Mixed representations of choice and outcome by GABA/glutamate cotransmitting neurons in the entopeduncular nucleus

Julianna Locantore[1], Yijun Liu[1], Jesse White[1], Janet Berrios Wallace[2], Celia C. Beron[2], Bernardo L. Sabatini[2], and Michael L. Wallace[1]

[1]Department of Anatomy and Neurobiology, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, USA

[2]Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical School, Boston, MA, USA

Abstract:

The basal ganglia (BG) are an evolutionarily conserved and phylogenetically old set of sub-cortical nuclei that guide action selection, evaluation, and reinforcement. The entopeduncular nucleus (EP) is a major BG output nucleus that contains a population of GABA/glutamate cotransmitting neurons (EP^{Sst+}) that specifically target the lateral habenula (LHb) and whose function in behavior remains mysterious. Here we use a probabilistic switching task that requires an animal to maintain flexible relationships between action selection and evaluation to examine when and how GABA/glutamate cotransmitting neurons contribute to behavior. We find that EP^{Sst+} neurons are strongly engaged during this task and show bidirectional changes in activity during the choice and outcome periods of a trial. We then tested the effects of either permanently blocking cotransmission or modifying the GABA/glutamate ratio on behavior in well-trained animals. Neither manipulation produced detectable changes in behavior despite significant changes in synaptic transmission in the LHb, demonstrating that the outputs of these neurons are not required for on-going action-outcome updating in a probabilistic switching task.

Introduction:

Animals select actions based on incoming sensory information, their current state, and past experience to achieve goals. Experiences modify behavior to promote the repetition of actions associated with positive outcomes and suppress those associated with bad outcomes. Nevertheless, it is also advantageous to maintain flexibility and adjust behavior if the environment changes and to exploit new opportunities as they arise. The basal ganglia (BG) are an evolutionarily ancient group of nuclei in the brain conserved in all vertebrates and crucial for goal-directed movements, including behavioral updating as a consequence of experience[1]. The BG are involved in both action repetition and exploration and consequently, defects in the BG contribute to human disorders ranging from Parkinson's and Huntington's disease, to drug addiction[2,3]. Despite the importance of BG to human behavior, how these evolutionarily conserved and phylogenetically old nuclei carry out these functions is not fully understood.

Neural activity from many areas of sensory, motor and limbic cortices converge onto the striatum, the main input structure of the BG[4]. The dorsal striatum modulates the output nuclei of the BG, the substantia nigra reticulata (SNR) and entopeduncular nucleus (EP), through two routes. The so-called direct pathway is formed by dopamine receptor type 1 (D1R) expressing striatal projection neurons (SPNs) that synapse onto output neurons in the SN and EP. The indirect pathway consists of dopamine receptor type 2 (D2R) expressing SPNs that innervate the globus pallidus externus (GPe), which projects to the SNR and EP. GPe also innervates the subthalamic nucleus (STN) which projects to SNR/EP. Canonically, the SNR and EP modulate motor output through their connections to cortically-projecting thalamic nuclei[3,4].

The function of EP is of particular interest because, whereas the EP in rodents (and globus pallidus internus (GPi) in primates) clearly has motor functions, it is distinct from SNR in that it projects to the lateral habenula

(LHb) and carries reward and sensory signals, implying additional limbic and associative functions[5,6]. Our previous work demonstrated that the mouse EP has at least two genetically defined cell-types that project to the LHb, one expresses Parvalbumin (*Pvalb*) and vGlut2 (*Slc17a6*) and is purely glutamatergic (excitatory), the other expresses Somatostatin (*Sst*), vGluT2 (*Slc17a6*), and vGaT (*Slc32a1*) and cotransmits both GABA and glutamate[7,8]. Here, we focus on activity patterns of cotransmitting EP^{Sst+} neurons in freely moving behavior and their involvement in ongoing action selection and outcome evaluation.

Cotransmitting neurons are increasingly recognized as important contributors to neural circuit function throughout the brain, but specifically manipulating one transmitter at a time to assess impacts on behavior has been challenging[9,10]. To determine the function of *Sst*+ GABA/glutamate cotransmitting EP neurons during behavior, we developed a probabilistic switching task where animals choose between two nose-poke ports that asymmetrically and probabilistically deliver water rewards. The task alternates the location (left or right) of the highly rewarded port every 50 rewards (referred to as a block transition), requiring the animal to remain flexible to maximize the number of rewards it receives. We find that animals are sensitive to changes in reward probability and accurately follow the location of the highly rewarded port following a block transition. EP^{Sst+} neurons are strongly engaged during this task and show bidirectional changes in activity during the choice and outcome periods of the task. We then test the requirement for ongoing cotransmission of both GABA and glutamate from EP^{Sst+} to the LHb for continued task performance. Additionally, we alter the GABA/glutamate ratio of cotransmission by genetically deleting vGlut2 from EP^{Sst+} neurons. Despite observing strong modulation of their activity during a trial, neither manipulation of synaptic release resulted in detectable changes in task performance and we conclude that EP^{Sst+} neurons are not required for ongoing trial-to-trial action-outcome evaluation in well-trained animals as assessed on a probabilistic switching task.

Results:

Changing reward probabilities affects performance on a probabilistic switching task

To examine the activities of EP^{Sst+} cotransmitting neurons during behavior we employed a dynamic, probabilistic switching task in mice modeled after behavioral paradigms shown to require basal ganglia circuitry[11–14]. Water restricted, freely-moving animals are placed in a behavioral arena with three nose-poke ports. A center poke initiates a trial, then the animal chooses to poke the left or right side ports to receive a water reward (~3uL, Figure 1a). Water rewards are delivered asymmetrically and probabilistically in a block structure such that once 50 rewards are gained, the reward probabilities are reversed (referred to as a block transition, Figure 1b, dotted vertical line). There is no cue that a transition to a new block has occurred; therefore, following a block transition, the probability that the animal chooses the highly rewarded port (p(high port)) drops dramatically. As the animal adjusts its choices, p(high port) gradually increases for the next 10-15 trials (Figure 1d). A well-trained animal will use its history of choices (left or right) and outcomes (rewarded or unrewarded) to guide future actions and is sensitive to block transitions (Figure 1b-e). Lights above the ports indicate when the center or side ports are active to assist training, but otherwise provide no information regarding the location of the highly rewarded port. Well-trained animals perform 350-500 trials in a forty-minute session with individual trials totaling 2-3 seconds with an enforced minimum inter-trial interval of 1 second (Figure 1a-c). The trial typically begins with a quick center port entry and exit, after which the animal must poke to the left or right side port within 8 seconds and lick the water spout (Figure 1c). If a reward is delivered the animal continues to lick and consume the reward (~1 sec). If the trial is unrewarded, the animal quickly returns to the center port to initiate a new trial (Figure 1c).

To test how different reward probabilities effected task performance we chose three pairs of reward probabilities ranging from more deterministic (90/10) to more stochastic (70/30) (Figure 1, red:90/10, blue:80/20, and green:70/30). We observed strong effects on several behavioral metrics (Figure 1d-i). First, following a block transition, p(high port) returned to pre-block transition levels more quickly (fewer trials) with

90/10 reward probabilities than with 80/20 or 70/30 (Figure 1d). This was quantified by fitting an exponential to the first 20 trials following a block transition and extracting the time constant ($tau_{p(high\ port)}$, Figure 1d inset and Figure 1-figure supplement 1a). Second, the probability of the mouse switching side port choice on consecutive trials (p(switch)) increased sharply following a block transition, and the maximum p(switch) was greatest for 90/10 reward probabilities and increased the fastest. Together, these measures indicate that the animals adapt their behavior most rapidly under these conditions (Figure 1e). Consequently, across all trials the probability that a trial is rewarded (p(reward)) is greatest in the 90/10 reward probability (mean:0.78, SEM:0.001), and progressively decreases with 80/20 and 70/30 conditions (Figure 1g). Given that a block transition occurs only after 50 rewards are gained, decreased p(reward) results in increased block lengths for 80/20 and 70/30 conditions relative to the 90/10 condition (Figure 1-figure supplement 1i).

Unlike p(reward) and p(high port), the average p(switch) across all trials does not depend on the reward probability conditions (Figure 1f-h). However, differences in switching behavior are revealed when we separated trials by the outcome on the previous trial. Across conditions, p(switch) following a rewarded trial is low (Figure 1i; 90/10: p(switch)=0.04), and following an unrewarded trial is high (Figure 1j; 90/10: p(switch)=0.35). However, the 70/30 condition shows p(switch) significantly lower than the 90/10 condition following both rewarded and unrewarded trials (Figure 1i-j). Therefore, the large impact changing reward probabilities has on p(switch) is revealed by considering the outcome of the previous trial (Figure 1e-j). But when considering all trials (Figure 1h), there is no difference in p(switch) across reward probabilities simply because there are far fewer rewarded trials in the 70/30 condition (Figure 1g). These findings prompted us to detail the impact of outcome and choice on p(switch) and we examined all possible combinations of trial history for two trials in the past (Figure 1-figure supplement 1d-e). For almost all types of trial history, the 70/30 condition had the lowest p(switch) indicating that the animal is much less likely to switch ports following an unrewarded trial in the 70/30 condition (Figure 1-figure supplement 1e) which likely contributes to decreased overall *p*(high port).

Finally, we used a linear model (termed Recursively Formulated Logistic Regression; RFLR), previously developed to describe the behavior of a mouse performing the probabilistic switching task, to examine if the animal's strategy changed with different reward probabilities[14,15]. In this model the next choice is based on evidence about the location of rewards, represented by the interaction between choice (left or right) and outcome (rewarded or unrewarded; Figure 1-figure supplement 1j). This variable decays over trials and is updated with new evidence from each new trial's choice and outcome with an additional bias towards or away from its most recent choice (Figure 1-figure supplement 1j). Thus, the parameters or the RFLR capture the tendency of an animal to repeat its last choice (alpha, α), the weight given to evidence about past choice and outcome (beta, β), and the time constant (tau, τ) over which the influence of choice and outcome history decays (Figure 1-figure supplement 1j-k). Importantly, the model performed equally well across reward probabilities as measured by negative log-likelihood of the fit (90/10= -0.25 SD=0.03; 80/20= -0.24 SD=0.03; 70/30=-0.25 SD=0.02) and accurately predicted mouse behavior (*p*(switch) and *p*(high port)) around block transitions (Figure 1-figure supplement 2). Modeling the different reward probabilities revealed that the most stochastic reward probability (70/30) had the greatest beta (β) and tau (τ) coefficients indicating that to accurately represent the animal's behavior the model needed to use evidence (from previous choice and outcome) accumulated from trials further in the past (Figure 1-figure supplement 1k-l). This strategy likely arises in conditions in which rewarded outcomes are more random (such as the 70/30 condition) and accounting for more past trials can improve the animal's chances at determining the location of the highly rewarded port.

EP^{Sst+} neurons respond during trial choice and outcome

To examine the activity of projection and genetically defined neuronal populations in the EP, we injected Cre-dependent GCaMP6f and tdTomato into EP and AAVretro Flp-dependent Cre into the LHb of a *Sst*-Flp mouse line. This resulted in GCaMP6f and tdTomato expression specifically in the *Sst*+ neurons of the EP without off-

target expression in surrounding areas (Figure 2a). We implanted a fiber optic above the EP and recorded EP^{Sst+} population calcium-mediated fluorescence changes during the probabilistic switching task using fiber photometry. We consistently observed dynamic changes in GCaMP6f mediated fluorescence while the animal was engaged in the task that were not present in the control (tdTomato) static fluorophore (Figure 2b and Figure 2 -figure supplement 1a). We aligned photometry signals to different behavioral events to examine how the EP^{Sst+} activity changed relative to different periods of a trial (Figure 2c). When we segregated trials by the direction the animal made its side-port choice (ipsi=same side as the recording site, contra=opposite side to the recording site) we observed large differences in EP^{Sst+} neuronal activity (Figure 2c). Following the center port entry (CE) a large rise in activity was present on ipsilateral trials not seen on contralateral trials (Figure 2c and e). This increase in activity was seen for all three reward probabilities tested (90/10, 80/20, and 70/30) and occurred while the animal was engaged in ipsiversive movements as similar increases were observed following side exit (SX) on contralateral trials as the animal was moving from the contralateral side port back to the center port (Figure 2-figure supplement 1c).

A large increase in EP^{Sst+} neuronal activity was also observed following side port entry (SE) on unrewarded trials for both contralateral and ipsilateral choices (Figure 2d). This was mirrored by a distinct decrease in fluorescence on rewarded trials following side port entry (Figure 2d). Increased EP^{Sst+} neuronal activity following an unrewarded outcome was partially due to the rapid withdrawal of the animal's snout following an unrewarded outcome however, differences in rewarded and unrewarded trials were still distinguishable when signals were aligned to side port exit indicating that these increases in EP^{Sst+} neuronal activity on unrewarded trials were a combination of outcome evaluation (unrewarded) and side port withdrawal occurring in quick succession (SX, Figure 2 -figure supplement 1d).

One hypothesis is that these outcome signals reflect reward prediction error[11], which implicitly reflects expectation (given that reward size does not change trial-to-trial). Under different reward probability conditions, the expected reward and corresponding error should scale; however, these patterns in response to rewarded and unrewarded trial outcomes were virtually identical on all reward probabilities tested (Figure 2e-f) indicating that they are unlikely to reflect changes in reward expectation. To further examine if reward prediction error (RPE) contributed to the changes in EP^{Sst+} neuronal activity observed following side port entry, we divided trials by whether the previous trial ($trial_{-1}$) was rewarded or unrewarded. For rewarded trials (both ipsilateral and contralateral), we observed a small effect of the previous trial outcome on EP^{Sst+} activity following side port entry (SE) (Figure 2 -figure supplement 1f-g). EP^{Sst+} activity on rewarded trials was increased when the previous trial was unrewarded, however this effect of trial history was not observed on unrewarded trials (Figure 2 -figure supplement 1f-g). Therefore, the bidirectional changes in EP^{Sst+} neuronal activity observed during the action evaluation period of a trial likely reflect a combination of outcome value and differential timing of movement sequences on rewarded and unrewarded outcomes. In sum, we saw two timepoints with differential activity during a trial: at trial initiation (CE), we found increased activity specifically during ipsiversive movements. Then, during outcome evaluation (SE), we found bidirectional modulation dependent on reward outcome.

Choice and outcome shape EP^{Sst+} activity

The movements of an animal during a trial of the probabilistic switching task are complex and occur in quick succession making it difficult to disambiguate which behavioral events may be associated with specific features of the simultaneously recorded neural signal (Figure 1c). Generalized linear models (GLMs) can be used to quantitatively determine which behavioral events explain the observed neural signal[14,16–18]. We defined a set of behavioral variables (such as the timing of rewards, port entries, etc.) as predictors for a GLM to fit the neural data (Figure 3a). For each behavioral variable the GLM assigns a kernel of time shifted beta (β) coefficients that represent the contribution of that variable to the neural signal (GCaMP6f fluorescence; Figure 3a and c). These kernels can then be convolved with the actual timing of behavior events in a trial and summed to create a "reconstructed" GCaMP6f signal which is compared to the actual (original) signal (Figure 3a, right).

To estimate the beta coefficients, the original dataset (90/10 reward probabilities, n=6 mice, 5 sessions/mouse) is divided into training (80%) and test (20%) datasets. We fit the model on the training data and evaluated it on the test data using mean squared error (MSE, which is the cost function minimized by the model), calculated by comparing the reconstructed neural signal and the actual (original) photometry signal (Figure 3a-b). We tested and compared multiple regularization methods (ridge, elastic net, ordinary least squares) across a large hyperparameter space and found that ridge regression performed most consistently when evaluated with MSE (MSE=0.80 SD= 0.001, R^2=0.19 SD=0.001, Figure 3d-e).

We then determined which behavioral variables contributed to GLM performance by omitting variables and examining the importance of each variable to the model performance as measured by a change in MSE (Figure 3e). We found that omitting reward variables decreased performance of the GLM (increased MSE), indicating that the neural signal cannot be entirely explained by the movements and port entries/exits of the animal during a trial (Figure 2e, "-Rew"). As we observed large differences in the neural signal during ipsilateral and contralateral movements (Figure 2c), we tested the requirement for choice direction on GLM performance by collapsing the ipsilateral and contralateral port entries into a single variable void of directionality but preserving event timing (e.g. SE_{Contra} and SE_{Ipsi} were combined and represented as a single "SE" variable). This resulted in a large drop in GLM performance (increased MSE) indicating that the direction of the side port choice (ipsi vs contra) was critical for accurate reconstruction of the neural signal (Figure 3e, "-Choice"). Omitting center port entry/exit together or individually also resulted in decreased GLM performance, but to a smaller degree than omission of choice direction (Figure 3e, "-Center"). The same pattern was true for side port entry/exit (Figure 3e, "-Side"). Together, testing the GLM performance revealed that both choice direction and reward were important for optimal model performance supporting an interpretation that EP^{Sst+} neurons signal both movement direction during a choice (ipsi vs contra) and reward aspects of a trial.

EP^{Sst+} neurons are not required for continued performance on a probabilistic switching task.
EP^{Sst+} neurons directly and exclusively project to the LHb a region principally implicated in evaluating negative outcomes of an action[7,19,20]. Photometry recordings from EP^{Sst+} neurons during behavior suggested that these neurons were actively engaged during both the action selection (ipsi vs contra side port) and outcome evaluation periods of a trial (Figure 2 and 3). We hypothesized that ablation of synaptic release from these neurons, thus blocking their ability to communicate with the LHb, would strongly impact the outcome evaluation phase of the task. We trained mice on the probabilistic switching task (90/10 reward probability) to reach predefined criteria where task performance was consistent over a week (See Methods, and Figure 4-figure supplement 1a-c). *Sst-Cre* mice were then injected with AAVs containing either Cre-dependent GFP (GFP, green, control) or Cre-dependent tetanus toxin light-chain which blocks synaptic vesicle fusion[21] (Tettx, red; Figure 4a). Mice continued daily sessions on the task for 3 weeks to allow for viral expression. Control animals showed no significant differences in behavioral performance after surgery, indicating that the surgery was well tolerated and resulted in no observable detrimental side effects (Figure 4-figure supplement 1 and 2). We quantified the number of Tettx expressing cells in the EP at the termination of behavioral testing as a percentage of the entire Sst+ population based on stereological estimates[22,23]. We found that our injections targeted 70±15.1% (mean±SD) of the EP^{Sst+} population (1070±230 neurons/animal, n=6 mice). In separate animals we functionally confirmed that 3 weeks of Tettx expression in EP^{Sst+} neurons were sufficient to block both optogenetically evoked IPSCs and EPSCs from EP^{Sst+} axons to LHb neurons (Figure 4-figure supplement 1j-k).

We then compared behavioral performance on the task before AAV injection to sessions collected 3 weeks post injection in both control and Tettx groups (Figure 4). Both groups performed well before and after viral injection, selecting the high reward port around a block transition similarly with no significant differences between groups (Figure 4b). Control and Tettx groups also showed no significant change in *p*(switch) around a block transition or following rewarded and unrewarded trials (Figure 4c-e) indicating that the sensitivity of the animal to detect the outcome of the previous trial and respond on subsequent trials was not significantly

perturbed. All other behavioral metrics (ITI duration, trial duration, p(reward), port bias, etc.) were unchanged between groups or when compared before and after AAV injection (Figure 4-figure supplement 1d-i). Consistent with our other measures the mouse behavioral strategy as assessed by the RFLR model was also unperturbed between groups (Figure 4-figure supplement 1m). Together these data indicate that ablation of both GABA and glutamate release from EP^{Sst+} neurons is not sufficient to result in profound behavioral performance changes in animals well trained on the probabilistic switching task despite strong modulation of EP^{Sst+} activity during a trial as reported by fiber photometry (Figure 2).

Genetic deletion of synaptic glutamate release from EP^{Sst+} neurons during the probabilistic switching task

EP^{Sst+} neurons simultaneously cotransmit both GABA and glutamate onto individual neurons in the LHb[7,8,24]. Although studies have suggested that the primary effect of EP^{Sst+} cotransmission in LHb is excitatory in vitro[24], the *in vivo* effects of EP^{Sst+} neurons on LHb are unknown. Additionally, the ratio of GABA/glutamate cotransmitted from EP^{Sst+} neurons has been shown to be plastic following exposure to environmental stressors and drugs of abuse possibly altering the net effect on LHb activity[24,25]. We reasoned that altering the ratio of GABA/glutamate cotransmission by genetic deletion of the vesicular glutamate transporter (vGluT2, *Slc17a6*) might have stronger effects on downstream LHb activity and associated behaviors than deleting both GABA and glutamate release together (Figure 4).

Similar to previous experiments, we trained *Sst-Cre* mice on the probabilistic switching task (90/10 reward probability) to criteria and then injected AAVs containing *SaCas9* with either control guide RNA for the ROSA26 locus (sgROSA) or guide RNA for the *Slc17a6* the gene encoding the vesicular glutamate transporter 2 (vGluT2) to permanently disrupt gene function[26] (Figure 5A). These animals were also injected with Cre-dependent oChief for post-hoc electrophysiological examination of synaptic transmission from EP^{Sst+} axons. Indeed, as confirmed after completion of behavioral experiments (5 weeks post injection), we observed near total loss of glutamatergic transmission from *Sst*+ axons as measured by voltage clamp recordings in LHb at -65mV (Figure 5-figure supplement 1j-k). Importantly, when LHb neurons were held at the reversal potential for AMPA receptors large optogenetically evoked IPSCs were revealed confirming that EP^{Sst+} neurons still made functional synaptic contacts with the LHb, but that they were now almost entirely GABAergic (Figure 5-figure supplement 1j-k).

We then compared behavioral performance on the task before AAV injection to sessions collected 3 weeks post injection in both ROSA26 (control) and *Slc17a6* (vGluT2 deletion) groups (Figure 5 and Figure 5-figure supplement 1a-c). Both groups performed well before and after viral injection, selecting the high reward port around a block transition similarly with no significant differences between groups (Figure 5b). ROSA26 and *Slc17a6* groups also showed no significant change in p(switch) around a block transition (Figure 5e). Examining p(switch) following different trial outcomes revealed that animals decreased their p(switch) following a rewarded trial following AAV injection, however this effect was similar between ROSA26 and *Slc17a6* groups (Figure 5c). Furthermore, p(switch) following an unrewarded outcome was not altered between groups (Figure 5d). We also examined p(switch) following various trial histories and combinations of choices and outcomes but did not observe any differences between groups indicating that the ability of the animal to detect the outcome of the previous trial and respond on subsequent trials was not perturbed (Figure 5- figure supplement 1l). All other behavioral metrics (ITI duration, trial duration, p(reward), port bias, etc.) were unchanged between groups or when compared before and after AAV injection (Figure 5-figure supplement 1 and 2). Consistent with our other measures the mouse behavioral strategy as assessed by the RFLR model was also unperturbed between groups (Figure 5-figure supplement 1m). We conclude that permanent deletion of glutamate release from EP^{Sst+} neurons effectively converts this normally cotransmitting population into a GABAergic neuronal population (Figure 5-figure supplement 1j-k). This, however is not sufficient to cause detectable behavioral changes in animals that are well trained on the probabilistic switching task (Figure 5).

Discussion

Here we described a probabilistic switching task in which mice use their history of choices and rewards to update and guide future actions[11–15]. We show that mice can detect block transitions when reward probabilities alternate sides and respond by an increase in switching between reward (side) ports. Animals are also sensitive to the combination of reward probabilities set in a session (90/10 vs 70/30) and modify their strategy. When the probability of receiving a reward becomes more stochastic (70/30 condition) they incorporate evidence over more trials in the past to inform future actions. As rewards become more stochastic, mice also take more trials to recover stable selection of the high-reward port following a block transition, switching between ports less frequently (Figure 1). Using fiber photometry, we show that populations of EP^{Sst+} neurons are strongly engaged during this task during both the trial choice and outcome epochs (Figure 2). This observation was then reinforced by a GLM that showed significant decreases in model performance when information about choice direction or choice outcome was omitted, indicating these are important predictive behavioral variables to reconstruct the photometry signal (Figure 3). We then tested the necessity of EP^{Sst+} neurons for continued performance on the probabilistic switching task. We found that permanent genetic blockade of synaptic release from these neurons did not result in detectable changes in task performance metrics (Figure 4). Finally, we tested if modifying the ratio of GABA and glutamate cotransmitted by EP^{Sst+} neurons had an impact on continued task performance by genetically deleting vGluT2, thereby strongly decreasing the amount of glutamate released at synapses. This manipulation did not result in significant changes in task performance across control and vGluT2 deleted groups (Figure 5). Together these data suggest that despite observing ongoing, task related activity in EP^{Sst+} neurons, cotransmission from these neurons to LHb was not required for continued task performance in well-trained animals.

Probabilistic switching task and basal ganglia circuits

Dynamic, probabilistic switching tasks have been used by many groups to examine how an animal uses its past experience to make its next choice[12–15,27–29]. While we focused on EP, other studies show that distributed circuits throughout the cortex, striatum and midbrain guide animals to flexibly choose actions in pursuit of rewards on a trial-to-trial basis and have helped inform models of reinforcement learning[30]. Dopamine signaling in the striatum is critical for optimal performance on this task as it can causally guide future choice, and may also underlie motivation during longer periods by integrating reward rate[13]. Dopamine release has also been shown to signal reward prediction error in these tasks which takes into account the reward history and expectation an animal has trial-to-trial[13,14].

Postsynaptic to dopamine release sites in striatum, SPNs are key mediators of “action value” in these tasks. Furthermore, in a task similar to the one described here, unilateral optogenetic stimulation of D1R SPNs in the dorsal medial striatum (DMS) immediately following center entry biased future choice in the contralateral direction indicating a causative role for SPNs in guiding choice behavior in a probabilistic switching task[12].

Other studies have examined groups of neurons directly downstream of D1R SPNs that are suggested to be involved in evaluating action outcomes[5,11,31]. These EP neurons receive convergent input from both striosomal (patch) and matrix D1R SPNs distributed throughout the striatum and project exclusively to the LHb, a region critical for the processing of “negative reward” signals[7,11,32]. In head-fixed classical conditioning tasks LHb-projecting EP neurons have been shown to increase their activity to punishments or reward omission and decrease their activity to rewards[5,11]. Interestingly, in a probabilistic switching task optogenetic stimulation of vGluT2+ EP neurons following side port entry, but not center entry, biased future choices away from side port paired with stimulation indicating that these neurons carried an “anti-reward’ signal[11]. An important consideration with the aforementioned studies is that behavioral changes resulting from phasic modulation of EP to LHb inputs rely on the combined action of *Sst*+ and an additional subset of purely glutamatergic (*Pvalb*+/*Slc17a6*+) EP neurons as both populations are targeted in these studies[11,31] (but also see[33]). These studies and others suggest a prominent role for various basal ganglia nuclei in distinct phases of a probabilistic switching task engaging circuitry in both the action selection (choice) and action evaluation (reward) epochs of

a trial. Notably, the observed effects of causal optogenetic manipulations in both striatum and EP depend on both reward history and previous choice, variables shown to be critical for animal performance[15,34].

Activity patterns of EP^{Sst+} neurons

Here we show the activity patterns of genetically defined EP^{Sst+} neurons during freely moving behavior. Activity of EP^{Sst+} neurons is robustly modulated on a trial-by-trial basis in the probabilistic switching task by both the direction of the choice (ipsilateral vs contralateral) and the outcome (reward vs unrewarded) of a trial (Figure 2 and 3). In contrast to thalamic projecting EP neurons which receive striatal input exclusively from the SPNs in the matrix compartment, EP^{Sst+} neurons receive input from both limbic associated "striosomes" (patches) and sensorimotor associated "matrix" subdivisions which may contribute to activity during outcome and choice epochs, respectively[7,11]. Notably, phasic changes in activity during ipsilateral and contralateral movements resemble those observed in substantia nigra pars reticulata (another output nucleus of the basal ganglia) during eye saccade tasks (i.e. increased/persistent activity for ipsilateral movements and decreased for contralateral movements)[35,36]. Our findings are also consistent with electrophysiological studies of individual LHb-projecting GPi neurons in primates that respond to both reward-related cues and sensory cues related to the direction of a target during an eye saccade task[5]. Reward related responses we observe during the outcome evaluation period of the task are consistent with those reported elsewhere, with phasic excitation following unrewarded outcomes and inhibition following reward[5,11]. In contrast to other reports using single neuron electrophysiological recordings of LHb projecting EP neurons, we did not observe bidirectional reward prediction error (RPE) coding with our photometry measurements[5,11] (Figure 2—figure supplement 1). Instead we observed a bidirectional value signal indicating whether or not a reward occurred and a small effect of trial history on rewarded trials only (Figure 2—figure supplement 1f-g). RPE like responses have been observed by recording dopamine release in the striatum during a probabilistic switching task similar to the one we describe here[14], therefore these features may only be present in a subset of EP^{Sst+} neurons, or are below our detection threshold with photometry recordings.

Despite strong engagement of EP^{Sst+} neuronal activity during the task, we were surprised to find that neither complete blockade of synaptic release or modification of the ratio of GABA/glutamate cotransmitted was sufficient to alter performance of well-trained animals on the task. These behavioral results may point to additional parallel circuits or rapid homeostatic plasticity in LHb which compensates altered EP^{Sst+} output during the gradual expression of viral constructs. Alternatively, activity in EP^{Sst+} neurons (and subsequent cotransmission in LHb) may be required as the animal learns the structure of the probabilistic switching task, but then no longer required in well-trained animals that have learned the sequence of actions needed for high behavioral performance, akin to those described for motor cortex or subregions of striatum[37,38]. Finally, the stochastic nature of behavior in this task may require higher power for differentiating effects than available in this set of experiments. For example, ablating EP^{Sst+} neurons may have effects on very small subsets of trial types that we haven't characterized due to insufficient statistical power (i.e. switch trials).

Functions for the EP in the probabilistic switching task

A sequence of training steps is used to instruct an animal to perform the probabilistic switching task we describe here (see Methods). Once they learn the progression of a trial (i.e. poke in center to begin then poke side for reward) we introduce a "block structure" where the high reward port switches sides following a pre-determined number of rewarded trials progressively growing to 50 reward blocks with side ports delivering rewards at 90% (high port) and 10% (low port) of trials (Figure 1). Prior to any photometry recording or synaptic manipulation of EP, animals must reach a predetermined ("expert") criterion and consistently perform at this level for several days (at least 5 consecutive sessions). There are subtle changes in performance such as decreased trial duration (see Figure 5-figure supplement 1) once the animals reach criterion, but by-in-large their performance has plateaued and stabilized. Critically, even though well-trained animals perform consistently, they do not perform the task habitually (i.e. they are sensitive to devaluation and will not perform if

they are not thirsty, data not shown). Also, well-trained animals continue to evaluate the previous choice and trial outcome to inform future decisions, engaging in trial-and-error action updating. However, expert animals have clearly mastered the sequence of actions required to move between ports and consume rewards. Perhaps our behavioral results demonstrate that long-term manipulations of EP^{Sst+} neurons do not affect continued performance on the task because this circuit is required for earlier stages of learning. Importantly, our results do not indicate that these neurons have no role in shaping initial task acquisition, particularly while the animals learn the location of rewards and the action sequence required to acquire them. Future studies should examine the function of EP^{Sst+} neurons in learning action/outcome associations prior to the crystallization of action sequences that lead to reward. Conversely, different populations of EP neurons that are not examined here, such as the thalamic projecting *Pvalb+/Slc32a1* population[7], may play an critical role for executing learned action sequences as seen in studies using a forelimb lever pressing task[39].

EP cotransmission and influence on LHb activity patterns

Major questions remain regarding how neurons in the LHb integrate and interpret signals from GABA/glutamate cotransmitting inputs from EP^{Sst+} neurons. EP^{Sst+} input appears to target a subregion of lateral LHb comprising the lateral and oval subregions[7,40,41]. In vitro cell-attached recordings from LHb show that most neurons respond to optogenetic stimulation of EP input with increases in spiking, however these recordings were performed in conditions where channelrhodopsin was expressed in all LHb projecting EP neurons possibly leading to a bias towards excitation[24,25]. Studies examining mPSCs or using minimal optogenetic stimulation of EP^{Sst+} axons have demonstrated that individual release sites and/or synaptic vesicles can cotransmit GABA and glutamate[8,24]. Using targeted optogenetic stimulation of multiple distinct EP^{Sst+} inputs onto a single LHb neuron, we found that the amplitudes of the EPSC and IPSC were correlated within a cell, but the ratio varied between cells. This indicates that when exclusively examining EP^{Sst+} inputs an individual LHb neuron may be excited or inhibited depending on the ratio set by the postsynaptic receptor composition[8]. Additional experiments need to examine how this diversity translates to an *in vivo* setting where the postsynaptic membrane potential is not clamped and could respond differently to cotransmission. Recent studies have demonstrated rapid, behaviorally induced plasticity in individual LHb neurons following a stressful tail-shock protocol[42]. Remarkably, LHb neurons change the sign of their responses (negative going to positive going) following sucrose reward delivery following stress[42]. It is tempting to speculate that GABA/glutamate cotransmitting synapses undergo plasticity to control LHb output by modifying the ratio of GABA/glutamate cotransmission under these or other environmental/behavioral changes[9].

Materials and Methods

Mice

The following mouse strains/lines were used in this study: C57BL/6J (The Jackson Laboratory, Stock # 000664), *Sst*-IRES-Cre (The Jackson Laboratory, Stock # 013044), *Sst*-IRES-Flpo (The Jackson Laboratory, Stock # 031629) and Pvalb-2A-Flp (The Jackson Laboratory, Stock # 022730). Animals were kept on a 12:12 reverse light/dark cycle under standard housing conditions. All procedures were performed in accordance with protocols approved by the Harvard Standing Committee on Animal Care or the Boston University Institutional Animal Care and Use Committee following guidelines described in the U.S. National Institutes of Health Guide for the Care and Use of Laboratory Animals.

Adeno-Associated Viruses (AAVs)

Recombinant AAVs used for fiber photometry measurements (AAV1-Syn-FLEX-GCaMP6f, AAV8-CAG-FLEX-tdTomato (Addgene #100833 and #51503, respectively), and AAVrg-Ef1a-fDIO-Cre), tetanus toxin experiments (AAV8-Syn-FLEX-TeLC-P2A-GFP was a gift from Dr. Fan Wang, AAV8-Syn-DIO-EGFP (Addgene # 135391 and #50457, respectively)) and *Slc17a6* knockout experiments (AAV1-CMV-FLEX-

SaCas9-sgSlc17a6 (Addgene #124847), AAV1- CMV-FLEX-SaCas9-sgROSA26 gift from Dr Larry Zweifel and AAV8-Ef1a-DIO-oChief-tdTomato (Addgene# 51094)) were commercially obtained from the Boston Children's Hospital Viral Core or directly from Addgene. Virus aliquots were stored at -80 °C, and were injected at a concentration of approximately 10^{11} or 10^{12} GC/ml.

Stereotaxic Surgeries

Adult mice were anesthetized with isoflurane (5%) and placed in a small animal stereotaxic frame (David Kopf Instruments). After exposing the skull under aseptic conditions, viruses were injected through a pulled glass pipette at a rate of 50 nL/min using a UMP3 microsyringe pump (World Precision Instruments). Pipettes were slowly withdrawn (< 100 µm/s) at least 10 min after the end of the infusion. Following wound closure, mice were placed in a cage with a heating pad until their activity was recovered before returning to their home cage. Mice were given pre- and post-operative subcutaneous ketoprofen (10mg/kg/day) or meloxicam (5mg/kg) and buprenorphrine XR (3.25mg/kg) as an analgesic and monitored daily for at least 4 days post-surgery. For fiber photometry experiments 200 µm diameter fibers (0.37NA Doric Lenses) with a stainless-steel ferrule were implanted ~200um above the injection site following the injection and adhered to the skull with cyanoacrylate glue and dental cement (C&B Metabond). Injection coordinates from Bregma for EP were -1.1mm A/P, 2.1mm M/L, and 4.2mm D/V and for LHb were -1.55mm A/P, 0.5mm M/L, and -2.85mm D/V. Injection volumes for specific anatomical regions and virus types were as follows EP: 250 nL (mix of GCaMP6f and tdTom.), 200 nL (TeLC or GFP), 400 nL (1:1 mix of SaCas9-sgRNA and oChief-tdTom) or (1:1 mix of TeLC and oChief-tdTom), LHb: 200 nL (fDIO-Cre).

Immunohistochemistry

Mice were deeply anesthetized with isoflurane and perfused transcardially with 4% paraformaldehyde in 0.1 M sodium phosphate buffer. Brains were post-fixed overnight, sunk in 30% (wt/vol) sucrose in phosphate buffered saline (PBS) and sectioned (50 µm) coronally (Freezing Microtome, Leica). Free-floating sections were permeabilized/blocked with 5% normal goat serum in PBS with 0.2% Triton X-100 (PBST) for 1 h at room temperature and incubated with primary antibodies at 4°C overnight and with secondary antibodies for 1 h at room temperature in PBST supplemented with 5% normal goat serum. Brain sections were mounted on superfrost slides, dried and coverslipped with ProLong antifade reagent containing DAPI (Molecular Probes). Primary antibodies used include: chicken anti-GFP (1:1000, A10262 Invitrogen) and rabbit anti-mCherry (1:500, Ab167453 Abcam). Alexa Fluor 594- and 488-conjugated secondary antibodies to chicken and rabbit (Invitrogen) were diluted 1:500. Whole sections were imaged with an Olympus VS120/200 slide scanning microscope. Occasionally, images were linearly adjusted for brightness and contrast using ImageJ software. All images to be quantitatively compared underwent identical manipulations.

Behavior apparatus, training, and task

The apparatus used for the behavior is as described previously[14,15] with the following modifications. Clear acrylic barriers 5.5 cm in length were installed in between the center and side ports to extend the trial time to aid in better behaviorally resolved photometry recordings (these were not in place for other behavior experiments Figures 4 and 5). Water was delivered in ~3 µL increments. Hardware and software to control the behavior box is available online:
https://github.com/HMS-RIC/TwoArmedBandit and https://edspace.american.edu/openbehavior/project/2abt/.

Mice were water restricted 1.2 ml per day prior to training and maintained at >80% initial body weight for the full duration of training and photometry. All training sessions were conducted in the dark under red light conditions. During the task a blue LED above the center port signals to the mouse to initiate a trial by poking in the center port. Blue LEDs above the side ports are then activated, signaling the mouse to poke in the left or right side port within 8 seconds. Side port reward probabilities are defined by custom

software (MATLAB) and ranged from 10%-90% depending on the experiment. Withdrawal from the side port ends the trial and begins a 1 second intertrial interval (ITI). An expert mouse can perform 300-700 trials in a 40 min session.

To train the mice to proficiency, they were subjected to incremental training stages. Each training session lasts for ~40 minutes, adjusted according to the mouse's performance. Mice progress to the next stage once they were able to complete at least 100 successful trials with at least a 75% reward rate. On the first day, they were habituated to the behavior box, with water being delivered from both side ports and triggered only by a side port poke. In the next stage, mice learned the trial structure – only a poke in center port followed by a side port poke delivers water. Then, the mice transitioned to learning the block structure, in which 50 rewarded trials on one side port triggers the reward probabilities to switch (block transition) we began probabilistic reward delivery at this stage (p_{High}=90%, p_{Low}=10%). For photometry experiments, mice performed trials in the presence of barriers in between the center and side ports. A series of transparent barriers of increasing size (small (3 cm), medium (4 cm), and long (5.5 cm)) aided in learning. Finally, the mice were subjected to fiber implantation. Following fiber implant surgeries, mice were retrained to achieve the same pre-surgery performance level. Recordings were performed 4 weeks after surgery to allow for stable viral expression levels as well as a consistent and proficient level of task performance from the mice.

For experiments where we manipulated synaptic release in EP^{Sst+} neurons (Figures 4-5) we trained mice (reward probabilities 90/10, no transparent barrier present) to the following criteria for the 5 days prior to virus injection: **1**) *p*(highport) per session was greater than or equal to 0.80 with a variance less than 0.003, **2**) *p*(switch) per session was less than or equal to 0.15 with a variance less than 0.001, **3**) the *p*(left port) was between 0.45-0.55 with a variance less than 0.005, and **4**) the animal performed at least 200 trials in a session. The mean and variance for these measurements was calculated across the five session immediately preceding surgery. The criterion were determined by comparing performance profiles in separate animals and chosen based on when animals first showed stable and plateaued behavioral performance. Following surgery, mice were allowed to recover for 3 days and then continued to train for 3 weeks during viral expression. Data collected during the 5 day pre-surgery period was then compared to data collected for 10 sessions following the 3 weeks allotted for viral expression (i.e. days 22-31 post-surgery).

Behavioral analysis and modelling (Recursively Formulated Logistic Regression (RFLR))

Several behavioral metrics were used to characterize performance in this task and evaluate the predictive model (RFLR) used to capture these behavioral patterns. We examined the trial-to-trial dynamics around a block transition using **1)** the probability of choosing the highly rewarded port (p(high port)) and **2)** the probability of choosing two different ports on subsequent trials (p(switch)) as a function of trial position within a block. To quantify differences across mice in p(switch) and the time course of p(high port) following a block transition, we used single value metrics of p(switch) max and the time constant of p(high port), $tau_{p(high\ port)}$, respectively. Smaller $tau_{p(high\ port)}$ indicates a more rapid (i.e fewer trials) recovery of stable selection of the new highly rewarded port, and a larger p(switch) max. indicates greater sensitivity of the behavior to the block transition.

The behavior was also modeled with the purpose of systematically characterizing normal and perturbed patterns of behavior across treatment groups. The above behavioral features are well captured by a recursively formulated logistic regression model (RFLR)[15], which requires three interpretable parameters to recapitulate mouse behavior. Given successful predictive accuracy across experimental conditions, we can inspect how the model captures changes in mouse behavior that result from neural perturbations. The RFLR predicts future choice via a weighted combination of choice history bias (i.e., perseveration, α), and a latent representation of evidence that gets updated by new action-outcome information on every trial (β) and decays across trials (τ). Maximum likelihood parameter estimates were found using the stochastic gradient descent optimization algorithm. Fits for α, β, τ were presented for each of the experimental groups. experimental groups. Given

comparable performance of the model across experimental conditions, comparison of parameter fits provides a method of evaluating consistency in the structure of the behavioral strategy, as defined by three parameters: a relative influence of choice perseveration, current evidence, and previous evidence (i.e., history). All additional details regarding RFLR runs are available in Jupyter Notebooks online at: https://github.com/celiaberon/2ABT_behavior_models

Fiber photometry

Fiber implants on the mice were connected to a 0.37 NA patchcord (Doric Lenses, MFP_200/220/900-0.37_2m_FCM-MF1.25, low autofluorescence epoxy), attached to a filter cube (FMC5_E1(465-480)_F1(500-540) _E2(555-570)_F2(580-680)_S, Doric Lenses). Excitation light from LEDs (Thorlabs) and was amplitude modulated at 167 Hz (470 nm excitation light, M470F3, Thorlabs; LED driver LEDD1B, Thorlabs) and 223 Hz (565 nm excitation light, M565F3, Thorlabs, LED driver LEDD1B, Thorlabs). The following excitation light power measured at the end of the patch cord were used: 470nm=50μW, 565nm=20μW. Signals from the photodetectors were amplified in DC mode with Newport photodetectors (NPM_2151_FOA_FC) and received by a Labjack (T7) DAC streaming at 2000 samples/sec. The DAC also received synchronous information about behavior events logged from the Arduino which controls the behavior box. The following events were recorded: center port entry and exit, side port entry and exit, lick onset and offset, and LED light onset and offset.

Photometry Analysis

The frequency modulated signals were detrended using a rolling Z-score with a time window of 1 minute (12000 samples). As the ligand-dependent changes in fluorescence measured *in vivo* are small (few %) and the frequency modulation is large (~100%), the variance in the frequency modulated signal is largely ligand independent. In addition, the trial structure is rapid with an average inter-trial interval of < 3 sec. Thus, Z-scoring on a large time window eliminates photobleaching without affecting signal. Detrended, frequency modulated signals were frequency demodulated by calculating a spectrogram with 1 Hz steps centered on the signal carrier frequency using the MATLAB ‘spectrogram’ function. The spectrogram was calculated in windows of 216 samples with 108 sample overlap, corresponding to a final sampling period of 54 ms. The demodulated signal was calculated as the power averaged across an 8 Hz frequency band centered on the carrier frequency. No additional low-pass filtering was used beyond that introduced by the spectrogram windowing. For quantification of fluorescence transients as Z-scores, the demodulated signal was passed through an additional rolling Z-score (1 min window). To synchronize photometry recordings with behavior data, center port entry timestamps from the Arduino were aligned with the digital data stream indicating times of center-port entries. Based on this alignment, all other port and lick timings were aligned and used to calculate the trial-type averaged data shown in all figures. The Z-scored fluorescence signals were averaged across trials, sessions, and mice with no additional data normalization. Statistical comparisons were made by measuring the mean z-scored fluorescence signal across a 500ms window immediately following a given behavioral event (CE, SE, SX, …) for all trials per mouse (n=6 mice).

Generalized linear model

Photometry recordings and behavioral data used for the GLM analysis (Figure 3) were collected from *Sst-Flp* mice as indicated with 6 sessions per mouse and ~ 500 trials/session. These data were aligned to behavioral events to create a predictive matrix X (of dimensions $N \times F$) and a response vector, $\mathbf{y}$ (of dimension N), where N is the number of samples recorded in a session and F is the number of behavioral “predictors” in the analysis. The predictors consisted of values 0 and 1 to indicate if a behavioral event (for example a center port entry) occurred in the time bin.

For each predictive matrix, a design matrix $\varphi(X)$ (of dimensions $N \times F\,(2T + 1)$) was constructed from T time shifts forward and backward (T = 20, 54 ms each) for each feature, allowing the GLM to fit coefficients that corresponded to time-based kernels for each of the predictive features in X. Data from the ITI period, in

which there are no task-relevant behavioral events, were excluded, and only data spanning shortly before center entry and after side-port exit were modelled. When initial and final time shifts spanned the boundary between two trials, the overlapped data were included twice (once in each of the trials on either side of the boundary) to ensure sufficient representation of each event in training and test datasets.

To evaluate the performance of the GLMs and determine which model and hyperparameter set was best, we performed a grid search across elastic net, ordinary least squares, and ridge regressions. For each model run, a 10-fold group shuffle split (GSS) by trial was applied to the training set to obtain cross-validated ranges for the MSEs, based on an 80–20 training/test split within each of the 10 GSS folds. Ridge regression (α=1) was determined to be the best model based on the lowest and least variable MSE score (0.80, SD=0.001). We then tested the effect of omitting behavioral variables on the GLM performance (Figure 3e) and re-fit the GLM with 5-fold GSS to obtain cross-validated ranges for the MSE values used in the box plots. For the model chosen (Ridge Regression), the algorithm minimizes an associated cost function with respect to the fitted coefficients as follows, where ***J*** is the cost function to be minimized, ***X*** is the design matrix (set of time-shifted behavioral events), ***y*** is the response vector (GCaMP6f), ***β*** is the set of fitted coefficients, $\|a\|^2_2$ is the sum of the squared entries in vector a, and ***α*** is the regularization parameter.

$$J(X, y) = \|y - X\beta\|_2^2 + \alpha \|\beta\|_2^2$$

All additional details regarding GLM runs are available in Jupyter Notebooks online at: https://github.com/mwall2017/sabatini-glm-workflow

Acute brain slice preparation

Brain slices were obtained from 50-150 day old mice (both male and female) using standard techniques. Mice were anesthetized by isoflurane inhalation and perfused transcardially with ice-cold artificial cerebrospinal fluid (ACSF) containing (in mM) 125 NaCl, 2.5 KCl, 25 $NaHCO_3$, 2 $CaCl_2$, 1 $MgCl_2$, 1.25 NaH_2PO_4 and 25 glucose (295 mOsm/kg). Cerebral hemispheres were removed, blocked and transferred into a slicing chamber containing ice-cold ACSF. Coronal slices of LHb (250 µm thick) were cut with a Leica VT1000s/VT1200s vibratome in ice-cold ACSF, transferred for 10 min to a holding chamber containing choline-based solution (consisting of (in mM): 110 choline chloride, 25 $NaHCO_3$, 2.5 KCl, 7 $MgCl_2$, 0.5 $CaCl_2$, 1.25 NaH_2PO_4, 25 glucose, 11.6 ascorbic acid, and 3.1 pyruvic acid) at 34°C then transferred to a secondary holding chamber containing ACSF at 34°C for 10 mins and subsequently maintained at room temperature (20–22°C) until use. All recordings were obtained within 4 hours of slicing. Both choline solution and ACSF were constantly bubbled with 95% O2/5% CO2.

Electrophysiology

Individual slices were transferred to a recording chamber mounted on an upright microscope and continuously superfused (4 ml/min) with room temperature ACSF. Cells were visualized through a 60X or 40X water immersion objective with infrared differential interference and epifluorescence to identify regions displaying the highest density of ChR2+ axons. Epifluorescence was attenuated and used sparingly to minimize ChR2 activation prior to recording. Patch pipettes (2–4 MΩ) pulled from borosilicate glass (Sutter Instruments) were filled with an internal solution containing (in mM) 135 $CsMeSO_3$, 10 HEPES, 1 EGTA, 3.3 QX-314 (Cl^- salt), 4 Mg-ATP, 0.3 Na-GTP, 8 Na2-Phosphocreatine (pH 7.3 adjusted with CsOH; 295 mOsm/kg) for voltage-clamp recordings. Membrane currents were amplified and low-pass filtered at 3 kHz using a Multiclamp 700B amplifier (Molecular Devices, Sunnyvale, CA), digitized at 10 kHz and acquired using National Instruments acquisition boards and a custom version of ScanImage [43] (available upon request or from https://openwiki.janelia.org/wiki/display/ephus/ScanImage) written in MATLAB (Mathworks, Natick, MA) or PClamp 11 (Molecular Devices). Electrophysiology data were analyzed offline in MATLAB and Clampfit. The approximate location of the recorded neuron was confirmed after termination of the recording using a 4X

objective to visualize the pipette tip, while referencing an anatomical atlas (Allen Institute Reference Atlas). For analyses in Figure S4-S5, the peak amplitude of PSCs measured were averaged across at least 10 trials. To activate oChief-expressing cells and axons, light from a 473 nm laser (Optoengine) was focused on the back aperture of the microscope objective to produce wide-field illumination of the recorded cell. For voltage clamp experiments, brief pulses of light (1 ms duration; 10 $mW \cdot mm^{-2}$ under the objective) were delivered at the recording site at 20 s intervals under control of the acquisition software.

Statistics

In Figure 1 and Figure 1 –figure supplement 1 we used one-way ANOVA with Tukey's post hoc test for multiple comparisons (p-values are designated as: *P<0.05, **P<0.01, ***P<0.001). In Figure 2 and Figure 2-figure supplement 1 we used paired t-tests for 2 groups or repeated measures ANOVA with Tukey's post hoc test for multiple comparisons for three groups. In Figure 4-5 and Figure 4-5 –figure supplements we used a two-way ANOVA with a Sidak's post hoc for multiple comparisons, when comparing before and after AAV injection and between control and Tettx groups. Students unpaired t-test was used for comparisons of EPSC/IPSC amplitudes.

Acknowledgements

The authors thank Emily Kraft and Julia Williams for assistance in behavioral training, James Levasseur for animal husbandry and genotyping, and Lillian Worth for administrative assistance. We thank Shay Neufeld for initial task development, box design, and behavioral analysis. We thank Jeffrey Markowitz for assistance developing the fiber photometry system and the members of the Sabatini and Wallace labs for helpful discussions and advice. The HMS Research Instrumentation Core (Ofer Mazor and Pavel Gorelik) were essential in the development of the behavioral boxes, PCB fabrication, and design of Ardunio/Matlab code. This work was supported by the Brain Behavior Research Foundation, the Whitehall foundation, NINDS R00NS105883, and NIMH R01MH133608, M.L.W. as well as the Howard Hughes Medical Institute and NINDS R01NS103226, B.L.S.

Additional Information

Author contributions

Julianna Locantore, Investigation, Methodology, Data Curation; Yijun Liu, Investigation, Methodology, Data Curation, Formal analysis; Jesse White, Investigation, Data Curation; Janet Berrios, Investigation, Data curation, Software; Celia C. Beron, Software; Bernardo L. Sabatini, Conceptualization, Resources, Supervision, Project Administration, Software; Michael L. Wallace, Investigation, Methodology, Data Curation Formal Analysis, Conceptualization, Resources, Supervision, Project Administration.

Ethics

The authors declare no financial or non-financial competing interests. All procedures were performed in accordance with protocols approved by the Harvard Standing Committee on Animal Care or the Boston University Institutional Animal Care and Use Committee following guidelines described in the U.S. National Institutes of Health Guide for the Care and Use of Laboratory Animals (HMS IACUC protocol #IS00000571; BU IACUC protocol #PROTO202100002). All surgery performed under isoflurane anesthesia.

References:

1. Graybiel, A.M., Aosaki, T., Flaherty, A.W., and Kimura, M. (1994). The basal ganglia and adaptive motor

control. Science (80-.). *265*.

2. Hyman, S.E., Malenka, R.C., and Nestler, E.J. (2006). Neural mechanisms of addiction: the role of reward-related learning and memory. Annu. Rev. Neurosci. *29*, 565–598. 10.1146/annurev.neuro.29.051605.113009.

3. Nelson, A.B., and Kreitzer, A.C. (2014). Reassessing models of Basal Ganglia function and dysfunction. Annu. Rev. Neurosci. *37*, 117–135. 10.1146/annurev-neuro-071013-013916.

4. Gerfen, C.R. (1992). The neostriatal mosaic: multiple levels of compartmental organization in the basal ganglia. Annu. Rev. Neurosci. *15*, 285–320. 10.1146/annurev.ne.15.030192.001441.

5. Hong, S., and Hikosaka, O. (2008). The globus pallidus sends reward-related signals to the lateral habenula. Neuron *60*, 720–729. 10.1016/j.neuron.2008.09.035.

6. Parent, A., and De Bellefeuille, L. (1982). Organization of efferent projections from the internal segment of globus pallidus in primate as revealed by fluorescence retrograde labeling method. Brain Res. *245*, 201–213.

7. Wallace, M.L., Saunders, A., Huang, K.W., Philson, A.C., Goldman, M., Macosko, E.Z., McCarroll, S.A., and Sabatini, B.L. (2017). Genetically Distinct Parallel Pathways in the Entopeduncular Nucleus for Limbic and Sensorimotor Output of the Basal Ganglia. Neuron *94*, 138--152.e5. 10.1016/j.neuron.2017.03.017.

8. Kim, S.A., Wallace, M.L., El-Rifai, M., Knudsen, A.R., and Sabatini, B.L. (2022). Co-packaging of opposing neurotransmitters in individual synaptic vesicles in the central nervous system. Neuron *110*, 1371-1384.e7. 10.1016/J.NEURON.2022.01.007.

9. Wallace, M.L., and Sabatini, B.L. (2023). Synaptic and circuit functions of multitransmitter neurons in the mammalian brain. Neuron *111*, 2969–2983. 10.1016/J.NEURON.2023.06.003.

10. Yao, Z., van Velthoven, C.T.J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., et al. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nat. 2023 6247991 *624*, 317–332. 10.1038/s41586-023-06812-z.

11. Stephenson-Jones, M., Yu, K., Ahrens, S., Tucciarone, J.M., van Huijstee, A.N., Mejia, L.A., Penzo, M.A., Tai, L.-H., Wilbrecht, L., and Li, B. (2016). A basal ganglia circuit for evaluating action outcomes. Nature *539*, 289–293. 10.1038/nature19845.

12. Tai, L.-H., Lee, A.M., Benavidez, N., Bonci, A., and Wilbrecht, L. (2012). Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. Nat. Neurosci. *15*, 1281–1289. 10.1038/nn.3188.

13. Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., Kennedy, R.T., Aragona, B.J., and Berke, J.D. (2016). Mesolimbic dopamine signals the value of work. Nat. Neurosci. *19*, 117–126. 10.1038/nn.4173.

14. Chantranupong, L., Beron, C.C., Zimmer, J.A., Wen, M.J., Wang, W., and Sabatini, B.L. (2023). Dopamine and glutamate regulate striatal acetylcholine in decision-making. Nat. 2023 6217979 *621*, 577–585. 10.1038/s41586-023-06492-9.

15. Beron, C.C., Neufeld, S.Q., Linderman, S.W., and Sabatini, B.L. (2022). Mice exhibit stochastic and efficient action switching during probabilistic decision making. Proc. Natl. Acad. Sci. U. S. A. *119*, e2113961119. 10.1073/PNAS.2113961119/SUPPL_FILE/PNAS.2113961119.SAPP.PDF.

16. Park, I.M., Meister, M.L.R., Huk, A.C., and Pillow, J.W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. Nat. Neurosci. *17*, 1395–1403. 10.1038/nn.3800.

17. Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H.J., Ornelas, S., Koay, S.A., Thiberge, S.Y., Daw, N.D., Tank, D.W., et al. (2019). Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. Nature *570*, 509–513. 10.1038/s41586-019-1261-9.

18. Driscoll, L.N., Pettit, N.L., Minderer, M., Chettih, S.N., and Harvey, C.D. (2017). Dynamic Reorganization

of Neuronal Activity Patterns in Parietal Cortex. Cell *170*, 986-999.e16. 10.1016/J.CELL.2017.07.021.

19. Proulx, C.D., Hikosaka, O., and Malinow, R. (2014). Reward processing by the lateral habenula in normal and depressive behaviors. Nat. Neurosci. *17*, 1146–1152. 10.1038/nn.3779.

20. Wang, D., Li, Y., Feng, Q., Guo, Q., Zhou, J., and Luo, M. (2017). Learning shapes the aversion and reward responses of lateral habenula neurons. Elife *6*. 10.7554/eLife.23045.

21. Zhang, Y., Zhao, S., Rodriguez, E., Takatoh, J., Han, B.-X., Zhou, X., and Wang, F. (2015). Identifying local and descending inputs for primary sensory neurons. J. Clin. Invest. *125*, 3782–3794. 10.1172/JCI81156.

22. Miyamoto, Y., and Fukuda, T. (2015). Immunohistochemical study on the neuronal diversity and three-dimensional organization of the mouse entopeduncular nucleus. Neurosci. Res. 10.1016/j.neures.2015.02.006.

23. Miyamoto, Y., and Fukuda, T. (2021). The habenula-targeting neurons in the mouse entopeduncular nucleus contain not only somatostatin-positive neurons but also nitric oxide synthase-positive neurons. Brain Struct. Funct. *226*, 1497–1510. 10.1007/S00429-021-02264-1/FIGURES/10.

24. Shabel, S.J., Proulx, C.D., Piriz, J., and Malinow, R. (2014). GABA/glutamate co-release controls habenula output and is modified by antidepressant treatment. Science (80-.). *345*, 1494–1498. 10.1126/science.1250469.

25. Meye, F.J., Soiza-Reilly, M., Smit, T., Diana, M.A., Schwarz, M.K., and Mameli, M. (2016). Shifted pallidal co-release of GABA and glutamate in habenula drives cocaine withdrawal and relapse. Nat. Neurosci. *19*, 1019–1024. 10.1038/nn.4334.

26. Hunker, A.C., Soden, M.E., Krayushkina, D., Heymann, G., Awatramani, R., and Zweifel, L.S. (2020). Conditional Single Vector CRISPR/SaCas9 Viruses for Efficient Mutagenesis in the Adult Mouse Nervous System. Cell Rep. *30*, 4303-4316.e6. 10.1016/J.CELREP.2020.02.092.

27. Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of Action-Specific Reward Values in the Striatum. Science (80-.). *310*, 1337–1340. 10.1126/science.1115270.

28. Parker, N.F., Cameron, C.M., Taliaferro, J.P., Lee, J., Choi, J.Y., Davidson, T.J., Daw, N.D., and Witten, I.B. (2016). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat. Neurosci. 2016 196 *19*, 845–854. 10.1038/nn.4287.

29. Bari, B.A., Grossman, C.D., Lubin, E.E., Rajagopalan, A.E., Cressy, J.I., and Cohen, J.Y. (2019). Stable Representations of Decision Variables for Flexible Behavior. Neuron *103*, 922-933.e7. 10.1016/J.NEURON.2019.06.001.

30. Cox, J., and Witten, I.B. (2019). Striatal circuits for reward learning and decision-making. Nat. Rev. Neurosci. *20*, 482–494. 10.1038/S41583-019-0189-2.

31. Shabel, S.J., Proulx, C.D., Trias, A., Murphy, R.T., and Malinow, R. (2012). Input to the lateral habenula from the basal ganglia is excitatory, aversive, and suppressed by serotonin. Neuron *74*, 475–481. 10.1016/j.neuron.2012.02.037.

32. Matsumoto, M., and Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. Nature *447*, 1111–1115. 10.1038/nature05860.

33. Lazaridis, I., Tzortzi, O., Weglage, M., Märtin, A., Xuan, Y., Parent, M., Johansson, Y., Fuzik, J., Fürth, D., Fenno, L.E., et al. (2019). A hypothalamus-habenula circuit controls aversion. Mol. Psychiatry 2019 249 *24*, 1351–1368. 10.1038/s41380-019-0369-5.

34. Bolkan, S.S., Stone, I.R., Pinto, L., Ashwood, Z.C., Iravedra Garcia, J.M., Herman, A.L., Singh, P., Bandi, A., Cox, J., Zimmerman, C.A., et al. (2022). Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. Nat. Neurosci. 2022 253 *25*, 345–357. 10.1038/s41593-022-01021-9.

35. Hikosaka, O., Takikawa, Y., and Kawagoe, R. (2000). Role of the basal ganglia in the control of

purposive saccadic eye movements. Physiol. Rev. *80*, 953–978. 10.1152/PHYSREV.2000.80.3.953/ASSET/IMAGES/LARGE/9J0300083007.JPEG.

36. Hikosaka, O., and Wurtz, R.H. (1985). Modification of saccadic eye movements by GABA-related substances. II. Effects of muscimol in monkey substantia nigra pars reticulata. https://doi.org/10.1152/jn.1985.53.1.292 *53*, 292–308. 10.1152/JN.1985.53.1.292.

37. Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A.L., Dhawale, A.K., Kampff, A.R., and Ölveczky, B.P. (2015). Motor cortex is required for learning but not for executing a motor skill. Neuron *86*, 800–812. 10.1016/J.NEURON.2015.03.024.

38. Reinhold, K., Iadarola, M., Tang, S., Kuwamoto, W., Sun, S., Hakim, R., Zimmer, J., Wang, W., and Sabatini, B.L. (2023). Striatum supports fast learning but not memory recall. bioRxiv, 2023.11.08.566333. 10.1101/2023.11.08.566333.

39. Dhawale, A.K., Wolff, S.B.E., Ko, R., and Ölveczky, B.P. (2021). The basal ganglia control the detailed kinematics of learned motor skills. Nat. Neurosci. 2021 249 *24*, 1256–1269. 10.1038/s41593-021-00889-3.

40. Wallace, M.L., Huang, K.W., Hochbaum, D., Hyun, M., Radeljic, G., and Sabatini, B.L. (2020). Anatomical and single-cell transcriptional profiling of the murine habenular complex. Elife *9*. 10.7554/eLife.51271.

41. Andres, K.H., During, M. Von, and Veh, R.W. (1999). Subnuclear organization of the rat habenular complexes. J. Comp. Neurol. *407*, 130–150. 10.1002/(SICI)1096-9861(19990428)407:1<130::AID-CNE10>3.0.CO;2-8.

42. Shabel, S.J., Wang, C., Monk, B., Aronson, S., and Malinow, R. (2019). Stress transforms lateral habenula reward responses into punishment signals. Proc. Natl. Acad. Sci. U. S. A. *116*, 12488–12493. 10.1073/PNAS.1903334116/SUPPL_FILE/PNAS.1903334116.SAPP.PDF.

43. Pologruto, T.A., Sabatini, B.L., and Svoboda, K. (2003). ScanImage: flexible software for operating laser scanning microscopes. Biomed. Eng. Online *2*, 13. 10.1186/1475-925X-2-13.

Figure Legends

Figure 1: Mice alter choices to changing reward probabilities in a probabilistic switching task
(A) Illustration of the animal movements and epochs (Trial Start, Choice, and Evaluation) of a single trial in the probabilistic two-port choice task. **(B)** A sample of a behavioral session showing periods when the highly rewarded port is on the left (white) and when it switches to the right (gray). The reward probabilities switch (dotted vertical lines) once 50 rewards are gained by the animal. Rewarded trials are represented by black circles and unrewarded trials are red circles, reward probabilities are 70/30. **(C)** Probability distributions of different behavioral events during rewarded (top) and unrewarded (bottom) trials to illustrate the timing of different events within a trial (one session (~500 trials), 90/10 rew. prob). CE=Center Entry, CX=Center Exit, SE=Side Entry, FL=First Lick, SX=Side Exit. **(D)** The probability of choosing the highly rewarded port (*p*(high port)) around a block transition (dotted vertical line) for different reward probabilities (black line = mean, shaded area=SEM). (Inset) Bar plot showing $tau_{p(high\ port)}$ (time constant) calculated from an exponential fit to the first 20 trials following a block transition for each animal (circles) (bar = mean, error bar= SEM) in the different reward probabilities. **(E)** The probability of choosing different side ports on consecutive trials (*p*(switch)) around a block transition (dotted vertical line) for different reward probabilities (black line = mean, shaded area=SEM). (Inset) Bar plot showing the maximum *p*(switch) in the 20 trials that follow a block transition for each animal (circles) (bar = mean, error bar= SEM) in the different reward probabilities. **(F)** The probability of choosing the highly rewarded port on all trials across reward probabilities (bar = mean, error bar= SEM). **(G)** The probability that a trial results in a reward across reward probabilities (bar = mean, error bar= SEM). **(H)** *p*(switch) across all trials for different reward probabilities (bar = mean, error bar= SEM). **(I)**

p(switch) for trials following a rewarded trial for different reward probabilities, percentages in bars represent the proportion of rewarded trials for each condition, also shown in **(G)** (bar = mean, error bar= SEM). **(J)** *p*(switch) for trials following an unrewarded trial for different reward probabilities, percentages in bars represent the proportion of unrewarded trials for each condition (bar = mean, error bar= SEM). For **D-J** n=9 mice, 8-10 sessions/mouse/rew. prob, ~550 trials/session.

Figure 2: Neural activity in EP^{Sst+} neurons encodes both choice and value.
(A) Viral injection location for specific infection of EP^{Sst+} neurons with GCaMP6f in a *Sst-Flp* mouse line and fiber implant location for photometry recording. **(B)** Fiber photometry recording of EP^{Sst+} neurons for individual trials during a behavioral session. Trials are aligned to center port entry (CE) and red dots indicated side port entry (SE). Only trials to the ipsilateral side (relative to the photometry recording) are shown and are divided by rewarded (left) and unrewarded (right) trials. **(C)** Averaged (±SEM) photometry signals across all mice aligned to center port entry (CE, top) or side port entry (SE, bottom) grouped by ipsilateral (green) and contralateral (magenta) choice (all sessions are at 90/10 reward probability, n=6 mice, 49 sessions, 20,355 trials). (right) Points represent the mean z-scored fluorescence per animal for the 500ms period immediately following the behavioral event, bars represent mean across animals ±SEM (all sessions are at 90/10 reward probability, n=6 mice, 49 sessions, 20,355 trials). **(D)** Averaged photometry (±SEM) signals across all mice aligned to side port entry (SE) grouped by rewarded (blue) or unrewarded (red) outcomes and divided by ipsilateral choice (top) or contralateral choice (bottom). (right) Points represent the mean z-scored fluorescence per animal for the 500ms period immediately following the behavioral event, bars represent mean across animals ±SEM (all sessions are at 90/10 reward probability, n=6 mice, 49 sessions, 20,355 trials). **(E)** Averaged (±SEM) photometry signals across different reward probabilities aligned to center port entry (CE) and divided by ipsilateral (top) and contralateral (bottom) choice. (right) Points represent the mean z-scored fluorescence per animal for the 500ms period immediately following the behavioral event, bars represent mean across animals ±SEM (90/10: n=6 mice, 49 sessions, 20,355 trials, 80/20: n=6 mice, 54 sessions, 27,433 trials, 70/30: n=6 mice, 49 sessions, 27,174 trials). **(F)** Averaged (±SEM) photometry signals across different reward probabilities aligned to side port entry (SE) and divided by rewarded (top) and unrewarded (bottom) outcomes. (right) Points represent the mean z-scored fluorescence per animal for the 500ms period immediately following the behavioral event, bars represent mean across animals ±SEM (90/10: n=6 mice, 49 sessions, 20,355 trials, 80/20: n=6 mice, 54 sessions, 27,433 trials, 70/30: n=6 mice, 49 sessions, 27,174 trials).

Figure 3: Generalized Linear Model of EP^{Sst+} neural activity during behavior.
(A) GLM workflow: behavioral variables are convolved with their kernels. Each time shift in the kernel consists of an independent β coefficient fit jointly by minimizing a cost function. The convolved signals are then summed to generate a reconstructed signal which can be directly compared to the original photometry trace. **(B)** The original dataset is divided into training and test datasets. The GLM is fit on the training data and evaluated on the test data using mean squared error (MSE). Following a grid search that compared multiple regularization types (ridge, elastic net, ordinary least squared) in combination with a large hyperparameter space, ridge regression (α=1) was found to give the smallest error following cross-validation. **(C)** Kernels for the behavioral variables included as features in the GLM. Behavioral predictors gave information regarding choice (Ipsi/Contra), reward and port entry and exit. **(D)** Average original (black) and reconstructed (green) photometry signals across trials aligned to behavioral events (solid line = mean, shaded area=SEM, R^2=0.19 SD=0.001, n=6 mice). **(E)** Box plots showing MSE for the full model (All) and models in which the indicated behavioral predictor(s) were omitted(-) for both the train (gray) and test (blue) datasets (Boxes represent the three quartiles (25%, 50%, and 75%) of the data and whiskers are 1.5*IQR, outliers are shown as dots, each model-run uses a different combination of data used for train/test split as illustrated in **B**).

Figure 4: Effects of permanent genetic silencing of synaptic release from EP^{Sst+} neurons on continued performance of a two-port choice probabilistic switching task.
(A) Viral injection location resulting in Cre-dependent expression of GFP (control) or tetanus toxin in EP^{Sst+}

neurons (green = Tettx-GFP, gray =DAPI). **(B)** The probability of choosing the highly rewarded port (*p*(high port)) around a block transition (dotted vertical line) for GFP (control, left) or Tettx (right) injected mice (gray= 5 days prior to AAV injection, green/red= days 21-30 post injection; black line = mean, shaded area=SEM). (Insets) Bar plot showing $tau_{p(high\ port)}$ (time constant) calculated from an exponential fit to the first 20 trials following a block transition for each animal (circles) (bar = mean, error bar= SEM) before and after AAV injection. **(C)** *p*(switch) for trials following a rewarded trial for GFP (green) and Tettx (red) injected animals (bar = mean, error bar= 95% CI). **(D)** *p*(switch) for trials following an unrewarded trial for GFP (green) and Tettx (red) injected animals (bar = mean, error bar= 95% CI). **(E)** The probability of choosing different side ports on consecutive trials (*p*(switch)) around a block transition (dotted vertical line) for GFP (control, left) or Tettx (right) injected mice (gray= 5 days prior to AAV injection, green/red= days 21-30 post injection; black line = mean, shaded area=SEM). (Insets) Bar plot showing the maximum *p*(switch) in the 20 trials that follow a block transition for each animal (circles) (bar = mean, error bar= SEM) before and after AAV injection. For **B-E** n=6 GFP control and n=6 Tettx mice, 5 sessions/mouse before AAV inj. and 10 session/mouse after AAV injection, GFP control= 15,120 trials before, 34,523 trials after; Tettx = 17,528 trials before, 32,761 trials after.

Figure 5: Effects of CRISPR Cas9 deletion of synaptic glutamate release from EP^{Sst+} neurons on continued performance of a two-port choice probabilistic switching task.

(A) Viral injection location resulting in Cre-dependent expression of oChief-tdTom+SaCas9-sgRNA for ROSA26 (control) or *Slc17a6* (vGlut2) in EP^{Sst+} neurons (red = tdTomato, gray =DAPI). **(B)** The probability of choosing the highly rewarded port (*p*(high port)) around a block transition (dotted vertical line) for sgROSA (control, left) or sg*Slc17a6* (right) injected mice (gray= 5 days prior to AAV injection, blue/orange= days 21-30 post injection; black line = mean, shaded area=SEM). (Insets) Bar plot showing $tau_{p(high\ port)}$ (time constant) calculated from an exponential fit to the first 20 trials following a block transition for each animal (circles) (bar = mean, error bar= SEM) before and after AAV injection. **(C)** *p*(switch) for trials following a rewarded trial for sgROSA26 (blue) and sg*Slc17a6* (orange) injected animals (bar = mean, error bar= 95% CI). **(D)** *p*(switch) for trials following an unrewarded trial for sgROSA26 (blue) and *sgSlc17a6* (orange) injected animals (bar = mean, error bar= 95% CI). **(E)** The probability of choosing different side ports on consecutive trials (*p*(switch)) around a block transition (dotted vertical line) for sgROSA26 (control, left) or *sgSlc17a6* (right) injected mice (gray= 5 days prior to AAV injection, green/red= days 21-30 post injection; black line = mean, shaded area=SEM). (Insets) Bar plot showing the maximum *p*(switch) in the 20 trials that follow a block transition for each animal (circles) (bar = mean, error bar= SEM) before and after AAV injection. For **B-E** n=10 sgROSA26 control and n=8 sg*Slc17a6* mice, 5 sessions/mouse before AAV inj. and 10 session/mouse after AAV injection, sgROSA26 control= 17,318 trials before, 39,710 trials after; sg*Slc17a6* = 13,520 trials before, 29,256 trials after.

Figure 1-figure supplement 1: *P*(switch) across different trial histories, additional behavioral metrics, and behavioral modeling using a recursively formulated logistic regression.
(A) The probability of choosing the highly rewarded port (*p*(high port), 90/10 rew. prob.) around a block transition (dotted vertical line) for individual mice (gray lines) and mean (black line = mean, shaded area=SEM). **(B)** The probability of choosing different side ports on consecutive trials (*p*(switch)) around a block transition (dotted vertical line) for individual mice (gray lines) and mean (black line = mean, shaded area=SEM). **(C)** *P*(switch) for left and right choices for an individual animal. Each dot represents a trial type (n=30 trial types) with a different history of choices and rewards for three trials prior. **(D)** Nomenclature for describing trial types with different reward histories (capital vs lowercase) and choice directions (right vs left). As animals show roughly symmetric *p*(switch) for left and right choices (See (**C**)) those trial types have been collapsed. **(E)** *p*(switch) for trial types segregated by reward history and choice direction across different reward probabilities, percentages above bars refer to the percentage of trials in each category for the different reward probabilities (bar = mean, error bar = 95% CI). **(F)** (left to right) Inter-trial interval, choice bias, and trial duration across different reward probabilities (bar = mean, error bar = 95% CI). **(G)** Violin plot showing the distribution of the number of trials completed in a ~40 min session for different reward probabilities (dots =

individual session, horizontal bar=median). **(H)** Trial duration following previously rewarded or unrewarded trials across different reward probabilities for "repeat" choices only (bar = mean, error bar = 95% CI). **(I)** Histogram showing the distribution of block lengths (number of trials prior to a block transition) for different reward probabilities. **(J)** The Recursively Formulated Logistic Regression (RFLR) model, which calculates the log odds of the mouse's next choice (Ψ_{t+1}) given its most recent choice (c_t) and a series of prior choices and rewards. c_t represents choice, r_t represents reward outcome on trial (t), relative to current trial i=0. α (alpha) is the weight of the most recent choice, β (beta) is the weight on the choice and reward outcome which decays exponentially across trials at a rate of τ (tau). **(K)** Summary of RFLR model coefficients across reward probabilities (coefficients highlighted in yellow in **(J)**), each dot represents an individual mouse (bar = mean, error bar = SEM, negative log-likelihood of fits were equivalent across reward probabilities; 90/10= -0.25 SD=0.03; 80/20= -0.24 SD=0.03; 70/30=-0.25 SD=0.02). **(L)** Exponential decay of choice and reward evidence (beta) for 8 trials in the past. Exponential fits made beta and tau coefficients observed for different reward probabilities and shown in **(K)**.

Figure 1-figure supplement 2: Model (RFLR) predictions for *p*(high port) and *p*(switch) around a block transition for different reward probabilities. (A) (top) The probability of choosing the highly rewarded port (*p*(high port), 90/10 rew. prob.) around a block transition (dotted vertical line). RFLR model predictions (grey) trained on 70% of the 90/10 trials and compared to the remaining 30% of trials (red), (black line = mean, shaded area=SEM). (Bottom) The probability of choosing different side ports on consecutive trials (p(switch)) around a block transition (dotted vertical line). RFLR model predictions (grey) trained on 70% of the 90/10 trials and compared to the remaining 30% of trials (red), (black line = mean, shaded area=SEM). **(B)** (top) The probability of choosing the highly rewarded port (*p*(high port), 80/20 rew. prob.) around a block transition (dotted vertical line). RFLR model predictions (grey) trained on 70% of the 80/20 trials and compared to the remaining 30% of trials (blue), (black line = mean, shaded area=SEM). (Bottom) The probability of choosing different side ports on consecutive trials (p(switch)) around a block transition (dotted vertical line). RFLR model predictions (grey) trained on 70% of the 80/20 trials and compared to the remaining 30% of trials (blue), (black line = mean, shaded area=SEM). **(C)** (top) The probability of choosing the highly rewarded port (*p*(high port), 70/30 rew. prob.) around a block transition (dotted vertical line). RFLR model predictions (grey) trained on 70% of the 70/30 trials and compared to the remaining 30% of trials (green), (black line = mean, shaded area=SEM). (Bottom) The probability of choosing different side ports on consecutive trials (p(switch)) around a block transition (dotted vertical line). RFLR model predictions (grey) trained on 70% of the 70/30 trials and compared to the remaining 30% of trials (green), (black line = mean, shaded area=SEM).

Figure 2-figure supplement 1: Alignment of photometry signals from EP^{Sst+} neurons to different behavioral events and comparisons accounting for reward history on 90/10 reward probability.
(A) Control fiber photometry recording of EP^{Sst+} neurons expressing static fluorophore tdTomato for individual trials during a behavioral session. Trials are aligned to center port entry (CE) and red dots indicate side port entry (SE). Only trials to the ipsilateral side (relative to the photometry recording) are shown and are divided by rewarded (left) and unrewarded (right) trials. **(B)** Averaged (±SEM) photometry signals across one mouse aligned to center port entry (CE, left) or side port entry (SE, right) for ipsilateral unrewarded trials. Traces show mean z-scored fluorescence intensity changes of simultaneously recorded from GCamp6f (green) and control fluorophore tdTomato (red). **(C)** Averaged (±SEM) photometry signals across all mice aligned to side port entry (SE), divided by unrewarded (left) or rewarded (right) outcome and grouped by ipsilateral (green) and contralateral (magenta) choice (90/10 reward probability, n=6 mice, 49 sessions, 20,355 trials). **(C)** Averaged (±SEM) photometry signals across all mice aligned to center port exit (CX, top) or side port exit (SX, bottom) grouped by ipsilateral (green) and contralateral (magenta) choice show similar changes during ipsiversive movements (90/10 reward probability, n=6 mice, 49 sessions, 20,355 trials). **(E)** Averaged (±SEM) photometry signals across all mice aligned to side port exit (SX) grouped by rewarded (blue) or unrewarded (red) outcomes. (right) Points represent the mean z-scored fluorescence per animal for the 500ms period immediately following the behavioral event, bars represent mean across animals ±SEM (all sessions are at

90/10 reward probability, n=6 mice, 49 sessions, 20,355 trials). **(F)** Ipsilateral trial-averaged (±SEM) photometry signals across all mice aligned to side entry (SE) divided by unrewarded (top) and rewarded (bottom) outcome, grouped by whether the previous trial (also ipsilateral) was rewarded (blue) or unrewarded (gray) plotted to examine if reward history impacts photometry signals. (right) Points represent the mean z-scored fluorescence per animal for the 500ms period immediately following the behavioral event, bars represent mean across animals ±SEM (all sessions are at 90/10 reward probability, n=6 mice, 49 sessions). **(G)** Contralateral trial averaged (±SEM) photometry signals across all mice aligned to side entry (SE) divided by unrewarded (top) and rewarded (bottom) outcome, grouped by whether the previous trial (also contralateral) was rewarded (blue) or unrewarded (gray) plotted to examine if reward history impacts photometry signals. (right) Points represent the mean z-scored fluorescence per animal for the 500ms period immediately following the behavioral event, bars represent mean across animals ±SEM (all sessions are at 90/10 reward probability, n=6 mice, 49 sessions). **(H)** Averaged (±SEM) photometry signals across all mice aligned to side entry (SE) divided by ipsi-rewarded (left) and contra-rewarded (right) trial types, grouped by whether the previous trial (opposite choice from current trial, i.e. "switch trials") was rewarded (blue) or unrewarded (gray) plotted to examine if reward and choice history impacts photometry signals (all sessions are at 90/10 reward probability, n=6 mice, 49 sessions).

Figure 4-figure supplement 1: Additional behavioral performance metrics before and after viral injection and electrophysiological validation of Tettx effects on GABA/glutamate cotransmission from EP^{Sst+} **neurons.**
(A) Trial duration (normalized to mean before AAV injection (day-5- day-1), red dashed line) across behavioral sessions (days) before and after AAV injection of GFP (black) or Tettx (gray) (n=6 GFP, n=6 Tettx animals, dots = mean, error bar = SEM). **(B)** *p*(high port) (normalized to mean before AAV injection(day-5- day-1), red dashed line) across behavioral sessions (days) before and after AAV injection of GFP (black) or Tettx (gray) (n=6 GFP, n=6 Tettx animals, dots = mean, error bar = SEM). **(C)** *p*(switch) (normalized to mean before AAV injection (day-5- day-1), red dashed line) across behavioral sessions (days) before and after AAV injection of GFP (black) or Tettx (gray) (n=6 GFP, n=6 Tettx animals, dots = mean, error bar = SEM). **(D-I)** Behavioral metrics for GFP (green) and Tettx (red) injected animals before (gray) and after (color) injection of AAV (dots = individual mice, bar = mean, error bars = 95% CI). **(J)** Sample whole-cell voltage-clamp recordings from lateral habenula (LHb) neurons clamped at either 0 mV (gray) or -65 mV (black) to isolate optogenetically evoked IPSCs or EPSCs, respectively, from oChief+ EP^{Sst+} axons. Sample traces on top are from a *Sst-Cre+* animal expressing oChief only in EP and bottom traces are from a *Sst-Cre+* animal expressing both oChief and Tettx, blue dashes represent the timing of the blue light pulse (1 ms duration). **(K)** Quantification of peak amplitude from optogenetically evoked IPSCs (top) and EPSCs (bottom) from oChief only (control, left) and oChief/Tettx (right) groups (n=8 cells control, 8 cells Tettx; bar = mean, error bar = SEM). **(L)** *p*(switch) for trial types divided by reward history and choice direction segregated into before injection (gray) and after injection (green=GFP, left; red=Tettx, right) (bar = mean, error bar = 95% CI). **(M)** Summary of RFLR model coefficients segregated into before injection (gray) and after injection (green=GFP; red=Tettx), each dot represents an individual mouse (bar = mean, error bar = SEM, negative log-likelihood of fits were equivalent across conditions; control= -0.22 SD=0.04; Tettx= -0.20 SD=0.05). **(N)** Trial duration following previously rewarded or unrewarded trials segregated into before injection (gray) and after injection (green=GFP; red=Tettx) (bar = mean, error bar = 95% CI). **A-I** and **L-N** n=6 GFP control and n=6 Tettx mice, 5 sessions/mouse before AAV inj. and 10 session/mouse after AAV injection, GFP control= 15,120 trials before, 34,523 trials after; Tettx = 17,528 trials before, 32,761 trials after.

Figure 4-figure supplement 2: Total number of trials per session and animal body weight changes before and after viral injection. (A) Body weight (normalized to mean before AAV injection) across behavioral sessions (days) before and after AAV injection of GFP (black) or Tettx (gray) (n=6 GFP, n=6 Tettx animals, dots = mean, error bar = SEM). **(B)** Total number of trials per session (normalized to mean before

AAV injection) across behavioral sessions (days) before and after AAV injection of GFP (black) or Tettx (gray) (n=6 GFP, n=6 Tettx animals, dots = mean, error bar = SEM).

Figure 5-figure supplement 1: Additional behavioral performance metrics before and after viral injection and electrophysiological validation of CRISPR-SaCas9 mediated deletion of *Slc17a6* (vGlut2) on GABA/glutamate cotransmission from EP^{Sst+} **neurons.**
(A) Trial duration (normalized to mean before AAV injection (day-5- day-1), red dashed line) across behavioral sessions (days) before and after AAV injection of sgROSA26 (black) or sg*Slc17a6* (gray) (n=10 sgROSA26, n=8 sg*Slc17a6* animals, dots = mean, error bar = SEM). **(B)** *p*(high port) (normalized to mean before AAV injection (day-5- day-1), red dashed line) across behavioral sessions (days) before and after AAV injection of sgROSA26 (black) or sg*Slc17a6* (gray) (n=10 sgROSA26, n=8 sg*Slc17a6* animals, dots = mean, error bar = SEM). **(C)** *p*(switch) (normalized to mean before AAV injection (day-5- day-1), red dashed line) across behavioral sessions (days) before and after AAV injection of sgROSA26 (black) or sgSlc17a6 (gray) (n=10 sgROSA26, 8 sg*Slc17a6* animals, dots = mean, error bar = SEM). **(D-I)** Behavioral metrics for sgROSA26 (blue) and sg*Slc17a6* (orange) injected animals before (gray) and after (color) injection of AAV (dots = individual mice, bar = mean, error bars = 95% CI). **(J)** Sample whole-cell voltage-clamp recordings from lateral habenula (LHb) neurons clamped at either 0 mV (gray) or -65 mV (black) to isolate optogenetically evoked IPSCs or EPSCs, respectively, from oChief+ EP^{Sst+} axons. Sample traces on top are from a *Sst-Cre+* animal expressing both oChief and sgROSA26 in EP and bottom traces are from a *Sst-Cre+* animal expressing both oChief and sg*Slc17a6* blue dashes represent the timing of the blue light pulse (1 ms duration). **(K)** Quantification of peak amplitude from optogenetically evoked IPSCs (top) and EPSCs (bottom) from sgROSA26 (control, left) and sg*Slc17a6* (right) groups (n=13 cells control, n=23 cells Tettx; bar = mean, error bar = SEM). **(L)** *p*(switch) for trial types divided by reward history and choice direction segregated into before injection (gray) and after injection (blue=sgROSA26, left; orange=sg*Slc17a6*, right) (bar = mean, error bar = 95% CI). **(M)** Summary of RFLR model coefficients segregated into before injection (gray) and after injection (blue=sgROSA26; orange=sg*Slc17a6*), each dot represents an individual mouse (bar = mean, error bar = SEM, negative log-likelihood of fits were equivalent across conditions; ROSA26= -0.25 SD=0.05; Slc17a6= -0.24 SD=0.05). **(N)** Trial duration following previously rewarded or unrewarded trials segregated into before injection (gray) and after injection (blue=sgROSA26; orange=sg*Slc17a6*) (bar = mean, error bar = 95% CI). **A-I** and **L-N** n=10 sgROSA26 control and n=8 sg*Slc17a6* mice, 5 sessions/mouse before AAV inj. and 10 session/mouse after AAV injection, sgROSA26 control= 17,318 trials before, 39,710 trials after; sg*Slc17a6* = 13,520 trials before, 29,256 trials after.

Figure 5-figure supplement 2: Total number of trials per session and animal body weight changes before and after viral injection. (A) Body weight (normalized to mean before AAV injection) across behavioral sessions (days) before and after AAV injection of sgROSA26 (black) or sgSlc17a6 (gray) (n=10 sgROSA26, n=8 sgSlc17a6 animals, dots = mean, error bar = SEM) **(B)** Total number of trials per session (normalized to mean before AAV injection) across behavioral sessions (days) before and after AAV injection of sgROSA26 (black) or sgSlc17a6 (gray) (n=10 sgROSA26, n=8 sgSlc17a6 animals, dots = mean, error bar = SEM)

Figure 1

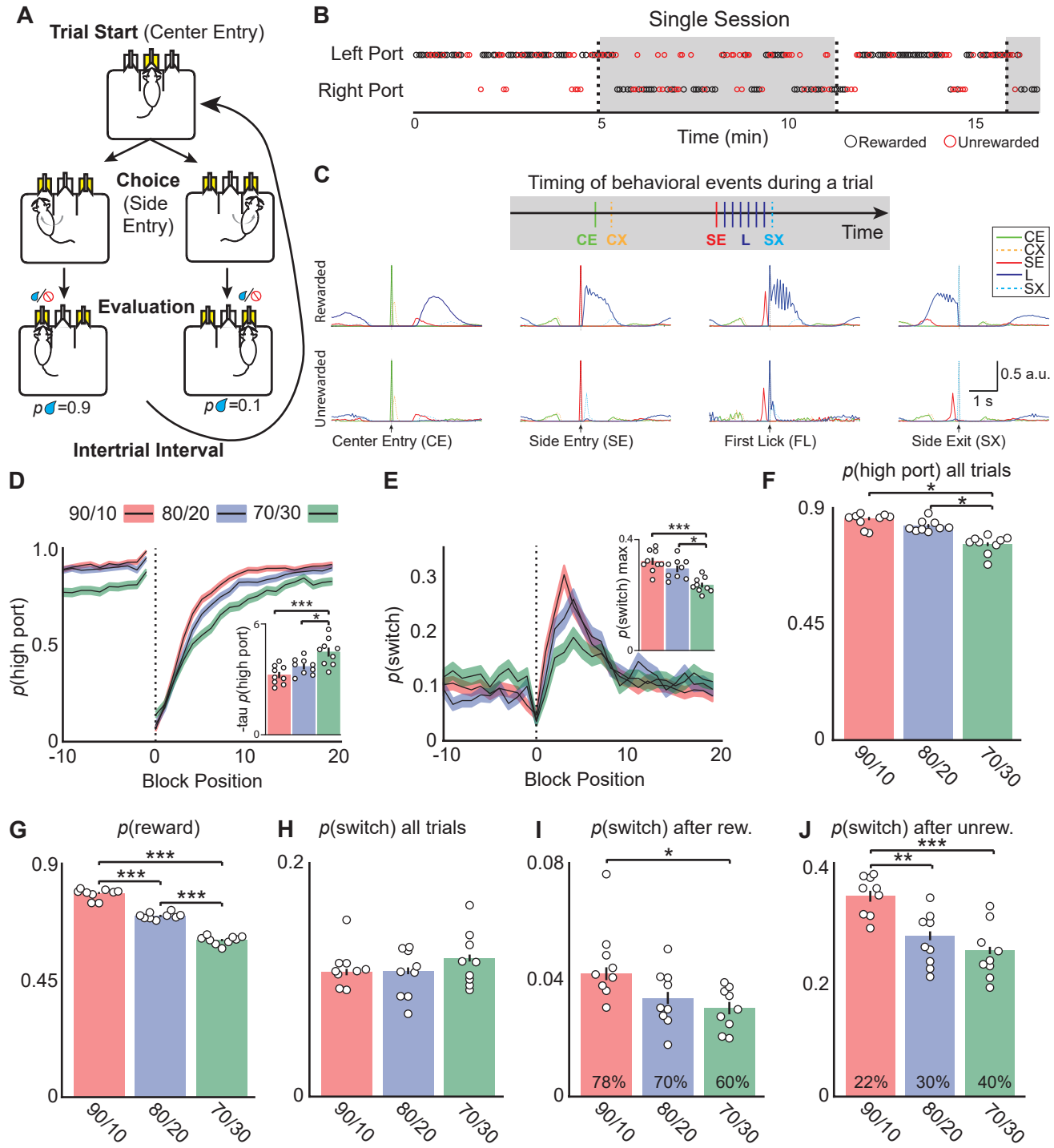

A
Trial Start (Center Entry)
Choice (Side Entry)
Evaluation
p=0.9
p=0.1
Intertrial Interval
B
Single Session
Left Port
Right Port
0
5
10
15
Time (min)
Rewarded
Unrewarded
C
Timing of behavioral events during a trial
CE
CX
SE
L
SX
Time
Rewarded
Unrewarded
0.5 a.u.
1 s
Center Entry (CE)
Side Entry (SE)
First Lick (FL)
Side Exit (SX)
D
90/10
80/20
70/30
p(high port)
Block Position
-tau p(high port)
E
p(switch)
Block Position
p(switch) max
F
p(high port) all trials
90/10
80/20
70/30
G
p(reward)
H
p(switch) all trials
I
p(switch) after rew.
78%
70%
60%
J
p(switch) after unrew.
22%
30%
40%

Figure 2

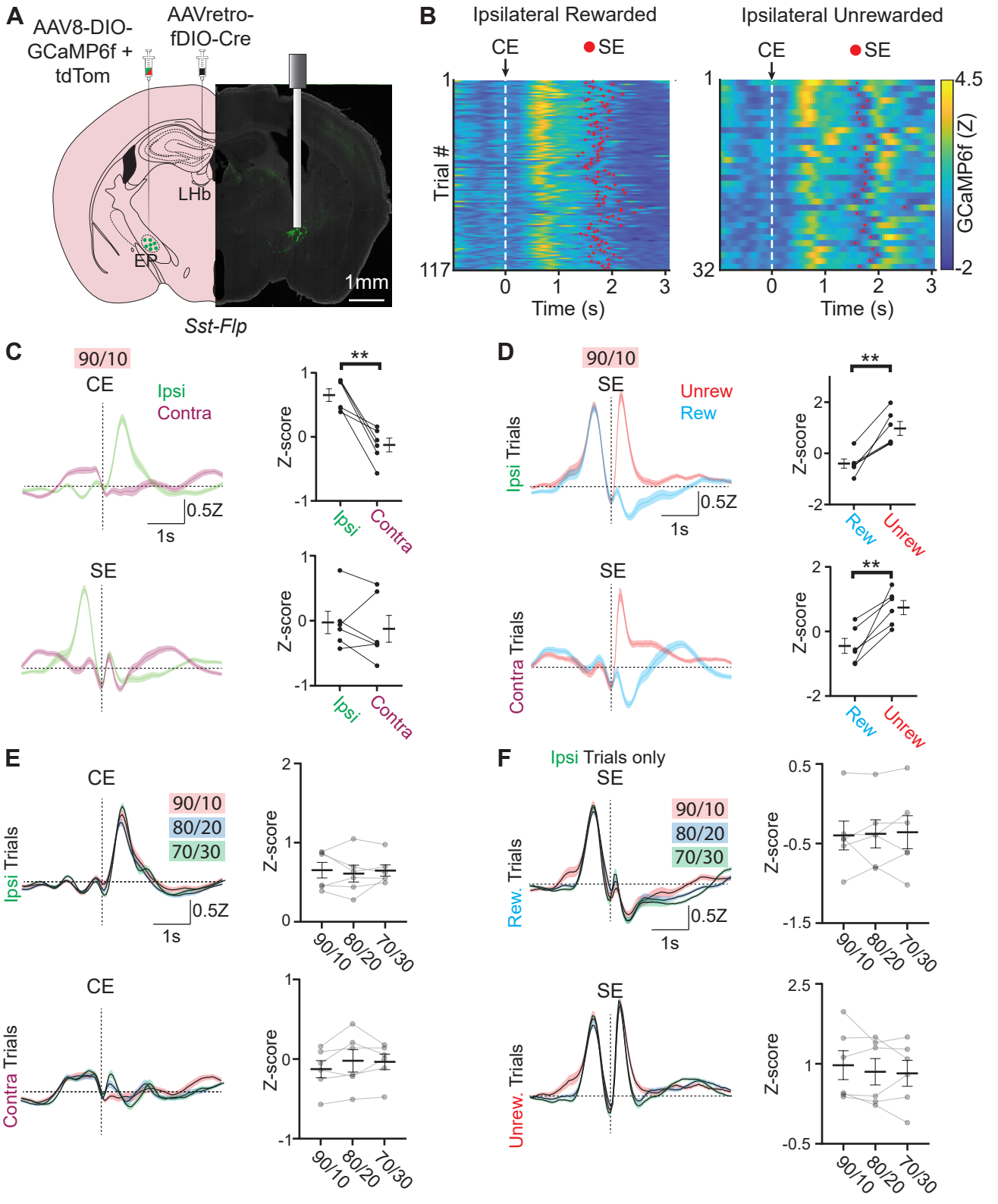

A
AAV8-DIO-GCaMP6f + tdTom
AAVretro-fDIO-Cre
LHb
EP
1mm
Sst-Flp
B
Ipsilateral Rewarded
Ipsilateral Unrewarded
CE
SE
Trial #
Time (s)
GCaMP6f (Z)
C
90/10
CE
Ipsi
Contra
0.5Z
1s
Z-score
Ipsi
Contra
SE
D
90/10
SE
Unrew
Rew
Ipsi Trials
Contra Trials
Rew
Unrew
E
CE
90/10
80/20
70/30
Ipsi Trials
Contra Trials
F
Ipsi Trials only
SE
Rew. Trials
Unrew. Trials

Figure 3

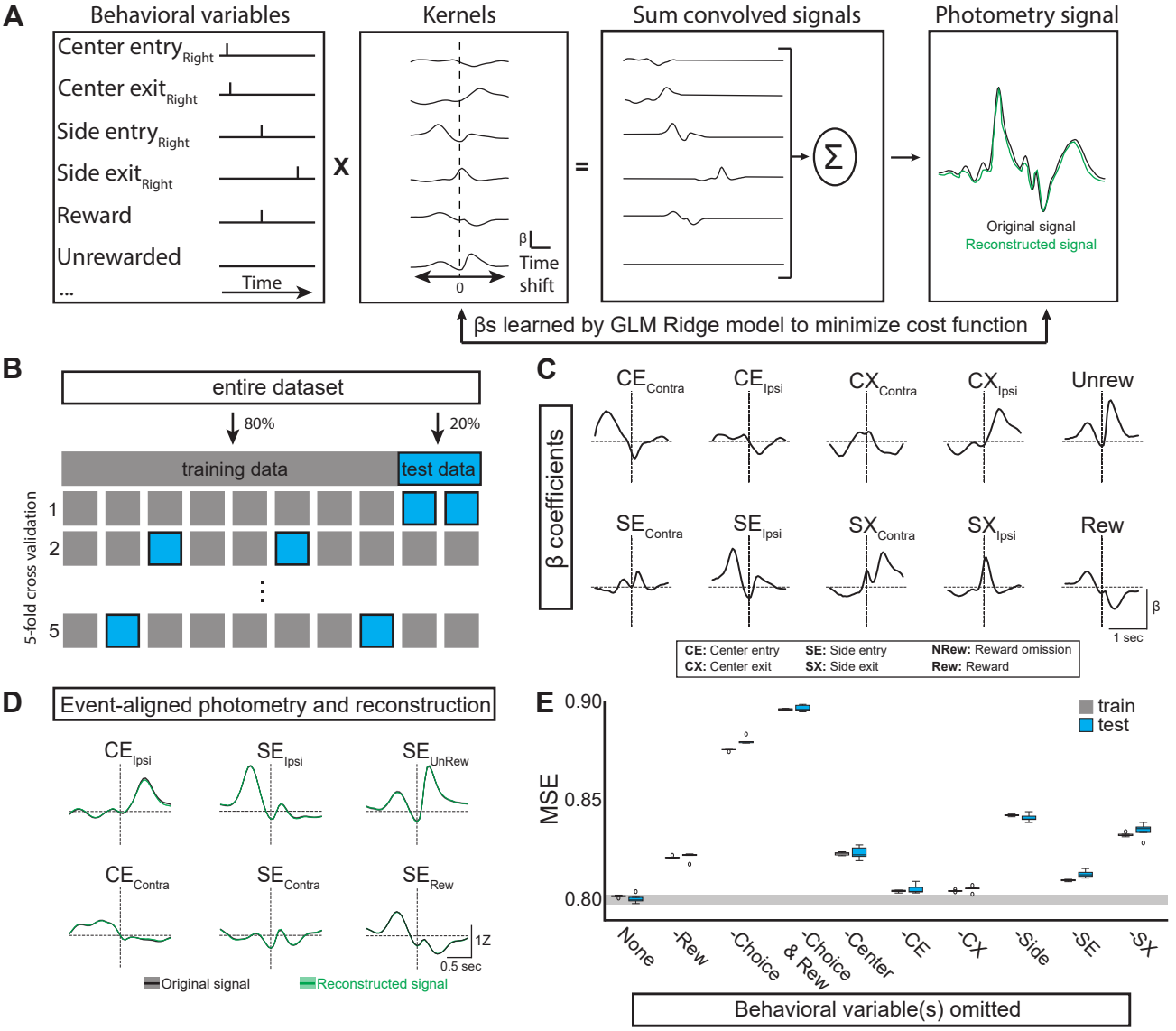

A
Behavioral variables
Center entry Right
Center exit Right
Side entry Right
Side exit Right
Reward
Unrewarded
...
Time
X
Kernels
β
Time shift
0
=
Sum convolved signals
Σ
Photometry signal
Original signal
Reconstructed signal
βs learned by GLM Ridge model to minimize cost function
B
entire dataset
80%
20%
training data
test data
5-fold cross validation
1
2
5
C
β coefficients
CE Contra
CE Ipsi
CX Contra
CX Ipsi
Unrew
SE Contra
SE Ipsi
SX Contra
SX Ipsi
Rew
β
1 sec
CE: Center entry
CX: Center exit
SE: Side entry
SX: Side exit
NRew: Reward omission
Rew: Reward
D
Event-aligned photometry and reconstruction
CE Ipsi
SE Ipsi
SE UnRew
CE Contra
SE Contra
SE Rew
1Z
0.5 sec
Original signal
Reconstructed signal
E
MSE
0.90
0.85
0.80
train
test
None
-Rew
-Choice
-Choice & Rew
-Center
-CE
-CX
-Side
-SE
-SX
Behavioral variable(s) omitted

Figure 4

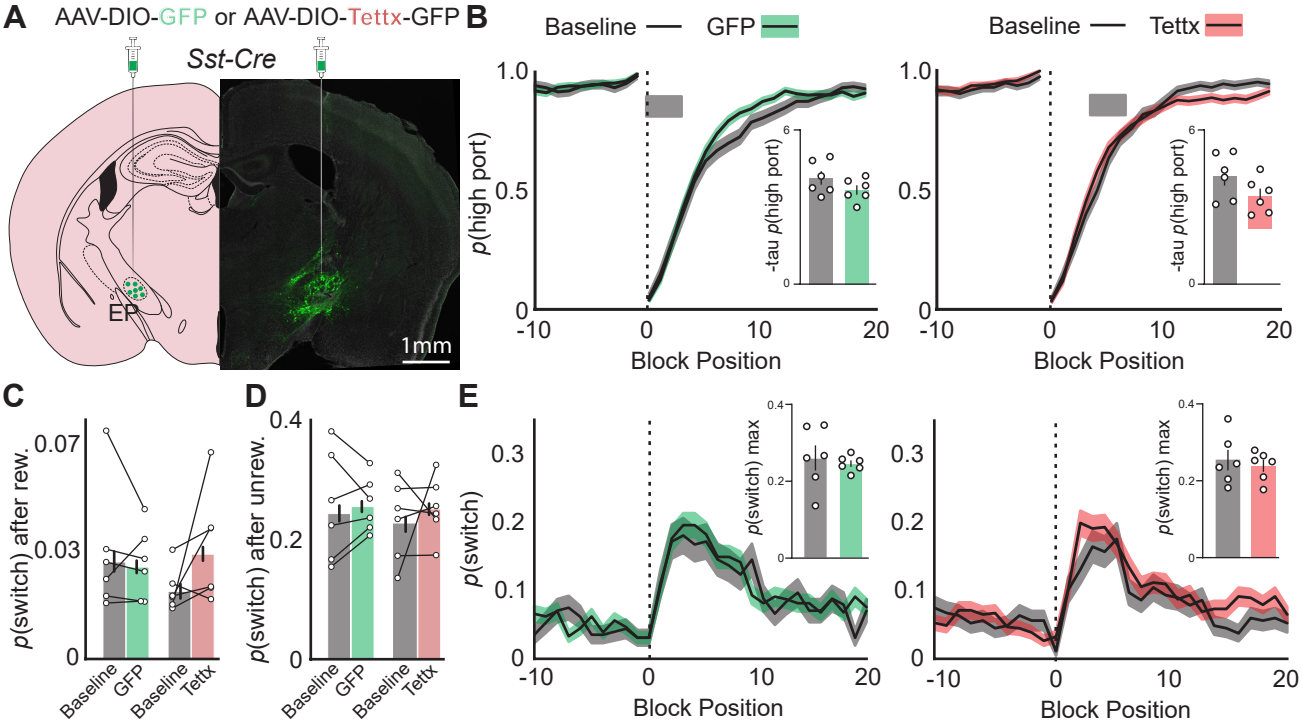
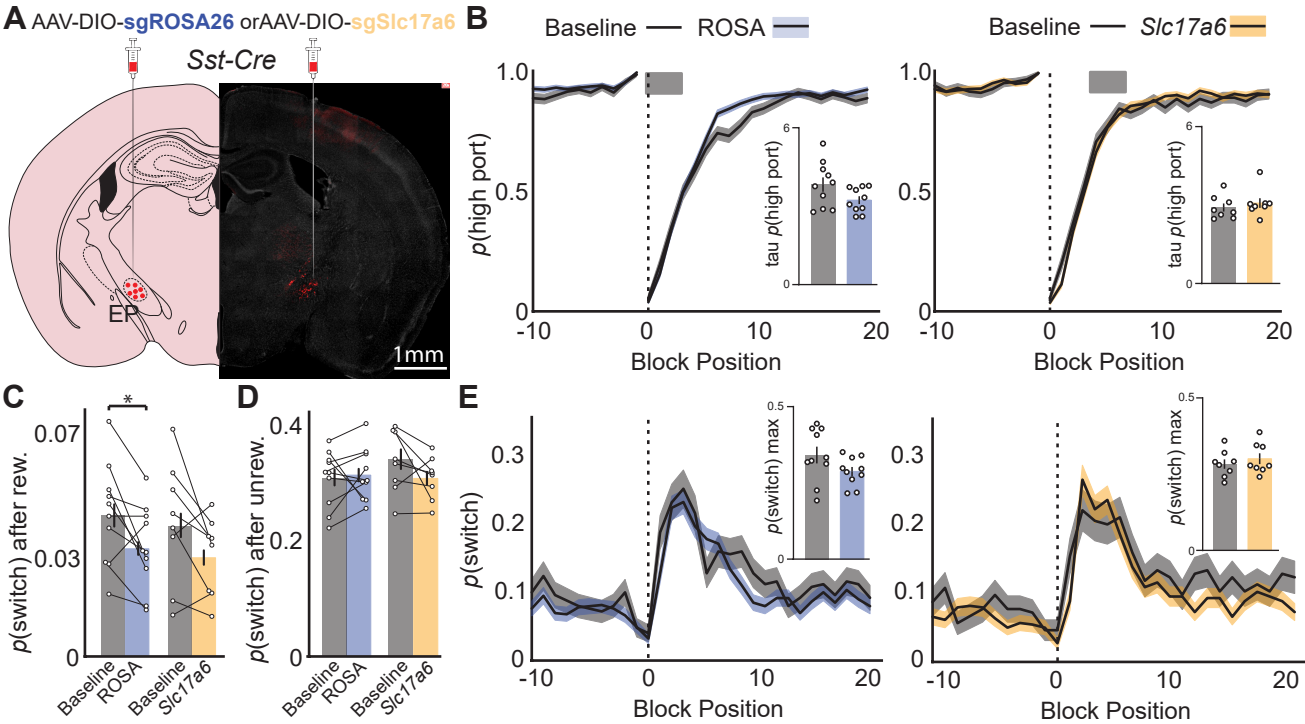

A
AAV-DIO-GFP or AAV-DIO-Tettx-GFP
Sst-Cre
EP
1mm
B
Baseline
GFP
Tettx
p(high port)
-tau p(high port)
Block Position
C
p(switch) after rew.
Baseline
GFP
Tettx
D
p(switch) after unrew.
E
p(switch)
p(switch) max

Figure 5


A
AAV-DIO-sgROSA26 orAAV-DIO-sgSlc17a6
Sst-Cre
EP
1mm
B
Baseline
ROSA
Baseline
Slc17a6
p(high port)
tau p(high port)
Block Position
C
p(switch) after rew.
Baseline
ROSA
Baseline
Slc17a6
D
p(switch) after unrew.
E
p(switch)
p(switch) max

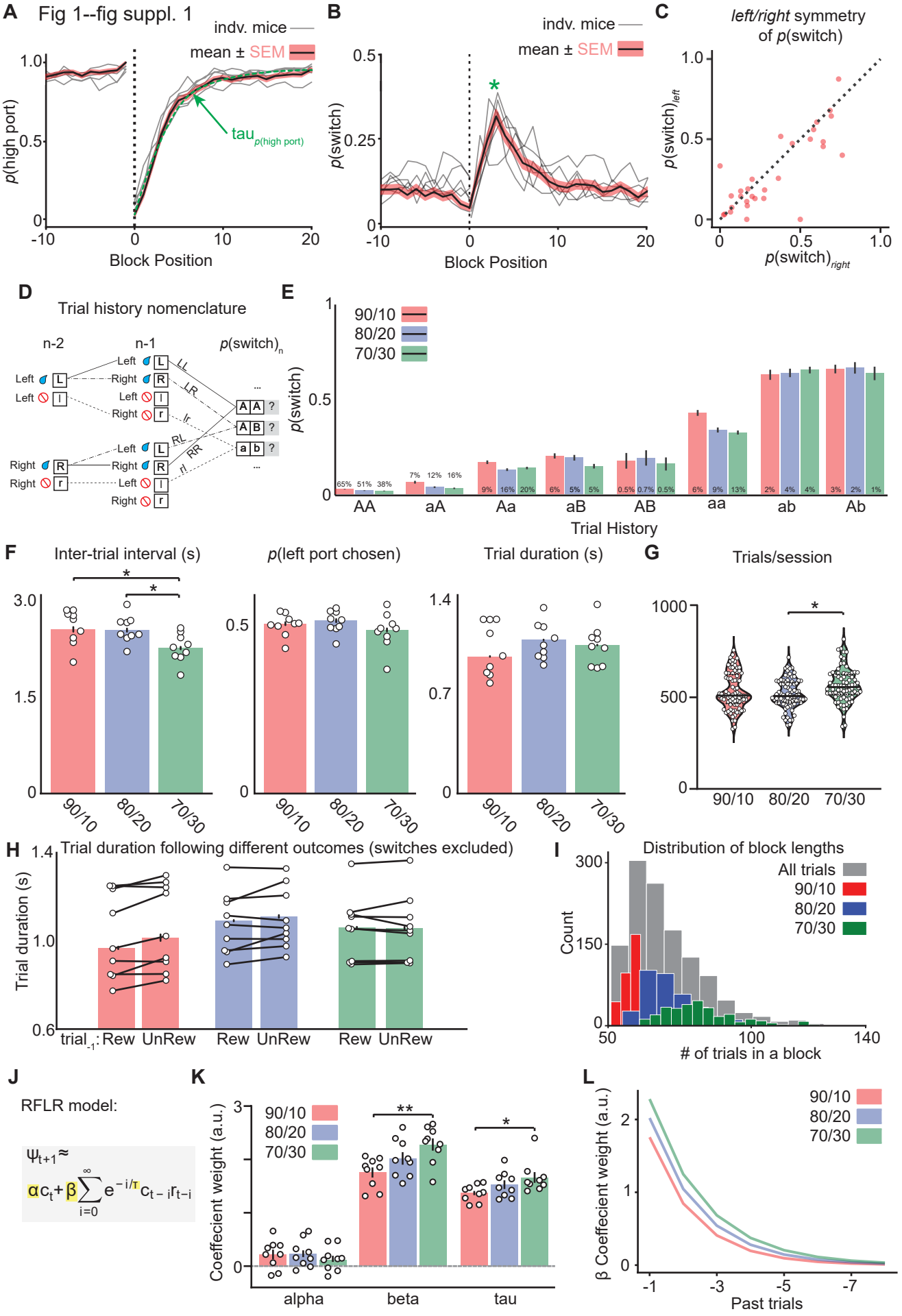


Fig 1--fig suppl. 1
A
indv. mice
mean ± SEM
p(high port)
tau p(high port)
Block Position
B
indv. mice
mean ± SEM
p(switch)
Block Position
C
left/right symmetry of p(switch)
p(switch)left
p(switch)right
D
Trial history nomenclature
n-2
n-1
p(switch)n
Left
Right
LL
LR
rr
RL
RR
rl
E
90/10
80/20
70/30
p(switch)
65% 51% 38%
7% 12% 16%
9% 16% 20%
6% 5% 5%
0.5% 0.7% 0.5%
6% 9% 13%
2% 4% 4%
3% 2% 1%
AA aA Aa aB AB aa ab Ab
Trial History
F
Inter-trial interval (s)
p(left port chosen)
Trial duration (s)
90/10 80/20 70/30
G
Trials/session
90/10 80/20 70/30
H
Trial duration following different outcomes (switches excluded)
Trial duration (s)
trial-1: Rew UnRew Rew UnRew Rew UnRew
I
Distribution of block lengths
All trials
90/10
80/20
70/30
Count
# of trials in a block
J
RFLR model:
$\Psi_{t+1} \approx \alpha c_t + \beta \sum_{i=0}^{\infty} e^{-i/\tau} c_{t-i} r_{t-i}$
K
90/10
80/20
70/30
Coeffecient weight (a.u.)
alpha
beta
tau
L
90/10
80/20
70/30
β Coeffecient weight (a.u.)
Past trials

Fig 1--fig suppl. 2

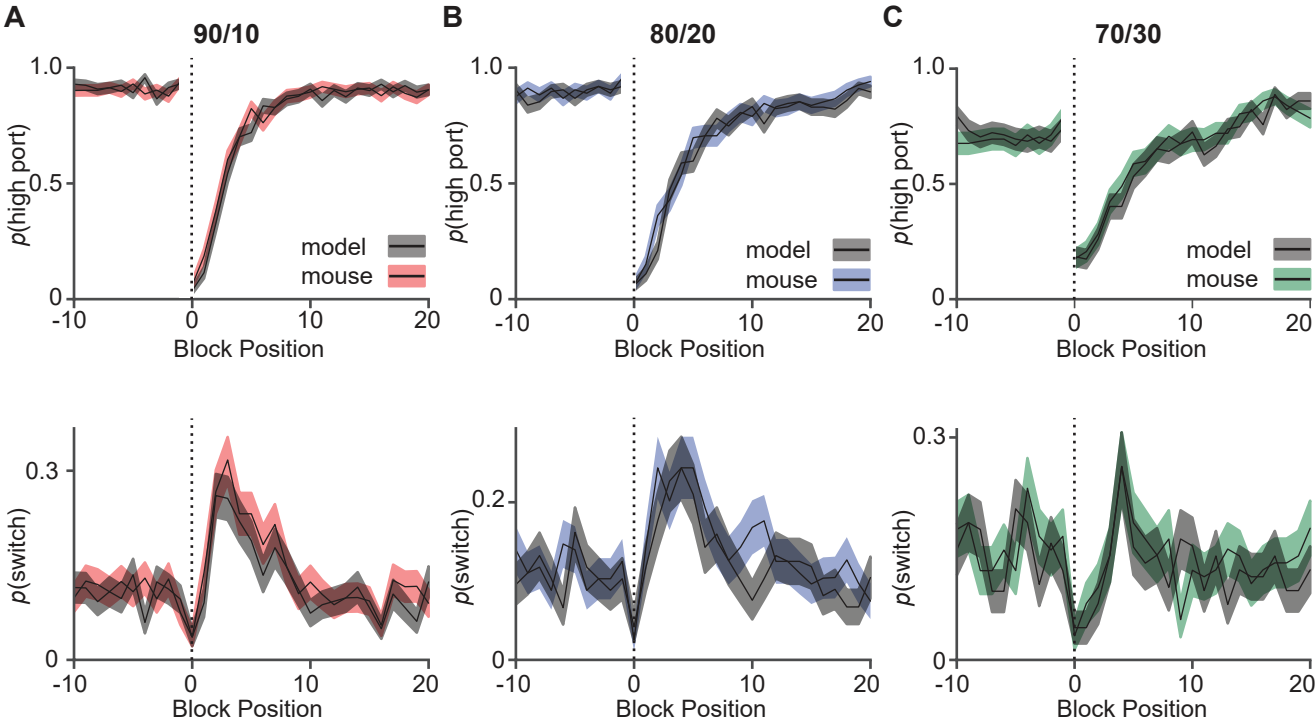
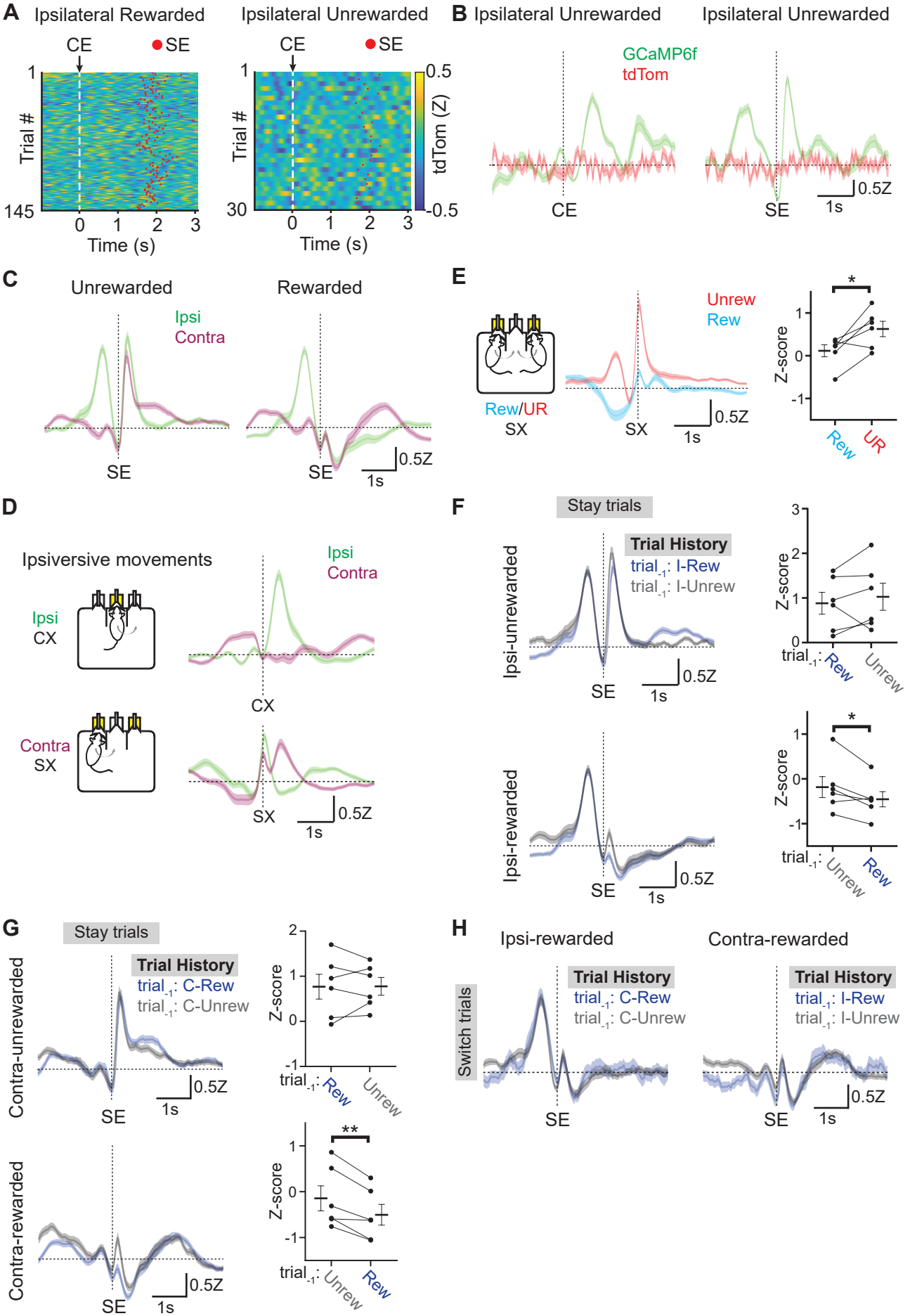


Fig 2 --fig suppl. 1



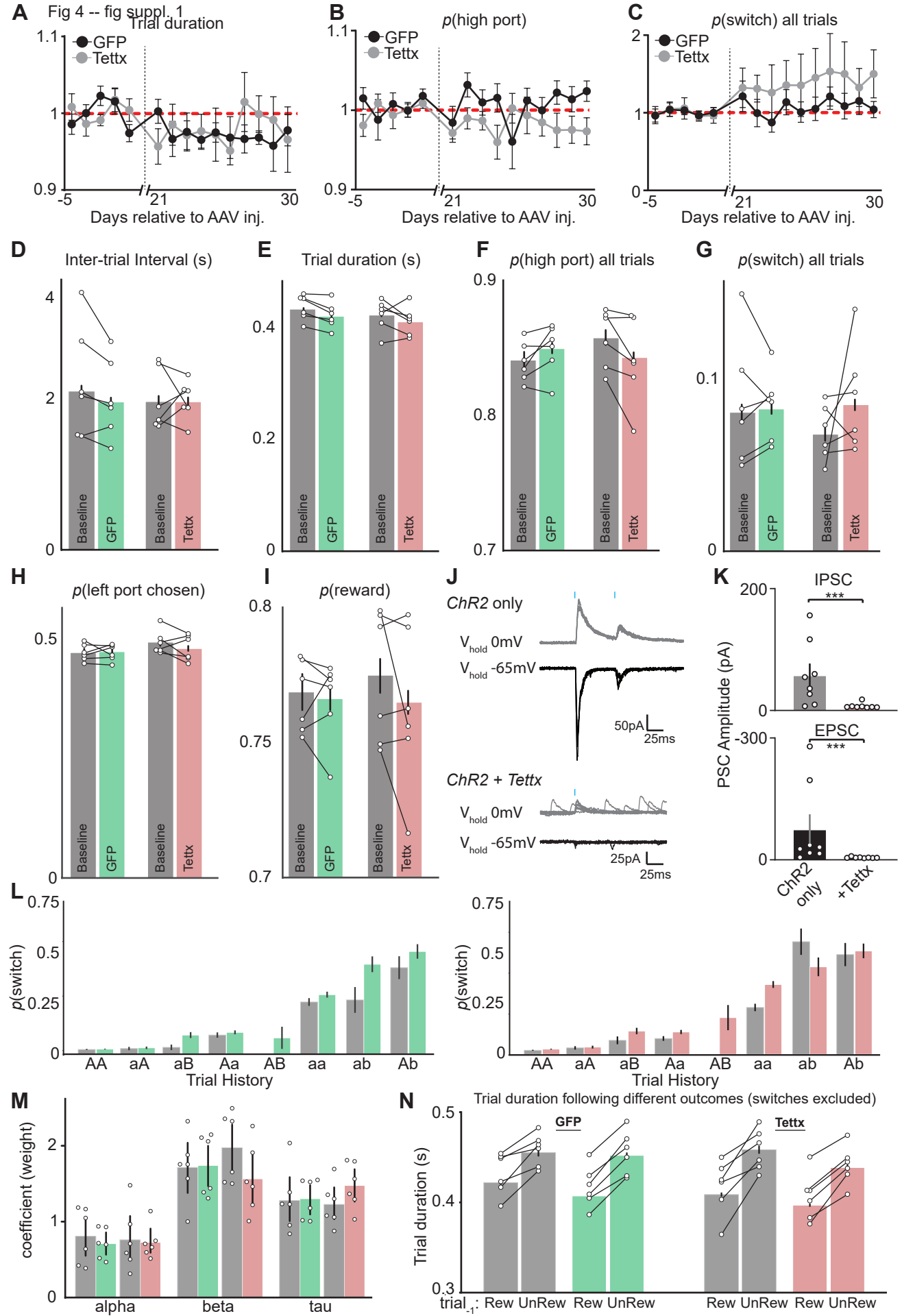
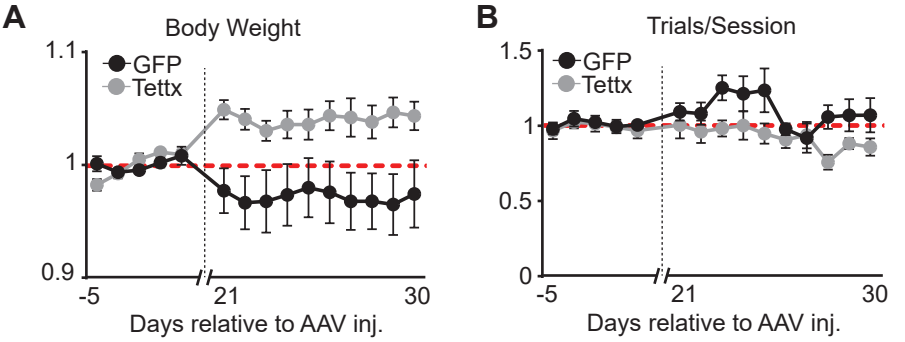
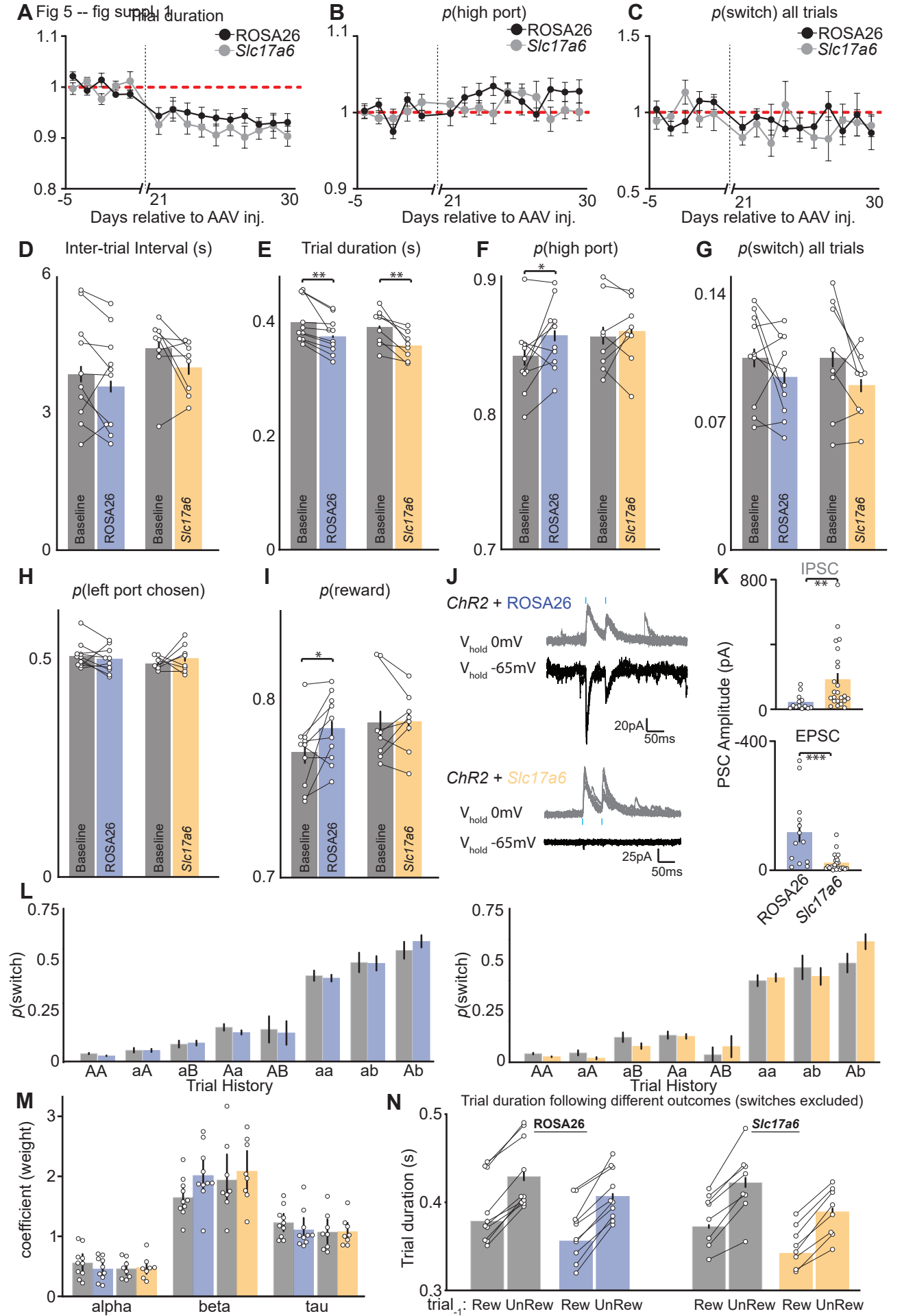


Fig 4 -- fig suppl. 1
A
Trial duration
GFP
Tettx
Days relative to AAV inj.
B
p(high port)
C
p(switch) all trials
D
Inter-trial Interval (s)
E
Trial duration (s)
F
p(high port) all trials
G
p(switch) all trials
Baseline
GFP
Tettx
H
p(left port chosen)
I
p(reward)
J
ChR2 only
ChR2 + Tettx
50pA
25ms
25pA
K
IPSC
EPSC
PSC Amplitude (pA)
ChR2 only
+Tettx
L
p(switch)
Trial History
AA
aA
aB
Aa
AB
aa
ab
Ab
M
coefficient (weight)
alpha
beta
tau
N
Trial duration following different outcomes (switches excluded)
Trial duration (s)
trial-1: Rew UnRew Rew UnRew Rew UnRew Rew UnRew

Fig 4--fig suppl. 2




A Fig 5 -- fig suppl 1
Trial duration
ROSA26
Slc17a6
Days relative to AAV inj.
B
p(high port)
C
p(switch) all trials
D
Inter-trial Interval (s)
E
Trial duration (s)
F
p(high port)
G
p(switch) all trials
Baseline
ROSA26
Slc17a6
H
p(left port chosen)
I
p(reward)
J
ChR2 + ROSA26
V hold 0mV
V hold -65mV
20pA
50ms
ChR2 + Slc17a6
25pA
K
IPSC
EPSC
PSC Amplitude (pA)
L
p(switch)
AA aA aB Aa AB aa ab Ab
Trial History
M
coefficient (weight)
alpha
beta
tau
N
Trial duration following different outcomes (switches excluded)
Trial duration (s)
trial-1: Rew UnRew Rew UnRew Rew UnRew Rew UnRew

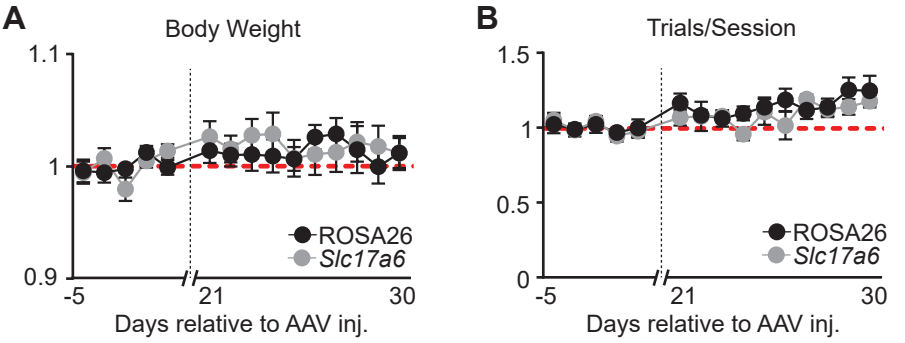



Fig 5--fig suppl. 2