Neural activity ramps in frontal cortex signal extended motivation during learning

Josue M. Regalado[1], Ariadna Corredera Asensio[1], Theresa Haunold[1], Andrew C. Toader[1], Yan Ran Li[1], Lauren A. Neal[1], and Priya Rajasethupathy[1,2*]

[1]Laboratory of Neural Dynamics & Cognition, The Rockefeller University, New York, NY 10065 USA
[2]Lead contact

* Correspondence: priya@rockefeller.edu

SUMMARY

Learning requires the ability to link actions to outcomes. How motivation facilitates learning is not well understood. We designed a behavioral task in which mice self-initiate trials to learn cue-reward contingencies and found that the anterior cingulate region of the prefrontal cortex (ACC) contains motivation-related signals to maximize rewards. In particular, we found that ACC neural activity was consistently tied to trial initiations where mice seek to leave unrewarded cues to reach reward-associated cues. Notably, this neural signal persisted over consecutive unrewarded cues until reward associated cues were reached, and was required for learning. To determine how ACC inherits this motivational signal we performed projection specific photometry recordings from several inputs to ACC during learning. In doing so, we identified a ramp in bulk neural activity in orbitofrontal cortex (OFC) -to-ACC projections as mice received unrewarded cues, which continued ramping across consecutive unrewarded cues, and finally peaked upon reaching a reward associated cue, thus maintaining an extended motivational state. Cellular resolution imaging of OFC confirmed these neural correlates of motivation, and further delineated separate ensembles of neurons that sequentially tiled the ramp. Together, these results identify a mechanism by which OFC maps out task structure to convey an extended motivational state to ACC to facilitate goal-directed learning.

INTRODUCTION

Animals must sustain an extended motivational state to achieve goal-directed learning. Imagine being hungry in the middle of a busy metropolis with no cellphone battery and no way of searching for the nearest restaurant. The feeling of hunger provides motivation to search for restaurant signs, scan menus, and contemplate what type of food to eat. If it is dinnertime and many restaurants are full, this motivational state (hunger) may persist for hours until a restaurant is selected. Thus, an animal's ability to carry out novel actions based on its desired goals is commonly referred to as goal-directed learning. This learning is of a more deliberate, informed nature than habitual learning, as they are sensitive to the current value of outcomes and can lead to a novel sequence of actions for a desired outcome[1–3].

Goal-directed learning often requires the ability to maintain an extended motivational state even in the midst of distracting and competing external variables[4,5]. This function has been long proposed to be carried out by the prefrontal cortex (PFC), as patients with PFC lesions struggle to perform tasks that require maintaining a motivational and goal-directed state, in the midst of competing sensory information, such as the Stroop task or the Wisconsin Card Sorting Task [6–11]. In particular, the anterior cingulate cortex (ACC) has been implicated in action selection over long timescales that are influenced by a variety of motivational factors, such as the value and effort required for each outcome[5,12–16]. For instance, when animals are given two choice options: one in which high-effort leads to high-rewards, and one in which low-effort leads to low-rewards, animals learn to exploit the high-effort, high-reward option[17,18]. Impairments to the ACC results in animals failing to accurately allocate motivation towards strategies that maximize reward[19,20]. Single-unit recordings from ACC have shown that neurons encode for choices that require effort with a higher-payoff, giving support for the hypothesis that this region is important for action-outcome associations and allocating resources for learning and for the maximization of reward over long timescales[13,21–23]. While the precise functions of ACC are still debated, its role in goal-directed learning is widely accepted[5,22,24–26].

To provide deeper mechanistic insight into how ACC encodes an extended motivational state to facilitate goal-directed learning, we sought to track how animals learn to adjust their behavior over days-long timescales to maximize reward when cue-reward contingencies change. We designed a task in which mice self-initiate trials and learn to associate cues with reward. Through neural activity recordings during behavior, we found that ACC neural activity was consistently tied to trial initiations where mice seek to leave unrewarded cues to reach a rewarded cue. Subsequently, by recording neural activity from inputs to ACC we identified a ramp in bulk activity in orbitofrontal cortex (OFC)-to-ACC projections as mice continuously existed unrewarded cues, peaking when they finally reached a rewarded cue, thus tracking an extended motivational state. Finally, cellular resolution imaging of OFC-to-ACC neurons identified populations of neurons that sequentially tile the observed bulk neural activity ramp across un-rewarded cue presentations. In particular, neurons that preferentially encoded reward cues, before learning, began to code for un-rewarded, cues after learning, including the motivation to exit these rooms to reach more reward-associated cues. Taken together, we identified a mechanism by which OFC neural activity ramps map out task structure and conveys an extended motivational state to ACC to enable goal directed learning.

RESULTS

ACC contains neural correlates of motivation during learning

We began by designing a learning task in which mice self-initiate trials and, upon brief cue exposure (an olfactory and auditory cue), learn to stop to collect a water reward (**Figure 1A,B; S1A,** Methods). We implemented this task in a head-fixed setting to enable hundreds of trials per session, and millisecond precision in tracking stimulus delivery and behavioral responses (**Figure 1A**). We used "time to initiate trials" as the primary measure of motivation, and "total reward obtained" as the primary measure of learning. Due to the self-paced nature of the task (**Figure 1B,C**), we found variation between our mice in how quickly they initiated trials and how many rewards they received per minute (**Figure 1C**). As expected, the faster mice initiated trials, the more rewards they obtained per minute, providing a strong correlation between motivation and learning (**Figure 1D**).

The anterior cingulate cortex (ACC) has been classically implicated in motivation and maximizing rewards, so we posited that ACC would contain motivation-related neural activity patterns in our task[5,27,28]. To test this hypothesis, we injected AAV1-CaMKII-GCaMP6f into the ACC and implanted fiber-optic cannulas to record bulk neural activity in ACC during behavior (**Figure 1E**). We observed strong neural responses in ACC that were tuned to reward delivery and trial initiations (**Figure 1F-G**). Notably, the ACC neural signal precedes speed onset in both cases, suggesting that ACC is not tracking speed but rather the motivation to initiate trials (**Figure 1F,G; Figure S1B**). To determine whether these motivation-related signals in ACC were required for learning, we performed hM4D(Gi)-based chemogenetic inhibition of ACC during a session of Clozapine N-oxide (CNO) injection (ACC inhibition session) versus a session of saline injection (control session) (**Figure 1H,I**). We found that ACC inhibition caused mice to have a significant increase in time to initiate trials (**Figure 1I**), which also resulted in a decreased number of rewards received per minute (**Figure 1J**). Furthermore, we found a small, but significant, decrease in speed during trial initiation (but not overall trial speed), suggesting that ACC inhibition might also impair vigor of movements during trial initiations (**Figure S1C**). Thus, we developed a self-paced behavioral task where mice learned cue-reward contingencies, and identified motivation-related signals in ACC that were required for learning and reward maximization.

ACC contains neural correlates of extended motivation during learning

We next sought to increase the motivational demand during learning. We thus extended our task by training mice to learn two sets of cue-outcome relationships, where one cue-set (olfactory + auditory) is associated with a sucrose water reward (hereafter referred to as "R" cues), whereas the other cue-set is associated with no reward ("N" cues). Since mice have been shaped to stop during cue presentations (**Figure 1**), it is now effortful for them to learn to continue running through the N cues so that they can reach more R cues, and thus maximize their total rewards in a session. Thus, motivation is assessed not only by "time to initiate trials after R cues", as before, but now also the more effortful measure of "time to initiate trials after N cues" (**Figure 2A**; see Methods). We measured overall learning through differences in their lick rates and speed during presentations, expecting progressive suppression of licking and increases in speed in the N cues compared to the R cues across days. Interestingly, we found that mice learned to suppress licking

in the N cues (**Figure 2A**; red arrows on day 2) much earlier than learning to increases speed in N cues (**Figure 2A**; red arrows on day 4; **Supplementary Video 1**). Across the cohort, on average, this increase in speed during N cues began as early as day 3, after they had learned to suppress their licking (day 2), as determined by speed and stop discrimination index (stop DI: % of stops in N – R / all trials), (**Figure 2B,C, S2A;** see Methods). Finally, there was also a significant correlation between stop DI and rewards obtained per minute, confirming that the development of this behavioral strategy is tied to reward maximization within a given training session (**Figure 2D**).

We next searched for neural correlates of motivation by recording bulk neural activity in ACC as mice performed this task, and aligning neural responses to behavioral frames, focusing on periods when mice learn to run during N cue presentations. As before (as in **Figure 1**), in this two cue-outcome relationship task, we again found that ACC continued to be active during reward delivery and during trial initiations (**Figure S2B,C**). Additionally, however, in this task we found that ACC began to significantly increase its activity, specifically during N cues, as early as T3, as mice exhibited a learned motivation to leave N cues to reach more R cues (**Figure 2B-D).** As a further confirmation of this result, we investigated ACC's activity during extended motivation across two consecutive N cues and found that ACC activity continued to remain high from the initial N cue presentation until an R cue was reached (**Figure 2E, S2D**). These neural responses, and in particular the dF/F difference in N vs R cues, positively correlated with the amount of reward obtained per minute, linking motivation related ACC activity to overall learning (**Figure 2D**). Importantly, in all cases, on a trial by trial basis, the neural signal preceded the behavioral ramp in speed (**Figure 2E**), and was present even if we restricted our analyses to cue presentations in which mice stopped (**Figure S2E**), suggesting a motivational rather than motor response. To further confirm this dissociation, we passively presented both sets of cues to the mice at the end of each training session. As expected, mice did not develop the motivation to run out of N cues (**Figure S2F**), and accordingly, the ACC neural activity was no longer different between N and R cues. These results together suggest that ACC encodes a motivation signal to initiate trials, and in particular corresponds to the behavioral measure of running during N cues to reach more R cues, thus facilitating goal-directed learning.

To test whether these motivation-related signals in ACC are required for learning, we expressed the inhibitory opsin stGtACR2 bilaterally in ACC and delivered light selectively when the mouse received R or N cues, for the first 6 days ("early") or last 4 days ("late") of training (**Figure 2F-G, S2H**). We found that early ACC inhibition prevented mice from learning to run out of N cues, even though they still learned to suppress their lick rates (**Figure 2G, S2G**). Late ACC inhibition had no effect on speed or lick rate behavior, as mice continued to run out during N cues while inhibition occurred, suggesting ACC activity does not broadly suppress speed (**Figure S2H**). All together, we identified an extended motivation signal in ACC that is required for learning and reward maximization.

Neural activity in orbitofrontal projections ramps until rewards are reached

The ACC receives projections from disparate regions across the brain that could facilitate the integration of value, internal state, and multisensory information, so we sought to identify how afferent projections may convey motivational signals to ACC during learning[29]. We injected

rgAAV-hSyn-Cre into ACC and injected cre-dependent GCaMP6f in the orbitofrontal cortex (OFC), anteromedial thalamus (AM), basolateral amygdala (BLA), locus coeruleus (LC), and implanted optical fibers above each region to record neural activity during learning in this task[30] (**Figure 3A**). We first characterized whether the previously observed ACC neural responses during reward delivery and trial initiations were present in any of the inputs to ACC (**Figure S3A**). We found that even before learning all projections responded significantly to rewards, and most (OFC_{ACC}, AM_{ACC}, and LC_{ACC}) increased their activity in anticipation of trial initiations (**Figure S3A**). Thus, motivation-related signals were broadly present in various projections to ACC.

We then searched for motivation-related neural responses that were specifically tied to learning. To do so, we aligned neural responses to trial initiations specifically during N cues, as mice learned to leave N cues to reach more R cues. We found that both OFC_{ACC} and AM_{ACC} had higher baseline activity during trial initiations after no-rewards (**Figure S3C-D**). To further understand this higher activity after no-rewards we analyzed sequences of "RNR" trials which contained reward, no-reward, and reward cues (**Figure 3B**). OFC_{ACC} and AM_{ACC} activity began rising during the N cue presentation and continued rising until an R cue was reached (black dotted arrow; **Figure 3B**). We quantified this motivational signal as a difference in pre-cue activity between N and R cues in RNR trial sequences across days and found that they emerged at the time of learning (~T3) and closely tracked performance of the learned behavior (T3-T6) (**Figure 3B**). To further build confidence in these results, we asked whether this continuous rise in OFC_{ACC} activity in RNR sequences would be further extended in RNNR sequences. Indeed, OFC_{ACC} activity continued ramping across two consecutive N trials, exhibiting higher pre-cue activity upon entering an R cue after two versus one N (black dotted line, **Figure 3C**).

To more directly determine whether this motivational ramp signal in OFC_{ACC} is tied to learning, we separated our mice into two groups, one that learned the task ("Learners", stop DI > 0.5 for at least 3 consecutive days) and one that did not learn ("Non-Learners") (**Figure 3D, S3D**). The Learners reached a high DI by T6, which persisted throughout the rest of training, whereas the "Non-Learners" only reached a significantly higher DI by T10 (**Figure 3D**). Both subsets of mice still learned to discriminate with licking at comparable rates (**Figure S3E**). When we compared OFC_{ACC} activity in a RNNR sequence of trials, we found that only Learners exhibited a significant ramp in neural activity from the first N cue to the final R cue presentation, which emerged coincidental with behavioral learning and persisted for the remaining days of training (**Figure 3E**). Together, we identify projection activity in OFC that ramps across N cues until an R cue is reached that is specifically tied to the development of a learned goal-directed behavior.

Orbitofrontal projection neurons tile unrewarded trials until rewards are reached

Given that we identified a ramp in OFC_{ACC} bulk neural activity during NNR sequences (**Figure 3**), we sought to determine whether a single persistently active population or a sequence of tiled neurons underlies this ramp. We thus performed real-time cellular resolution imaging of OFC projections to ACC by injecting rgAAV-hSyn-Cre into ACC and cre-dependent GCaMP6f in OFC (**Figure 4A**). We implanted a GRIN lens above OFC and imaged the region under a 2-photon microscope as mice performed the learning task (**Figure 4A**). We focused our analysis on days where behavioral learning emerged (**Figure 4B**), and on NNR trial sequences. We found individual neurons that were uniquely active across the first N, second N, or R cue, thereby tiling the sequence

of NNR trials (**Figure 4C-D**). We further found that an increasing number of neurons were active along the sequence of NNR trials and most prominently before learning (**Figure 4E-F, S4A-C**). Thus, collectively, as an ensemble, these neurons ramp consecutive N cues and peak upon reaching R cues, providing underlying mechanism for the previously observed photometry results.

To determine how these NNR ensembles facilitate learning we tracked the same population of neurons "before" and "after" learning **(Stop DI > 0.4; Figure 4G, S4D)**. We identified an ensemble of neurons that were uniquely responsive to R cues preceded by 2 N cues, before learning, and characterized their responses after learning. Interestingly, these neurons were no longer responsive to R-cue-onset but rather to pre-R-cue activity, which then became progressively more responsive to the preceding N-cue-onset, aligning with the learned behavioral transition of mice leaving N cues to reach R cues (**Figure 4G, H**). To determine whether OFC_{ACC} activity ramps were required for learning, we optogenetically inhibited these projections bilaterally and delivered light only on R or N cues. We then specifically assessed whether previous trial history affected behavioral responses on the current cue condition (**Figure 4I, J)**. Interestingly, while both mCherry control and OFC_{ACC} inhibition cohorts could increase their speed during N cues following an R-cue, OFC_{ACC} mice were impaired in doing so if the N cue was followed by an N-cue (**Figure 4I, J, S4E)**. Taken together, these data demonstrate that ensembles of neurons progressively tile the OFC motivational ramp, and that the initial reward responsive neurons become progressively linked to trial initiations during unrewarded cues thus effectively linking actions to outcomes to maximize rewards (**Figure 4K**).

DISCUSSION

In this study, we developed a self-paced cue-outcome learning task to determine how mice extend their motivational state to maximize reward over long timescales. We identify the ACC as broadly critical to maximizing reward in our task, especially as mice learn to run out of unrewarded cues. We found that upstream inputs to ACC from OFC sustain a ramp-like increase in activity through consecutive unrewarded cues until mice reach rewarded cues. Cellular resolution imaging of OFC projection neurons revealed ensembles of neurons that tile the motivational ramp, and a progressive shift in ensemble tuning during learning such that neurons initially encoding for reward become progressively linked to motivated actions, i.e., trial initiations to reach more rewards. We therefore present a model where OFC contains neurons that increasingly link reward to motivated behaviors, conveying a motivational ramp to ACC, to facilitate learning and reward maximization (**Figure 4E, 4K**).

The orbitofrontal cortex has been implicated in guiding adaptive, flexible behavior by signaling information about future outcomes[31–35]. One view sees OFC's function as encoding for the value of the outcomes of events, with various neural correlates having been found for value-guided behavior. Another view sees OFC's function more as building a model of the causal relationships between events, which may or may not entail value-assessments, into a cognitive map [36]. Indeed, OFC neurons have been found to encode sensory-sensory associations even prior to any kind of learning[37]. A way to link both perspectives into a single account has been to view value and a cognitive map as occurring along a spectrum, where inferring value onto outcomes hinges upon a map that is created. We have found that mice learn to run out during N cues to more quickly reach R cues, thereby acquiring more rewards over a training session. This behavior can be viewed as both value-guided, as the mouse suppresses their lick rate during N cues, and also requiring a mental model of the environment, as running occurs with the expectation of reaching R cues in the future. Indeed, the pseudorandom trial structure ensures that N cues will be presented no more than two times in a row, such that after two N cues an R cue is guaranteed (see Methods). We thus parsimoniously position OFC as functioning in model-based behaviors, and in the accurate planning of actions based on the learned transition structure of a task[38].

A combined map of task space and value-information carried by OFC could be used to inform downstream regions, such as ACC, for adjusting behavior. The ACC has been shown to carry information necessary for switching or staying with current behaviors during decision-making and learning in order to maximize rewards and minimize threats or punishments[5,27,28]. We posit that ACC reads information from OFC about task structure and value to perform computations relevant to allocating behavioral control. We have seen this through our findings that ACC is important for learned behaviors associated with maximizing rewards in our self-paced learning task. We compare the decision to run during N cues to a foraging decision to leave a patch to find alternative options, and ACC's importance in the development of this behavior is reminiscent to signals previously described at the time a foraging decision is reached[39,40].We found inhibition of ACC activity affected the development of running during N cues, effectively diminishing an animals' ability to strategy switch[20,41–43].

Here we focused on learning as a systems-level process guided by top-down signals of task structure, value, and behavioral control. A more synaptic-level approach into how ACC integrates

information from upstream regions during learning could reveal important micro-circuit computations, structural changes associated with learning[44], and potential mechanisms underlying seconds-long behavioral timescale learning rules [45].

MATERIALS AND METHODS

Key resources table

REAGENT or REASOURCE	SOURCE	IDENTIFIER
Virus strains		
AAV1-CaMKIIa-GCaMP6f	Upenn Vector Core	Addgene #100834
AAV1-CAG-FLEX-GcaMP6f	Douglas Kim	Addgene #100835
AAV1-CKIIa-stGtACR2-FusionRed	Ofer Yizhar	Addgene #105669
AAV9-CaMKIIa-hM4D(Gi)-mCherry	Bryan Roth	Addgene #50477
AAV9-CaMKIIa-mCherry	Bryan Roth	Addgene #114469
rgAAV-hSYN-Cre	James Wilson	Addgene #105553
AAV1-hSyn1-SIO-stGtACR2	Ofer Yizhar	Addgene #105677
AAV9-hSyn-DIO-mCherry	Bryan Roth	Addgene #50459

Mice

All procedures were done in accordance with guidelines derived from and approved by the Institutional Animal Care and Use Committees (protocol #22087-H) at The Rockefeller University. Animals used were 8-10 weeks-old naive male C57BL/6J mice (Jackson Laboratory, Strain #000664) at the time of surgery. Mice were group housed (3-5 per cage) with ad libitum food and water, unless mice were water restricted for behavioral assays, in which case they were given 1 mL water per day. Body weight was monitored daily to ensure it was maintained above 80% of the pre-restriction measurement. Surgical procedures and viral injections were carried out in mice under protocols approved by Rockefeller University IACUC and were performed in mice anesthetized with 2% isoflurane using a stereotactic apparatus (Kopf).

Surgical Procedures

Puralube vet ointment was applied to the eyes and 0.2mg/kg meloxicam was administered intraperitoneally using a 1mL syringe. Hair from the scalp was trimmed, and the area was sterilized using povidone-iodine swabs and subsequently ethanol swabs. An incision covering the anteroposterior extent was made to allow access to the skull. Injection sites were accessed using a dental drill which made 0.5mm holes through the skull. All virus was injected using a 35G beveled

needle in a 10ul NanoFil Sub-Microliter Injection syringe (World Precision Instruments) controlled by an injection pump (Harvard Apparatus) at a rate of 100nl/min. After all viral delivery, an additional 5-10 mins delay was applied to avoid backflush before slowly removing the injection needle. Animals that required cannulas or GRIN lenses were implanted immediately following viral injection. Following surgery, mice were allowed to recover in a single housed cage for up to 12 hours, and were given meloxicam tablets. Mice were typically housed for three weeks to allow for adequate expression before behavioral testing or histology.

Viral injections:

- In retrograde tracing experiments, mice were unilaterally injected in ACC (A/P +1.0, M/L, ±0.35, D/V -1.4) with rgAAV-CAG-tdT at a volume of 500 nl (1.0 x 10^{13} vg/mL).
- For fiber photometry experiments, 1ul of AAV1-CaMKIIa-GCaMP6f (UPenn Viral Core, diluted to 5 x 10^{12} vg/mL) and rgAAV-hSYN-Cre (1.20 x 10^{13} vg/mL) was injected into ACC, and AAV1-CAG-FLEX-GcaMP6f (5.0 x 10^{12} vg/mL) was injected into OFC (A/P: 2.5, M/L: ±1.0, D/V: -2.5), AM (A/P: -0.75, M/L: ±0.5, D/V: -3.55), BLA (A/P: -1.23, M/L: 2.75, D/V: -4.7) and LC (A/P: -5.45 M/L: ±0.85, D/V: -3.7). One week after virus injection, mice were unilaterally implanted with 1.25 mm ferrule-coupled optical fibers (0.48 NA, 400 μm diameter, Doric Lenses) cut to the desired length so that the implantation site is ~0.2 mm dorsal to the injection site.
- For cellular imaging, rgAAV-hSYN-Cre (1.20 x 10^{13} vg/mL) was injected into ACC and AAV1-CAG-FLEX-GcaMP6f (5.0 x 10^{12} vg/mL) was injected into OFC.
- For optogenetic inhibition of ACC, AAV1-CaMKIIa-stGtACR2 (1 x 10^{13} vg/mL) was injected into ACC bilaterally. For controls, AAV1-CaMKIIa-mCherry (7 x 10^{12} vg/mL) was injected.
- For chemogenetic inhibition of ACC, AAV9-CaMKIIa-hM4D(Gi) (1 x 10^{13} vg/mL) was injected into ACC bilaterally. For controls, AAV9-CaMKIIa-mCherry (1 x 10^{13} vg/mL) was injected bilaterally in either region.
- For optogenetic inhibition of OFC-ACC projections, rgAAV-hSYN-Cre (1.20 x 10^{13} vg/mL) was injected into ACC bilaterally and either AAV1-hSyn1-SIO-stGtACR2 (1.50 x 10^{13} vg/mL) or AAV9-hSyn-DIO-mCherry (9.0 x 10^{12} vg/mL) for controls was injected bilaterally into OFC.

Cannula implants

One week after viral injections, mice undergoing photometry or optogenetic experiments were implanted with fiber optic cannulas (Doric Lenses). For photometry, mice were unilaterally implanted with 1.25 mm ferrule-coupled optical fibers (0.48 NA, 400 μm diameter, Doric Lenses) cut to the desired length so that the implantation site is ~0.2 mm dorsal to the injection site. For optogenetics, mice were implanted bilaterally with 1.25mm cannulas (0.22 NA, 200um diameter, Doric Lenses). In both cases, cannula implants were slowly lowered using a stereotaxic cannula holder (Doric) at a rate of 1 mm/min until it reached the implantation site, 0.2 mm dorsal to the injection site. In the case of bilateral AM optogenetic inhibition, one cannula was implanted at a 10-degree angle laterally to the skull in order to prevent stereotactic hindrance. Optic glue (Edmund Optics) was then used to seal the skull/cannula interface and a custom titanium headplate was glued to the skull using adhesive cement (Metabond).

GRIN lens implants:

Immediately following viral injections, mice undergoing calcium imaging were implanted with gradient-index (GRIN) lens(es). An incision covering the anteroposterior extent was made, and the skin overlying the skull was cleared. The skull was then cleared and textured using a scalpel. Using a dental drill, 1mm diameter holes were made at stereotactically determined sites of implantation. Site of drilling was immediately covered using chilled Ringers solution, and using a sterile 28 G x 1.2" insulin syringe and low-pressure vacuum suction, the underlying dura was removed. GRIN lenses (1.0 mm diameter, 4.38mm length, 0.5 NA from GRINTECH (NEM-1 00-25-1 0-860-5-0.5p)) were wrapped in a 1.25mm wide custom length stainless steel sleeve (McMaster, catalog # 5560K46) using optic glue, made to cover only the part of the lens held external to the brain. With a 0.5mm burr (Fine Science Tools) attached to a stereotaxic cannula holder, the GRIN was slowly lowered into the brain at a rate of 1mm/min, ending 0.2mm dorsal to the injection site. The skull was constantly flushed with chilled 1x PBS. Every time the lens moved 0.8 mm more ventral, it was temporarily retracted 0.4 mm dorsally at the same rate, before continuing down again. We found this especially helpful to maximize the number of observed cells when imaging in deep regions. The skull-sleeve connection was then sealed with glue, and further secured with adhesive cement. A custom titanium headplate was glued to the skull using adhesive cement. Immediately following surgery, mice were injected with 0.2mg/kg dexamethasone subcutaneously to reduce inflammation.

Histology

Animals were deeply anesthetized with 5% isoflurane before transcardial perfusion with ice-cold PBS and 4% paraformaldehyde in 0.1M PB. Brains were then post-fixed by immersion for ~24 hours in the perfusate solution followed by 30% sucrose in 0.1M PB at 4°C. The fixed tissue was cut into 40 µm coronal sections using a freezing microtome (Leica SM2010R), free-floating sections were stained with DAPI (1:1000 in PBST), and mounted on slides with ProLong Diamond Antifade Mountant (Invitrogen). Images were taken on a Nikon Inverted Microscope Eclipse Ti-E with a 4x/0.2 NA objective lens. Whole-slide-images were stitched with NIS-Elements imaging software and further analyzed in ImageJ and MATLAB.

Virtual Reality System

We used a custom-built virtual reality environment, modified from a previously reported version[46]. In brief, a 200-mm-diameter styrofoam ball was axially fixed with a 6-mm-diameter assembly rod (Thorlabs) passing through the center of the ball and resting on 90° post holders (Thorlabs) at each end, allowing free forward and backward rotation of the ball. Mice were head-fixed in place above the center of the ball using a headplate mount. Virtual environments were designed in the virtual reality MATLAB engine ViRMEn[47]. The virtual environment was displayed by back-projection onto white fabric stretched over a clear acrylic hemisphere with a 14-inch diameter placed ~20 cm in front of the center of the mouse. The screen encompasses ~220° of the mouse's field of view and the virtual environment was back-projected onto this screen using a Vamvo Ultra Mini Portable projector. The rotation of the styrofoam ball was recorded by an optical computer mouse (Logitech) that interfaced with ViRMEn to transport the mouse through the virtual reality environment. A National Instruments Data Acquisition (NIDAQ) device was used to send out TTL pulses to trigger the CMOS camera, laser for optogenetics, and the various Arduinos controlling tones, odors, airpuff, lick ports. Additionally, the NIDAQ recorded the capacitance changes of the lick port when licking occurred and the CMOS camera exposures to align lick rate and neural recording/imaging to trial events.

Behavioral shaping

Starting approximately 3 weeks after surgery, mice were put on a restricted water schedule, receiving 1 mL of water in total per day. Body weight was monitored daily to ensure it was maintained above 80% of the pre-restriction measurement.

After a week of water deprivation, mice were habituated to the styrofoam ball for 2 days by receiving their 1 mL of water per day in head-fixed condition. Then mice were put onto a linear track (vertical gray bars) where water release was contingent on running a short time to trigger the onset of cues (an odor and tone) where they received 5 seconds of water delivery. Over the course of a session, and in subsequent days, the duration needed to run increased. Once mice could run on the ball for 2 seconds, we introduced a condition to stop during cue onset to trigger water delivery. Over the course of a session, and in subsequent days, the duration needed to stop increased. If a mouse took longer than 10 minutes to receive their 1 mL of water on a given day, the duration needed to run and/or stop to get water was repeated on the following day until they could reliably walk on the ball for water under 10 minutes. Once all mice from a cohort were able to run for 1 second, stop during cues for 3 seconds, and complete at least 80% of initiated trials, training began.

Behavioral task

In the final version of the task that was used during all experiments, mice ran down a virtual linear track to trigger contextual cues used to predict the outcome they will receive (~4ul of sucrose water or no water) if they stop. At the beginning of the linear track, mice self-initiated trials by running (speed > 10cm/s) down a virtual linear track for 1 second. Olfactory and auditory cues would then be presented for 3 seconds. The auditory cues consisted of 5 KHz or 9 KHz tones outputted by a thin plastic speaker (Adafruit) and olfactory cues consisting of α-pinene or octanol were diluted with mineral oil to 10% and released by a custom-built olfactometer. Both auditory and olfactory cues were outputted by Arduino code under the control of ViRMEN code. The cues for reward were a 5KHz tone and alpha-pinene while the cues for no-reward were 9KHz tone and octanol. Outcome onset would happen under the condition that a mouse dropped their speed below 10cm/s for at least 1 second before the end of the cues. If the mouse failed to stop for at least 1 second, they would be immediately placed at the start of the linear track and would need to run for 1 second to trigger the next trial start. The outcomes consisted of free access to 10% sucrose water presented by a lickometer (reward) or no water (no-reward), alongside another presentation of contextual cues, for 3 seconds. Sucrose water output were controlled by Arduino code under the control of ViRMEN code. After the outcome zone mice were transported to the beginning of the linear track to start the next trial. The order of reward and no-reward cue was pseudo-randomly predetermined through code so as to not lead to more than 2 of the same cues presented in a row.

Performance in the task was assessed by average speed and average anticipatory lick rate (during the 3 seconds of cue presentation) for all reward and no-reward trials in a given session. Prior to training, mice were given a “pre-exposure” session where they were exposed to each set of cues, with tap water given upon outcome trigger in both. They were then given 10 days of training (referred to as T1-T10). Each mouse was given 15 minutes on the ball for each training session, and supplemental water was given to each mouse if they failed to drink 1ml during a session.

Behavioral analysis

For behavioral experiments, we quantified several variables within a given session per mouse. We determined how long it took for mice to initiate trials based on how long (s) it took for their speed to be above 1cm/s for over 1 second after a reward, their speed (cm/s) during trial initiations, the percentage of times they stopped (ie, their speed (cm/s) was below 1 for at least 1 second by the 3rd second after cue onset) after cue onset within a given session, and how many rewards they received per minute (total rewards per session/minutes in a session). For analysis in **Figure 1J** we calculated time to initiate trial on a per trial basis, and rewards per minute on a per minute basis.

We also assessed learning by calculating a stop and normalized lick rate difference, which we refer to as the stop and lick discrimination index (DI). The DI was calculated as follows:

$$\text{Stop DI} = \frac{\text{stops in reward cues - stops in no-reward cues}}{\text{stops in reward cues + stops in no-reward cues}}$$

$$\text{Lick DI} = \frac{\text{mean lick rate in reward cues - mean lick rate in aversive cues}}{\text{mean lick rate in reward cues + mean lick rate in aversive cues}}$$

A DI of 1 therefore indicates perfect discrimination, while a DI of 0 indicates chance performance. For all sessions, stops were assessed by whether they triggered the reward or no-reward out period of the trial (ie, their speed (cm/s) was below 1 for at least 1 second by the 3rd second after cue onset) and lick rate was calculated in the window of time 3 seconds after the onset of the cues. Repeated measures ANOVA with Tukey's post hoc test was used to assess learning by comparing to discrimination during pre-exposure. We separated out cohorts of mice in **Figure 3** based on how well they discriminated with stops. We determined "Learner" mice by seeing if their stop DI reached above 0.5 for at least 3 consecutive days by Training Day 10, and "Non-Learner" mice as those mice who did not.

Chemogenetic inhibition of ACC

For chemogenetic silencing experiments, we injected AAV9-CaMKIIa-hM4D(Gi) (or AAV9-CaMKII-mCherry for controls) bilaterally into ACC. For a week prior to behavioral shaping, mice were habituated to handling and intraperitoneal injections of saline. A solution of clozapine N-oxide (CNO) was prepared at a concentration of 0.5 mg/mL, and mice were injected at a dosage of 5 mg/kg. Behavioral experiments were conducted 45 minutes after injection.

Optogenetic inhibition of ACC

Mice were injected with AAV1-CaMKII-stGtACR2 bilaterally in ACC, while control cohorts were injected with AAV1-CaMKII-mCherry. Cannulas were implanted directly above the injection sites. After three weeks, mice underwent shaping as described above, then moved onto training. For inhibition during training, light from a 473nm laser (15 mW at fiber tip) was delivered through a mono fiber optic patch cord for 3 seconds (cue zone followed by reinforcement zone) upon the animal entering the cue zone, throughout the duration of training (~15 minutes).

In Vivo Multi Site Photometry Recordings

Photometry Setup:

A custom multi-fiber photometry setup was built as previously[30] with some modifications that were incorporated to increase signal to noise, detailed below. Excitation of the 470 nm (imaging)

and 405 nm (isosbestic control) wavelengths were provided by LEDs (Thorlabs M470F3, M405FP1) which are collimated into a dichroic mirror holder with a 425 nm long pass filter (Thorlabs DMLP425R). This is coupled to another dichroic mirror holder with a 495 nm long pass dichroic (Semrock FF495-Di02-25x36) which redirects the excitation light on to a custom branching fiberoptic patchcord of five bundled 400 mm diameter 0.22NA fibers (BFP(5)_400/430/1100-0.48_3m_SMA-5xMF1.25, Doric Lenses) using a 10x/0.5NA Objective lens (Nikon CFI SFluor 10X, Product No. MRF00100). GCaMP6f fluorescence from neurons below the fiber tip in the brain was transmitted via this same cable back to the mini-cube, where it was passed through a GFP emission filter (Semrock FF01-520/35-25), amplified, and focused onto a high sensitivity sCMOS camera (Prime 95b, Photometrics). The multiple branch ends of a branching fiberoptic patchcord were used to collect emission fluorescence from 1.25mm diameter light weight ferrules (MFC_400/430-0.48_ZF1.25, Doric Lenses) using a mating sleeve (Doric SLEEVE_ZR_1.25). The excitation was alternated between 405nm and 470nm by a custom made JK flip flop which takes the trigger input from the sCMOS and triggers the two excitation LEDs alternatively. Bulk activity signals were collected using Photometrics data acquisition software, Programmable Virtual Camera Access Method (PVCAM).

Photometry Recordings:
While mice performed the self-paced contextual learning VR task we recorded bulk calcium signals from five regions: ACC, OFC, AM, BLA, and LC simultaneously. Mice shown in **Figures 1-2** with ACC recordings also contained OFC, AM, BLA, and LC recordings, which we compile and show all together in **Figure 3**. We recorded at 18 Hz with excitation wavelengths alternating between 470 nm and 405nm, capturing calcium dependent and independent signals respectively, resulting in an effective frame rate of 10 Hz.

Data Processing:
For analysis, the images captured by the CMOS camera were post-processed using custom MATLAB scripts. Regions of interest were manually drawn for each fiber to extract fluorescence values throughout the experiment. The 405-nm reference trace was scaled to best fit the 470-nm signal using least-squares regression. The normalized change in fluorescence (dF/F) was calculated by subtracting the scaled 405-nm reference trace from the 470-nm signal and dividing that value by the scaled 405-nm reference trace. The true baseline of each dF/F trace was determined and corrected by using the MATLAB function *msbackadj*, estimating the baseline over a 200-frame sliding window, regressing varying baseline values to the window's data points using a spline approximation, then adjusting the baseline in the peak range of the dF/F signal.

Bulk neural responses:
The adjusted calcium signals from photometry were aligned to task events (for example, cue onset, reward, trial initiation, etc) in ViRMEn by time-stamping behavioral frames captured through the NIDAQ. Photometry signals from all animals from a given region were Z-scored across the entire session. The mean regional responses to task variables (**Figures 1, 2, 3**), is the mean of these aligned Z-scored signals across all animals, with s.e.m. calculated across all recorded trials. We then sought to quantify the difference in mean average activity patterns observed in response to each cue presentation. To calculate the differential response to reward and trial initiation portions of the task (**Figure 1**), we calculated the time it took for dF/F activity to rise above 1std and speed

to rise above 1cm/s (**Figure 1F**) or dF/F activity to rise above 0.5std and speed to rise above 2cm/s (**Figure 1G**). We zeroed the dF/F activity to the start of reward or 2 seconds prior to trial initiation.

We then sought to quantify the difference in temporal divergence activity patterns observed in reward or no-reward cue presentation. To calculate the differential response to cue onset we calculated the mean Z-scored signal from 0 to 3 seconds after cue onset (**Figure 2**). We also quantified the difference in pre-cue activity along sequences of trials with 1 or 2 no-reward cues to identify ramps in neural activity between reward cues (**Figure 3**). We zeroed the pre-cue activity of all trials within a given sequence to the activity at the time of the first reward cue and calculated the mean z-scored signal between 2 second before to 0 seconds before cue onset.

In-vivo cellular resolution calcium imaging

Imaging setup:

For imaging in **Figure 4**, mice were imaged throughout training in 15 min sessions per day. Volumetric imaging was performed using a resonant galvanometer two-photon imaging system (Bruker), with a laser (Insight DS+, Spectra Physics) tuned to 920 nm to excite the calcium indicator, GCaMP6f, through a 16×/0.8 water immersion objective (Nikon) interfacing with an Gradient Refractive Index (GRIN) lens through a few drops of distilled water. Prior to each session, mice were headfixed and each GRIN lens was carefully cleaned with 70% ethanol. Fluorescence was detected through GaAs photomultiplier tubes using the Prairie View 5.4 acquisition software. Black dental cement was used to build a well around the implant to minimize light entry into the objective from the projector. High-speed z-stacks were collected in the green channel (using a 520/44 bandpass filter, Semrock) at 512 × 512 pixels covering each x–y plane of 800 × 800 mm over a depth of ~150 µm (30 µm apart) by coupling the 30 Hz rapid resonant scanning (x–y) to a z-piezo to achieve ~3.1 Hz per volume. Average beam power measured at the objective during imaging sessions was between 20–40 mW. An incoming TTL pulse from ViRMEn at the start of behaviour enabled time-locking of behavioural epochs to imaging frames with millisecond precision.

Source extraction:

Calcium imaging data for **Figure 4** was acquired by Prairie View 5.4 acquisition software and subsequently processed using the Suite2p toolbox[48]. Motion correction, ROI detection and neuropil correction were performed as described. Cell identification was verified by manually validating every extracted source. Cell registration across sessions for **Figure 4** was performed with a combination of custom scripts and existing packages (Cell Reg[49]).

Calculation of single cell dF/F and transient identification:

For each cell detected via automated source extraction, a normalized ΔF/F was calculated and individual Ca2+ transients were identified as previously described[46]. Briefly, ΔF/F was defined as: $(F - F_{baseline})/F_{baseline}$, where F is the raw output ("F") from the suite2p algorithm, and where $F_{baseline}$ is the baseline fluorescence, calculated as the mean of the fluorescence values for a given cell, continuously acquired over a 20 s sliding window to account for slow time-scale changes observed in the fluorescence across the recording session. For all analysis, this dF/F was then normalized by z-scoring the entire time series across a session. To identify statistically significant transients, we first calculated an estimate of the noise for each cell using a custom MATLAB script, with a previously described method[50,51]. In essence, we identified the limiting noise cutoff level for a

given cell using time periods that are unlikely to contain neural events, and then defined a transient as significant if it reached above at least 3σ of this estimated noise level. A custom MATLAB script using the function "findpeaks" was used to identify any remaining obvious transients not identified by this method (typically when multiple transients occurred in rapid succession). Additional specifications required transients to persist above this noise level for at least 300 ms (roughly twice the duration of the half-life decay time of GCaMP6f). The transient duration was defined as the first and last frames where the dF/F reached 3σ. The value of dF/F was set to zero outside the duration of every identified transient to minimize effects of residual background fluorescence.

Single cell cue tuning:
To calculate the tuning of an individual cell in anticipation to or during reward or no-reward cues, we z-scored the trial averaged the activity on a given neuron across all the cue presentations for a given reward or no-reward trial. A cell was considered tuned if the magnitude of its trial averaged z-scored activity was at least 0.75 between 3 seconds before or after cue onset. To find tuning for cues based on previous trial history, we pre-selected cues that were preceded by specific combinations of trials.

Transient time analysis:
To calculate the transient times of an individual cell tracked before and after learning, we first preselected cells that are tuned to a particular trial sequence (such as NNR). We then identified the frame when the dF/F value first rises above the noise threshold (see Calculation of single cell dF/F and transient identification) 7 seconds before or after cue onset (such as for the R cue in an NNR-tuned cell). We took all the transients for any given cell across all the trials in a session, in case a single cell fired more than one transient.

Statistical Analysis
Sample sizes were selected based on expected variance and effect sizes from the existing literature, and no statistical methods were used to determine sample size a priori. Prior to experiments being performed, mice were randomly assigned to experimental or control groups. The investigator was blinded to all behavioral studies. Data analyses for calcium imaging (in vitro and in vivo datasets) were automated using MATLAB scripts. Statistical tests were performed in MATLAB 2017a, 2021b, or Graphpad Prism.

ACKNOWLEDGMENTS

We thank Nakul Yadav for discussions surrounding 2-photon imaging and resulting analysis. We are grateful to Dr. Vanessa Ruta, Dr. Gaby Maimon, Dr. Zachary Mainen, and Rajasethupathy lab members for helpful discussions related to various aspects of the study. We thank Alessandra Bonito Oliva for help with project management and discussions/comments on the manuscript. This work was supported by an NSF Graduate Research Fellowship and HHMI Gilliam fellowship awarded to JMR, and grants from the Searle and Klingenstein foundations and from the National Institutes of Health under award number DP2AG058487 to PR.

AUTHOR CONTRIBUTIONS

JMR and PR conceived the study and designed the experiments. ACA characterized the VR task and performed stereotactic surgeries for multi-fiber photometry. TH performed the chemogenetics experiment. ACT performed cell tracking analysis. YRL and LN performed optogenetics experiments. JMR and PR wrote the manuscript with input from all authors. PR supervised all aspects of the work.

DECLRATION OF INTERESTS

The authors declare no competing financial interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. One or more of the authors of this paper received support from a program designed to increase minority representation in science.

References

1. Balleine, B.W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology *37*, 407–419. 10.1016/S0028-3908(98)00033-1.

2. Tolman, E.C. (1948). Cognitive maps in rats and men. Psychological Review *55*, 189–208. 10.1037/h0061626.

3. Pezzulo, G., van der Meer, M.A.A., Lansink, C.S., and Pennartz, C.M.A. (2014). Internally generated sequences in learning and executing goal-directed behavior. Trends Cogn Sci *18*, 647–657. 10.1016/j.tics.2014.06.011.

4. Miller, E.K., and Cohen, J.D. (2001). An Integrative Theory of Prefrontal Cortex Function. Annual Review of Neuroscience *24*, 167–202. 10.1146/annurev.neuro.24.1.167.

5. Shenhav, A., Botvinick, M.M., and Cohen, J.D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron *79*, 217–240. 10.1016/j.neuron.2013.07.007.

6. Stroop, J.R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology *18*, 643–662. 10.1037/h0054651.

7. Yuan, P., and Raz, N. (2014). Prefrontal cortex and executive functions in healthy adults: a meta-analysis of structural neuroimaging studies. Neurosci Biobehav Rev *42*, 180–192. 10.1016/j.neubiorev.2014.02.005.

8. D'Esposito, M., and Postle, B.R. (2015). The cognitive neuroscience of working memory. Annu Rev Psychol *66*, 115–142. 10.1146/annurev-psych-010814-015031.

9. Milner, B. (1963). Effects of Different Brain Lesions on Card Sorting: The Role of the Frontal Lobes. Archives of Neurology *9*, 90–100. 10.1001/archneur.1963.00460070100010.

10. Pardo, J.V., Pardo, P.J., Janer, K.W., and Raichle, M.E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. Proc. Natl. Acad. Sci. U.S.A. *87*, 256–259. 10.1073/pnas.87.1.256.

11. Shallice, T., and Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. Brain *114 (Pt 2)*, 727–741. 10.1093/brain/114.2.727.

12. Hauber, W., and Sommer, S. (2009). Prefrontostriatal circuitry regulates effort-related decision making. Cereb. Cortex *19*, 2240–2247. 10.1093/cercor/bhn241.

13. Hillman, K.L., and Bilkey, D.K. (2010). Neurons in the Rat Anterior Cingulate Cortex Dynamically Encode Cost-Benefit in a Spatial Decision-Making Task. Journal of Neuroscience *30*, 7705–7713. 10.1523/JNEUROSCI.1273-10.2010.

14. Wallis, J.D., and Kennerley, S.W. (2011). Contrasting reward signals in the orbitofrontal cortex and anterior cingulate cortex. Ann. N. Y. Acad. Sci. *1239*, 33–42. 10.1111/j.1749-6632.2011.06277.x.

15. Cowen, S.L., Davis, G.A., and Nitz, D.A. (2012). Anterior cingulate neurons in the rat map anticipated effort and reward to their associated action sequences. J. Neurophysiol. *107*, 2393–2407. 10.1152/jn.01012.2011.

16. Shenhav, A., Cohen, J.D., and Botvinick, M.M. (2016). Dorsal anterior cingulate cortex and the value of control. Nature Neuroscience *19*, 1286–1291. 10.1038/nn.4384.

17. Walton, M.E., Bannerman, D.M., Alterescu, K., and Rushworth, M.F.S. (2003). Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. J Neurosci *23*, 6475–6479. 10.1523/JNEUROSCI.23-16-06475.2003.

18. Schweimer, J., Saft, S., and Hauber, W. (2005). Involvement of catecholamine neurotransmission in the rat anterior cingulate in effort-related decision making. Behavioral Neuroscience *119*, 1687–1692. 10.1037/0735-7044.119.6.1687.

19. Amiez, C., Joseph, J.P., and Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. Cereb Cortex *16*, 1040–1055. 10.1093/cercor/bhj046.

20. Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., and Rushworth, M.F.S. (2006). Optimal decision making and the anterior cingulate cortex. Nat Neurosci *9*, 940–947. 10.1038/nn1724.

21. Monosov, I.E., Haber, S.N., Leuthardt, E.C., and Jezzini, A. (2020). Anterior cingulate cortex and the control of dynamic behavior in primates. Curr Biol *30*, R1442–R1454. 10.1016/j.cub.2020.10.009.

22. Holroyd, C.B., and Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. Trends Cogn Sci *16*, 122–128. 10.1016/j.tics.2011.12.008.

23. Hillman, K.L., and Bilkey, D.K. (2012). Neural encoding of competitive effort in the anterior cingulate cortex. Nat Neurosci *15*, 1290–1297. 10.1038/nn.3187.

24. Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., and Cohen, J.D. (2001). Conflict monitoring and cognitive control. Psychol Rev *108*, 624–652. 10.1037/0033-295x.108.3.624.

25. Heilbronner, S.R., and Hayden, B.Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. Annu Rev Neurosci *39*, 149–170. 10.1146/annurev-neuro-070815-013952.

26. Rushworth, M.F.S., Kolling, N., Sallet, J., and Mars, R.B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Curr Opin Neurobiol *22*, 946–955. 10.1016/j.conb.2012.04.011.

27. Monosov, I.E. (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. Nat Commun *8*, 134. 10.1038/s41467-017-00072-y.

28. Kolling, N., Wittmann, M.K., Behrens, T.E.J., Boorman, E.D., Mars, R.B., and Rushworth, M.F.S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. Nature Neuroscience *19*, 1280–1285. 10.1038/nn.4382.

29. Fillinger, C., Yalcin, I., Barrot, M., and Veinante, P. (2017). Afferents to anterior cingulate areas 24a and 24b and midcingulate areas 24a′ and 24b′ in the mouse. Brain Structure and Function *222*, 1509–1532. 10.1007/s00429-016-1290-1.

30. Kim, C.K., Yang, S.J., Pichamoorthy, N., Young, N.P., Kauvar, I., Jennings, J.H., Lerner, T.N., Berndt, A., Lee, S.Y., Ramakrishnan, C., et al. (2016). Simultaneous fast measurement of circuit dynamics at multiple sites across the mammalian brain. Nature Methods *13*, 325–328. 10.1038/nmeth.3770.

31. Rudebeck, P.H., Saunders, R.C., Prescott, A.T., Chau, L.S., and Murray, E.A. (2013). Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. Nat Neurosci *16*, 1140–1145. 10.1038/nn.3440.

32. Montague, P.R., and Berns, G.S. (2002). Neural economics and the biological substrates of valuation. Neuron *36*, 265–284. 10.1016/s0896-6273(02)00974-1.

33. Mainen, Z.F., and Kepecs, A. (2009). Neural representation of behavioral outcomes in the orbitofrontal cortex. Curr Opin Neurobiol *19*, 84–91. 10.1016/j.conb.2009.03.010.

34. Rich, E.L., and Wallis, J.D. (2016). Decoding subjective decisions from orbitofrontal cortex. Nat Neurosci *19*, 973–980. 10.1038/nn.4320.

35. Padoa-Schioppa, C., and Conen, K.E. (2017). Orbitofrontal Cortex: A Neural Circuit for Economic Decisions. Neuron *96*, 736–754. 10.1016/j.neuron.2017.09.031.

36. Behrens, T.E.J., Muller, T.H., Whittington, J.C.R., Mark, S., Baram, A.B., Stachenfeld, K.L., and Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. Neuron *100*, 490–509. 10.1016/j.neuron.2018.10.002.

37. Sadacca, B.F., Wied, H.M., Lopatina, N., Saini, G.K., Nemirovsky, D., and Schoenbaum, G. (2018). Orbitofrontal neurons signal sensory associations underlying model-based inference in a sensory preconditioning task. eLife *7*, e30373. 10.7554/eLife.30373.

38. Drummond, N., and Niv, Y. (2020). Model-based decision making and model-free learning. Current Biology *30*, R860–R865. 10.1016/j.cub.2020.06.051.

39. Blanchard, T.C., and Hayden, B.Y. (2014). Neurons in dorsal anterior cingulate cortex signal postdecisional variables in a foraging task. J Neurosci *34*, 646–655. 10.1523/JNEUROSCI.3151-13.2014.

40. Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. Nature Neuroscience *14*, 933–939. 10.1038/nn.2856.

41. Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R.F., Dayan, P., and Costa, R.M. (2021). The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection. Neuron *109*, 149-163.e7. 10.1016/j.neuron.2020.10.013.

42. Sarafyazd, M., and Jazayeri, M. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. Science *364*, eaav8911. 10.1126/science.aav8911.

43. Tervo, D.G.R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., and Karpova, A.Y. (2014). Behavioral Variability through Stochastic Choice and Its Gating by Anterior Cingulate Cortex. Cell *159*, 21–32. 10.1016/j.cell.2014.08.037.

44. Peters, A.J., Liu, H., and Komiyama, T. (2017). Learning in the Rodent Motor Cortex. Annual Review of Neuroscience *40*, 77–97. 10.1146/annurev-neuro-072116-031407.

45. Bittner, K.C., Milstein, A.D., Grienberger, C., Romani, S., and Magee, J.C. (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. Science *357*, 1033–1036. 10.1126/science.aan3846.

46. Rajasethupathy, P., Sankaran, S., Marshel, J.H., Kim, C.K., Ferenczi, E., Lee, S.Y., Berndt, A., Ramakrishnan, C., Jaffe, A., Lo, M., et al. (2015). Projections from neocortex mediate top-down control of memory retrieval. Nature *526*, 653–659. 10.1038/nature15389.

47. Aronov, D., and Tank, D.W. (2014). Engagement of Neural Circuits Underlying 2D Spatial Navigation in a Rodent Virtual Reality System. Neuron *84*, 442–456. 10.1016/j.neuron.2014.08.042.

48. Pachitariu, M., Stringer, C., Dipoppa, M., Schröder, S., Rossi, L.F., Dalgleish, H., Carandini, M., and Harris, K.D. (2017). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. 061507. 10.1101/061507.

49. Sheintuch, L., Rubin, A., Brande-Eilat, N., Geva, N., Sadeh, N., Pinchasof, O., and Ziv, Y. (2017). Tracking the Same Neurons across Multiple Days in Ca2+ Imaging Data. Cell Rep *21*, 1102–1115. 10.1016/j.celrep.2017.10.013.

50. Toader, A.C., Regalado, J.M., Li, Y.R., Terceros, A., Yadav, N., Kumar, S., Satow, S., Hollunder, F., Bonito-Oliva, A., and Rajasethupathy, P. (2023). Anteromedial thalamus gates the selection and stabilization of long-term memories. Cell *186*, 1369-1381.e17. 10.1016/j.cell.2023.02.024.

51. Yadav, N., Noble, C., Niemeyer, J.E., Terceros, A., Victor, J., Liston, C., and Rajasethupathy, P. (2022). Prefrontal feature representations drive memory recall. Nature *608*, 153–160. 10.1038/s41586-022-04936-2.

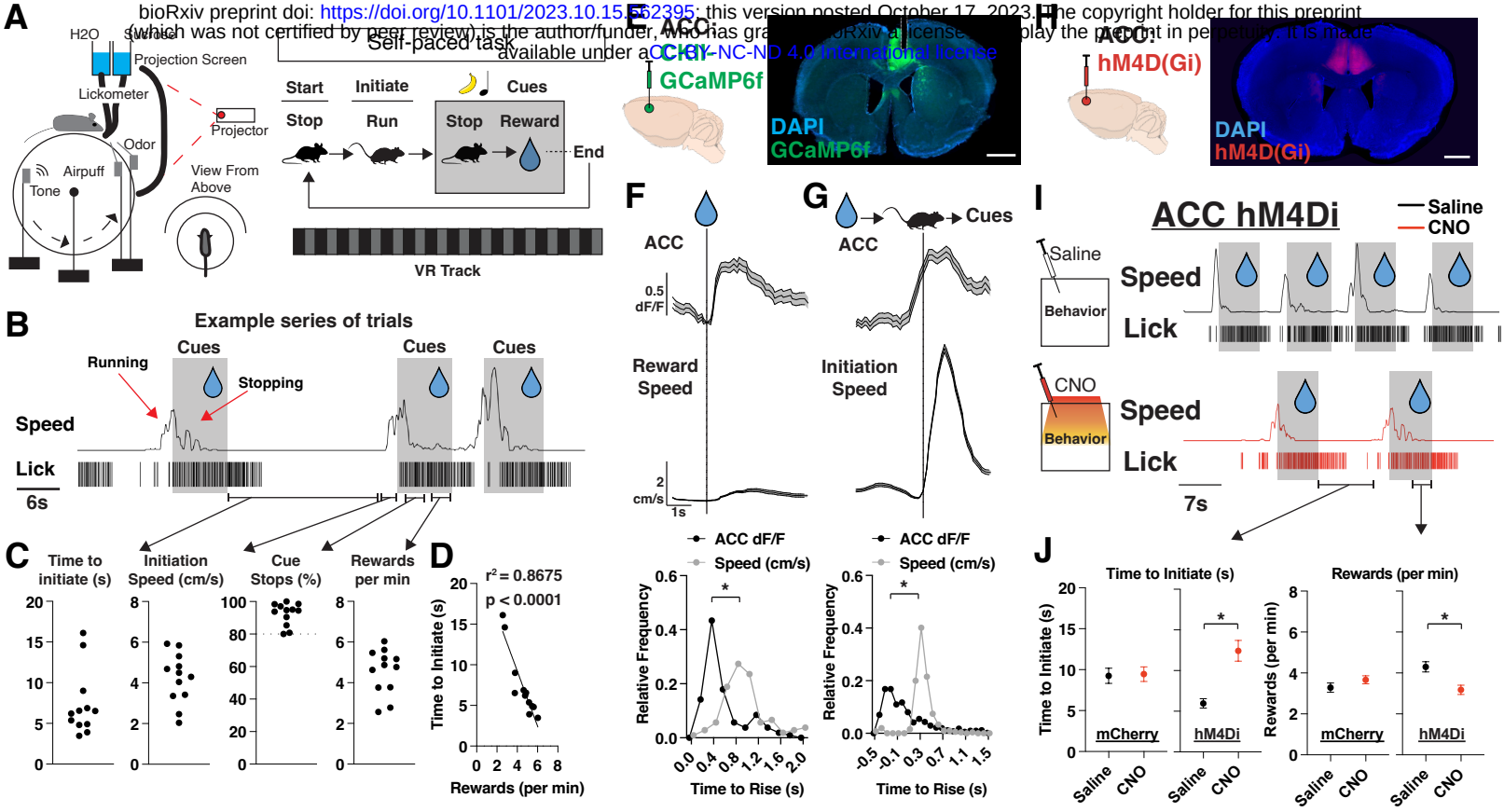

Figure 1. Neural activity in ACC signals a motivational state to obtain reward

(A) Schematic of virtual reality experimental setup and trial structure. A mouse initiates a trial by running to trigger the onset of cues (olfactory and auditory). After cue onset, a mouse stops to collect a water reward, which ends the trial (see Methods).

(B) Representative traces of speed and licks from one mouse during a session, with shaded portions corresponding to when cues are on. Red arrows correspond to periods when mice are running to trigger cue onset or stopping to trigger water delivery. Black arrows correspond to sections of a session where we can quantify time to initiate trials, initiation speed, cue stops, and rewards.

(C) Quantification per mouse of time to initiate a trial (far left; seconds), initiation speed (left; cm/s), % trials in which a stop occurred during cue presentation (right), and rewards received per minute. Individual data points shown (N=12 mice).

(D) Scatter plots of the mean time (s) to initiate a trial plotted alongside rewards received per minute per mouse (N=12 mice). Individual data points shown, with a best fit line, represented by the solid line in the figure. r2=0.8675 and p<0.0001 are determined by linear regression.

(E) Left: bulk neural activity recording experimental design. GCaMP6f was injected into the anterior cingulate cortex (ACC) and neural activity was recorded on a fiber photometry setup (see Methods). Right: Brain histology from a representative mouse showing DAPI in blue, GCaMP6f in green and photometry cannula implantation in ACC (dotted white lines). Scale bar: 1mm.

(F) Top: Trial average plots of ACC activity (z-scored dF/F) and speed (cm/s) aligned to reward onset. Data are mean (solid line) ± s.e.m (shaded area). Bottom: Relative frequency plots of the time (s) for ACC dF/F or speed to rise above 1 std or 1 cm/s during rewards, respectively (N=105 trials across 12 mice). *p<0.05, paired t-test between time to rise (s) between ACC and speed. Data is the frequency of values across time.

(G) Same as F, but for trial initiations. (N=510 trials across 12 mice).

(H) Injection strategy for DREADDS-based chemogenetic inhibition of ACC during self-paced task. Coronal section from an animal virally injected with AAV1-Cam-Kii-hM4D(Gi) in ACC. DAPI is shown in blue and hM4D(Gi) in red. Scale bar: 1mm.

(I) Representative traces of speed and licks from one mouse during the task on a day with saline (top) or CNO (bottom) administration 45 minutes prior to a session, with shaded portions corresponding to when cues are presented.

(J) Left: Quantification of time (s) to initiate trial (left) across saline and CNO sessions in mCherry-control mice (N=188 trials across 6 mice) and hM4D(Gi)-DREADDs mice (N=215 trials across 4 mice). Right: same as left but for rewards received per minute in mCherry-control mice (N=60 minutes across 6 mice) and hM4D(Gi)-DREADDs mice (N=40 minutes across 4 mice). p=0.8707 for mCherry and *p<0.05 for hM4Di (time to initiate), p=0.2073 for mCherry and *p<0.05 for hM4Di (rewards per min), unpaired t-test between saline and CNO sessions per group. Data are mean ± s.e.m.

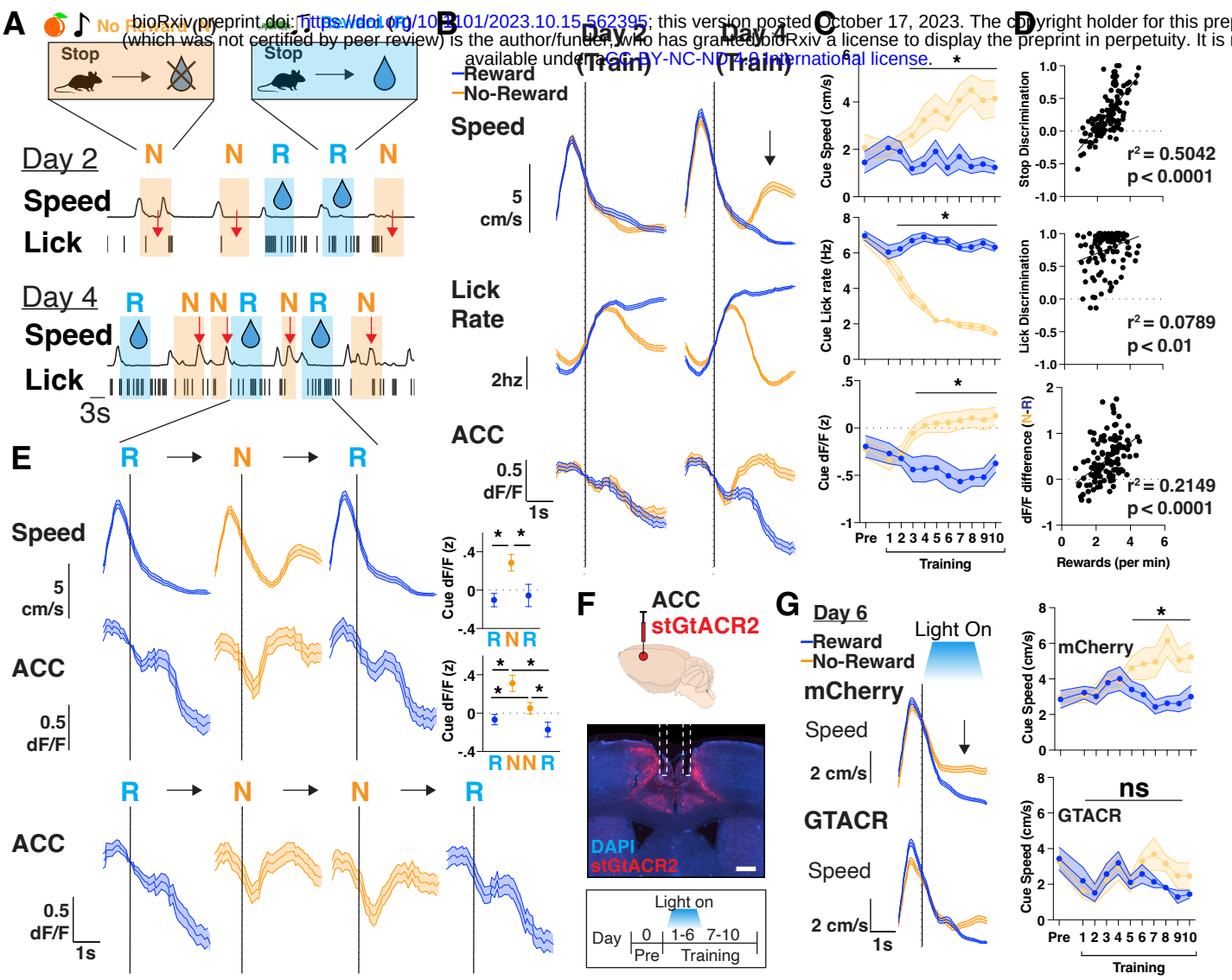

Figure 2. Neural activity in ACC scales to match an increased motivational state during learning

(A) Top: Schematic of training where mice learn to associate stopping to one set of cues with no water reward ("N") or with water reward ("R"). Bottom: Representative traces of speed and licks from one mouse during a session on Training Day 2 and Day 4, with shaded portions corresponding to when a reward cues (R, blue) or no-reward cues (N, orange) is presented. Red arrow denotes the suppression of licks on Day 2, and rise in speed during no-reward cues on Day 4.

(B) Trial averaged speed (cm/s; top), lick rate (Hz; middle) and ACC activity (dF/F z-scored; bottom) aligned to cue presentation across day 2 and 4 of training, separated by reward and no-reward cues (blue vs orange). Black arrow signifies rise in speed after no-reward cue presentation. N=12 mice. Data are mean (dark line) with s.e.m. (shaded area).

(C) Quantification of average cue speed (cm/s; top), lick rate (Hz; middle) and ACC activity (dF/F z-scored; bottom) across training, separated by reward and no-reward cues (blue vs orange). N=12 mice in each group, data are mean ± s.e.m. *p<0.05, paired t-test between reward and no-reward.

(D) Scatter plots of rewards per minute vs stop discrimination (top), lick discrimination (middle), or dF/F difference (bottom) for each mouse throughout training (N=120 data points, 12 mice per each of 10 days). Data are individual points with best fit line. r2 and p values are shown, as determined by linear regression.

(E) Top: Trial averaged speed (cm/s) and ACC activity (dF/F z-scored) aligned to cue presentation across 3 trials consisting of a reward, no-reward, and reward cue (RNR). Bottom: Trial averaged ACC activity (dF/F z-scored) aligned to cue presentation across 4 trials consisting of a reward, no-reward, no-reward and reward cue (RNNR). Right: Quantification of average cue dF/F activity across RNR and RNNR trial sequences. N=12 mice. *p<0.05, one-way repeated measured ANOVA with post-hoc Tukey's multiple comparison test. Data are mean (dark line) with s.e.m. (shaded area) or data are mean ± s.e.m (right).

(F) Top: Injection strategy for stGtACR2-based optogenetic inhibition of ACC during training. Middle: Brain histology from a representative mouse showing DAPI in blue, stGtACR2 in red and photometry cannula implantation in ACC. Scale bar: 1mm. Bottom: optogenetic inhibition was targeted to days 1-6 of training and mice were allowed to continue training for days 7-10.

(G) Left: Trial averaged plots of speed (cm/s) aligned to cue entry on T6 for mCherry controls and GTACR inhibition mice, separated by reward or no reward cues. Right: Quantification of mean speed during cue presentations. N=8 mice for mCherry, 4 for GTACR early inhibition. *p<0.05, paired t-test.

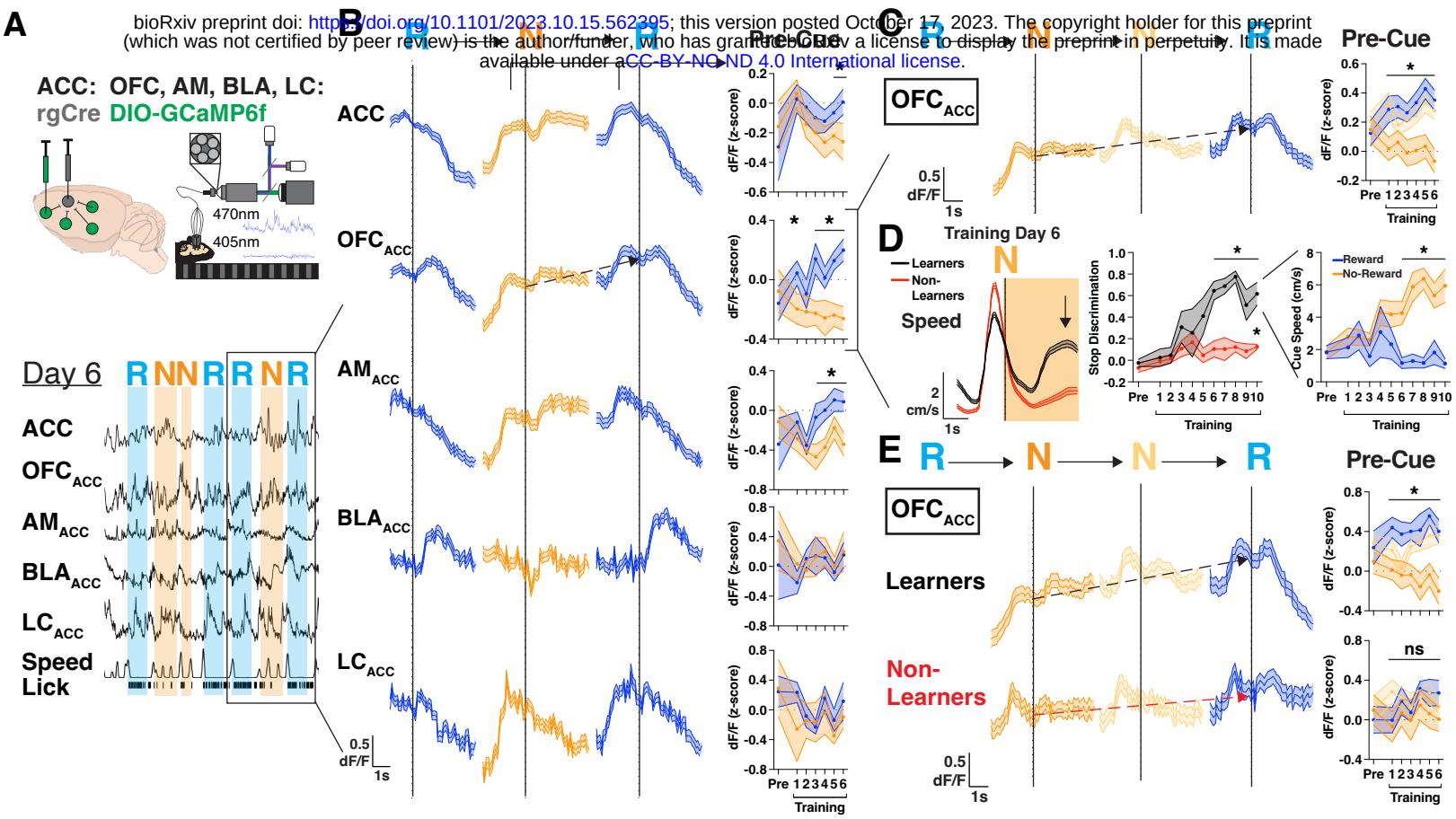

Figure 3. Mice with extended motivational states during learning display neural activity ramps in OFC

(A) Injection strategy and fiber-based photometry setup to record bulk GCaMP6f of projections to ACC from OFC_{ACC} (orbitofrontal cortex), AM_{ACC} (anteromedial thalamus), BLA_{ACC} (basolateral amygdala), or LC_{ACC} (locus coeruleus). Representative traces for a single mouse showing traces for each region dF/F, speed, and licks. Shaded portions are shown corresponding to when a reward cues (R, blue) or no-reward cues (N, orange) are presented.

(B) Left: trial averaged bulk GCaMP6f dF/F of ACC, OFC_{ACC}, AM_{ACC}, BLA_{ACC}, and LC_{ACC} during a sequence of trials on T6 including reward, no-reward, and reward cues (RNR). Black arrows denote the rise in pre-cue activity from N cue to the following R cue in the RNR sequence. Right: quantification of pre-cue activity for the N cue and following R cue. Data are mean (solid line) ± s.e.m (shaded area). N=19, 12, 5, 4 mice, data are mean (solid line) ± s.e.m (shaded area), *p<0.05, paired t-test between N vs R cues.

(C) Left: trial averaged bulk GCaMP6f dF/F of OFC_{ACC} during a sequence of trials including reward, two no-reward, and reward cues (RNNR). Red arrows denote the rise in pre-cue activity from first N cue to the last R cue in the RNNR sequence. Right: quantification of pre-cue activity for the first N cue, second N cue and last R cue. Data are mean (solid line) ± s.e.m (shaded area). N=19 mice, data are mean (solid line) ± s.e.m (shaded area), *p<0.05, one-way repeated measures ANOVA with post-hoc Tukey's multiple comparison test.

(D) Left: speed (cm/s) for "Learner" (black; reached a DI > .5 for 3 consecutive days) or "Non-Learner" (red) mice on training day 6 aligned to no-reward cue onset. Middle: discrimination index for each group of mice throughout training. Right: speed during reward and no-reward cues for "Learner" mice. N=7 ("Learner") and 9 ("Non-Learner") mice. Data are mean (solid line) ± s.e.m (shaded area), *p<0.05, unpaired t-test between Learner and Non-Learner DI (middle), paired t-test between reward and no-reward cues (right).

(E) Left: trial averaged bulk GCaMP6f dF/F of OFC_{ACC} during a sequence of trials including reward, two no-reward, and reward cues (RNNR). Black arrows denote the rise in pre-cue activity from first N cue to the last R cue in the RNNR sequence. Red arrows denote the absence of this ramp in Non-Learner mice. Right: Quantification of pre-cue activity for the first N cue, second N cue and last R cue. Data are mean (solid line) ± s.e.m (shaded area). N=7 ("Learners") and 9 ("Non-Learner") mice, data are mean (solid line) ± s.e.m (shaded area), *p<0.05, one-way repeated measures ANOVA with post-hoc Tukey's multiple comparison test.

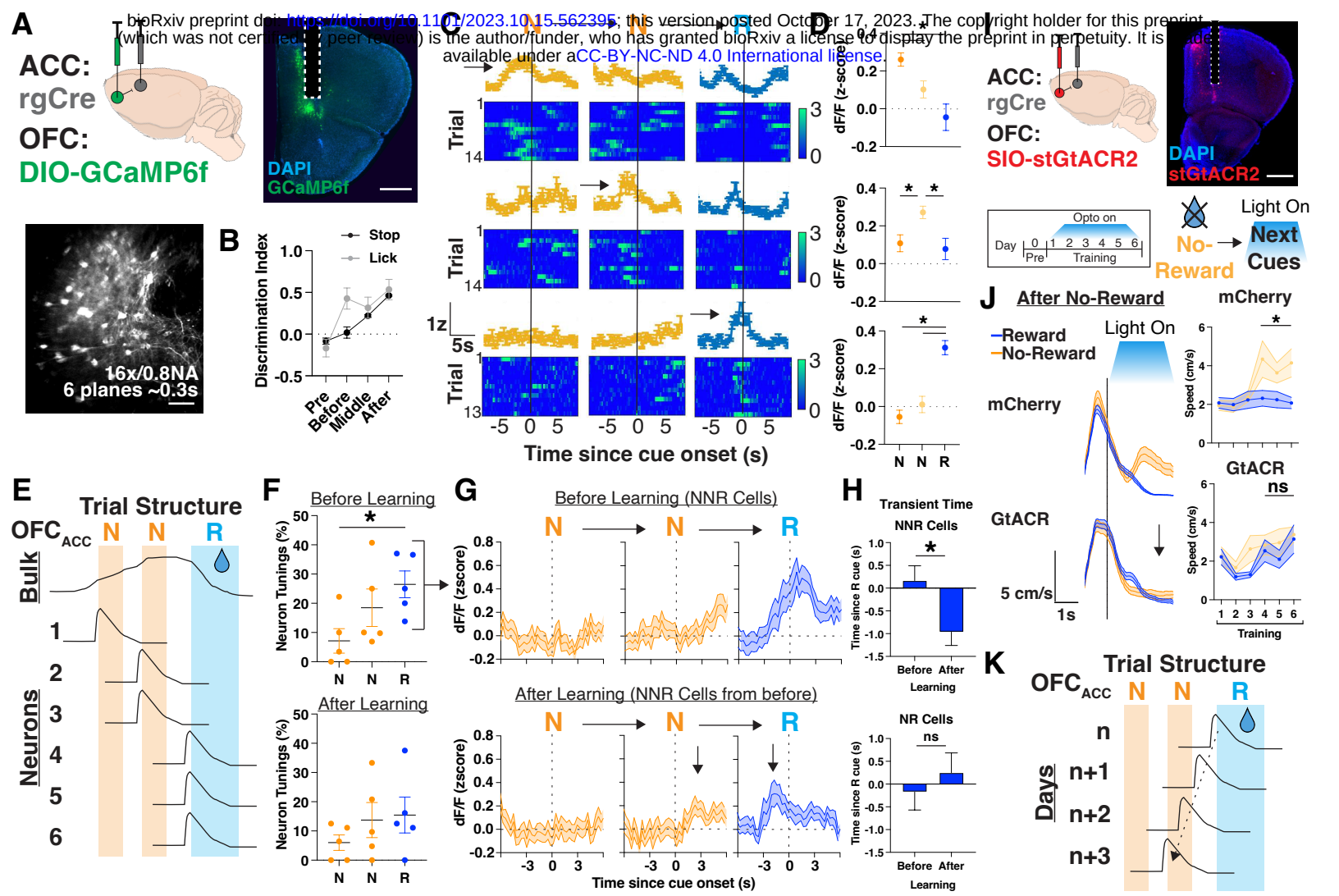

Figure 4. Orbitofrontal cortex projection neurons tile sequences of trials with no-rewards

(A) Injection strategy (top left), histology (top right; scale bar, 1mm) and z-projection images of two-photon recording (bottom left; mean over time; scale bars, 200 μ m) of GCaMP expressing OFC projection neurons with GRIN implants. Bottom right: sequence of trials with z-scored dF/F for individual neurons, with shaded portions corresponding to when a reward cues (R, blue) or no-reward cues (N, orange) are presented. Red arrow denotes a dF/F transient occurring after 2 consecutive N cues.

(B) Stop (black) or lick (grey; see methods) discrimination index on the first day stop DI reaches > 0.4 (”after”) and the two previous days (”before” and “middle”). N=5 mice.

(C) Representative neurons with tunings (std > 0.75 for 3 seconds prior to or after cue presentation) to separate cues in an NNR trial sequence. Trial averaged activity of a N (top), NN (middle), and NNR (bottom) neuron with heat map showing individual trial responses.

(D) Quantification of neurons tuned to separate cues within an NNR trial sequence and their activity to all other cues. N=17 (N), 18 (NN), 32 (NNR) cells out of 115 cells in total. *p<0.05, one-way repeated measures ANOVA with post-hoc Tukey’s multiple comparison test.

(E) Schematic of OFC_{ACC} bulk activity based on Figure 3 results and potential single neuron findings that tile a sequence of trials with two no-rewards followed by a reward cue presentation (NNR).

(F) Percentage of neurons tuned to different cues in an NNR trial sequence before (top) or after (bottom) training. N=5 mice. *p<0.05, one-way repeated measures ANOVA with post-hoc Tukey’s multiple comparison test.

(G) Ensemble average plots of neurons tuned to R cues after 2 consecutive N cue presentations (NNR cells) before learning (top) and their activity after learning (bottom). Black arrows denote the rise in activity prior to R cues after learning. N=18 NNR cells out of 81 cells tracked across days.

(H) Quantification of transient time (s) since R cue onset for neurons tracked across days. N= 132, 170 transient events before and after learning across 18 NNR cells and 105, 59 transient events before and after learning across 12 NR cells. *p<0.05, unpaired t-test.

(I) Left: Injection strategy for stGtACR2-based optogenetic inhibition of OFC_{ACC} during training. Optogenetic inhibition was targeted to training for 6 days. Right: Brain histology from a representative mouse showing DAPI in blue, stGtACR2 in red and photometry cannula implantation in ACC. Scale bar: 1mm.

(J) Left: mean animal speed (cm/s) aligned to cue zone entry after no-reward on T6 for mCherry control or GtACR mice. Black arrow signifies lack of speed increase during N cues. Right: quantification of mean change speed in cue zone after no-reward, assessed separately for each cue presentation. N=10 mice for mCherry and 13 mice for GtACR, *p<0.05, paired t-test.

(K) Schematic of reward-responsive OFC projection neurons becoming increasingly active during no reward cues that precede reward cues over days.

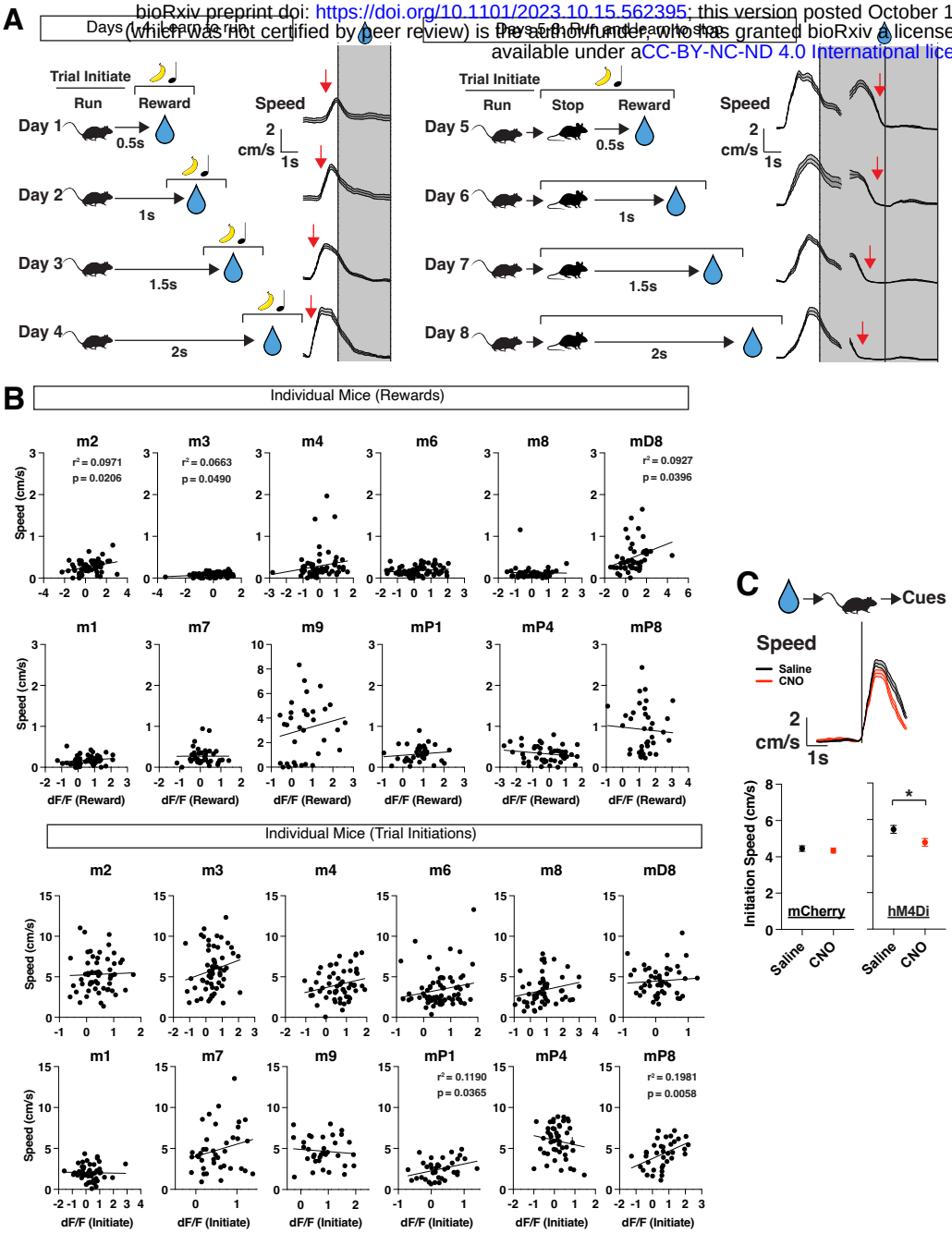

Figure S1. Task shaping and speed related differences between mice and during ACC inhibition
(A) Left: schematic of behavioral shaping. Mice were shaped to run (>1 cm/s) for increasing durations over the course of 4 days to obtain rewards. Trial averaged plots of speed along the days of shaping (N=3 mice). Red arrow denotes the increasing duration needed to run to trigger rewards. Right: Days where mice decrease their speed (<1cm/s; stop) during cues for rewards. Red arrows denote the increasing duration to stop to trigger rewards. Data are mean (solid line) ± s.e.m (shaded area).
(B) Scatter plots of speed (cm/s) and ACC dF/F during reward (top) or trial initiations (bottom) for individual mice. Individual data points shown, with a best fit line, represented by the solid line in the figure. r2 and p values, as determined by linear regression, are shown for each mouse that had a p<0.5.
(C) Left: Trial average plots of speed (cm/s) aligned to trial initiation for saline-administered day (black) or CNO (red; N=4 mice). Data are mean (solid line) ± s.e.m (shaded area). Quantification of speed during trial initiation for mCherry-control mice (N=187, 214 trials across 6 mice) and hM4D(Gi)-DREADDs mice (N=166, 120 trials across 4 mice). p=0.5692 for mCherry and *p=0.0217 for hM4Di, unpaired t-test between saline and CNO sessions per group. Data are individual points, with mean ± s.e.m.

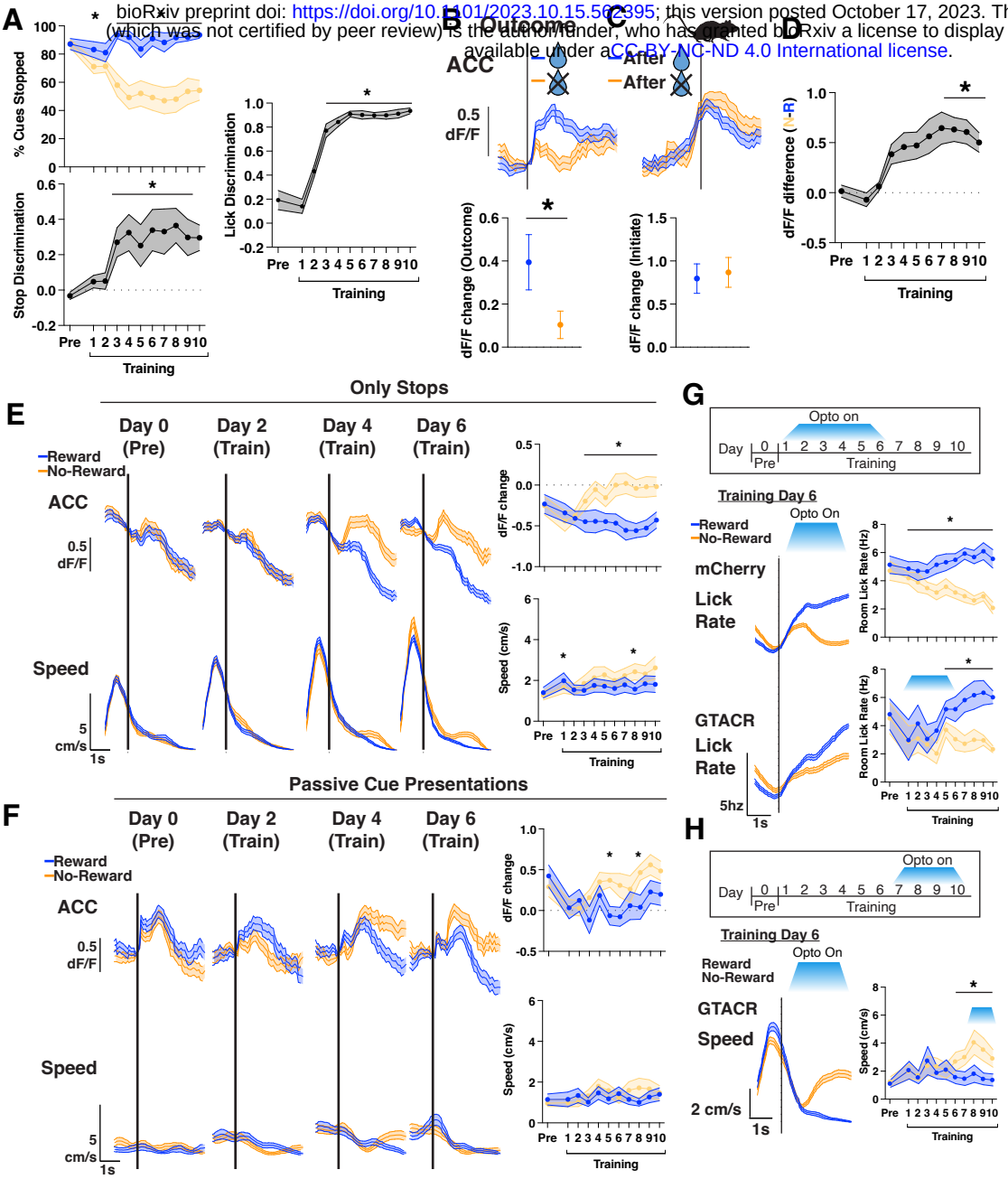

Figure S2. Lick rate discrimination, ACC learning signal controls, and ACC inhibition lick rate learning

(A) Left: Percentage of cue presentations with stops, separated by reward and no-reward cues (blue vs orange) and quantification of stop discrimination index (see Methods) across training. Right: Quantification of lick discrimination index (see Methods) across training. N=12 mice, data are mean ± s.e.m. *p<0.05, paired t-test between reward and no-reward each day, or one-way repeated measures ANOVA with post-hoc Tukey's multiple comparison test between each training day and preexposure.

(B) Top: Trial average plots of ACC activity (z-scored dF/F) and speed (cm/s) aligned to outcome onset, separated by reward or no-rewards. Data are mean (solid line) ± s.e.m (shaded area). Bottom: Quantification of the mean dF/F and speed during outcome. N=12 mice in each group, data are mean ± s.e.m. *p<0.05, paired t-test between reward and no-reward each day.

(C) Same as B, but for trial initiations after each outcome.

(D) Quantification of ACC dF/F difference between reward and no-reward cues (see Methods) across training. N=12 mice, data are mean ± s.e.m. *p<0.05, one-way repeated measures ANOVA with post-hoc Tukey's multiple comparison test between each training day and preexposure.

(E) Left: mean dF/F in ACC (top) and speed (bottom) aligned to cue onset for cue presentations in which stops occurred (blue vs orange). N=12 mice. Data are mean (dark line) with s.e.m. (shaded area). Right: Quantification of mean change in dF/F and speed in cue zone, assessed separately for each cue presentation. N=12 mice in each group, data are mean ± s.e.m. *p<0.05, paired t-test between reward and no-reward each day.

(F) Same as E, but for passive presentation of the reward and no-reward cues.

(G) Top: optogenetic inhibition was targeted to days 1-6 of training and mice were allowed to continue training for days 7-10. Bottom left: Trial averaged plots of lick rate (hz) aligned to cue onset on T6 for mCherry controls and GTACR inhibition mice, separated by reward or no reward cues. Bottom right: Quantification of mean lick rate during cues. N=8 mice for mCherry, 4 for GTACR early inhibition. *p<0.05, paired t-test.

(H) Top: optogenetic inhibition was targeted to days 7-10 of training. Bottom: Trial averaged plots of speed (cm/s) aligned to cue entry on T10 for GTACR inhibition mice, separated by reward or no reward cues. Quantification of mean speed during cues. N=8 mice for mCherry, 4 for GTACR late inhibition. *p<0.05, paired t-test. N=4 mice for GTACR late inhibition. *p<0.05, paired t-test.

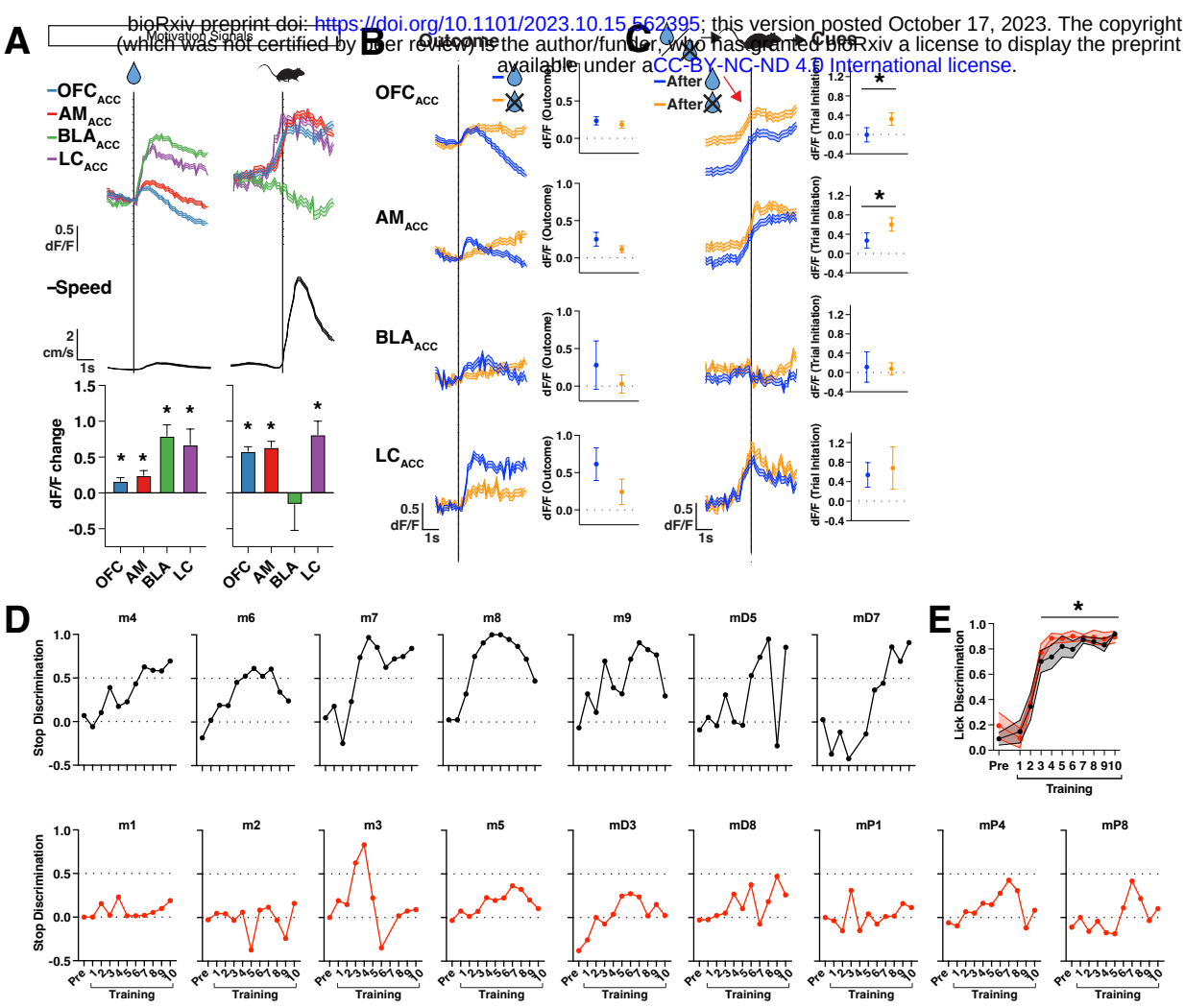

Figure S3. Motivation signals in bulk projection activity, and behavior of learners

(A) Top: Trial average plots of each projection activity (z-scored dF/F) and speed (cm/s) aligned to reward onset (left) or trial initiations (right). Data are mean (solid line) ± s.e.m (shaded area). Bottom: Quantification of the mean dF/F change during reward (left) or trial initiation (right; N=19, 12, 5, 4 mice). Data are mean ± SEM.

(B) Left: Trial average plots of projection activity (z-scored dF/F) aligned to outcome onset, separated by reward or no-rewards. Data are mean (solid line) ± s.e.m (shaded area). Right: Quantification of the mean dF/F during outcome. N=19, 12, 5, 4 mice, data are mean ± s.e.m.

(C) Same as C, but for trial initiations after each outcome. *p<0.05, paired t-test between reward and no-reward each day.

(D) Plots of stop discrimination across preexposure or training for each mouse in the “Learner” group (top) or “Non-Learner” group (bottom). Dashed line is shown for 0 and 0.5 stop discrimination.

(E) Lick discrimination between the “Learner” or “Non-Learner” mice. *p<0.05, one-way ANOVA between training days and preexposure, with post-doc Tukey’s multiple comparison test.

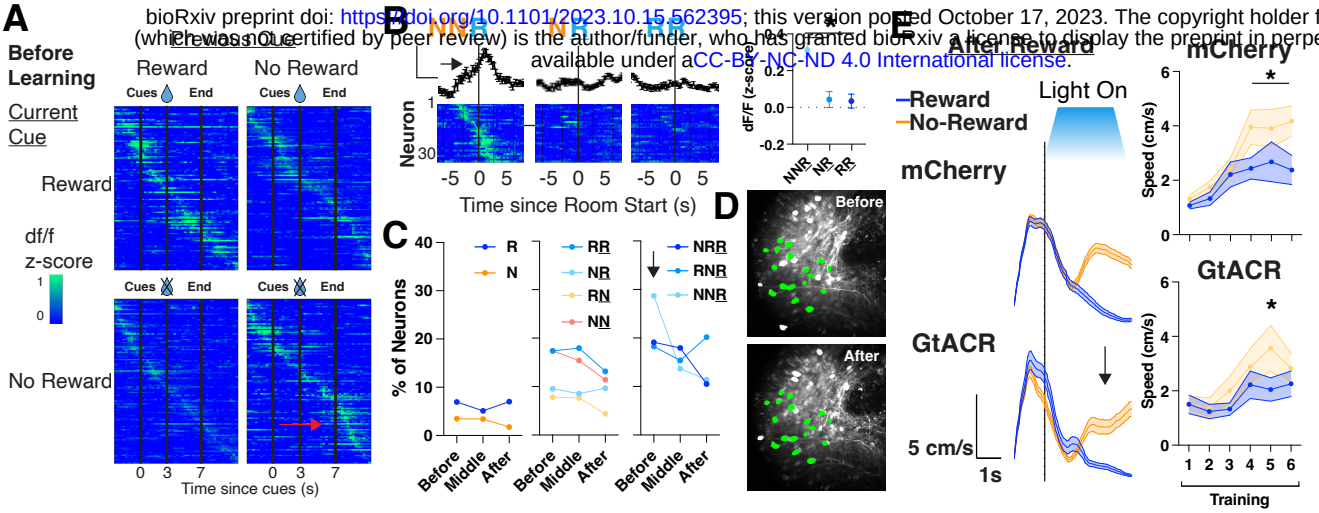

Figure S4. Neuron tunings to NNR task structure and inhibition of OFC_{ACC} neurons.

(A) Mean z-scored dF/F for every recorded cell, aligned to cue onset separated by current cue and previous cue presentation (N=115 cells across 5 mice). Red arrow denotes the increased response after no-reward cues when the previous cue was no-reward.

(B) Left: mean population activity (top; z-scored dF/F) and individual neuron trial averaged activity (bottom) for NNR neurons during R cues preceded by 2 N, 1 N, or 1R cue presentations. Right, quantification of NNR tuned neurons' response to R cues preceded by 2 N, 1 N, or 1 R cue presentations (N=32 cells). *p<0.05, one-way repeated measures ANOVA with post-hoc Tukey's multiple comparison test.

(C) Percentage of neurons tuned (std > 0.75 3 seconds before or after cue onset) to R or N cues in total (left) or based on what cues preceded it one trial before (middle) or across two trials before (right). N=115, 117, 114 cells for "before", "middle", and "after" days. Black arrow denotes the higher percentage of neurons that are tuned to R cues after 2 N cues (NNR cells) before learning.

(D) Neural sources for a single field of view. Sources captured throughout learning are highlighted in green.

(E) Left: mean animal speed (cm/s) aligned to cue zone entry after reward on T6 for mCherry control or GtACR mice. Black arrow signifies speed increase during N cues. Right: quantification of mean change speed in cue zone after reward, assessed separately for each cue presentation. N=10 mice for mCherry and 13 mice for GtACR, *p<0.05, paired t-test.

Supplementary Video 1

Behavior during learning. Playback speed: 2x. Shown here is a representative mouse learning to stop in cues that predict reward (blue walls) and run throughout consecutive cue presentations that predict no-reward (yellow walls). Displayed trial sequence order is reward, no-reward, no-reward, and reward (RNNR).