A corticostriatal circuit updates subjective beliefs about latent task states

Margaret L. DeMaegd[1], David Hocker[1], Harsha Gurnani[1], Mitzi Adler-Wachter[1], Jonny Schindler[1], Shannon S. Schiereck[1], Cristina Savin[1,2], Christine M. Constantinople[1]*

[1] Center for Neural Science, New York University; New York, NY 10003.
[2] Center for Data Science, New York University; New York, NY 10003.

*Corresponding author. E-mail: constantinople@nyu.edu.

Beliefs about states of the world profoundly impact decision-making and learning, but little is known about how neural circuits represent and update beliefs. We performed projection-specific recordings and perturbations from neurons in the orbitofrontal cortex (OFC) projecting to the intermediate or rostral caudate putamen (CPi/CPr) in rats performing a task with hidden reward states. Stimulating OFC→CPi neurons biased rats' beliefs towards high reward states. Recordings from optogenetically-tagged OFC→CPi neurons showed that they encoded evidence for high reward states, and evidence encoding was shaped by local inhibition. Finally, projection-specific perturbations disrupted encoding of hidden states within OFC via long-range cortico-basal ganglia-thalamic loops. These findings reveal the circuit implementation of a core cognitive computation, updating subjective beliefs about abstract latent states of the environment.

Main Text

Inference is a core computation that is important for many cognitive functions, including perception, motor control, and decision-making (*1*). In partially observable environments, state inference is required for forward mental simulations, planning, or goal-directed decision-making (*2–4*). Uncertainty about the current state can powerfully impact behavior, including by dynamically adjusting the rate at which animals learn from new experiences (*5–9*). Moreover, aberrant belief updating is a hallmark of schizophrenia and other neuropsychiatric disorders (*10–12*). While the orbitofrontal cortex (OFC) has been implicated in inferring values or partially observable task contingencies (*13–18*), how the connectivity patterns and dynamics of synaptic inputs in downstream circuits mediate computations for inference is unclear. Here, we uncover the circuit mechanisms by which OFC neurons that target the intermediate caudate putamen (CPi) represent and causally update beliefs about latent task states in rats performing a rich decision-making task.

Results

OFC neurons target CPi and CPr

We trained rats to perform a temporal wagering task in which they revealed how much they valued different amounts of water rewards based on how long they were willing to wait for them (*8*) (Fig. 1A). When rats initiated trials, the frequency of an auditory cue indicated the volume of water reward offered on that trial (5, 10, 20, 40, or 80 μL). At the offset of the cue, the reward was assigned randomly to one of the two side ports, indicated by an LED. The rat could wait for an uncued and unpredictable delay to obtain the reward, or could opt out by poking in the other side port at any time to start a new trial. On 15-25% of trials, rewards were withheld to assess how long rats were willing to wait for them (catch trials), providing a behavioral measure of their subjective value of the reward.

Rats experienced hidden states, or blocks of trials in which they were presented with low (5, 10, or 20 μL) or high (20, 40, or 80 μL) volumes of water rewards, interleaved with mixed blocks which offered all rewards (Fig. 1B). Wait times were sensitive to the hidden reward states (Fig. 1C,D). Rats' behavior is well-accounted for by a model in which they compare the value of waiting to a threshold that reflects the opportunity cost, or average reward in each block (*8*, *18*, *19*) (Fig. 1E). The model uses Bayes' rule to infer the most likely block, and implements a fixed, block-specific wait time threshold based on that state inference. We note that the shape of the wait time curves somewhat differs between the rats and the model, due to the model's assumption that subjective value reflects the binary logarithm of the reward volume. We view this as a qualitative difference, as the focus of the present study is on contextual sensitivity, or the modulation of wait times by the hidden reward states. In that respect, the inferential model reproduces myriad aspects of behavior (*8*, *18*) (Fig. 1F,G). We previously found that inference in this task requires the OFC, and inactivating OFC impairs rats' ability to update their beliefs about the block (*18*). Here, we sought to characterize how different OFC projection neuron populations contribute to belief updating.

Cortical innervation of the striatum, the input structure of the basal ganglia, is extensive and topographic. Mesoscale projection mapping studies in the mouse have divided the caudate putamen (CP) into rostral extreme, rostral, intermediate, caudal, and caudal extreme subregions (*20*, *21*). Subregions can be further segmented into "communities" based on cortical and thalamic innervation patterns (*20*, *22*). To characterize OFC innervation of CP subregions in the rat, we injected AAV-CB7-mCherry or AAV-hSyn-mCherry in the lateral OFC (Fig. 1H,I). Consistent with past studies of rat OFC, we targeted lateral orbitofrontal cortex (LO) and the medial aspects of the agranular insular cortex (AI) (*18*, *23–28*). We identified terminal fields in the CPr and CPi (Fig. 1J). At the level of the CPr, lateral OFC axons innervated the intermediate ventral community, consistent with ventrolateral OFC projections in the mouse (*20*). At the level of CPi,

we observed innervation of the ventrolateral community. This is in contrast with findings in the mouse, which have reported OFC innervation of dorsomedial subregions, although anterograde injections in those studies were more medial, and appeared more anterior than our injections based on anatomical landmarks (*20*, *21*). It is also possible that species differences between mice and rats contribute to this discrepancy.

We next injected the retrograde fluorescent tracers CTB-Alexa 488 or CTB-Alexa 647 specifically in the intermediate ventral CPr or ventrolateral CPi (Fig. 1K,L). Quantification of retrogradely-labeled cell bodies in OFC revealed that the lateral LO and AI were enriched with neurons that projected to the CPr, and this projection class straddled the putative boundary between these regions. In contrast, neurons projecting to CPi tended to be localized in LO and more medial orbital regions. We next sought to characterize the function of OFC→CPr and OFC→CPi neurons in the temporal wagering task.

OFC→CPi neurons bias beliefs

We bilaterally injected a retrograde Cre virus in the ventrolateral CPi and a Cre-inducible excitatory opsin (ChR2.0) in the OFC, and implanted a tapered fiber optic among OFC→CPi axon terminals in the striatum (Fig. 1M). In optogenetic sessions, we delivered blue light (465nm) for 500 ms aligned to the time of reward delivery on 40% of trials. This manipulation caused rats to wait less time for rewards in each of the blocks (Fig. 1N,O). This optogenetic effect persisted over trials, as light was delivered at the time of reward, but the behavioral effect was apparent on trials in which rats opted out and received no reward, which are by definition different trials from the stimulation trials. Notably, the optogenetic effect size depended on the block, with the largest effect in low blocks, an intermediate effect in mixed blocks, and the weakest effect in high blocks (Fig. 1P). The block-dependent effects were supported by a Friedman test indicating within subject differences, and by the relative fraction of reward offers with significant wait time

differences in each of the blocks. In contrast, optogenetically stimulating OFC→CPr neurons at the time of reward failed to change how long rats were willing to wait for rewards compared to control sessions (fig. S 1A), indicating that these corticostriatal pathways are functionally dissociable.

We sought to use our behavioral model to generate a hypothesis about the cognitive process that was disrupted by optogenetic stimulation. The model estimates the most likely block ($P(B_t \mid R_t)$) by combining the likelihood of observing the current reward in that block ($P(R_t \mid B_t)$) with the prior belief about the block ($P(B_t)$), where the prior is recursively updated over trials. We found that if we implemented a biased prior, such that the model is biased to believe it is in a high block and against believing it is in a mixed or low block, this reduces how long the model is willing to wait, with the largest change in low blocks, intermediate change in mixed blocks, and weakest change in high blocks (Fig. 1Q,R). Intuitively, the biased prior induces mistaken inferences in the model, increasing the probability that it will use the opportunity cost or wait time threshold associated with a high block. Because the opportunity cost of the low block is the most different from a high block, the strongest effect on wait times is in low blocks, and intermediate in mixed blocks. Even an ideal Bayesian observer would make mistaken inferences on about 10% of trials (*8*); the biased prior eliminates mistaken inferences in high blocks, thereby also modestly reducing wait times in those blocks.

As an additional test, we simulated the behavior of the model with an optimal versus biased prior, and visualized the changes in wait times at different block transitions. Because the wait times are modulated by both the reward volumes and the blocks, we first z-scored the wait times for each reward independently, before pooling z-scored wait times over different reward volumes, to isolate contextual effects from the effects of reward volume on wait times (*18*). The prior that was biased towards high blocks made the model especially resistant to accepting that it had transitioned from a mixed into a low block, predicting that wait times should be slower to change

specifically at this transition but not others (Fig. 1S). To test this prediction, we compared the rats' z-scored wait times in different bins of five trials at transitions from mixed into low blocks, for control and optogenetic sessions. Consistent with the model, wait times in optogenetic sessions were slower to change following transitions into low blocks (Fig. 1T). Similar effects were not observed at other block transitions, consistent with the model prediction (fig. S 2).

We tested whether other model parameters could reproduce these patterns of behavioral results (fig. S 3). Simulating an additive gain to the opportunity cost associated with each block failed to reproduce the block-dependent changes in wait times or slower behavioral changes at mixed-to-low block transitions. Simulating a change in time-varying reward expectations (i.e., the reward hazard rate), or biased perception of rewards towards large reward volumes both could reproduce block-dependent changes in wait times but failed to recapitulate slower behavioral changes at mixed-to-low block transitions. These other model parameters were thus unable to reproduce all the patterns of behavioral changes produced by optogenetic stimulation of OFC→CPi neurons. We therefore hypothesized that optogenetically stimulating OFC→CPi neurons at the time of reward induced biased beliefs that persisted over trials, such that rats thought they were more likely to be in a high block and less likely to be in the other blocks.

OFC→CPi neurons encode high block evidence

We next sought to characterize the task-related dynamics of OFC→CPi and OFC→CPr neurons. We used a viral strategy to express ChR2.0 in these projection neurons, implanted a fiber optic in the CPi or CPr, and a Neuropixels probe in the OFC (Fig. 2A). We recorded neural activity from the probe as rats performed the task, and at the end of the behavioral session, we antidromically stimulated OFC→CP axons, to identify projection-specific neurons on the basis of collision tests (*29*). At a precise delay, orthodromic and antidromic action potentials should 'collide' and, due to refractory sodium channels on either side of the collision, cancel (Fig. 2B,C). Leveraging

our high-throughput behavioral training approach which yielded large pools of trained subjects for experiments, we recorded from 95 OFC neurons projecting to the CPr or CPi, identified from collision tests.

On average, OFC→CPi neurons appeared to encode evidence in favor of high blocks at the time of reward delivery (Fig. 2D), consistent with our optogenetic manipulation of OFC→CPi neurons in which stimulation biased beliefs towards high blocks. Specifically, they tended to have higher firing rates for 40/80 μL compared to 5/10 μL. We computed the mean unsigned discriminability index (d') for trials offering 40/80 μL versus 5/10 μL in mixed blocks (Methods). OFC→CPi neurons exhibited significantly higher d' values than OFC→CPr neurons, or the overall untagged OFC population (Fig. 2H), particularly at the time of reward. We trained a logistic regression decoder to decode whether rewards in mixed blocks were large (40/80 μL) or small (5/10 μL) based on the firing rates of OFC→CPi neurons, OFC→CPr neurons, or the total OFC population (Fig. 2I). Decoder performance was higher for OFC→CPi compared to OFC→CPr neurons, and comparable to the entire OFC population, which included many thousands of neurons, some of which presumably projected to the CPi.

40 and 80 μL are different reward volumes but provide identical evidence for high blocks, and 5 and 10 μL similarly provide identical evidence for low blocks (and against high blocks). The fact that the mean firing rates of OFC→CPi neurons were similar for 40/80 μL and 5/10 μL suggests that these neurons encode evidence for updating beliefs about the blocks, as opposed to encoding reward volumes *per se*. We wondered what mechanism might convert information about graded reward volumes into categorical evidence for belief updating. Notably, there was heterogeneity in the sign of evidence encoding in the OFC→CPi projection population (Fig. 2E). Visualizing the signed d' for large versus small rewards across all projection neurons revealed a bias towards positive d' values, that is, neurons with higher firing rates for large rewards. However, many neurons exhibited negative d' values, or higher firing rates for small rewards. We separated

neurons whose mean d' values in the [0 1s] window after reward were in the bottom (negative d') or top (positive d') quintile of d' values. On average, the positive d' neurons appeared to encode 40/80 μL categorically, but exhibited more graded encoding of the smaller rewards [5 10 20 μL] (Fig. 3A, left). The opposite was true for the negative d' neurons: they encoded 5/10 μL categorically but exhibited more graded encoding of larger rewards [20 40 80 μL] (Fig. 3B, left). This held if we quantified discriminability for 40 μL *versus* 80 μL and 5 μL *versus* 10 μL in these populations (Fig. 3A,B right). To control for potential motor correlates due to shorter drinking times for smaller rewards, we restricted this analysis to a window during which rats were drinking for even the smallest volumes, the [0 1.5s] window after reward. For both groups, preferred rewards that elicited higher firing rates were less discriminable than non-preferred rewards that elicited lower firing rates.

Given that preferred rewards elicit higher firing rates, we hypothesized that this effect might reflect a saturating non-linearity, potentially implemented at the circuit level. To test this hypothesis, we measured the maximum firing rate of each neuron at other task events, including the offer cue and when the side light turned on indicating the reward port. We found that most neurons had higher firing rates at these other task epochs, compared to the reward epoch (Fig. 3C-E), indicating that the saturating firing rate for preferred rewards is not a cell-intrinsic maximum, but likely reflects recurrent circuit dynamics including local inhibition. Consistent with this hypothesis, while OFC→CPi neurons were faithfully driven by 10Hz optogenetic stimulation, they showed reduced ability to follow 20Hz and 40Hz pulse trains (Fig. 3F). This ceiling likely reflects recruitment of inhibitory interneurons, which are well-entrained by higher frequency pulse trains. We speculate that such a circuit mechanism transforms graded representations of reward volumes, which are apparent for non-preferred rewards that elicit lower firing rates, in the neurons' linear dynamic range, into nonlinear categorical representations of evidence for belief updating. Heterogeneity in the sign of evidence encoding (i.e., the fact that some neurons have positive d' values

and others have negative d' values) enables categorical encoding of both 5/10 and 40/80.

Taken together, these findings suggest that neurons in the CPi integrate evidence for updating beliefs about blocks from OFC→CPi neurons at the time of reward. On average, the projection population coding axes are biased towards having high activity for high block evidence, which explains why stimulating this projection biased rats' beliefs towards high blocks.

Decodable belief distributions

Based on the responses of the OFC→CPi neurons, and our optogenetic results, belief updating appears to occur at the time of reward. We next wondered if the OFC→CPi neurons might represent subjective beliefs about the blocks, perhaps at other times in the trial. The posterior distribution over blocks is a discrete probability distribution with three values on each trial, one for the posterior probability of each block. Since the probabilities must sum to one, being able to decode any two of those probabilities allows one to estimate the third, i.e., know the full probability distribution (Fig. 4A).

Because the block probabilities are highly correlated with the reward offers in this task, we computed the average firing rate of OFC→CPi neurons for trials offering 20 μL, with similar MAP values but different full posterior distributions. Specifically, we selected trials where the most likely block was a mixed block, with a high probability (>= 0.8), and the second most likely block was either a high or a low block with a low probability (<0.05). Given that the reward (20 μL) and most likely block (mixed) are all identical on these trials, any differences in firing rates would be consistent with reflecting the full posterior distribution. Indeed, there was a significant difference in the average firing rates of OFC→CPi neurons across these conditions, particularly during the delay as animals waited for rewards (Fig. 4B). This is the time when their beliefs about the block could impact the behavioral policy of how long to wait. The difference after reward was trending but did not reach significance, although we note that this could reflect the limited number

of trials used for this analysis. Using all trials, the firing rates of individual OFC→CPi neurons often seemed to reflect the most likely block (i.e., maximum a posteriori or MAP estimate), and the second most likely block, both before and after reward delivery (Fig. 4C,D).

We used a decoding approach to determine whether downstream neurons in the striatum would be able to decode the first and second most likely block from this projection class. Most optogenetically-tagged neurons were not recorded simultaneously, so we constructed pseudo-populations for a logistic regression decoder, and evaluated performance using leave-one-out cross validation. It was possible to decode the first and second most likely block with high accuracy from this projection class (using separate decoders), indicating that in principle the CPi could estimate the full posterior distribution over blocks (Fig. 4E).

These results suggest that in addition to updating beliefs about the blocks at the time of reward, OFC→CPi neurons can in principle convey the full belief distribution over blocks to downstream neurons in the CPi, including as rats wait for rewards.

OFC→CPi neurons impact OFC block encoding

We previously identified neural correlates of inferred states in OFC at the single neuron and population levels (*18*). We therefore hypothesized that optogenetically updating beliefs about inferred states by stimulating OFC→CPi neurons might disrupt representations of inferred states in OFC. To test this hypothesis, we recorded population dynamics in OFC while optogenetically stimulating the terminals of OFC→CPi projection neurons (Fig. 5A). We used a supervised dimensionality reduction method, demixed PCA (*30*), to identify latent factors that reflect encoding of specific task variables, including time relative to different task events, the offered reward volume, and the hidden reward state or block (Fig. 5B). We performed neurophysiological recordings for two successfully completed blocks without any perturbations, to estimate the demixed principal components in the absence of perturbations. We identified latent factors or demixed

components from these control blocks. While recording from the same neurons, we then performed optogenetic stimulation on 40% of trials while animals received water rewards. The first two optogenetic blocks were identical to the previous two control blocks, and then the alternation between high and low blocks resumed. We projected neural activity from both conditions onto the components derived from the control blocks.

Due to the requirements of implementing demixed PCA, we separately compared blocks offering large rewards in high and mixed blocks (20, 40, 80 μL), and small rewards in low and mixed blocks (5, 10, 20 μL). Projecting activity from control blocks onto the component that most strongly encoded the hidden states showed differential dynamics across blocks but not for different rewards, as expected. However, following optogenetic stimulation of OFC→CPi terminals in the striatum, representations of blocks in OFC was degraded (Fig. 5C,E). We quantified this over sessions by comparing the average difference between the projection for the different blocks (Methods). In contrast, projecting activity onto the component that most strongly encoded reward volumes showed similar encoding for both control and optogenetic blocks (Fig. 5D,F). This was especially true for large rewards, which may produce stronger inputs that make the circuit dynamics more robust to perturbations. Stimulating OFC→CPi terminals at the time of reward disrupted population-level representations of hidden states in OFC, but largely spared encoding of rewards.

Optogenetic perturbations of axon terminals could impact OFC dynamics via back-propagating action potentials, or via long-range cortical-basal ganglia-thalamic loops. Surprisingly, stimulation of axon terminals produced relatively weak changes in the somatic spiking of optogenetically-tagged OFC→CPi neurons identified from collision tests, eliciting less than a single extra spike on stimulation compared to control trials (Fig. 5G). We speculate that these weak effects may be due to cortical inhibition that contributes to the saturating nonlinearity we previously documented at the time of reward (Fig. 3), or collisions, since this projection class

was often active at the time of reward/stimulation. These results suggest that OFC→CPi perturbations more likely impacted population-level cortical representations of inferred states through long-range cortical-basal ganglia-thalamic loops.

Discussion

OFC in rodents is required to infer values based on high-order associations (*13*), and to infer states with different task contingencies (*14–16*). Studies in rats and monkeys have found that OFC is important for learning when task contingencies are partially observable, potentially reflecting a key role in state inference (*14*, *17*, *31*, *32*). OFC is thought to support confidence judgments in perceptual decision-making (*33*, *34*), suggesting a role in computing or evaluating subjective belief distributions over perceptual stimuli. Our results extend these findings and show that (1) rat OFC supports belief updating (*18*), consistent with OFC playing a privileged role in learning (*27*, *35*, *36*), (2) projection-specific OFC→CPi neurons instantiate the belief updating computation, and their activation produces persistent effects that last over trials, (3) belief updating propagates through cortico-basal ganglia thalamic loops and impacts OFC representations of inferred states, and (4) local (presumably inhibitory) recurrent mechanisms convert graded representations of rewards into categorical representations of evidence for belief updating. The last observation suggests that OFC inhibitory interneurons may be a key locus of plasticity as animals learn to infer latent task structure from sensory features.

Tasks requiring the integration of noisy sensory evidence in favor of a binary choice (*37*, *38*), or motor timing tasks (*39*), have shown that the CP integrates cortical inputs over time on single trials until a bound or threshold is reached, and a movement is initiated indicating the subject's decision. For evidence accumulation, sensory evidence favors one of two mutually exclusive motor acts: saccade in the direction of moving dots, orient to the side with more auditory clicks. In these cases, it is conceptually straightforward that different CP neurons might support different

movements, and integrate sensory inputs in favor of those movements (*40*).

While we also find that the CP is the integrator in our task, our results differ from this previous work in several key respects. First, we find that the CP integrates block evidence across trials, as opposed to within them, raising the intriguing question of what neural mechanisms can support integration on across-trial timescales. Second, evidence is integrated to compute an abstract belief about latent states, as opposed to one of two mutually exclusive motor acts. Because the reward and opt-out ports randomly change on each trial in our task, striatal neurons cannot support integration in favor of a particular left/right choice or orienting movement. Instead, they appear to integrate block evidence to compute subjective beliefs about latent states. The behavioral policy (how long to wait on each trial), is then conditioned on the reward offer, subjective beliefs about states, and the side of the reward and opt-out ports on that trial. How this is implemented at the circuit level is an important future direction.

Uncertainty in sensory and motor domains has been studied for continuous variables, and experimental evidence has supported various neural coding schemes in these domains, including probabilistic population codes (*41–44*), distributed distributional codes (*45*), and neural sampling codes (*46–49*). However, none of these traditional coding schemes are applicable to dynamic inference of categorical (as opposed to continuous) variables. While previous studies have shown that dopamine neurons can reflect uncertainty about binary categorical variables (*50*, *51*), the simple nature of this problem makes it impossible to dissociate representation of the mean from that of the full distribution. Our work is consistent with studies showing that uncertainty is represented in higher order cortical regions (*34*, *52–54*), but provides novel insight into how specific projection classes explicitly represent probability distributions over discrete, categorical variables for dynamic inference of latent states.

We found that OFC→CPi neurons encode block evidence at the time of reward delivery, to a greater extent than OFC→CPr neurons and the broader OFC population. The average tuning

of OFC→CPi neurons at the time of reward approximates the likelihood of encountering each reward in a high block ($P(R_t \mid B_t = high)$), suggesting that this information may be combined with representations of prior beliefs in downstream neurons in the CPi. Downstream neurons integrate the spikes from OFC→CPi neurons to update beliefs about latent reward states, which are reflected in OFC (*18*), likely via long-range basal ganglia-thalamocortical loops. Notably, during the delay period, OFC→CPi neurons reflected the full belief distribution over blocks, indicating that neurons in the CPi in principle can access the full posterior distribution over latent states as animals wait for rewards. In contrast, the CPr may not be behaviorally engaged by this task, as we did not identify a consistent behavioral effect from perturbing OFC→CPr neurons, or specific task variables encoded by that population.

The recent proliferation of large-scale recording methods in neuroscience has led to a paradigm shift that emphasizes the analysis of neural populations over the activity of single neurons (*55*, *56*). We highlight two points raised by our findings. First, most cognitive tasks are complex and involve multiple algorithmic computations (e.g., working memory, state inference, deliberation between alternatives). In the absence of perturbations, which computation is actually supported by the piece of tissue under study is often unclear. In our recent study, persistent inactivation of OFC with muscimol impaired rats' ability to update their beliefs about the blocks (*18*). This is consistent with our present findings, in which transiently stimulating OFC→CPi neurons did update rats' beliefs, but in a biased, directional way (towards high blocks), due to the tuning of the OFC→CPi projection neurons, which on average have higher firing rates for high block evidence. Perturbations were critical for generating the hypothesis that OFC→CPi neurons instantiate the belief updating computation. It is not clear that we would have inferred this function from recordings alone.

Second, while population decoding approaches implicitly assume that downstream circuits can access all neurons (or some random subset of them), it is unclear if downstream circuits

uniformly sample from the population, or receive a circumscribed subset of response types. We found that OFC→CPi neurons encode specific task variables, and downstream circuits in the CPi do not uniformly sample from the entire population, consistent with past studies of OFC projections to the striatum (*23*). While the overall OFC population may have exhibited heterogeneous and mixed encoding of task variables, downstream neurons in the CPi received largely "unmixed" information about block evidence at the time of reward. Our findings show that the responses of single neurons can be highly informative of their function, especially in the context of an algorithmic hypothesis grounded in causal perturbations.

References

1. K. Doya, *Bayesian brain: Probabilistic approaches to neural coding* (MIT press, 2007).
2. N. D. Daw, A. C. Courville, D. S. Touretzky, Representation and timing in theories of the dopamine system. *Neural computation* **18**, 1637–1677 (2006).
3. A. D. Redish, S. Jensen, A. Johnson, Z. Kurth-Nelson, Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychological review* **114**, 784 (2007).
4. S. J. Gershman, D. M. Blei, Y. Niv, Context, learning, and extinction. *Psychological review* **117**, 197 (2010).
5. T. E. Behrens, M. W. Woolrich, M. E. Walton, M. F. Rushworth, Learning the value of information in an uncertain world. *Nature neuroscience* **10**, 1214–1221 (2007).
6. M. R. Nassar *et al.*, Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience* **15**, 1040–1046 (2012).
7. J. T. McGuire, M. R. Nassar, J. I. Gold, J. W. Kable, Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* **84**, 870–881 (2014).
8. A. Mah, S. S. Schiereck, V. Bossio, C. M. Constantinople, Distinct value computations support rapid sequential decisions. *Nature communications* **14**, 7573 (2023).
9. A. Mah, C. E. Golden, C. M. Constantinople, Dopamine transients encode reward prediction errors independent of learning rates. *Cell reports* **43** (2024).
10. R. Dudley, P. Taylor, S. Wickham, P. Hutton, Psychosis, delusions and the "jumping to conclusions" reasoning bias: a systematic review and meta-analysis. *Schizophrenia bulletin* **42**, 652–665 (2016).
11. R. Jardri, S. Duverne, A. S. Litvinova, S. Denève, Experimental evidence for circular inference in schizophrenia. *Nature communications* **8**, 14218 (2017).
12. H. Stuke, H. Stuke, V. A. Weilnhammer, K. Schmack, Psychotic experiences and overhasty inferences are related to maladaptive learning. *PLoS computational biology* **13**, e1005328 (2017).
13. J. L. Jones *et al.*, Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science* **338**, 953–956 (2012).
14. R. C. Wilson, Y. K. Takahashi, G. Schoenbaum, Y. Niv, Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81**, 267–279 (2014).
15. P. Vertechi *et al.*, Inference-based decisions in a hidden state foraging task: differential contributions of prefrontal cortical areas. *Neuron* **106**, 166–176 (2020).
16. A. Banerjee *et al.*, Value-guided remapping of sensory cortex by lateral orbitofrontal cortex. *Nature* **585**, 245–250 (2020).

17. M. F. Rushworth, M. P. Noonan, E. D. Boorman, M. E. Walton, T. E. Behrens, Frontal cortex and reward-guided learning and decision-making. *Neuron* **70**, 1054–1069 (2011).

18. S. S. Schiereck *et al.*, The orbitofrontal cortex updates beliefs for state inference. *Neuron* **114**, 507–520 (2026).

19. D. Hocker, C. M. Constantinople, C. Savin, Compositional pretraining improves computational efficiency and matches animal behaviour on complex tasks. *Nature Machine Intelligence* **7**, 689–702 (2025).

20. H. Hintiryan *et al.*, The mouse cortico-striatal projectome. *Nature neuroscience* **19**, 1100–1114 (2016).

21. U. Chon, D. J. Vanselow, K. C. Cheng, Y. Kim, Enhanced and unified anatomical labeling for a common mouse brain atlas. *Nature communications* **10**, 5067 (2019).

22. B. J. Hunnicutt *et al.*, A comprehensive excitatory input map of the striatum reveals novel functional organization. *elife* **5**, e19103 (2016).

23. J. Hirokawa, A. Vaughan, P. Masset, T. Ott, A. Kepecs, Frontal cortex neuron types categorically encode single decision variables. *Nature* **576**, 446–451 (2019).

24. C. M. Constantinople *et al.*, Lateral orbitofrontal cortex promotes trial-by-trial learning of risky, but not spatial, biases. *Elife* **8**, e49744 (2019).

25. A. M. Wikenheiser, M. P. Gardner, L. E. Mueller, G. Schoenbaum, Spatial representations in rat orbitofrontal cortex. *Journal of Neuroscience* **41**, 6933–6945 (2021).

26. R. Basu *et al.*, The orbitofrontal cortex maps future navigational goals. *Nature* **599**, 449–452 (2021).

27. K. J. Miller, M. M. Botvinick, C. D. Brody, Value representations in the rodent orbitofrontal cortex drive learning, not choice. *Elife* **11**, e64575 (2022).

28. J. L. Romero-Sosa, A. Yeghikian, A. M. Wikenheiser, H. T. Blair, A. Izquierdo, Neural coding of choice and outcome are modulated by uncertainty in orbitofrontal but not secondary motor cortex. *Nature Communications* **16**, 8931 (2025).

29. P. Bishop, W Burke, R Davis, Single-unit recording from antidromically activated optic radiation neurones. *The Journal of physiology* **162**, 432 (1962).

30. D. Kobak *et al.*, Demixed principal component analysis of neural population data. *elife* **5**, e10989 (2016).

31. T. A. Stalnaker *et al.*, Orbitofrontal neurons infer the value and identity of predicted outcomes. *Nature communications* **5**, 3926 (2014).

32. J. D. Wallis, Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* **30**, 31–56 (2007).

33. A. Lak *et al.*, Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* **84**, 190–201 (2014).

34. P. Masset, T. Ott, A. Lak, J. Hirokawa, A. Kepecs, Behavior-and modality-general representation of confidence in orbitofrontal cortex. *Cell* **182**, 112–126 (2020).

35. M. P. Gardner *et al.*, Processing in lateral orbitofrontal cortex is required to estimate subjective preference during initial, but not established, economic choice. *Neuron* **108**, 526–537 (2020).

36. K. M. Costa *et al.*, The role of the lateral orbitofrontal cortex in creating cognitive maps. *Nature neuroscience* **26**, 107–115 (2023).

37. M. M. Yartsev, T. D. Hanks, A. M. Yoon, C. D. Brody, Causal contribution and dynamical encoding in the striatum during evidence accumulation. *Elife* **7**, e34929 (2018).

38. T. Doi, Y. Fan, J. I. Gold, L. Ding, The caudate nucleus contributes causally to decisions that balance reward and uncertain visual information. *Elife* **9**, e56694 (2020).

39. Z. Yang, M. Inagaki, C. Gerfen, L. Fontolan, H. K. Inagaki, The frontal cortex adjusts striatal integrator dynamics for flexible motor timing. *bioRxiv*, 2024–06 (2024).

40. P. Znamenskiy, A. M. Zador, Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature* **497**, 482–485 (2013).

41. W. J. Ma, J. M. Beck, P. E. Latham, A. Pouget, Bayesian inference with probabilistic population codes. *Nature neuroscience* **9**, 1432–1438 (2006).

42. M. Jazayeri, J. A. Movshon, Optimal representation of sensory information by neural populations. *Nature neuroscience* **9**, 690–696 (2006).

43. S. Deneve, Bayesian spiking neurons I: inference. *Neural computation* **20**, 91–117 (2008).

44. J. M. Beck *et al.*, Probabilistic population codes for Bayesian decision making. *Neuron* **60**, 1142–1152 (2008).

45. M. Sahani, P. Dayan, Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural computation* **15**, 2255–2279 (2003).

46. J. Fiser, P. Berkes, G. Orbán, M. Lengyel, Statistically optimal perception and learning: from behavior to neural representations. *Trends in cognitive sciences* **14**, 119–130 (2010).

47. C. Savin, S. Denève, Spatio-temporal representations of uncertainty in spiking neural networks. *Advances in neural information processing systems* **27** (2014).

48. G. Orbán, P. Berkes, J. Fiser, M. Lengyel, Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron* **92**, 530–543 (2016).

49. R. M. Haefner, P. Berkes, J. Fiser, Perceptual decision-making as probabilistic inference by neural sampling. *Neuron* **90**, 649–660 (2016).

50. C. K. Starkweather, B. M. Babayan, N. Uchida, S. J. Gershman, Dopamine reward prediction errors reflect hidden-state inference across time. *Nature neuroscience* **20**, 581–589 (2017).

51. S. J. Gershman, N. Uchida, Believing in dopamine. *Nature Reviews Neuroscience* **20**, 703–714 (2019).

52. A. Kepecs, N. Uchida, H. A. Zariwala, Z. F. Mainen, Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**, 227–231 (2008).

53. R. Kiani, M. N. Shadlen, Representation of confidence associated with a decision by neurons in the parietal cortex. *science* **324**, 759–764 (2009).

54. T. A. Langlois, J. A. Charlton, R. L. Goris, Bayesian inference by visuomotor neurons in the prefrontal cortex. *Proceedings of the National Academy of Sciences* **122**, e2420815122 (2025).

55. S. Saxena, J. P. Cunningham, Towards the neural population doctrine. *Current opinion in neurobiology* **55**, 103–111 (2019).

56. R. B. Ebitz, B. Y. Hayden, The population doctrine in cognitive neuroscience. *Neuron* **109**, 3055–3068 (2021).

57. M. W. Khaw, P. W. Glimcher, K. Louie, Normalized value coding explains dynamic adaptation in the human valuation process. *Proceedings of the National Academy of Sciences* **114**, 12696–12701 (2017).

58. W. L. Conte, H. Kamishina, R. L. Reep, The efficacy of the fluorescent conjugates of cholera toxin subunit B for multiple retrograde tract tracing in the central nervous system. *Brain Structure and Function* **213**, 367–373 (2009).

59. T. Z. Luo *et al.*, An approach for long-term, multi-probe Neuropixels recordings in unrestrained rats. *Elife* **9**, e59716 (2020).

60. G. Paxinos, C. Watson, *The rat brain in stereotaxic coordinates: hard cover edition* (Elsevier, 2006).

Acknowledgments

We thank Roozbeh Kiani and members of our labs for comments on the manuscript and helpful discussions. We thank research technicians in the Constantinople lab for animal training.

Funding: This work was supported by F32MH132176 (MLD), an NSF CAREER Award (CMC), NIH grant R01DA065418 (CMC and CS), and a McKnight Scholars Award (CMC).

Author Contributions

Conceptualization: MLD, CS, CMC

Methodology: MLD, DH, HG

Investigation: MLD, JS, MAW, DH, HG, SSS

Visualization: MLD, DH, HG

Funding acquisition: MLD, CS, CMC

Project administration: CS, CMC

Supervision: CS, CMC

Writing – original draft: MLD, CMC

Writing – review editing: MLD, DH, HG, CS, CMC

Competing Interests

The authors declare that they have no competing interests.

Data, code, and materials availability

The data generated in this study will be deposited in a Zenodo database upon publication. Code used to analyze all data and generate figures will be available at https://github.com/constantinoplelab/published/tree/main upon publication.

Supplementary materials

Materials and Methods

Rats

593 Long-evans rats (*Rattus norvegicus*) between the ages of 6 and 24 months were used in this study (367 males, 226 females). Rats were water restricted to motivate them to perform the task. At the end of the day, rats were given 20 minutes of access to ad libitum water. Following behavioral sessions on Friday until mid-day Sunday, rats received ad libitum water. Animal use procedures were approved by the New York University Animal Welfare Committee (UAWC #2021-1120) and carried out in accordance with National Institute of Health standards.

Behavior training and task logic

Rats were trained in incremental stages of task complexity in a high-throughput behavioral facility using a computerized training protocol as described previously (*8*).

LED illumination from the central nose port indicated that the animal could initiate a trial by poking its nose in that port. Upon trial initiation the center LED turned off. While in the central nose port, rats needed to maintain center fixation for a duration drawn uniformly from [0.8, 1.2] seconds. During the fixation period ("Offer cue"), a tone played from both speakers, the frequency of which indicated the volume of the offered water reward for that trial [1, 2, 4, 8, 16 kHz, indicating 5, 10, 20, 40, 80 μL rewards for male rats and 4, 8, 16, 32, 64 μL for female rats]. Female rats consumed less water due to their spaller body size and performed substantially fewer trials than the males when offered the same rewards as the males. For analysis, the reward volumes were treated as equivalent to the corresponding values for the male rats. Following the fixation period, one of the two side LEDs was illuminated, indicating that the reward might be delivered at that port ("Reward port on"); the side was randomly chosen on each trial. This

event also initiated a variable and unpredictable delay period, which was randomly drawn from an exponential distribution with mean = 2.5 seconds. The reward port LED remained illuminated for the duration of the delay period, and rats were freely behaving during this period. When reward was available, the reward port LED turned off ("Reward announced"), and rats could collect the offered reward by nose poking in that port ("Reward delivered"). The rat could also choose to terminate the trial at any time by nose poking in the opposite, un-illuminated side port ("Opt Out"), after which a new trial would immediately begin. On a proportion of trials (15-25%), the delay period would only end if the rat opted out (catch trials). If rats did not opt-out within 100 seconds on catch trials, the trial would terminate.

The trials were self-paced: after receiving their reward or opting out, rats were free to initiate another trial immediately. However, if rats terminated center fixation prematurely, they were penalized with a white noise sound and a time out penalty, typically lasting 2 seconds ("Violation"). The trial following a premature fixation break offered the rats the same amount of reward in order to dis-incentivize premature terminations for small volume offers.

We introduced semi-observable, hidden-states in the task by including uncued blocks of trials with varying reward statistics (*57*): high and low blocks, which offered the highest three or lowest three rewards respectively, were interspersed with mixed blocks, which offered all volumes. Unless otherwise noted, there was a hierarchical structure to the blocks, such that high and low blocks alternated after mixed blocks (e.g., mixed-high-mixed-low, or mixed-low-mixed-high). Each session always started with a mixed block.

In a subset of sessions used to test the impact of OFC→CPi pertubation on cortical representation of inferred states we modified the hierarchical structure to the blocks, such that the first two blocks of the session were repeated twice and then began to alternate between high and low block (e.g. mixed-high-mixed-high-mixed-low). These sessions varied day-by-day so that the animal not experience one block more often than the other.

In both cases blocks transitioned after 40 successfully completed trials. Because rats prematurely broke fixation on a subset of trials, in practice, blocks varied in number of trials.

Rats were considered to have achieved task proficiency when their time to opt-out exhibited linear sensitivity to the offered reward volume and time to opt-out for the 20 μL offer in low blocks was longer than in high blocks for at least 3 consecutive days. Rats that satisfied these criteria underwent surgeries specific to experiments. After recovering from surgeries, rats were retrained until they achieved the same pre-surgery performance which typically took one session.

Behavioral model

We made use of the behavioral model previously established to describe rats' behavior in the temporal wagering task (*8*, *18*). In brief, the model is adapted from Lak et al. (*33*) which described wait time (WT) in terms of the value of the environment (or opportunity cost of continuing to wait for the reward to arrive), the delay distribution, and the catch trial probability. Given an exponential delay distribution we predict the wait time as:

$$WT = D\tau\left(\frac{C}{1-C} \times \frac{R-\kappa\tau}{\kappa\tau}\right)$$

where τ is the time constant of the exponential delay distribution, C is the the probability of reward, R is the reward volume on the current trial, κ is the opportunity cost, and D is a scaling parameter. The decision to opt out depends on the opportunity cost, κ, which is stable within trials but can vary across trials and hidden reward states, i.e. blocks. The inferential model has three discrete opportunity cost parameters associated with each block (κ_{low}, κ_{mix}, κ_{high}). On each trial, the model chose the κ associated with the most probable block given the recent reward history according to Baye's theorem, described by:

$$P(B_t|R_b) \propto P(R_t|B_t)P(B_t)$$

where B_t is the block and R_t is the reward on trial t. The likelihood, $P(R_t \mid B_t)$ is the probability of the reward for each block. For example:

$$P(R_t|B_t = Low) = \begin{cases} \frac{1}{3}, & for\ R_t = 5, 10, 20\ \mu L \\ 0, & for\ R_t = 40, 80\ \mu L \end{cases}$$

To calculate the prior, $P(B_t)$, we marginalized over the previous block and used the previous estimate of the posterior:

$$P(B_t) = \sum_{B_{t-1}} P(B_t|B_{t-1})P(B_{t-1}|R_{t-1})$$

The "hazard rate," $P(B_t \mid B_{t-1})$ incorporates knowledge of the task structure, including the block length and block transition probabilities. For example:

$$P(B_t = Low|B_{t-1}) = \begin{cases} 1 - H_0, & for\ B_{t-1} = Low \\ H_0, & for\ B_{t-1} = Mix \\ 0, & for\ B_{t-1} = High \end{cases}$$

where H_0 = 1/40 to reflect the block length. Including H_0 as an additional free parameter has been shown previously to not improve the performance of the inferential model on held-out test data (*8*), so H_0 was treated as a constant term. See (*8*) for additional discussion of the hazard rate.

Biased behavioral model

For the model with a biased prior, we included a term that was added to the probability of being in a high block, and subtracted from the low and mixed block probabilities as in:

$$P(B_t = \text{Low}) = P(B_t = \text{Low}) - bias \tag{6}$$

$$P(B_t = \text{Mix}) = P(B_t = \text{Mix}) - bias \tag{7}$$

$$P(B_t = \text{High}) = P(B_t = \text{High}) + bias \tag{8}$$

After the additive bias was introduced, negative beliefs were set to zero and the block probabilities were normalized to values between zero and one. The bias used for the model simulations was set to 0.04.

In separate models we modulated the opportunity cost of all blocks by increasing the κ for

each block by 0.15, or modulated the time constant of the time-varying value function, τ, by reducing its value from 2.5 to 2.0. In a third model we replaced the offered reward, R_t, with 80 μL on 40% of trials, to simulate a perturbation that altered the perception of reward volume.

Surgery

Rats were anaesthetized (isofluorane at 3% for induction and 1.5-2.5% for maintenance) and placed in a stereotaxic frame. In experiments in which hardware was attached to the skull, skull screws were inserted to secure implants before proceeding with surgeries specific to experiments. Target coordinates relative to the skull surface at bregma for the OFC were: AP: 3.7 mm, ML: ± 2.9mm, DV: 5 mm, the CPi: AP: -0.2 mm, ML: ± 3.5 mm, DV: 4.5 mm, and CPr: AP: 1.2 mm, ML: ± 2.5 mm, DV: 6.9 mm, SNr: AP: -5.0, -5.3, -5.7 mm, ML: ± 2.3 mm, DV: 8.4, 8.3, 8.2 mm, and GPe: AP: -1.5 mm, ML: ± 3.3 mm, DV: 6.8 mm.

Anatomical experiments

For anterograde tracing of OFC projection neurons, rats were injected intracranially with 100 nL of pENN.AAV.CB7.CI.mCherry.WPRE.RBG (AAV9) (AddGene, #105544-AAV9) in the OFC. Viral expression was given 2 weeks to mature before animals were sacrificed.

For retrograde tracing of OFC projection neurons, rats were injected with 120 nL of Alexa fluorophore 657 conjugated cholera toxin subunit B (ThermoFisher Scientific, C34778) in either the CPi or CPr. CTB was diluted at 0.5% (wt/vol) in PBS (*58*). Following the injection, the burr hole with filled with Kwik-Sil (WPI), and the skin was sutured. The CTB was given 10 days to spread retrogradely to somata.

598 **Optogenetic stimulation experiments**

599 For optogenetic manipulations and optotagging protocols, rats were injected with 500 nL of
600 retrograde-serotype Cre virus ("Cre_{rg}", AAVrg-pgk-Cre, AddGene #24593) in the CPi or CPr,
601 and Cre-dependent channelrhodopsin ("ChR2", AAV9-EF1a-double floxed-hChR2, AddGene
602 #20298) in the OFC. 100 nL of virus was injected at a time at a rate of 10 nL sec^{-1}.

603 Rats used for optogenetic manipulation, but not electrophysiology, were then immediately
604 bilaterally implanted with a tapered fiber optic (0.66 NA, 1 mm active length, 200 μM core,
605 Optogenix) in the CPi or CPr such that the emitting tip of the fiber was within the target region.
606 The fiber was secured to the skull surface using dental cement and Dentin and then the headcap
607 was sealed with dental acrylic. Rats were allowed to recover for a minimum of 5 days. Rats
608 performed the task for at least 2 weeks before any LED stimulation was delivered during the task
609 to allow for cells to fully express the virus.

610 **Electrophysiological experiments**

611 Rats used for optotagged electrophysiological experiments underwent two distinct surgeries;
612 one to inject viruses and a second to implant physiological hardware after the viruses were fully
613 expressed. We found that chronically implanted Neuropixels probes could be damaged or signals
614 could degrade sufficiently before the ChR2 virus was fully expressed. Following viral injections
615 as described for optogenetic manipulation, all injection site burr holes were sealed with Kwik-sil
616 (WPI) and the skin sutured. Viruses were given 2-6 weeks to express. During this period rats
617 were first given 5 days to recover, and then were returned to performing the task daily to ensure
618 their behavior remained stable.

619 The day before the implantation surgery, a Neuropixels probe assembly was prepared by
620 mounting the probe (Neuropixels probe 1.0 or 2.0) onto a custom-designed 3D printed piece
621 based on other chronically implanted headcaps (*59*) and staining with DiI. Open Ephys built-

in self tests were run on the probe assembly prior to surgery to ensure all probe shanks were functional prior to implantation.

During the second surgery, fiber optics were implanted bilaterally in the CPi or CPr, and the Neuropixels probe was implant using the probe assembly in the OFC. A silver ground wire that shorted the ground and reference connection on the probe was inserted in the cerebellum and fixed with Dentin (Parkell). Craniotomies were covered with Kwik-Cast (WPI) and gaps between the probe assembly and the skull were sealed with sterile petroleum jelly. The probe and fibers were then fixed to the skull with Metabond (Parkell) and Absolute Dentin (Parkell), and sealed with Ortho-jet dental acrylic (Patterson Dental Comparny). Bubble wrap was placed around the implant to protect it from damage. Rats were allowed to recover for a minimum of 5 days before returning to performing the task and electrophysiological recording.

Histology

Following completion of each experiment, rats were deeply anesthetized and transcardially perfused with phosphate buffer saline (PBS) and formalin. Brains were post-fixed with formalin for 3 days and sectioned at 50 μm slices on a Leica VT1000 vibratome.

Immunohistochemistry

To confirm ChR2 expression, ChR2-conjugated EYFP signal was amplified using antibody labeling. Sections were treated with 1% hydrogen peroxide in 0.01 M PBS for 30 min, washed with 0.01 M PBS, and blocked in 0.01 M PBS containing 0.05% sodium azide and 1% bovine serum albumin (PBS-BSA-Azide) for 30 minutes prior to treating with primary antibodies. EFPF was amplified with Rabbit anti-GFP (Thermo Fisher Scientific A11122, 1:2,000, lot #2083201) as primary antibody and Goat anti-Rabbit AlexaFluor 488 IgG (Thermo Fisher Scientific A11008, 1:200, lot #2147635) as secondary antibody. Primary and secondary antibodies were made in PBS-BSA-Azide and primary antibodies additionally contained 0.2% Triton-x. Sections were

646 incubated in primary antibody overnight at room temperature, washed with PBS for 15 min, and
647 treated with secondary antibody for 1 hour in the dark. After washing with PBS for 15 min,
648 sections were mounted on glass slides using Prolong Gold (Vector Laboratories).

649 **Imaging**

650 For electrophysiology and optogenetic experiments, sections were imaged with an Olympus
651 VS120 Virtual Slide Microscope to verify probe and fiber optic locations, and ChR2 expression.
652 Neuropixels probe tracks were reconstructed from post-mortem histology, and the location of
653 individual recording channels relative to areal boundaries was estimated. Channels that were
654 estimated to be outside of OFC or agranular insula (AI) were excluded from further analysis.

655 For anatomical experiments, injection sites were confirmed using 10x magnification on an
656 Olympus VS120 Virtual Slide Microscope. Light intensity was controlled across all sections.
657 Cell bodies were imaged at 10x and 20x magnificantion using a Leica DM6 CS Confocal with
658 standardized parameters developed to maximize signal-to-noise ratios. For each animal 3 sections
659 near the target location of AP: Bregma +3.7 mm were selected by anatomical markers and imaged.
660 Brain sections were then manually aligned to standardized brain sections described in Paxinos et.
661 al. (*60*) to reduce variability caused by brain size, or stretching/shrinking during processing or
662 mounting, and enable averaging across animals. OFC projection neurons somata were manually
663 labeled and counted using the ImageJ cell counting pluggin which tracks number and location.
664 Cell number was averaged within 250 μm^2 bins tiling the brain section and then further averaged
665 across the 3 sections imaged for each animal. We then calculated the mean cell number in each
666 tile across animals.

Electrophysiological recordings

Neuropixels data acquisition

Spike data was acquired from 384 channels at 30 kHz using Neuropix-PXI hardware and OpenEphys (https://github.com/open-ephys). To synchronize physiological recordings with behavior and optogenetic LED pulses, TTL pulses were recorded simultaneously using NI-DAQmx. NI-DAQmx and Neuropix-PXI recordings were aligned using a six digit temporal barcode generated by an Arduino Zero and simultaneously sent to both recordings.

Spike sorting

Spike data were preprocessed using Kilosort 2.5. After preprocessing, clusters that were identified from Kilosort as single-units were manually inspected using the Phy2 Python package (https://github.com/cortex-lab/phy, no longer maintained). Clusters identified by Kilosort as "good" units with unreliable waveform shape, autocorrelogram, or spike amplitude time course were manually classified as multi-unit or noise and removed from further analysis. Units were further curated using a custom MATLAB script. Units with greater than 1% inter-spike intervals less than 1 ms, firing rates less than 1 Hz, or completely silent for more than 5% of the total recording were excluded. To convert spikes to firing rates, spike counts were binned in 50 or 200 ms bins where described in the text.

Optogenetic identification of projection specific neurons

Optogenetic tagging of projection specific neurons was performed after the rat had completed its 90 minute behavioral session to prevent optogenetic stimulation influencing the rat's behavior. Blue light (465-nm PlexBright LEDs, Plexon) was delivered in 20 trains of 10 pulses at 10, 20, 40 Hz with 1, 3, 5 ms pulse duration, and 100, 200, 300 mA current applied to the LED to elicit action potentials in OFC projection neuron axon terminals. Optically tagged neurons were

identified using 10 Hz, 3 ms, 200 mA, pulses. During the optotagging protocol rats were left in the behavioral chamber without a task to perform so that stimulation could not affect task performance on subsequent days.

Neurons which reliably generated at least one action potential (i.e. with response to a greater number of LED pulses than two standard deviations from the mean of all OFC cells) within the same 2 ms window following LED onset and with a latency less than 20 ms were further assessed for action potential collisions. These criteria reduced the number of cells assessed for collisions from 7697 to 251 cells. We manually assessed LED pulse 'trials' in which the neuron elicited an action potential before the median latency of the first action potential after LED onset. If a spike occurred before the median latency, as well as at the expected delay following LED onset, the neuron failed the collision test, and was not included as a direct projection neuron.

Optogenetics experiment

After recovering from surgery, rats were habituated to being tethered until they achieved pre-surgery performance which typically took 1-2 sessions. Following habituation, control sessions (no LED) and stimulation sessions (LED on on 40% of trials) were alternated every 5 days for a minimum of 12 sessions total per rat. Blue LED modules (465 nm, 1.9-2.3 mW, PlexBright LEDs, Plexon) were connected to a commutator and 200 μM core patch cords that were tethered to the fiber implants of rats. The connections between the LEDs, patch cords, and fiber implants were sealed with black electric tape or black shrink tubes to block light leak. Continuous blue light (465-nm PlexBright LEDs, Plexon) was delivered randomly on 40% of trials for 500 ms during reward delivery. Optical power was measured daily prior to the animal performing the task and the current applied to the LED was modulated to maintain an average of 10 mW during a 1 s continuous pulse (Thorlabs, PM100D2).

Quantification and statisical analyses

Exact p-values were reported if greater than 10^{-20}. For p-values smaller than 10^{-20} , we reported $p << 0.001$

Detrending wait times

To account for modest, gradual increases in wait times over the course of the behavioral session, and remove potential effects of motivation and satiety from the wait times, we performed the following detrending procedure for all wait time analyses. First, we removed outlier wait times that were one standard deviation above the pooled-session mean. We fit a regression line to the mean wait time as a function of trial number, averaging over all sessions, using MATLAB's regress.m function. We then subtracted the regression line from the wait times on each session, adding back the offset parameter so that wait time units were in seconds and not zero-meaned.

Wait time sensitivity to reward blocks

For all analyses, we removed wait times that were one standard deviation above the pooled-session mean. Without thresholding, the contextual effects are qualitatively similar, but the wait time curves are shifted upwards because of outliers that likely reflect inattention or task disengagement. When assessing whether a rat's wait time differed by blocks, we compared each rat's wait time on catch trials offering 20 μL in high and low blocks using a non-parametric Wilcoxon rank-sum test, given that the wait times are roughly lognormally distributed.

To determine the effect of optogenetic stimulation on wait times, we normalized the rat's wait time to their mean wait time for 20 μL trials in control session mixed blocks. Because each animal underwent both control and optogenetic sessions, we used Wilcoxon signed-rank test to quantify change in wait time with optogenetic manipulations. To measure difference across the entire block caused by optogenetic stimulation, we averaged the normalized wait times for all

volumes in each block, and subtracted the mean wait time for each block in control sessions from the mean wait time for each block in optogenetic stimulation sessions. We used a Friedman test to assess for difference across all blocks, and when appropriate used Wilcoxon signed-rank tests to test individual comparisons.

Block transition dynamics

To assess behavioral dynamics around block transitions, for each rat, we detrended wait times over the course of the session, and then z-scored wait-times for all catch trials. We then z-scored wait times for each volume in each time bin (trial relative to block transition), before averaging over all volumes. We smoothed the average curve for each rat using a 5-point causal filter, before averaging over rats.

Because rats receiving optogenetic stimulation had a limited number of trials, we did not always have wait times for each volume at each time bin relative to block transition, so we combined trials into 5 or 10 trial windows around block transitions. We then compared wait times within these windows in control and optogenetic sessions using a Friedman test, and Wilcoxon signed-rank test for individual comparisons when appropriate.

Discriminability index

To measure encoding discriminability used a d' metric which used the following equation:

$$d' = \frac{\mu_{small} - \mu_{large}}{\sqrt{0.5(\sigma_{small}^2 + \sigma_{large}^2)}} \tag{9}$$

where μ_{small} represented the mean firing rate on smaller reward volume mixed block trials and μ_{large} represented the mean firing rate on larger reward volume mixed block trials, and $0.5(\sigma_{small}^2 + \sigma_{large}^2)$ is the mean variance.

We then subtracted the mean d' of a label-shuffled control. Shuffles were computed 100 times and maintained a balance of low and high reward trial, and used the 200 ms time bins to

758 generate the shuffle, as this was the bin-width used for computing the reward-aligned firing rates.
759 We indicate d' values greater than the 95th percentile of the label-shuffled d' for that population
760 with a larger line width in each figure. We then only compared d' values between different
761 projection classes at time points where the projection specific neurons' d' was greater than the
762 95th percentile of the label-shuffled d' for that population. We denote time points at which the d'
763 of OFC→CPi or OFC→CPr differed from the untagged OFC population using a nonparametric
764 permutation test with lines above the data.

765 **Decoding reward volume**

766 We used firing rates from different projection classes to decode reward volume on each trial.
767 We decoded low volumes (5/10 μL) or high volumes (40/80 μL) on mixed block trials using
768 neural data in 6 ms time bins. We trained logistic regression decoders (SciKit-Learn) using a 2-
769 fold cross validation where within each fold a decoder was fit to a training set using a no penalty,
770 saga solver. We report the average performance for decoding held-out test data for each fold.

771 **Decoding most likely and second most likely block**

772 We used neural responses from optotagged OFC→CPi neurons to separately classify each
773 trial's most likely block and second most likely block on each trial. These class labels were
774 generated from a Bayes optimal model's estimation of block probabilities (*8*). We only decoded
775 from rewarded trials, and used neural data averaged over 200 ms time bins as the regressors.
776 We took a pseudo-population approach: data across multiple sessions from non-simultaneously
777 recorded neurons was used, and for each "pseudo trial" that was decoded, we chose a trial each
778 from neuron's recordings that shared the same class label.

779 The data sampling procedure was as follows: we performed 20-fold cross validation, where
780 for each fold we randomly sampled neurons (with replacement) with correct class labels to obtain
781 pseudo populations for 100 pseudo trials as a test set for each condition (i.e., 100 low blocks, 100

mixed blocks, 100 high blocks). For the test set, we generated those 100 pseudo trials from 20 distinct trials for each condition (i.e, 20 distinct low block trials, 20 distinct mixed block trials, 20 distinct high block trials). The same procedure was performed for a training set of 500 pseudo trials for each class label, but the neural recordings from these 60 withheld trials for testing data were not used to ensure that they were never seen by the decoder. Decoding was trained with Matlab's 'fitmnr' function for every 200 ms time bins from [-1,3]s around reward delivery, and testing accuracy was averaged over the withheld test data set. This process was performed 20 times, and reported accuracies are averaged over all 20 folds. One recording session did not contain enough trials of each type to adequately build the testing data set, which forced us to omit 3 neurons from the decoding. Thus, 57 neurons were used in the decoder.

In order to build a null distribution, we shuffled the trial labels of the optotagged recordings, and performed the same analysis using 20-fold cross validation. We performed the shuffle 20 times, and reported the results as the distribution over fold-averaged test accuracies. Finally, we also performed the analysis using data from all OFC units. In order to balance the statistics of neurons used across OFC→CPi projectors and all OFC, for each of the 20-fold regressions we randomly sampled 57 neurons, and then performed the regression as above. To assess if OFC→CPi significantly decoded most likely block and second most likely block more than the general OFC population, we performed Wilcoxon signed-rank tests in each 200 ms time bin, and assessed significance at $p < 0.05$.

Demixed principal component analysis

To assess the impact of projection specific manipulations on population-level cortical representations of inferred states we used demixed principal component analysis (*30*) (dPCA, software available at: http://github.com/machenslab/dPCA). This aims to decompose the data into interpretable latent components that explain variance in the data. We used reward volume, block

identity, and time (task event aligned temporal dynamics), as categorical predictors of neuronal activity. Because dPCA requires that each variable is present in each category we were unable to compare high and low blocks directly as high and low blocks include non-overlapping reward volumes. Therefore, compared mixed to low blocks and mixed to high blocks for each session. To ensure the rat experienced a high/low block in both the control and stimulation phases of a single session, we modified the task design such that on alternating days the animal experienced a mixed and high or low block with the LED turned off ("Opto off") and those two blocks repeated once the LED was turned on for 40% of trials ("Opto on"). After the first two blocks of the Opto on phase, the blocks resumed alternation between high and low blocks interspersed with mixed blocks as in the standard task. For analysis, we required that all sessions had at least 2 trials of each reward volume in each block and binned spikes into 50 ms time bins to generate firing rates.

For each session, we first generated a weight matrix and demixed principal components (dPCs) for each session using the activity during the first two blocks (mixed and low, or mixed and high) in the session during the Opto off phase. We projected neuronal activity onto the top block or reward dPCs (i.e., the block and reward PCs that explained the most variance) to generate PSTHs of activity for each category. To generate PSTHs of the neuronal activity during Opto on, we projected neuronal activity recorded during the Opto on phase using the weight matrix and principal components estimated from control trials.

To compare across sessions, we calculated the mean difference in block or reward encoding during the Opto off and Opto on phases across all sessions. Specifically, we computed the mean difference between the preferred block (which ever had the higher firing rate) and nonpreferred block for each reward volume (20, 40, 80 μL for high vs. mixed block comparisons, and 5, 10, 20 μL for low vs. mixed block comparisons) in the Opto off and Opto on phase for each session and then averaged these values across all sessions. To determine changes in reward encoding, we found the mean difference between the highest and lowest volumes (20 and 80 μL for mixed vs.

high and 5 and 20 μL for mixed vs. low) in the Opto off and Opto on phase for each session and then averaged these values across all sessions.

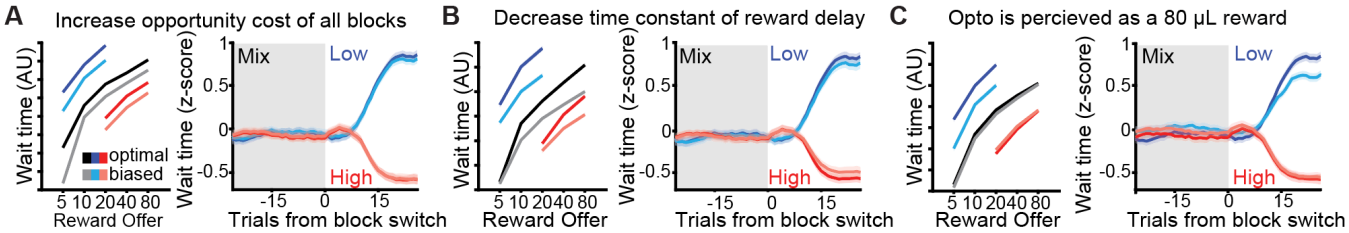

Fig. S 1: **OFC→CPr activations does not affect wait times A.** Schematic to target OFC→CPr neurons for LED stimulation during reward delivery. **B.** Wait time for an example, and the population of implanted rats. Wilcoxon signed-rank test for difference between control and opto wait times, mixed block: 5 μL $p = 0.0781$, 10 μL $p = 0.0781$, 20 μL $p = 0.2969$, 40 μL $p = 0.4688$, 80 μL $p = 0.4688$, low block: 5 μL $p = 0.5781$, 10 μL $p = 0.0781$, 20 μL $p = 0.8125$, high block: 20 μL $p = 0.1562$, 40 μL $p = 0.3750$, 80 μL $p = 0.9375$. **C.**Mean difference between opto and control session wait times averaged across all volumes for each block, Friedman's test $p = 0.7807$.

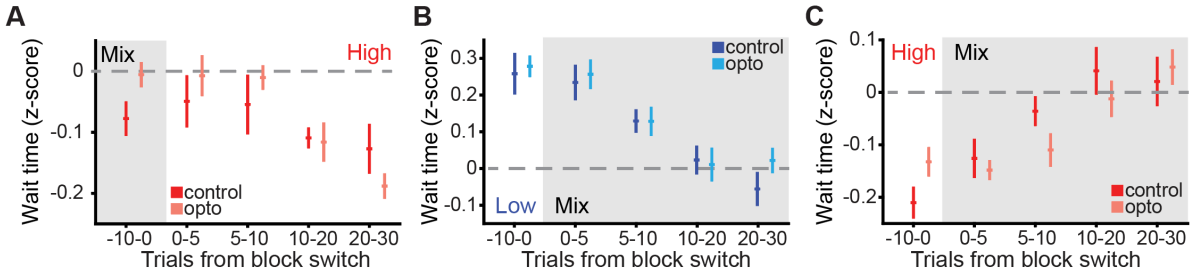

Fig. S 2: **Rat wait times at other transitions are unaffected by stimulation**. **A.** Mean wait time across all animals in control and opto sessions around mixed to high block transitions grouped into 5 trial bins. Friedman test, $p = 0.5867$). **B.** Mean wait time across all animals in control and opto sessions around low to mix block transitions grouped into 5 trial bins. Friedman test, $p = 0.7353$). **C.** Mean wait time across all animals in control and stimulation sessions around high to mix block transitions grouped into 5 or 10 trial bins. Friedman test, $p = 0.7803$).

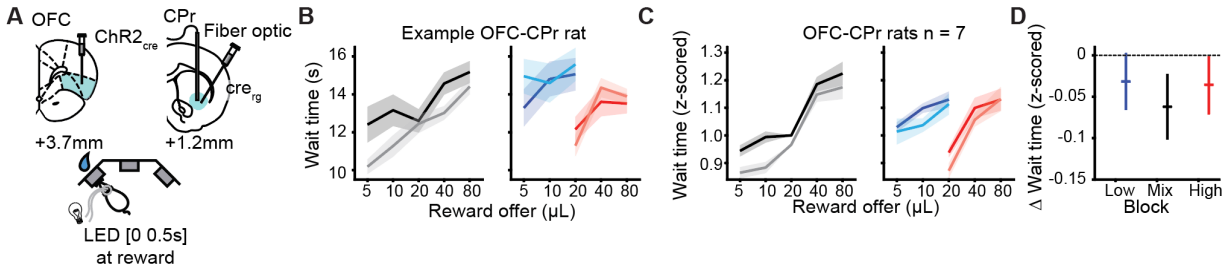

Fig. S 3: **Alternative model biases do not replicate animal behavior. A.** Model in which the opportunity cost of all blocks was increased decreases wait times, but does not modulate transitions. **B.** Model in which the time constant of the reward delay decreases. **C.** Model in which the reward offer is instantiated as 80 μL regardless of the block or actual offered reward.

	OFC-CPi p-values				OCF-CPr p-values			
Time point	Offer cue	Reward Port on	Reward announced	Reward delivered	Offer cue	Reward Port on	Reward announced	Reward delivered
-1	0.34	0.199	<0.001	0.015	<0.001	<0.001	<0.001	0.052
-0.9	0.006	0.165	0.043	0.289	<0.001	<0.001	<0.001	0.003
<0.001	<0.001	<0.001	0.194	0.034	0.003	<0.001	<0.001	<0.001
-0.7	<0.001	<0.001	0.002	0.002	<0.001	<0.001	<0.001	0.16
-0.6	0.142	<0.001	0.077	0.43	<0.001	<0.001	<0.001	0.01
-0.5	0.119	<0.001	0.001	<0.001	0.007	<0.001	<0.001	<0.001
-0.4	0.007	0.001	<0.001	<0.001	<0.001	0.237	0.006	<0.001
-0.3	0.002	0.032	0.05	<0.001	<0.001	<0.001	<0.001	<0.001
-0.2	<0.001	0.415	0.301	<0.001	0.007	<0.001	0.036	<0.001
-0.1	<0.001	0.091	0.003	<0.001	<0.001	<0.001	<0.001	<0.001
0	<0.001	0.254	<0.001	0.367	0.053	<0.001	<0.001	<0.001
0.1	<0.001	<0.001	<0.001	0.42	0.003	<0.001	<0.001	0.046
0.2	0.472	0.097	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
0.3	0.374	0.168	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
0.4	0.013	0.097	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
0.5	<0.001	0.252	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
0.6	<0.001	0.166	<0.001	<0.001	<0.001	<0.001	0.038	<0.001
0.7	0.059	<0.001	<0.001	<0.001	<0.001	<0.001	0.04	<0.001
0.8	0.03	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	<0.001
0.9	0.084	0.022	<0.001	<0.001	<0.001	0.262	0.003	<0.001
1	0.001	<0.001	<0.001	<0.001	<0.001	0.057	<0.001	<0.001
1.1	0.027	<0.001	<0.001	0.003	<0.001	0.137	<0.001	<0.001
1.2	0.011	<0.001	0.035	0.51	<0.001	<0.001	<0.001	<0.001
1.3	0.31	0.006	0.203	<0.001	<0.001	<0.001	<0.001	<0.001
1.4	0.268	0.056	0.092	<0.001	0.001	0.002	<0.001	<0.001
1.5	<0.001	0.098	0.044	<0.001	<0.001	0.001	<0.001	<0.001
1.6	<0.001	0.08	0.161	0.002	0.003	<0.001	0.026	<0.001
1.7	<0.001	0.254	0.001	<0.001	<0.001	<0.001	<0.001	<0.001
1.8	0.003	0.315	<0.001	<0.001	<0.001	<0.001	0.002	<0.001
1.9	0.23	0.114	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2	<0.001	0.466	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2.1	0.006	0.078	0.03	<0.001	0.008	<0.001	<0.001	<0.001
2.2	0.08	<0.001	<0.001	<0.001	<0.001	0.324	<0.001	<0.001
2.3	0.026	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
2.4	0.076	<0.001	<0.001	<0.001	<0.001	0.039	<0.001	<0.001
2.5	0.143	0.147	<0.001	0.026	0.002	<0.001	<0.001	<0.001
2.6	0.001	0.021	0.018	<0.001	0.173	<0.001	<0.001	<0.001
2.7	<0.001	0.39	0.001	<0.001	<0.001	<0.001	<0.001	<0.001

2.8	<0.001	0.151	0.005	<0.001	0.085	<0.001	<0.001	0.365
2.9	<0.001	0.049	<0.001	0.006	<0.001	<0.001	0.001	0.034
3	0.039	0.068	<0.001	0.001	0.5	0.294	0.051	0.457

table S1. Non-parametric permutation test p values for each time point comparing the discriminability of large and small reward volumes between a sample-number matched untagged OFC population and OFC→CPi neurons and OFC→CPr neurons.

Time	p value
-1	0.313463065
-0.8	0.232225592
-0.6	0.910817249
-0.4	0.550291942
-0.2	0.736875371
n0	0.39053287
0.2	0.000389774
0.4	0.000162856
0.6	8.84492E-05
0.8	8.85746E-05
1	8.85746E-05
1.2	8.85746E-05
1.4	8.85746E-05
1.6	8.85746E-05
1.8	8.84492E-05
2	0.000103346
2.2	0.881292709
2.4	0.537866329

table S2. Wilcoxon signed-rank test p values for each time point comparing decoding accuracy of the most likely block from OFC→CPi neurons to the sample-number matched OFC population.

Time	p value
-1	0.073138069
-0.8	0.851924595
-0.6	0.708905309
-0.4	0.808250596
-0.2	0.076617923
0	0.003185143
0.2	0.627446033
0.4	0.001507287
0.6	0.000103346
0.8	8.85746E-05
1	8.85746E-05
1.2	8.85746E-05
1.4	0.000390231
1.6	0.005111049
1.8	0.000120422
2	0.000103346
2.2	0.000120422
2.4	0.015240061

table S3. Wilcoxon signed-rank test p values for each time point comparing decoding accuracy of the second most likely block from OFC→CPi neurons to the sample-number matched OFC population.

Time point	p-value	Time point	p-value
-1	0.378	1.05	0.289
-0.95	0.323	1.1	0.294
-0.9	0.292	1.15	0.232
-0.85	0.386	1.2	0.11
-0.8	0.315	1.25	0.097
-0.75	0.388	1.3	0.143
-0.7	0.243	1.35	0.159
-0.65	0.233	1.4	0.245
-0.6	0.344	1.45	0.218
-0.55	0.506	1.5	0.466
-0.5	0.263	1.55	0.406
-0.45	0.151	1.6	0.255
-0.4	0.145	1.65	0.187
-0.35	0.108	1.7	0.267
-0.3	0.066	1.75	0.088
-0.25	0.039	1.8	0.108
-0.2	0.023	1.85	0.334
-0.15	0.022	1.9	0.388
-0.1	0.024	1.95	0.306
-0.05	0.099	2	0.252
0	0.158	2.05	0.464
0.05	0.266	2.1	0.264
0.1	0.184	2.15	0.205
0.15	0.241	2.2	0.236
0.2	0.228	2.25	0.402
0.25	0.249	2.3	0.236
0.3	0.235	2.35	0.292
0.35	0.33	2.4	0.381
0.4	0.388	2.45	0.41
0.45	0.464	2.5	0.456
0.5	0.41	2.55	0.401
0.55	0.5	2.6	0.244
0.6	0.411	2.65	0.301
0.65	0.311	2.7	0.234
0.7	0.124	2.75	0.327
0.75	0.208	2.8	0.316
0.8	0.127	2.85	0.451
0.85	0.123	2.9	0.392
0.9	0.174	2.95	0.426

0.95	0.33	3	0.454
1	0.256		

table S4. Non-parametric permutation test p values for each time point comparing encoding of the second most likely block on 20 μL trials when the probability of the trial being a mix block is high.

Time	p value	Time	p value
-1	0.186	1.05	<0.001
-0.95	0.084	1.1	<0.001
-0.9	0.044	1.15	<0.001
-0.85	0.022	1.2	0.001
-0.8	0.004	1.25	0.001
-0.75	0.005	1.3	0.004
-0.7	0.004	1.35	<0.001
-0.65	0.001	1.4	0.004
-0.6	0.012	1.45	0.005
-0.55	0.017	1.5	0.009
-0.5	0.015	1.55	0.021
-0.45	0.022	1.6	0.023
-0.4	0.008	1.65	0.007
-0.35	0.003	1.7	0.012
-0.3	0.002	1.75	0.009
-0.25	0.018	1.8	0.003
-0.2	0.014	1.85	0.01
-0.15	0.065	1.9	0.005
-0.1	0.065	1.95	0.002
-0.05	0.092	2	0.002
0	0.029	2.05	0.018
0.05	0.013	2.1	0.029
0.1	0.009	2.15	0.026
0.15	0.002	2.2	0.027
0.2	<0.001	2.25	0.018
0.25	<0.001	2.3	0.016
0.3	0.003	2.35	0.01
0.35	0.002	2.4	0.004
0.4	0.003	2.45	0.002
0.45	<0.001	2.5	0.002
0.5	0.001	2.55	0.001
0.55	0.003	2.6	0.001
0.6	0.015	2.65	<0.001
0.65	0.004	2.7	<0.001
0.7	0.006	2.75	0.001
0.75	0.008	2.8	0.002
0.8	0.004	2.85	0.003
0.85	0.004	2.9	<0.001
0.9	0.004	2.95	0.001
0.95	<0.001	3	0.001

1	<0.001		

845 **table S5.** Non-parametric permutation test p values for each time point comparing block
846 encoding of high and mix blocks with and without optogenetic stimulation.

Time	p value	Time	p value
-1	0.39	1.05	0.463
-0.95	0.486	1.1	0.466
-0.9	0.491	1.15	0.477
-0.85	0.219	1.2	0.496
-0.8	0.139	1.25	0.49
-0.75	0.105	1.3	0.501
-0.7	0.046	1.35	0.46
-0.65	0.024	1.4	0.486
-0.6	0.108	1.45	0.463
-0.55	0.141	1.5	0.482
-0.5	0.098	1.55	0.461
-0.45	0.072	1.6	0.425
-0.4	0.089	1.65	0.372
-0.35	0.087	1.7	0.261
-0.3	0.204	1.75	0.204
-0.25	0.312	1.8	0.284
-0.2	0.321	1.85	0.326
-0.15	0.361	1.9	0.447
-0.1	0.401	1.95	0.351
-0.05	0.238	2	0.303
0	0.321	2.05	0.305
0.05	0.22	2.1	0.31
0.1	0.191	2.15	0.322
0.15	0.206	2.2	0.359
0.2	0.095	2.25	0.385
0.25	0.089	2.3	0.44
0.3	0.124	2.35	0.444
0.35	0.225	2.4	0.456
0.4	0.206	2.45	0.396
0.45	0.249	2.5	0.39
0.5	0.262	2.55	0.39
0.55	0.365	2.6	0.368
0.6	0.396	2.65	0.332
0.65	0.428	2.7	0.303
0.7	0.441	2.75	0.331
0.75	0.484	2.8	0.31
0.8	0.411	2.85	0.387
0.85	0.403	2.9	0.375
0.9	0.39	2.95	0.3
0.95	0.451	3	0.218

1	0.414		

847. **table S6.** Non-parametric permutation test p values for each time point comparing re-
848 ward encoding between 20 and 80 μL with and without optogenetic stimulation.

Time	p value	Time	p value
-1	0.344	1.05	0.039
-0.95	0.377	1.1	0.052
-0.9	0.327	1.15	0.015
-0.85	0.314	1.2	0.016
-0.8	0.319	1.25	0.006
-0.75	0.358	1.3	<0.001
-0.7	0.223	1.35	<0.001
-0.65	0.201	1.4	0.003
-0.6	0.222	1.45	0.003
-0.55	0.166	1.5	0.002
-0.5	0.102	1.55	0.005
-0.45	0.123	1.6	0.006
-0.4	0.192	1.65	0.004
-0.35	0.224	1.7	0.001
-0.3	0.202	1.75	0.002
-0.25	0.333	1.8	<0.001
-0.2	0.299	1.85	<0.001
-0.15	0.227	1.9	<0.001
-0.1	0.108	1.95	<0.001
-0.05	0.049	2	0.001
0	0.031	2.05	<0.001
0.05	0.014	2.1	<0.001
0.1	0.003	2.15	<0.001
0.15	0.002	2.2	<0.001
0.2	0.003	2.25	<0.001
0.25	0.007	2.3	<0.001
0.3	0.002	2.35	<0.001
0.35	0.002	2.4	<0.001
0.4	0.001	2.45	<0.001
0.45	0.004	2.5	<0.001
0.5	0.005	2.55	<0.001
0.55	0.01	2.6	<0.001
0.6	0.018	2.65	<0.001
0.65	0.03	2.7	<0.001
0.7	0.031	2.75	<0.001
0.75	0.021	2.8	<0.001
0.8	0.021	2.85	<0.001
0.85	0.023	2.9	<0.001
0.9	0.033	2.95	0.004
0.95	0.036	3	<0.001

1	0.095		

table S7. Non-parametric permutation test p values for each time point comparing block encoding of low and mix blocks with and without optogenetic stimulation.

Time	p value	Time	p value
-1	0.446	1.05	0.15
-0.95	0.118	1.1	0.167
-0.9	0.055	1.15	0.096
-0.85	0.034	1.2	0.077
-0.8	0.087	1.25	0.12
-0.75	0.107	1.3	0.14
-0.7	0.094	1.35	0.172
-0.65	0.203	1.4	0.139
-0.6	0.231	1.45	0.112
-0.55	0.063	1.5	0.112
-0.5	0.144	1.55	0.095
-0.45	0.047	1.6	0.072
-0.4	0.065	1.65	0.07
-0.35	0.345	1.7	0.057
-0.3	0.457	1.75	0.033
-0.25	0.45	1.8	0.048
-0.2	0.312	1.85	0.046
-0.15	0.298	1.9	0.076
-0.1	0.401	1.95	0.061
-0.05	0.442	2	0.077
0	0.436	2.05	0.072
0.05	0.184	2.1	0.047
0.1	0.063	2.15	0.03
0.15	0.018	2.2	0.022
0.2	0.026	2.25	0.027
0.25	0.037	2.3	0.035
0.3	0.058	2.35	0.04
0.35	0.103	2.4	0.036
0.4	0.28	2.45	0.027
0.45	0.26	2.5	0.029
0.5	0.203	2.55	0.016
0.55	0.225	2.6	0.007
0.6	0.235	2.65	0.009
0.65	0.194	2.7	0.016
0.7	0.248	2.75	0.025
0.75	0.329	2.8	0.009
0.8	0.364	2.85	0.017
0.85	0.308	2.9	0.045
0.9	0.241	2.95	0.055
0.95	0.284	3	0.053

1	0.224		

table S8. Non-parametric permutation test p values for each time point comparing reward encoding between 5 and 20 µL with and without optogenetic stimulation.

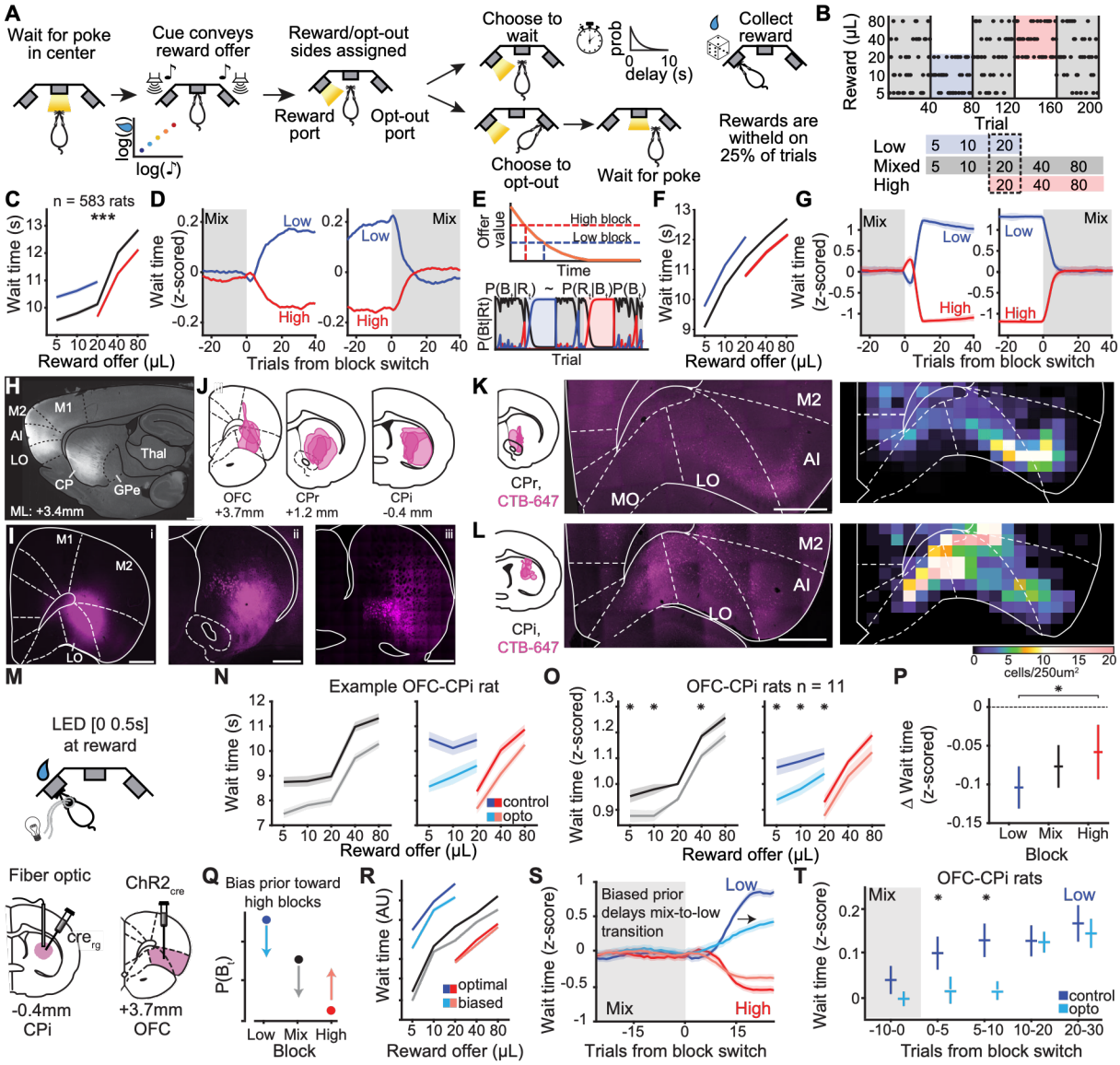

Figure 1: **OFC→CPi neurons bias beliefs about hidden states**. **A.** Schematic of behavioral paradigm. **B.** Block structure of the task **C.** Mean (+/- sem) wait time on catch trials by reward in each block across all rats. $p << 0.001$, Wilcoxon signed-rank test. **D.** Mean wait times at block transitions away from (left) or into (right) mixed blocks. **E.** Schematic of inferential behavioral model that compares a time-varying value function to a block-specific expectation of average reward to determine how long to wait. **F.** Mean wait times from the behavioral model. **G.** Behavioral model wait times around block transitions. **H.** Sagittal section of a rat brain showing anterograde fluorescent labeling of OFC projections throughout the striatum. Abbrev. M1 - primary motor cortex, M2 - secondary motor cortex, AI - agranular insula, LO - lateral orbitofrontal cortex, CP - caudate putamen, GPe - globus palladas externus, Thal - thalamus. **I.** Coronal section of an example brain at the **i.** OFC injection site Bregma +3.7 mm, **ii.** Bregma 1.2 mm consistent

with CPr, and **iii.** Bregma -0.2 mm consistent with CPi. Scale bar = 1 mm. **J.** Overlay of fluorescently labeled regions in N=3 rats. **K.** Overlay of injection sites for retrograde fluorescent labeling of OFC→CPr neurons (left), an example coronal section of the frontal cortex showing labeled somata (middle), and quantification of mean cell density across all animals N = 4 (right). **L.** Retrograde fluorescent labeling of OFC→CPi neurons as in K. **M.** Schematic of viral and implant sites to target OFC→CPi neurons for optogenetic stimulation at reward delivery. **N.** Mean (+/- sem) wait times of an example rat on sessions in which no LED stimulation was applied ("control", darker colors) and session in which LED stimulation was applied ("opto", lighter colors). Data is separated into mixed blocks (left), and low and high blocks (right). **O.** Mean z-scored wait time across all rats. Wilcoxon signed rank test, mixed block: 5 μL $p = 0.0244$, 10 μL $p = 0.0098$, 20 μL $p = 0.0830$, 40 μL $p = 0.0322$, 80 μL $p = 0.0830$, low block: 5 μL $p = 0.0049$, 10 μL $p = 0.0098$, 20 μL $p = 0.0420$, high block: 20 μL $p = 0.3652$, 40 μL $p = 0.2783$, 80 μL $p = 0.1016$. **P.** Mean difference between opto and control session wait times averaged across all volumes for each block, Friedman test indicating within subject differences $p = 0.0021$, Wilcoxon signed-rank post hoc comparison between low and high block optogenetic modulation, low vs. mix $p = 0.1607$, low vs. high $p = 0.0084$, high vs. mix $p = 0.1266$ **Q.** Schematic of the biased prior applied to a single trial where the model's belief favors a low block; the prior biases the model towards high block beliefs and disfavors low and mixed block beliefs. **R.** Comparison of wait times for behavioral model with an optimal or biased prior. **S.** Model predictions of wait times around block transitions. Biased prior delays wait time changes at mixed→low block transitions. **T.** Mean wait time across all animals in control and opto sessions around mixed→low block transitions grouped into 5 trial bins. Friedman test, $p < 0.0247$, Wilcoxon signed-rank post hoc test, trials [-10 0] $p = 0.1602$, trials [0 5] $p = 0.0337$, trials [5 10] $p = 0.0161$, trials [10 20] $p = 0.4492$, trials [20 30] $p = 0.4155$.

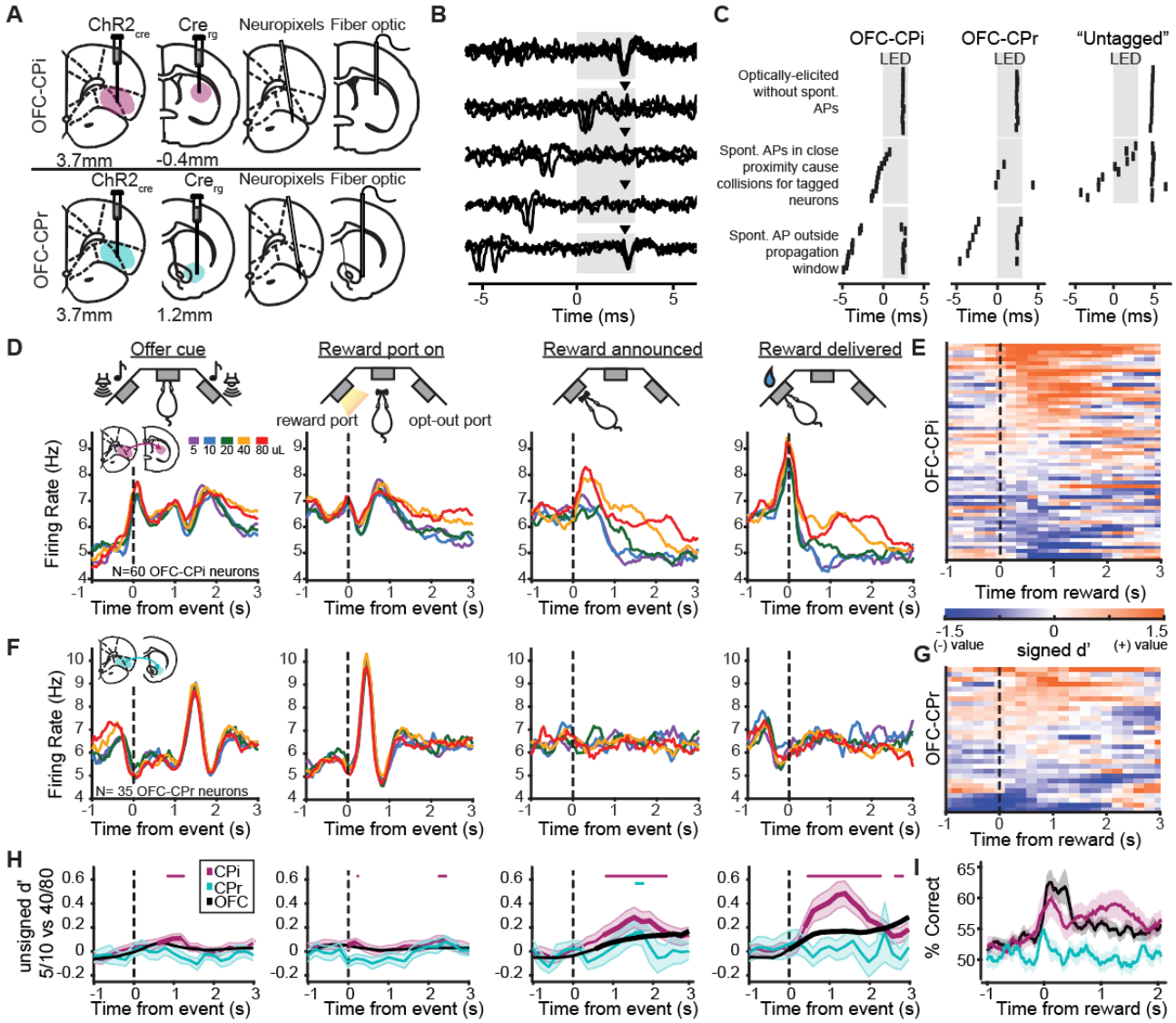

Figure 2: **Task-related dynamics of projection neurons A.** Schematic of viral and implant strategies to target projection-specific neurons. **B.** Waveform overlays of spikes from an example OFC→CPi neuron aligned to LED onset and sorted by the timing of the last action potential prior to the LED onset. Arrows indicate the expected timing of LED elicited spikes. **C.** Raster plots of spikes from three example cells recorded in the OFC which either pass collision tests and are therefore projection neurons to the CPi and CPr (left and middle respectively) or do not pass the collision test are therefore labeled "untagged." **D.** Mean PSTH of OFC→CPi neurons aligned to each task event, for each reward offer in mixed blocks. **E.** Heatmap of signed discriminability index (d') for trials offering 40/80 versus 5/10 µL in mixed blocks for each OFC→CPi neuron aligned to reward delivery. **F-G.** Same analyses as in D,E for OFC→CPr neurons. **H.** Mean (+/- sem) unsigned discriminability index (d') for trials offering 40/80 versus 5/10 µL in mixed

blocks for OFC→CPi (pink), OFC→CPr (teal), and untagged OFC neurons (black). Bolded traces indicate time points where the unsigned d' is larger than the 95th percentile of the label-shuffled data for that neuronal population. Time points with significant differences ($p < 0.05$, non-parametric permutation test) between OFC→CPi or OFC→CPr and untagged OFC neurons is depicted in matching colors above each PSTH and actual p values for both tests are listed in table S1. **i.** Logistic decoding of reward volume (40/80 μL or 5/10 μL) in mixed blocks.

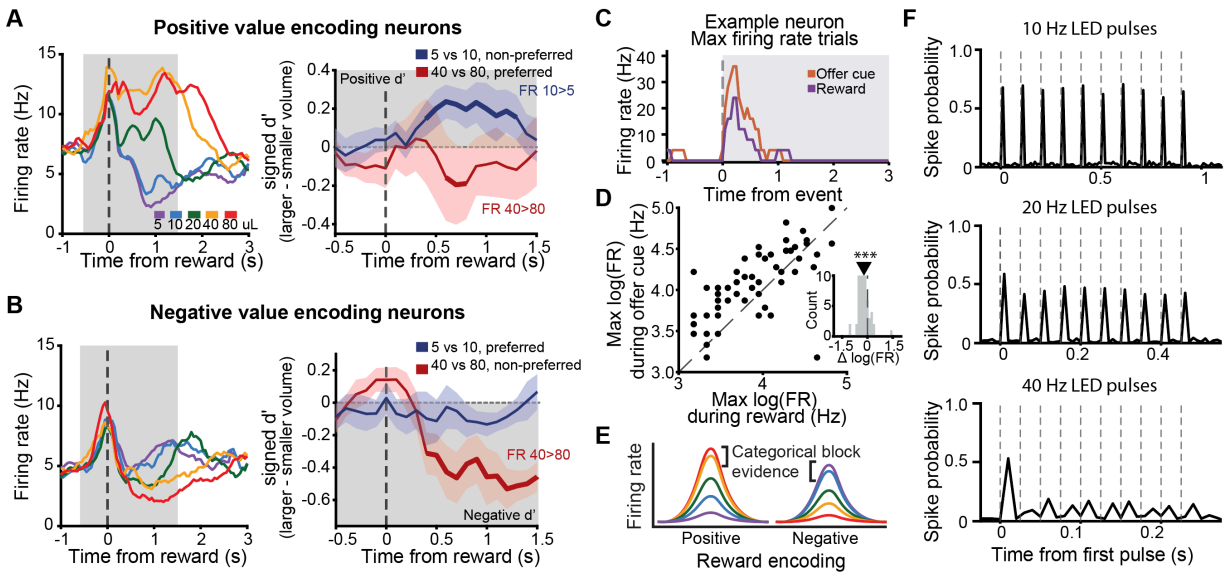

Figure 3: **Saturating nonlinearity supports categorical encoding of block evidence. A.** PSTH of positive value neurons sorted by reward volume and aligned to reward delivery (left). Mean (+/- sem) d' between the neurons' two preferred volumes (reward volumes with the highest firing rate responses; 40 vs. 80 µL) and non-preferred volumes (reward volumes with the lowest firing rate responses; 5 vs. 10 µL). Bolded traces indicate time points where discriminability is greater than the 95th percentile of the label-shuffled data. Grey box indicates d' values consistent with signed d' for large vs. small rewards (i.e., positive value coding). **B.** Same as panel A but for negative value neurons. Grey box indicates d' values consistent with signed d' for large vs. small rewards (i.e., negative value coding). **C.** PSTH of an example OFC→CPi neuron on the single trials with the highest firing rate during offer cue (orange) and reward delivery (purple). **D.** Maximum firing rate at offer cue and reward delivery. Inset shows a histogram of the difference in maximum firing rate at the two events (firing rate at reward - firing rate at offer cue). Mean difference denoted by triangle above the histogram. Wilcoxon signed-rank test, $p = 3.86 \times 10^{-5}$. **E.** Schematic of saturating non-linear reward encoding that provides categorical block evidence. **F.** Probability of a spike over time aligned to the first LED pulse in a 10 pulse train at 10 Hz (top), 20 Hz (middle), 40 Hz (bottom) for an example neuron. Dashed lines indicate the onset of each LED pulse in the train.

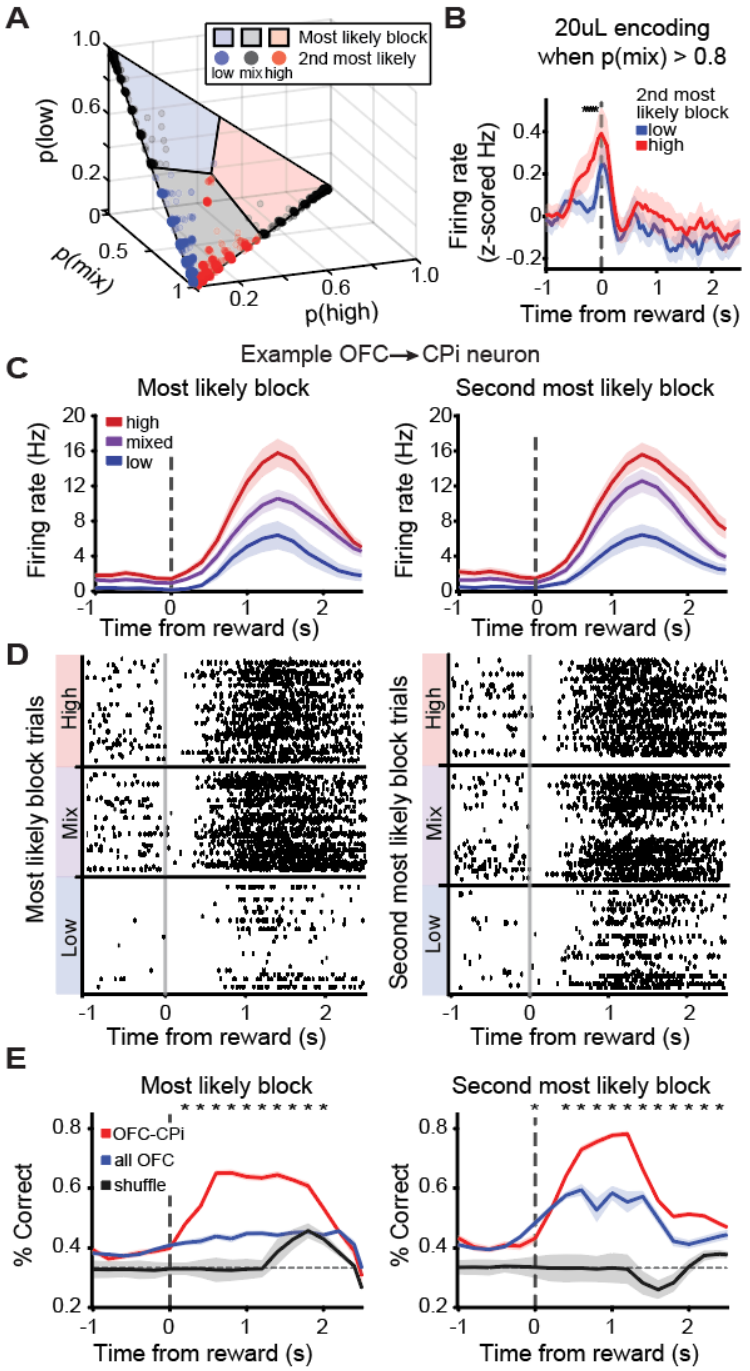

Figure 4: **OFC→CPi neurons convey the full belief distribution over blocks**

A. Distribution of most likely and second most likely block from OFC→CPi recording sessions. Block probabilities for each rewarded trial were estimated from a Bayesian belief model of the task (*8*), and are plotted in 3D probability space. The colored regions of the simplex correspond to the most likely block for that trial, and dot colors convey second most likely block. Red = high block, black = mixed block, blue = low block. **B.** Mean (+/- sem) PSTH of OFC→CPi activity conditioned on 20 μL trials in a mixed block in which the behavioral model indicated that the probability of a mixed block was high (P(mix) > 0.8) and the second most likely block was high (P(high) < 0.05, red) or low (P(low) < 0.05, blue). Asterisks indicate significant differences depending on the second most likely block (non-parametric permutation test, $p < 0.05$, actual p values listed in table S4). **C.** Conditional PSTHs aligned to reward delivery for an example projection neuron that encodes the most likely block (left) and the second most likely block (right). **D.** Trial-by-trial activity of the example neuron in C aligned to reward delivery and sorted by the most likely block (left) or second most likely block identity (right). **E.** Decoding accuracy of OFC→CPi neurons to classify (left) most likely block and (right) second most likely block. Colored lines represent different populations (red = OFC→CPi, blue = all OFC, black = shuffle distribution built by shuffling trials labels of OFC→CPi data. Error bars for red and blue curves are sem, and error bars for the shuffle distribution are 1.96 standard deviations calculated over 20 shuffles. Significant differences in decoding accuracy between OFC→CPi neurons and a sample-number matched untagged OFC population via Wilcoxon signed-rank test ($p < 0.05$) is denoted above each timepoint (actual p values listed in tables S2 and S3).

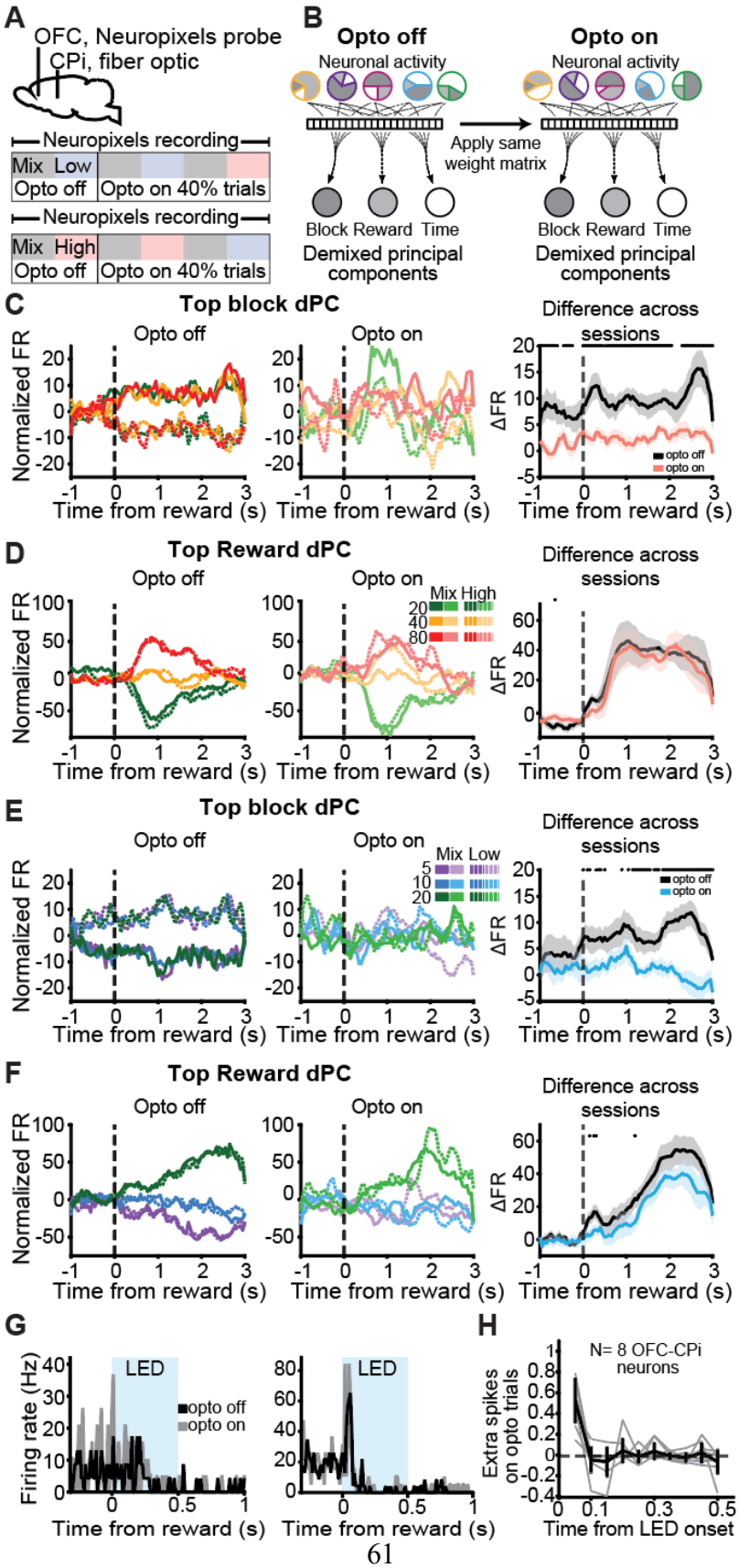
A
OFC, Neuropixels probe
CPi, fiber optic
Neuropixels recording
Mix Low
Opto off
Opto on 40% trials
Neuropixels recording
Mix High
Opto off
Opto on 40% trials
B
Opto off
Neuronal activity
Opto on
Neuronal activity
Apply same weight matrix
Block Reward Time
Demixed principal components
Block Reward Time
Demixed principal components
C
Top block dPC
Opto off
Opto on
Difference across sessions
Normalized FR
ΔFR
opto off
opto on
Time from reward (s)
D
Top Reward dPC
Opto off
Opto on
Mix High
20
40
80
Difference across sessions
Normalized FR
ΔFR
Time from reward (s)
E
Top block dPC
Opto off
Opto on
Mix Low
5
10
20
Difference across sessions
opto off
opto on
Normalized FR
ΔFR
Time from reward (s)
F
Top Reward dPC
Opto off
Opto on
Difference across sessions
Normalized FR
ΔFR
Time from reward (s)
G
LED
opto off
opto on
Firing rate (Hz)
Time from reward (s)
H
N= 8 OFC-CPi neurons
Extra spikes on opto trials
Time from LED onset

Figure 5: **OFC→CPi perturbations impact cortical representations of inferred states**
A. Schematic of implantation sites (top) and task structure (bottom). To ensure we could record the activity of identical neurons with and without stimulation, LED stimulation began after the second complete block. On a given day the rat experienced a mixed block and either a high or low block with and without optogenetic stimulation. After the first four blocks, the block identity alternated as in the standard task. **B.** Schematic of demixed PCA strategy. A weight matrix was learned based on neuronal activity during the first two control blocks,"Opto off," to produce demixed principal components which explain task variables of interest: block, reward, and time. The weight matrix was applied to the same identified neurons whose activity was recorded while the rat received optogenetic stimulation at OFC→CPi axons during reward delivery. **C.** PSTH of neuronal activity projected into the top block demixed principal component without (left) and with (middle) optogenetic stimulation from an example session in which the rat experienced mix and high blocks. Mean (+/- sem) change in firing rate between mix and high blocks during blocks without stimulation (black) and with stimulation (red) across all animals and sessions (N= 3 animals, n = 15 sessions). Asterisks denote a significant difference in encoding with and without stimulation ($p < 0.05$, non-parametric permutation test). **D-F.** Analyses as in C for the top reward dPC for high/mixed blocks (D), top block dPC for low/mixed blocks (E), or top reward dPC for low/mixed blocks (F). Actual p values for each plot are listed in tables S5-8. **G.** PSTH of OFC→CPi neurons at reward on trials with and without LED stimulation. **H.** The number of extra spikes recorded from OFC→CPi neurons during reward delivery on trials with LED stimulation, compared to control trials.