval using those three limb markers. We considered time intervals around button presses ( $-2$  s,  $+2$  s) and computed the average limb speed and firing rate of all the dIPFC neurons of monkey M in non-overlapping time-bins of 200 ms width. The pupil diameter and the eye velocity were computed with the same method as in ref. 26 **(b)** For comparison, the correlation coefficient of the reward predictors with firing rates of dIPFC

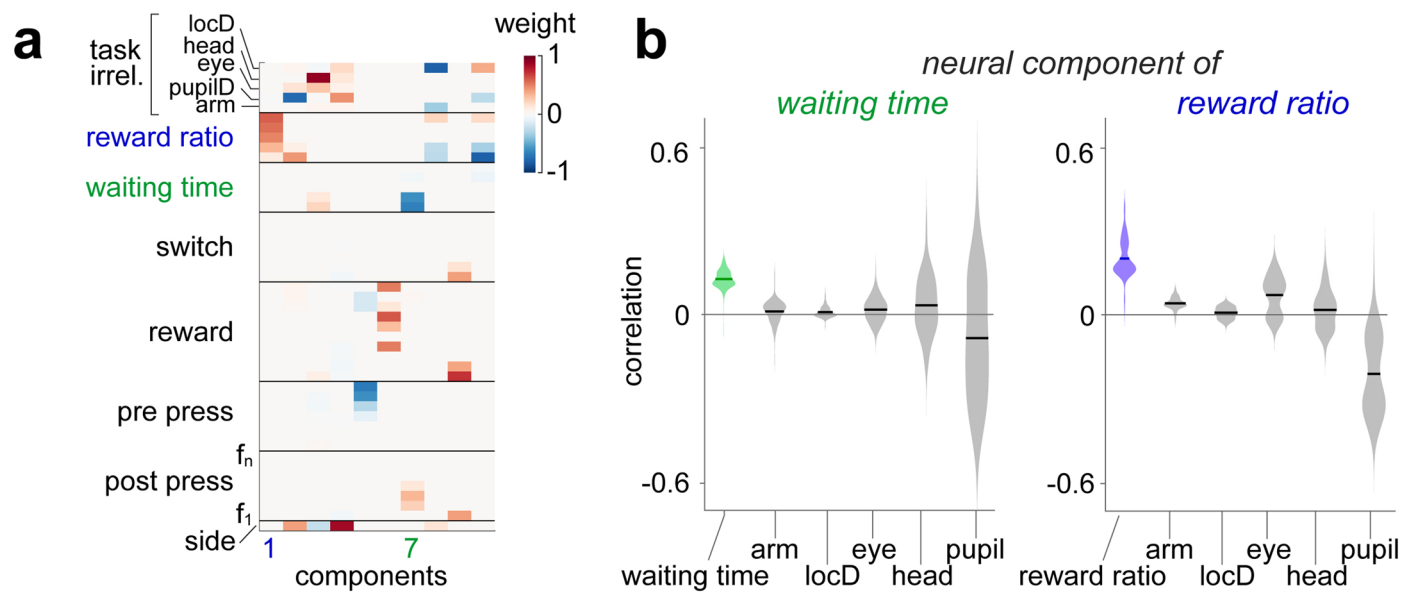
neurons for monkeys G, T, and M **(c)** Magnitude of movements of the head, arm, and eye are similar whether the animal chooses to stay (horizontal axes, arbitrary units; 5 sessions, 238–336 presses per session) or switch locations (vertical axes, arbitrary units; 5 sessions, 3–27 presses per session) around the time of a button press. Points depict mean movement magnitudes, and widths of shaded ellipses indicate standard errors of the means along each axis. The top row shows movements after presses, and the bottom row shows movements before presses. **(d)** Prediction of monkey M choices, cross-validated using 1000 sub-samples of presses (80% of presses used for training, 20% for testing). The prediction performance was calculated as the area under the curve of the output of logistic regression (as in Fig. 5c), except that the predictors were the task-irrelevant variables, eye, body, and head movements, and pupil diameter (left) or reward predictors (right). The mean (black line) and the Gaussian-smoothed distribution (gray shades) across 1000 sub-sampled test data are shown.



**Extended Data Fig. 6 | An encoding model for reconstructing the activity of individual neurons using the continuously evolving experimental variables.**

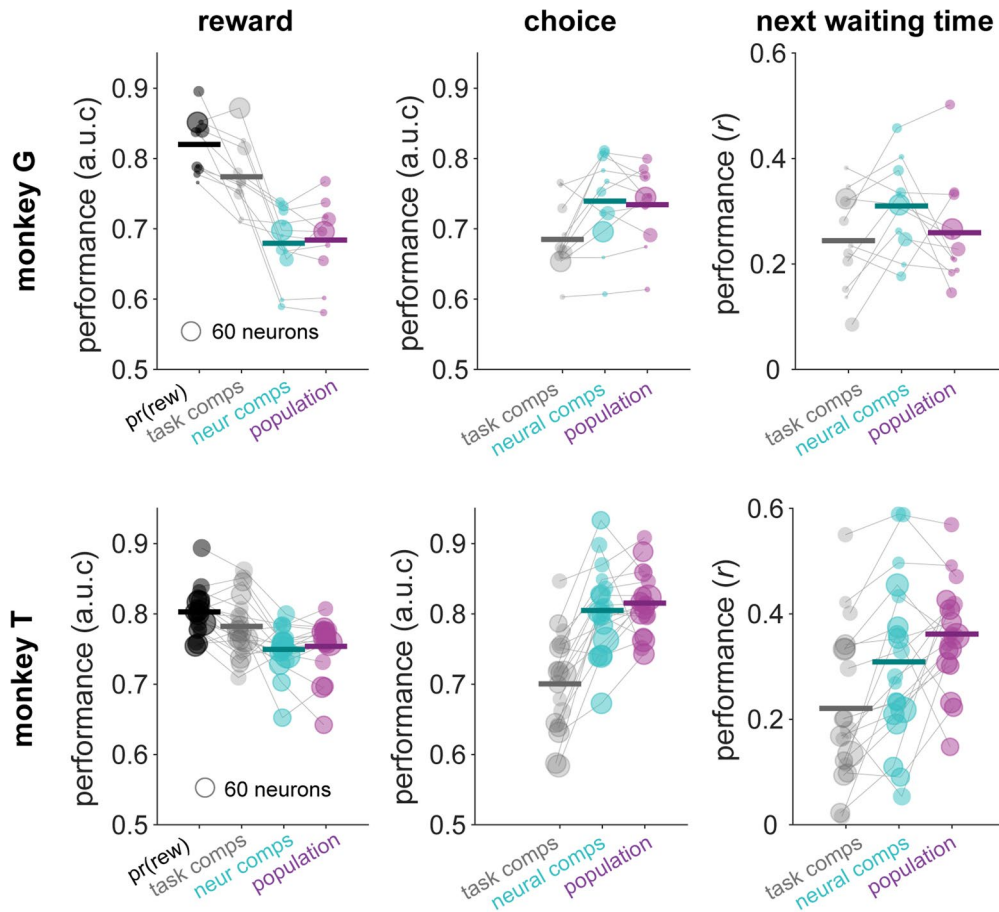
(a) A generalized linear model to reconstruct continuously evolving firing rates of individually recorded neurons. Rates are predicted from combinations of the task variables passed through a set of basis functions. The basis functions were pulse-shaped temporal delay filters for press, reward, and choice events. (The time of the reward and the choice events were assumed to match the time of the press after which the reward was delivered or the choice was made.) The basis functions for continuously evolving task variables (waiting time, reward ratio, and 2-dimensional location) were instantaneous power functions with powers of  $\frac{1}{2}$ , 1, 2, 3, and 5. Altogether, 51 predictors were made using these 6 task variables. The model was fit to the training data using a Gaussian likelihood function; the trained model was used to reconstruct the neural activity for held-out testing data. (b) The improvement in the performance of the model when either one or both reward predictors were used alongside the other task variables in (a). Improvement was calculated as the percentage increase in the correlation

between the recorded and reconstructed activities. While the waiting time improved the model performance for the entire population by 5% ( $p < 10^{-3}$ ), this improvement was insignificant for the reward ratio ( $p = 1$ ) at the level of individually recorded neurons. (c) The reconstructed and recorded firing rates averaged across 6 neurons in a sample session. The neurons were selected from 60 simultaneously recorded neurons in this session by clustering them using the correlation matrix of their reconstructed activities (inset), then choosing all neurons in a sample cluster (bracket in the inset) to show here. (d) The correlation coefficient between each reward predictor and the reconstructed neural activity averaged across neurons in the same cluster *vs.* the number of neurons in the cluster. The vertical line shows a cluster size of 5, which was the lower bound of the cluster size for the clusters that were included in Fig. 4c. Across sessions, the average firing rates of clusters of  $\geq 5$  neurons were positively correlated with the waiting time ( $p = 0.004$ , *top*) and negatively correlated with the reward ratio ( $p = 0.004$ , *bottom*).

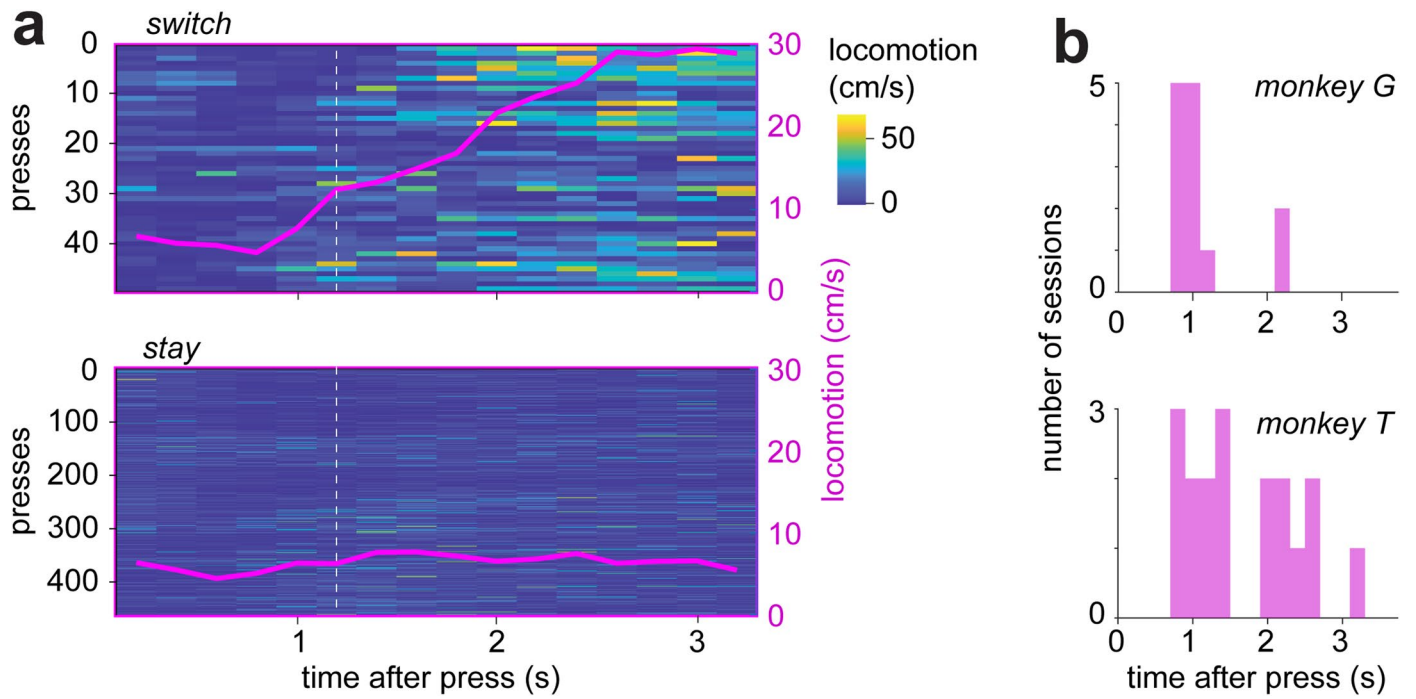


**Extended Data Fig. 7 | Extending the canonical correlation analysis from Fig. 4 to include eye and limb movements of monkey M.** (a) The weight of the contribution of each task variable in the 10 first canonical components, sorted in descending order of the components' correlations between the task and neural spaces. The indices of the components representing waiting time and reward ratio are color coded. (b) *Left:* Pearson correlation coefficient between

component 7, the neural component, best associated with the waiting time, and the true waiting time, as well as between this component and any of the task-irrelevant variables. *Right:* same as left but for neural component 1 instead of component 7, and reward ratio instead of waiting time. The mean (lines) and the Gaussian-smoothed distribution (shades), pooled across held-out chunks of all sessions are shown.



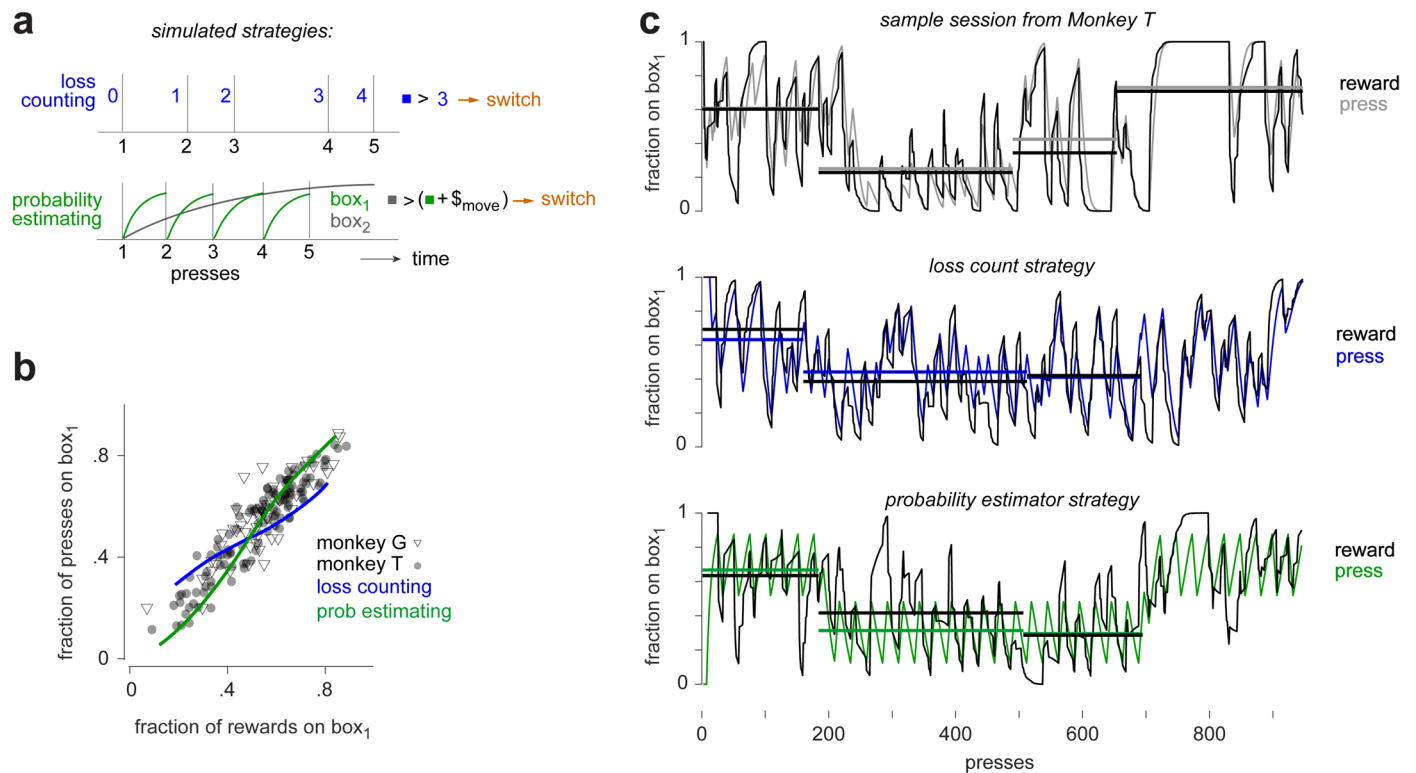
**Extended Data Fig. 8 | Prediction of rewards, choice, or next waiting time from task components, neural components, and the entire simultaneously recorded population.** same as Fig. 5b–d, but separated for monkeys G and T.



**Extended Data Fig. 9 | Detecting the time of the switches using the locomotion data.** (a) The magnitude of locomotion in each 200 ms time bin, separated for switch and stay trials. Trial averages are shown as solid yellow lines.

The white dashed line shows when the average locomotion speed in switch trials is greater than 3 SD of the magnitude of locomotion in stay trials. (b) Average switch time in each session.





**Extended Data Fig. 10 | Analysis of the matching behavior for monkeys G and T and two simulated agents.** (a) Illustration of foraging strategies for two simulated agents. The ‘loss counting’ agent switches to the other box when the *loss count* exceeds a threshold drawn from a Gaussian distribution ( $\lambda = 2.66$ ,  $\sigma = 1.9$ ). The ‘probability estimator’ agent switches to the other box when the probability of reward availability on the other box exceeds the probability of reward availability in the current box by a fixed switching cost<sup>1</sup>. The inter-press times were drawn from a random geometric distribution for both agents. The parameters of these strategies, namely the *loss count* distribution, the inter-press time distribution, and the switching time were estimated from the behavior of the monkeys. Each agent was simulated for 100 rewards for each set of reward

schedules for box<sub>1</sub> and box<sub>2</sub>. The variable interval reward schedules spanned the range between VI-5 and VI-50 in steps of 1 s and were drawn independently for each box. (b) Matching behavior of two monkeys and two simulated agents: the fraction of behavioral presses at box 1 is approximately proportional to the fraction of rewards at box 1 (26 sets of schedules for monkey G and 59 sets of schedules for monkey T are shown). Curves show spline fits for the matching behavior for the simulated agents. (c) Dynamic matching for a sample session of Monkey T with 3 sets of reward schedules: VI15-VI25, VI25-VI15, and VI15-VI25 again. We compared two simulated agents of panel a. The reward and press rates were calculated locally using a causal half-Gaussian filter<sup>8</sup>.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** The neural data was collected using the the 'central' software (BlackRock Microsystems). The behavioral data and video data was collected using custom code in Matlab 2015a. The eye position and pupil size data was collected using the hardware and software provided by ISCAN Inc.

**Data analysis** The neural data was pre-processed using offline sorter v4 (Plexon inc.). The behavioral, neural and part of the video material was processed using costum code in Matlab 2015a-2020b. The rest of the video material was processed using deepLabCut (open source; version 2.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The pre-processed data used for this study is available at <https://doi.org/10.6084/m9.figshare.24762996.v1>. The raw data will be available upon request.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	not applicable
Reporting on race, ethnicity, or other socially relevant groupings	not applicable
Population characteristics	not applicable
Recruitment	not applicable
Ethics oversight	not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We conducted this study with two monkeys in order to achieve robustness for the behavioral and electrophysiological results despite possible differences in the behavioral strategies between monkeys (all results were consistent across animals). We limited the number of monkeys to two for the main results to meet the requirements of lab animal use regulations that requires minimizing the number of animals in each study. Choosing the sample size of two is typical in electrophysiological studies in monkeys. To provide additional data on eye position, movements and pupil size we used control data from a third monkey which was implanted in the same cortical area as part of another study.
Data exclusions	For the continuous-time analysis, we exclude time bins that were > 60 s away from any press. For the press-locked analysis, we also excluded presses that were made less than 2 s after the previous press to avoid mixing in the press-locked neural activity.
Replication	The experiments were repeated up to 20 times in each animal. The experimental equipment is commercially available and the custom made software and codes for reproduction of the analyses are available upon request.
Randomization	Not applicable to this study. Two subjects were not categorized or randomly assigned. The same experiments were repeated in both animals.
Blinding	Since no group allocation was done in this study, blinding was not required.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Two adult male rhesus monkeys ( <i>Macaca mulatta</i> ) were used for the main results. monkey G: 15 kg, 9 years old; monkey T: 12 kg, 9 years old). An additional adult male rhesus monkey ( <i>Macaca mulatta</i> ; Monkey M, 10 kg, 11 years old) was used for the control analysis.
Wild animals	We did not include wild animals.
Reporting on sex	Due to the limited number of animals used for the main study (n=2), no reporting on sex or gender was possible.
Field-collected samples	We did not include field collected samples.
Ethics oversight	All experiments were performed under protocols approved by The University of Texas at Houston Animal Care and Use Committee (AWC) and the Institutional Animal Care and Use Committee (IACUC) for the University of Texas Health Science Center at Houston (UTHealth).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>