Behavioral/Cognitive

Prefrontal Cortex Representation of Learning of Punishment Probability During Reward-Motivated Actions

David S. Jacobs and Bita Moghaddam

Behavioral and Systems Neuroscience Program and the Department of Behavioral Neuroscience, Oregon Health and Science University, Portland, Oregon 97239

Actions executed toward obtaining a reward are frequently associated with the probability of harm occurring during action execution. Learning this probability allows for appropriate computation of future harm to guide action selection. Impaired learning of this probability may be critical for the pathogenesis of anxiety or reckless and impulsive behavior. Here we designed a task for punishment probability learning during reward-guided actions to begin to understand the neuronal basis of this form of learning, and the biological or environmental variables that influence action selection after learning. Male and female Long-Evans rats were trained in a seek-take behavioral paradigm where the seek action was associated with varying probability of punishment. The take action remained safe and was followed by reward delivery. Learning was evident as subjects selectively adapted seek action behavior as a function of punishment probability. Recording of neural activity in the mPFC during learning revealed changes in phasic mPFC neuronal activity during risky-seek actions but not during the safe take actions or reward delivery, revealing that this region is involved in learning of probabilistic punishment. After learning, the variables that influenced behavior included reinforcer and punisher value, pretreatment with the anxiolytic diazepam, and biological sex. In particular, females were more sensitive to probabilistic punishment than males. These data demonstrate that flexible encoding of risky actions by mPFC is involved in probabilistic punishment learning and provide a novel behavioral approach for studying the pathogenesis of anxiety and impulsivity with inclusion of sex as a biological variable.

Key words: anxiety; decision making; fiber photometry; impulsivity; mPFC; sex as a biological variable

---

Significance Statement

Actions we choose to execute toward obtaining a reward are often associated with the probability of harm occurring. Impaired learning of this probability may be critical for the pathogenesis of anxiety or reckless behavior and impulsivity. We developed a behavioral model to assess this mode of learning. This procedure allowed us to determine biological and environmental factors that influence the resistance of reward seeking to probabilistic punishment and to identify the mPFC as a region that flexibly adapts its response to risky actions as contingencies are learned.

Introduction

Actions executed toward obtaining a desired outcome are often associated with varying risk of harmful consequences. For example, driving a car to go to a restaurant for dinner (reward) is associated with the probability of getting into a car accident (punishment). This probability increases if driving in a blizzard and increases even further if the driver is drunk. In these

contexts, the desired outcome (or reward) is certain after the successful execution of the action. What changes is the probability that a punishment may occur during action execution. Importantly, the punishment probability associated with an action is often learned. In the above example, one is either taught, or learns by experience, the hazards of driving in bad weather or while drunk. After learning, computation of this probability is fundamental to making the optimal decision to execute, or to inhibit, reward-guided actions. Abnormalities in this computation may lead to an exaggerated assessment of punishment risk, which is a hallmark of anxiety disorders, or to attenuated calculation of this risk, which may be associated with reckless behavior or impulsivity (Bechara et al., 2002; Hartley and Phelps, 2012; Ersche et al., 2016; Vanderschuren et al., 2017; Jean-Richard-Dit-Bressel et al., 2018).

We sought to design a model for punishment probability learning during reward-guided actions with two aims in mind: (1) to begin to understand the neuronal basis of this form of
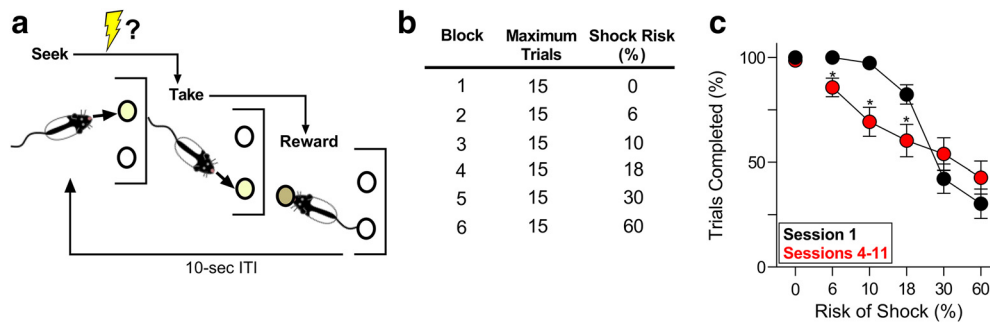
Figure 1. Outline and characterization of learning in the probabilistic punishment task. **a**, Schematic of a given trial. **b**, Structure of a given session. **c**, Evidence of learning demonstrated by behavioral changes in Session 1 compared with Sessions 4–11.

learning and (2) to characterize the biological and environmental variables that influence the decision to execute, or to resist, reward-directed actions after learning. We posited that neuronal ensembles in the mPFC, a region extensively implicated in post-learning risky decision-making (St Onge and Floresco, 2010; Orsini et al., 2018) and punishment representation (Pascoli et al., 2015), are dynamically involved in punishment probability learning. Our choice to focus on the mPFC was further supported by a large literature implicating subregions of mPFC in fear conditioning (Baeg et al., 2001; Corcoran and Quirk, 2007) or tasks that assess the impact of punishment on rewarded action when alternative reinforcing outcomes are possible (B. T. Chen et al., 2013; Friedman et al., 2015; Orsini et al., 2018).

Guided by previous work (Azrin et al., 1963; Pelloux et al., 2007; Simon et al., 2009; Park and Moghaddam, 2017a,b), we designed a task for rats where actions taken toward obtaining the same reward were associated with changes in the risk of punishment. The focus on obtaining the same reward was critical because choosing different rewarding outcomes is not always an option in the real world and, in the case of addictive disorders, may no longer be salient or viable (Volkow et al., 2003). The task used a chained schedule of reinforcement where an initial "seek" action preceded a "take" action, which then led to reward delivery. Seek and take actions were operationally similar, but punishment (mild shock) probability was introduced, using an ascending design, only contingent on the seek action. Learning was quantified by changes in trial completion and seek action latency as a function of punishment risk.

Using fiber photometry to measure real-time changes in neuronal calcium activity, we find that learning is associated with changes in the phasic response of mPFC neuronal activity during the seek action but not during the take action, reward delivery, or shock. After learning, the variables that influenced behavior included reinforcer and punisher value, pretreatment with the common anxiolytic diazepam, and sex. The influence of sex as a biological variable was explored in detail to reveal critical similarities and differences on how risk of punishment is integrated into reward-guided actions.

Materials and Methods

Subjects
Adult Long-Evans rats, pair-housed on a reverse 12 h:12 h light/dark cycle, were used. All experimental procedures were performed during the rodents' dark (active) cycle. Subjects were run in several cohorts with equal representation of males and females in each cohort.

For task characterization, 28 adult rats (14 male, 14 female) were obtained from Charles River at postnatal day 50–55. About a week after

arrival, they were handled and food restricted to 14 g/d. The food restriction was monitored throughout the study to maintain their weight at 90% of free feeding weight, with the target weight increasing by 5 g/wk. Training began at postnatal day 65–69, at which time males and females on average weighted, 278 and 208 g, respectively. For fiber photometry experiments, 8 adult rats (4 male, 4 female) bred in house were used (male 330–380 g, female 230–242 g; older than postnatal day 86 at time of recording). All experimental procedures were approved by the Oregon Health and Science University Institutional Animal Use and Care Committee and were conducted in accordance with National Institutes of Health's *Guide for the care and use of laboratory animals*.

Overview of experimental design
The probabilistic punishment task paradigm is depicted in Figure 1a, b. For fiber photometry (n = 2–4 per cohort) and task characterization studies (n = 12–16 per cohort), separate cohorts with equal male-female representation were used to ensure replicability of performance. All cohorts were trained to learn the task similarly. After learning, two of the cohorts used for task characterization were treated differently as follows: one cohort (n = 16) was tested with shock intensity adjusted for body weight (1 mA/kg, 300 ms); another cohort (n = 12) was tested in the task after diazepam treatment followed by satiation, behavior-titrated shock intensity, shock extinction, and progressive ratio (PR) after a washout period (see below).

Surgery for fiber photometry
Viral infusion surgery. Before task training, subjects were injected with a virus (AAV8-hSyn-GCaMP6s-P2A-tdTomato, Oregon Health and Science University Vector Core, 5^{e13} ng/mL) to allow for pan-neuronal expression of fluorescent calcium indicator GCaMP6s in the prelimbic mPFC as well as a non–calcium-dependent fluorophore tdTomato. The coexpression of tdTomato allows for a motion artifact control signal to be used to correct GCaMP signals in noisy environments with rodents (Soares et al., 2016; Matias et al., 2017; Babayan et al., 2018; Menegas et al., 2018). Rats were anesthetized with isoflurane and placed in a stereotaxic apparatus. Following an incision and topical application of lidocaine, craniotomy was performed to lower a 10 µL syringe (Hamilton) for virus infusion into the mPFC. Two injections were made (325 nL/site at 50 nL/min) at the coordinates AP +2.7, ±ML 0.65, DV −2.5 and −3.5 mm (from dura) with the most ventral injection always performed first. A microcontroller (World Precision Instruments) was used for the injections. Virus was allowed to diffuse for 5 min after the most ventral injection. The needle was slowly raised, and the second injection was performed and allowed 12 min to diffuse. After this, the needle was removed, the incision was stapled, and animals were given a 5 mg/kg injection of carpofen subcutaneously.

Fiber implant surgery. After allowing at least 6 weeks for virus expression and stabilization, subjects were implanted with an optical fiber (400 µm core) aimed at the prelimbic region of the mPFC (AP +2.7, ±ML 0.65, DV −3.3 mm from dura) using the surgical procedures outlined above, with the exception that three additional bore holes were made for three skull screws, which surrounded the craniotomy of the

mPFC. The optical fiber was slowly lowered and was glued to the skull with light-curing epoxy (Tetric N-flow, Ivoclar Vivadent). Subjects were given 5 mg/kg of carpofen and allowed about a week to recover with ad libidum access to food before behavioral testing and habituation to recording patchord began.

Apparatus

Operant chambers (Coulbourn Instruments) were used for behavioral testing. They included two nose poke holes, which could be illuminated, on one wall located 2 cm above a grid floor. The grid floor was connected to a shock generator. The food trough was on the opposite wall and was used to dispense 45 mg sucrose pellets (Bio-Serv) and detect food trough entries. Chambers contained infrared activity monitors (Coulbourne Instruments) located on the roof of the chamber. Graphic State software (version 3.03 and 4, Coulbourn Instruments) running on a windows PC was used for programming the task. For fiber photometry experiments, the operant chamber had an opening in the top of the box to permit entry for the recording patchcord.

Chain schedule training

After 1 d of habituation to the operant box and food trough (60 min, pellet dispensed every 45 s on average), subjects began chained schedule training. Subjects were first trained to respond on the "take" nose poke under a fixed ratio 1 (FR1) for 45 mg sucrose pellets. Daily sessions lasted until 90 pellets were delivered or 90 min elapsed. This phase of training lasted 6 d. Subjects were then trained to respond on the "seek" nose poke. A response on the seek nose poke (first link of the chain) resulted in extinguishing of the seek nose poke light concurrent with a 750 ms delay. The take nose poke was illuminated next, Completion of a FR1 on the take nose poke (second link of the chain) resulted in extinguishing the take nose poke light and food delivery and food trough illumination. Subjects were required to retrieve the pellet to initiate the 10 s intertrial interval (ITI). After the ITI, the seek nose poke was illuminated and a new trial began. Responding during the ITI was recorded but had no scheduled consequences. The side of seek and take nose pokes were counterbalanced across subjects. After the completion of 90 trials or 90 min, the session was terminated. All subjects were given 4 d of training and were moved to no-shock baseline procedures.

No-shock baseline procedures

This procedure began after subjects reliably learned the chained reinforcement schedule. The schedule of reinforcement was identical to that in previous training with the exception that the 90 trials were broken into six 15 trial blocks that were 15 min in length. Each block began by a 3 min time-out period, where all lights were extinguished, followed by a 12 min response period where subjects could earn up to 15 pellets. The nose poke light remained on until the subject made a response on the illuminated nose poke or until the end of the 12 min response period. If the 12 min response period ended before completion of the 15 possible trials, then lights were extinguished, and the subject moved to the next block. If subjects completed 15 trials before the 12 min response period elapsed, all lights were extinguished and responding produced no programmed consequences for the duration of the 12 min response period. Thus, these sessions served as a control to verify that subjects learned and could complete the sequential actions in a blockwise manner without punishment. These control sessions are hereafter referred to as "no-shock" sessions. After subjects performed this procedure for 4 d, they began the probabilistic punishment task.

Probabilistic punishment task procedures

After no-shock baseline procedures, footshock contingencies were introduced for the probabilistic punishment task. The reinforcement schedule was identical to that used in the no-shock baseline procedure, but now each block was accompanied by an increase in probability of a mild footshock (0.25 mA, 300 ms) immediately after the seek action. As was the case in no-shock procedures, subjects could complete up to 15 trials in a block. A trial ended on reward retrieval after completion of the seek and take action sequence or after the 12 min response period elapsed in the

absence of action execution. To prevent generalization of the shock to other blocks, the risk of the seek action contingent footshock increased with each successive block in the same ascending order for each session (0%, 6%, 10%, 18%, 30%, 60%). To assess learning of probabilistic punishment, we performed the task for 12 consecutive sessions. When two-way ANOVA of trial completion in the last five consecutive sessions in either sex revealed no main effect of session or interaction as described by (Simon et al., 2009) the performance was considered stable.

The behavioral procedure was optimized for fiber photometry. We increased the delay between the seek action and take cue illumination to 1.5 s to account for the relatively slow offset of GCaMP6s activity (Decay Time $t_{1/2} = 1$ s for 1 action potential; T.W. Chen et al., 2013), added a 1 s delay between the take action and reward delivery, changed the ITI to 15 s to increase the number of samples when normalizing the control signal, and decreased the task to four blocks (0%–18% risk of shock, increasing in quarter log units) to allow for 20 trials per block rather than 15. Along with decreasing the number of blocks, we reduced the inter-block intervals to 2 min to shorten the task to 56 min. This shortened task length was done to mitigate photobleaching of the fluorophores from continuous light exposure. Behavior was considered stable after a minimum of 4 sessions and when individual trial completion was within 25% of a 3 d mean for 3-consecutive sessions.

Fiber photometry systems and recording procedures

Two commercially available fiber photometry systems, Neurophotometrics Model: FP3001 and Tucker-Davis Technologies RZ5 were used. For Neurophotometrics, recording ($n = 2$) was accomplished by providing both 470 nm and 560 nm excitation light through the 400 μm core patchcord to the mPFC for GCaMP6 and tdTomato signals, respectively. LEDs were reflected through a dichroic mirror and onto a 20× Olympus objective. Excitation power was measured at 240–260 μW at the tip of the patch cord. Emission at 510–530 and 630–660 nm, from 470 and 560 nm excitation light, respectively, was split with an image splitting filter and captured via a high quantum efficiency CMOS Pointgrey camera. Recordings were performed using bonsai open source software (Lopes et al., 2015) and recorded at 41 Hz.

For Tucker-Davis Technologies recording ($n = 5$), excitation light was emitted from a 465 and 560 nm LED (Doric Instruments), sinusoidally modulated at 220 and 310-Hz, respectively, and controlled through an LED driver interfacing with the Tucker-Davis Technologies RZ5 processor running Synapse software. Excitation light was passed through a 400 μm core patchcord connected to a dual-fluorescence mini-cube (Doric Instruments). Light intensity at the tip of the patchcord was started at 10 μW but adjusted on an individual basis to optimize comparable levels of GCaMP and tdTomato signal and prevent photodetector clipping. This resulted in a range of 1–10 μW for light intensity for these subjects. GCaMP and tdTomato emission (500-540 nm and 580-680 nm, respectively) were collected back through the patchcord to dichroic mirrors and bandpass filters within the Doric minicube. Fluorescence was converted to voltage through two femtowatt detectors (Newport 2151). Synapse software demodulated fluorescence signals in real time at 1 kHz with a 6 Hz low pass filter.

For both systems, time stamps of behavioral events were collected by 5 V TTL pulses that were read into an Arduino interfaced with bonsai software or the same RZ5 processor in the Neurophotometrics and Tucker-Davis Technologies systems, respectively, to allow for aligning calcium activity with specific behavioral events in the task. Following behavioral training, but before shock contingency exposure, subjects were well acclimated to connection of the recording patchcord to assure changes in behavior were not due to distraction from the recording setup. The recording fiber was prebleached once over 12 h and for 30 min before recording sessions (constant illumination of both LEDs at ~300 μW power). To prevent slippage of patchcord connector from the implant, the THOR ADAL3 connector was used instead of a standard ceramic ferrule. Subjects were connected to a dummy fiber for nonrecording days, on control no-shock and probabilistic punishment task sessions, which mirrored the recording fiber but did not emit any light. Recording was performed at 2 or 3 time points: in the third no-shock session before probabilistic punishment ($n = 5$), at the first probabilistic

punishment session when subjects first experience the footshock contingencies (Session 1; $n = 7$), and after behavior had stabilized, i.e., the task was learned (Sessions 5–8; $n = 7$).

Body weight and behavior-titrated shock
Subjects in one cohort ($n = 16$) were required to achieve stable performance (as indicated above) using a footshock, which was titrated based on body weight (1.0 mA/kg, 300 ms) (Cooper et al., 2014; Orsini et al., 2016). In another cohort ($n = 12$), shock intensity was later titrated for each subject until animals showed comparable levels of action suppression of ~50% trial completion for the session (behavior-titrated shock). This was done by increasing or decreasing the shock intensity by ~0.05 mA until stable behavior was acquired (three consecutive sessions within 25% of the session mean). This procedure allowed for comparisons of how punishment probability affects reward seeking when the shock intensity produced action suppression in all subjects.

Diazepam testing
Injectable diazepam (Pfizer/Hospira) at a concentration of 5 mg/mL was assessed. Sterile saline (0.9% NaCl) was used for control injections. Diazepam (1.0 and 2.0 mg/kg) or saline was administered intraperitoneally 10 min before operant sessions with all injections given at a volume of ≤0.5 mL/kg. Doses of the same drug were separated by at least 1 d contingent on subjects performing within 25% of baseline (prediazepam) levels or after reestablishment of stable behavior (mean overall trials completed for three consecutive sessions within 25%).

Satiety testing
Subjects had 22 h of unlimited access to standard laboratory chow before a probabilistic punishment task session.

Shock threshold testing
Procedures were performed similar to previously published methods (Söderpalm and Engel, 1988). Subjects were placed in the chamber under red light for 15 min with no scheduled consequences on day 1. On day 2, after 3 min of acclimation to the operant chamber, a footshock was applied about every 40 s (contingent on all four paws being on the shock grid). An ascending intensity (0, 0.05, 0.06, 0.08, 0.1, 0.13, 0.16, 0.2, 0.3 mA) was used until the subject responded to the stimulus with a flinch, defined as a sudden rearward jerk immediately after shock administration.

Extinction of shock-suppressed behavior
Animals were tested for three sessions on probabilistic punishment task behavior using the behavior-titrated shock intensities to ensure that behavior remained stable. They were then tested in extinction sessions where no footshock punishment was administered during the task.

Progressive Ratio behavior
PR was assessed after extinction of shock-suppressed behavior in accordance to previously published methods (Richardson and Roberts, 1996). Briefly, completion of a fixed ratio on what was previously the take nose poke resulted in a food pellet. The fixed ratio increased according to the following algorithm, response ratio = $[(5 \times e^{0.2n}) - 5]$, where n is the number of reinforcers earned for a given session. PR sessions ran for 5 h or until 45 min elapsed without the completion of a ratio. The last completed ratio was considered the subjects' breakpoint. PR sessions were run for 6 d, and all subjects reached a stability criterion of two consecutive sessions with a breakpoint within two step sizes.

Open field testing
In Cohort 1, animals were tested on the open field 3 d after food restriction but before any operant training. The open field consisted of a gray box (36 inch × 36 inch) with gray walls (16 inch height). Subjects were placed in the center of the open field and allowed 10 min to explore the field. Zone entries as well as total distance traveled was monitored by camera and analyzed using Smart software (version 3.0, Harvard Apparatus). Dependent measures were percent time in the inner region, percent time in the outer region, and locomotor activity (total distance traveled).

Task characterization data and statistical analysis
We assessed both no-shock and probabilistic punishment task sessions. Trial completion was measured as the percentage of completed trials (of the 15 possible) for each block, whereas action latencies were defined as time from nose poke cue onset to action execution. Group mean values for each risk block or comparing risk and no-risk blocks are presented as mean ± SEM in all figures. In addition to assessing trial completion over each punishment risk, we analyzed overall trial completion to determine whether subgroups emerged that were differentially sensitive to punishment based on an 80%/30% trial completion split where >80% completion was resistant, 30%–80% completion was moderate, and <30% completion was sensitive to punishment (Gabriel et al., 2019). For action latency measures, data were only included in analyses if the subjects completed two or more trials for a given block. Because the lack of some latency data from subjects not completing any trials complicates the ability to perform repeated-measures ANOVA, latency behavior in risk associated trials was collapsed across the five blocks with risk of shock (6%–60%) to yield values for behavioral indices of action latency changes when punishment risk was present versus the no risk (0% risk) block.

Statistical procedures used either an ANOVA or mixed-effects model. Three-way ANOVA was used with factors of risk blocks, sex, and session type and followed up with two-way or one-way ANOVA where appropriate. Because of smaller sample sizes in pharmacological, extinction, titration, and satiety procedures, we assessed these manipulations with two-way ANOVA using factors of risk block and manipulation or treatment. Tests were done with the Greenhouse-Geisser correction for sphericity violation where appropriate. Activity counts during the ITI and during the shock period (i.e., the 300 ms period during which the shock was administered) were also quantified, and activity during blocks was collapsed and compared using two-tailed t tests. *Post hoc* comparisons were performed using the Bonferroni correction. An α level of 0.05 was used for all tests. Behavior data files were processed using custom-written scripts in Python (versions 2.7 and 3.0), and all statistical analyses were performed in GraphPad Prism (version 8) or R (version 3.6.1, ez package).

Behavioral modeling. Behavioral modeling of punishment probability dependent changes in trial completion was performed by fitting a sigmoid using a four-parameter logistic regression equation (4PLR) with the least-squares method to the three stability days when shock was titrated for behavioral output. The 4PLR used the following equation:

$$Y = d + (a - d)/(1 + 10^a((c - X) \times b))$$

Where Y is the percent of trials completed, X is the log risk of shock, a is the top of the asymptote, constrained to be less than or equal to 100, d is the bottom asymptote, c is the X value associated with a 50% decrease in behavior, and b is the measurement of steepness of the curve. To assess the integration of increasing probabilistic punishment, we quantified the slope of the linear portion of the sigmoid between high and low action by fitting a linear trendline to the bend points of the sigmoid. Briefly, we used the estimates for the top and bottom asymptote (a and d, respectively) and applied the following formula:

$$Upper = (a - d)/(1 + k) + d$$

$$Lower = (a - d)/(1 + 1/k) + d$$

Where k is a constant equal to 4.6805 (Sebaugh and McCray, 2003). A linear trendline was then fit to the upper and lower values to yield a slope for the linear portion of the sigmoid. Modeling was performed in GraphPad Prism (version 8).

Fiber photometry data analysis
Signals from the 465 (GCaMP6) and 560 (tdTomato) streams were processed in Python (version 3) using custom-written scripts; 465 and 560 streams were broken up based on the start and end of a given trial (for a given trial n: start of the ITI of $trial_{n-1}$ to end of the ITI of $trial_n$). This was done to fit the control 560 signal to the 465 signal on a trial × trial basis using a least-squares linear fit (numpy polyfit function in Python), as fitting the control signal to the entire session recording can be difficult when high amounts of motion are present, as in the current task or if

bleaching rates are different between fluorophores (Soares et al., 2016; Matias et al., 2017). The fitted 560 signal was then subtracted from the corresponding 465 signal to yield the change in fluorescent activity ($\Delta F/F = 465$ signal – fitted 560 signal/fitted 560 signal) that is corrected for non–calcium-dependent motion artifact and photobleaching from extended light exposure.

To assess whether changes in calcium activity were present after behavioral action execution (seek action and take action) and to normalize activity changes based on basal fluorescence, perievent z scores were computed by comparing the $\Delta F/F$ after the behavioral action to the 1.5 s baseline $\Delta F/F$ before execution. For example, the changes in $\Delta F/F$ following the seek action were compared with mean of the $\Delta F/F$ 1.5–0.02 s before the seek action. Because data from Neurophotometrics were sampled at 41 Hz, we downsampled the Tucker-Davis Technologies signals to 41 Hz as well for graphical purposes using Fourier method (scipy library in Python). We further separated punished (i.e., shock) trials from unpunished trials, to investigate differential activity that was seen during punishment administration and during anticipation of, but no actual administration of, punishment. To quantify positive or negative changes in calcium activity following action execution, we performed net area under the curve (AUC) analysis (trapezoidal method) for 1 s after the seek action or take action and for the 2 s after reward delivery. These values were analyzed using mixed-effects model with factors session and risk block. *Post hoc* Bonferroni corrections were used when comparing sessions with other sessions or comparing with the 0% risk block. To investigate individual differences, we performed correlational analyses between behavioral and mPFC activity changes before and after learning (i.e., between Session 1 and Session 5–8). To account for sex differences in behavior, we normalized behavior changes by taking the change in trial completion ($Trials_{Session\ 5-8} - Trials_{Session1}$) and z-scoring it for each sex. Thus, more negative values reflect subjects who showed greater decreases in resistance to probabilistic punishment. For mPFC activity, we took the difference between risk block z-score AUCs for seek actions in the corresponding sessions ($AUC_{Session\ 5-8} - AUC_{Session1}$). After verifying no violations in normality with the Shapiro-Wilk test, we performed two-tailed Pearson correlations for punished and unpunished trials. All statistical tests were performed with an α level of 0.05 in GraphPad Prism (version 8).

Histology and imaging
Viral expression and fiber placements were verified after behavioral testing. Subjects were anesthetized with chloral hydrate (400 mg/kg) and transcardially perfused with 0.01 M PBS followed by 4% PFA. Brains were removed and postfixated in PFA for 36 h before being placed in 20% sucrose solution and stored at 4°C. Forty-micron brain slices were collected on a cryostat (Leica Microsystems) and preserved in 0.05% phosphate-buffered azide. Brain slices were mounted to slides and coverslipped with Vectashield antifade mounting medium. An Apotome.2 microscope (Carl Zeiss) was used to image brain slices for GFP (Carl Zeiss Filter set 38: 470 nM excitation/525 nM emission) and tdTomato (Carl Zeiss Filter Set 43: 545 nM excitation/605 nM emission) to validate expression of both fluorophores in cells near the fiber tip. Fiber placement and extend of viral expression was verified according to Paxinos and Watson (1998).

Excluded data
Behavioral data from 4 individual subjects' sessions were excluded due to feeder malfunctions. Fiber photometry data from 1 male were excluded from fiber photometry experiments due to injection of a differing GCaMP6-expressing viral construct from the other subjects, complicating his comparison with other subjects. This subject was included for behavioral analysis. Trials where the optical fiber patch cord fell off during action periods and needed to be reconnected were also excluded.

Results
Learning of probabilistic punishment task
Guided by other tasks (Pelloux et al., 2015; Park and Moghaddam, 2017a), our task used a chained schedule of

reinforcement involving two sequential actions. Rats first executed a "seek" instrumental action in one nose poke followed by a "take" instrumental action in a second nose poke, which then led to reward delivery (Fig. 1a). The "seek" action was punished by delivering of a mild footshock after rats completed the action. The probability of the "seek" action being punished escalated in a blockwise manner throughout a single session (Fig. 1b).

Rats first learned to perform the sequential actions without punishment, which are designated as no-shock sessions, for at least four sessions. To validate task learning and to further mirror other procedures assessing risky choice, we determined stable probabilistic punishment task behavior after introduction of footshock punishment by identifying when five consecutive sessions produced no significant effect of session nor session × risk interaction in a two-way ANOVA for each cohort (Simon et al., 2009). These methods determined that stable behavior was observed in Sessions 4–11 (range for all cohorts; main effect of session or session × risk block interaction: F values < 1.97, p values > 0.13). After assessing the first 11 probabilistic punishment task sessions, we noted that task behavior differed in Session 1, when animals were first learning of the shock contingency, compared with Sessions 4–11. In Session 1, there was an overall resistance to probabilistic punishment in the 6%-18% risk blocks that decreased after subjects learned the task in Sessions 4–11 (session × risk block interaction: $F_{(3.07,83.05)} = 18.04$, $p < 0.01$; *post hoc p* values < 0.021; Fig. 1c).

Characterization of probabilistic punishment task after learning
Collapsed data for the last two no-shock sessions and the probabilistic punishment task sessions when performance under shock risk was stable as determined by ANOVA are shown in Figure 2a–e. As noted earlier, trial completion decreased as a function of punishment probability (risk block × session type interaction: $F_{(2.3,60.04)} = 32.27$, $p < 0.0001$), with significant decreases for all risk blocks in probabilistic punishment (i.e., shock) sessions compared with the corresponding block in the no-shock sessions (*post hoc p* values < 0.029). Inspection of these data at the individual level revealed considerable between-subject variability in punishment resistance, with subjects showing anywhere from complete punishment resistance to little. Dividing subjects based on trial completion into punishment-resistant (>80% trial completion), moderate (30%–80% trial completion), and sensitive (<30% trial completion) subgroups resulted in 15 of 28 resistant, 8 of 28 moderate, and 5 of 28 sensitive subjects (Fig. 2a, right).

The suppressive effects of probabilistic punishment were observed during the latency to complete the "seek" action, that is, the risky action (risk × session type interaction: $F_{(1,26)} = 27.9$, $p < 0.001$, Fig. 2b). Increased seek action latency was observed in the risk blocks of probabilistic punishment sessions compared with the corresponding blocks in no-shock sessions (Fig. 2b, right; *post hoc p* < 0.001). Overgeneralization of shock risk to the 0% risk block (i.e., first block) was not observed (*post hoc p* $= 0.99$ vs no-shock). Of note, variability increased at higher risk blocks because fewer subjects completed more than one trial (21 of 28 subjects). Subjects also demonstrated anticipation of footshock as the latency to complete the first "seek" action of a block increased with punishment risk compared with no-shock sessions (risk × session type interaction: $F_{(1,26)} = 29.08$, $p < 0.001$; Fig. 2c) and was also specific to blocks with a risk of shock (*post hoc p* < 0.01). Seek actions were followed by a small (<1 s) but significant increase in latency to complete the take action in
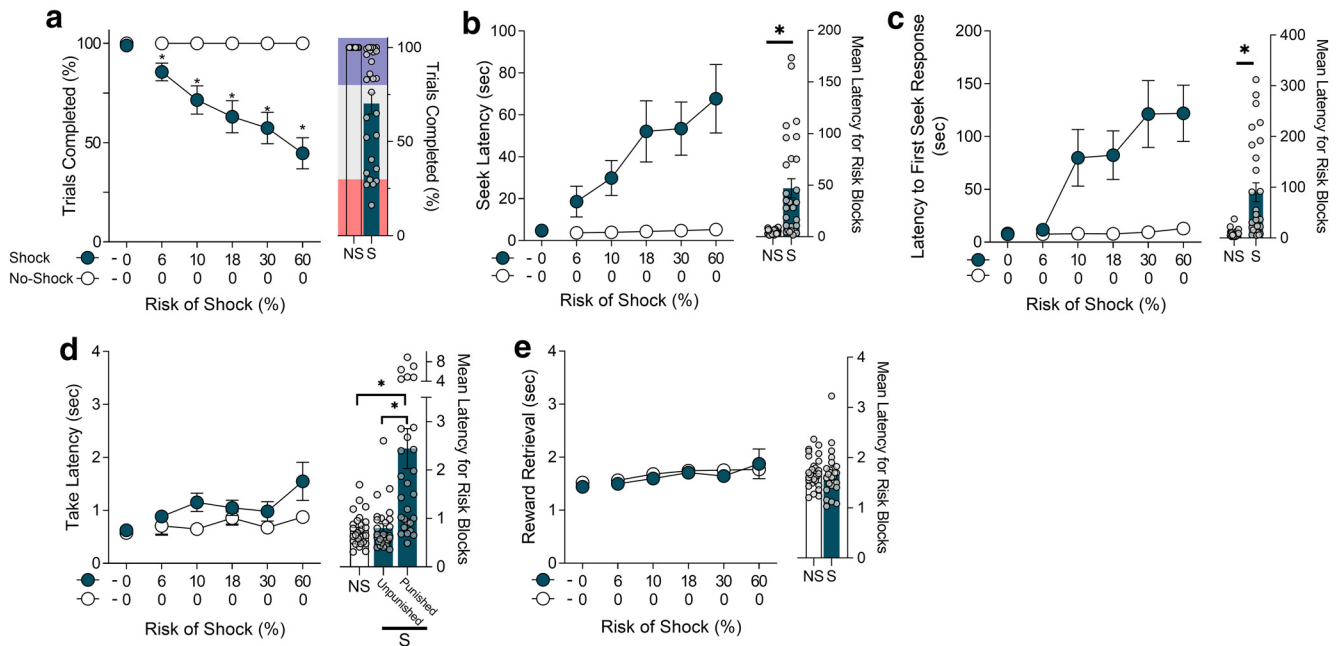
Figure 2. Assessing the effects of probabilistic punishment on task behaviors after learning. *a*, The number of trials completed in each block decreased on average in the probabilistic punishment task sessions where shock risk was present (S, blue) but not in no-shock sessions (NS, white). Subgroups with differing levels of punishment resistance emerged. Blue area represents punishment-resistant. Gray area represents moderate. Red area represents punishment-sensitive (right). *b*, *c*, Increasing risk of shock was also associated with an increase in latency to complete the risky "seek" action during and at the start of risk blocks. *d*, The take response modestly but significantly increased with risk of shock in punished trials. *e*, Latency to retrieve the food reward was not affected by shock risk. NS, No-shock; S, shock. *$p < 0.05$ versus no-shock. $n = 28$ for trials completed and for mean risk block latencies. Gray symbols represent individual data points.

probabilistic punishment sessions (risk × session type interaction: $F_{(1,26)} = 9.45$, $p = 0.005$, Fig. 2*d*). Because in some trials, the take action is operationally preceded by footshock, we further investigated whether the increase in take action latency is related to receiving a footshock. A one-way ANOVA was used to compare take action latency for take actions preceded by footshock (punished) and no footstock (unpunished) in punishment risk sessions with take actions from the corresponding blocks (i.e., blocks 2-6) in no-shock sessions. This analysis revealed that take action latency increases seen in probabilistic punishment sessions were related to receiving the footshock punishment (main effect of trial type: $F_{(1.07,29.1)} = 17.3$, $p < 0.01$). Take action latency in punished trials was increased compared with the no-shock sessions and the unpunished trials of punished sessions (*post hoc* p values ≤ 0.001), whereas unpunished trial latencies were comparable with that of the no-shock sessions (*post hoc* $p = 0.99$, Fig. 2*d*, right).

Reward retrieval latency was not influenced by risk of shock (risk × session type interaction: $F_{(1,26)} = 0.31$, $p = 0.58$; Fig. 2*e*) but modestly increased in later blocks compared with earlier blocks regardless of whether shock risk was present (main effect of block: $F_{(1,26)} = 27.6$, $p < 0.001$). This suggests a lack of overgeneralization of punishment to the context.

In one cohort, we also assessed innate anxiety in the open field before task training to assess whether individual patterns exploratory behavior would be associated with learned punishment-related behavior in the probabilistic punishment task ($n = 16$; Table 1). Individual variability in punishment resistance was not associated with exploratory behavior in the inner or outer zones of the open field or overall locomotor activity. Similarly, increases in seek latency during risk blocks were not associated with increased time spent in the inner or outer zones of the open field, nor with activity as assessed through locomotor activity.

Table 1. Pearson correlation (*r*) values of open field behavior with probabilistic punishment task measures[a]

	Trial completion	Seek latency	Take latency	Reward retrieval
OF inner	−0.17	0.18	−0.06	−0.21
OF outer	0.03	−0.09	0.01	0.24
OF activity	−0.49	0.39	0.32	0.38

OF, Open field. All *p* values > 0.06 uncorrected.

PFC activity represents learning of probabilistic punishment

The mPFC is implicated in risky choice, the representation of aversive stimuli, and action inhibition (B. T. Chen et al., 2013; Friedman et al., 2015; Pascoli et al., 2015; Orsini et al., 2018; Verharen et al., 2019). Importantly, neurons in mPFC are sensitive to punishment risk as well as to the experience of a stressor or punisher (for review, see McEwen and Morrison, 2013; Park and Moghaddam, 2017b). Little is known, however, if the mPFC flexibly encodes probabilistic punishment during learning. Thus, we hypothesized that, in our task, the mPFC processes risky action differently when probabilistic punishment is a factor. We used fiber photometry to record mPFC calcium activity in rats performing an optimized version of the probabilistic punishment task (see Materials and Methods; Fig. 3*a*), which focused on the risk blocks that were different between Session 1 and after learning (Fig. 1*c*). Fiber photometry, compared with spike recording, provided the advantage of being able to record the mPFC response during footshock (i.e., punished trials).

Viral expression was apparent in the mPFC and fiber placements were located within the prelimbic region (Fig. 3*b*). Fiber placement in 1 subject was on the borderline of the prelimbic/infralimbic region. Inspection of the data, however, revealed that the patterns of activity for this subject were similar to the rest of the group (data not shown). Using an individualized method of
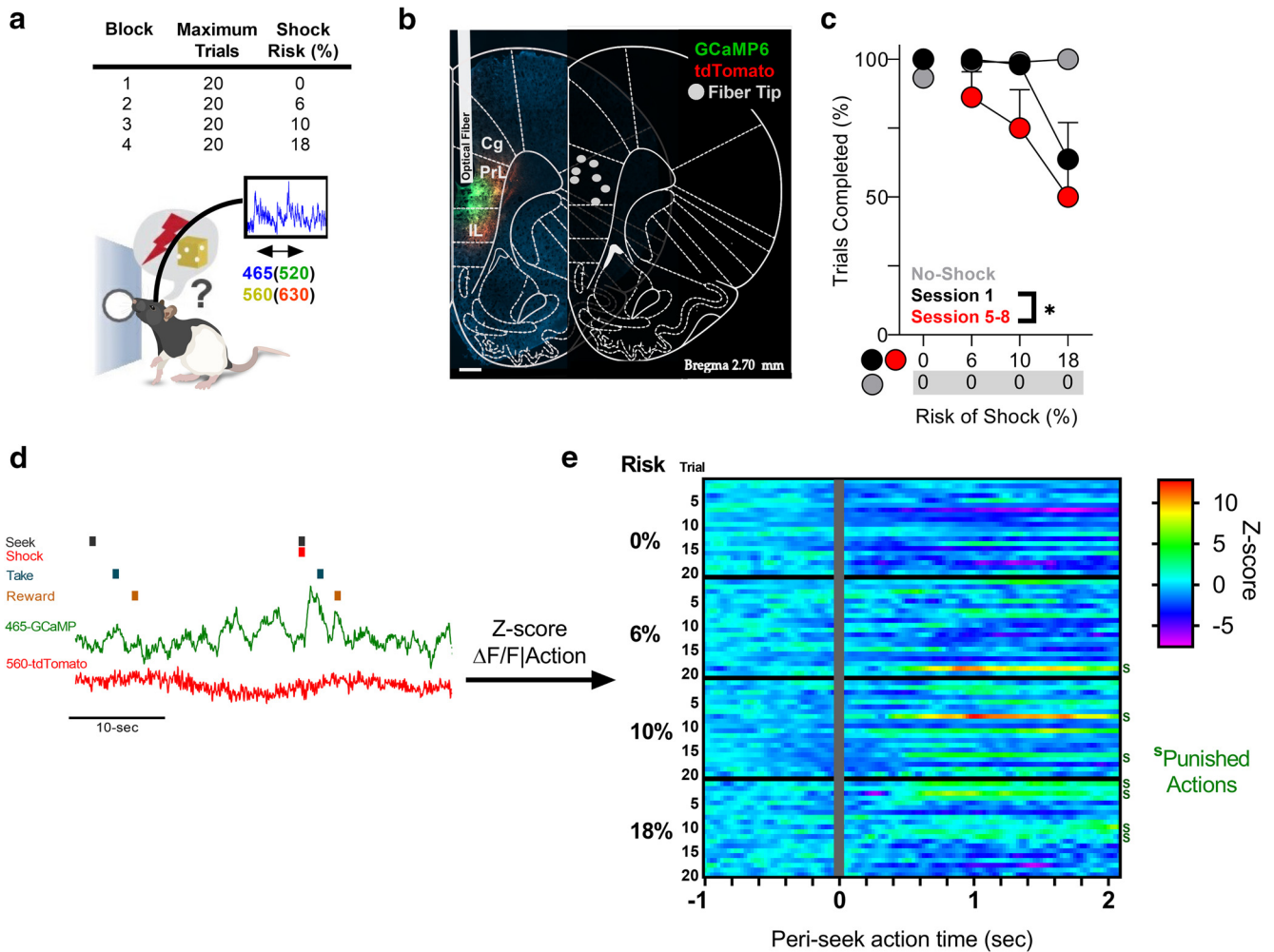
Figure 3. Recording neuronal calcium activity with GCaMP6 in mPFC during the probabilistic punishment task. *a*, Subjects performing a modified version of the probabilistic punishment task while neuronal calcium activity in the mPFC was recorded via GCaMP6s. *b*, GCaMP6s (green) and tdTomato (red) were readily expressed in the mPFC, with optical fibers (gray dots) targeting the prelimbic region of the mPFC. Scale bar, 500 µm. *c*, More trials were completed before learning (Session 1, black) compared with after learning (Session 5–8, red). *d*, A representative raw trace depicting GCaMP6s and tdTomato signals. *e*, Representative z-score calculations surrounding the seek action. Each row is a trial with time 0 being the seek action time. Green S indicates punished (i.e., shocked) trials. *p < 0.05. n = 8.

determine task learning (see Materials and Methods), we found that performance was stable by Session 5–8. After learning, punishment resistance decreased in Session 5–8 compared with Session 1 (Fig. 3c, main effect of session: $F_{(1,7)} = 5.95$, $p = 0.04$), mirroring the behavior seen in the full version of the probabilistic punishment task at the corresponding 0%–18% risk blocks (Fig. 1c). Increasing shock risk resulted in decreases in trial completion (main effect of risk block: $F_{(1.2,8.8)} = 9.42$, $p = 0.01$; *post hoc* $p = 0.002$ at 18% risk). While no risk block × session interaction was present ($F_{(1.2,8.44)} = 1.614$, $p = 0.24$), it is apparent that the differences between Session 1 and Session 5–8 were driven by the 6%-18% risk blocks as all subjects completed 100% of the trials when risk was 0% for Session 1 and 5–8.

After z-scoring calcium activity based on behavioral action, we noticed robust elevations in the mPFC following shock administration (Fig. 3d,e). This led us to divide trials into punished and unpunished trials focused around the seek action (i.e., the action with a risk of shock). Unpunished seek actions revealed a different response depending on session (main effect of session: $F_{(1.56,9.4)} = 5.9$, $p = 0.026$). We observed a decrease in activity during seek action execution in Session 1 that was similar

to those observed in the no-shock session (*post hoc* $p = 0.40$, Fig. 4a,b). After learning (Session 5–8), the magnitude of decrease during seek action execution was attenuated (*post hoc* $p < 0.01$ vs no-shock, $p < 0.01$ vs session 1; Fig. 4c, right). Near-significant changes in calcium activity with increasing blocks were also observed (main effect of risk block: $F_{(1.58,9.5)} = 3.7$, $p = 0.069$).

In contrast to the decrease in mPFC activity during seek action execution when animals did not receive shock, the same action executed when shock was received produced a large increase in mPFC activity (main effect of risk block [footshock]: $F_{(1.52,9.1)} = 6.5$, $p = 0.02$). This increase was not different before or after task learning (paired *t* test: $t_{(6)} = 0.89$, $p = 0.40$; Fig. 4d).

Analysis of individual differences in behavioral and neural activity changes before and after task learning (i.e., Session 1 and Session 5–8) revealed a significant negative correlation (Pearson's $r = -0.79$, $p = 0.03$, two-tailed) between the magnitude of decrease in punishment resistance (behavioral change) with increases in the seek action mPFC activity state for risky, unpunished trials (Fig. 4e, left). While individual differences in mPFC responsivity change to punishment were observed, these differences were not associated with behavioral changes (Pearson's $r = -0.12$, $p = 0.80$, two-tailed; Fig. 4e, right).
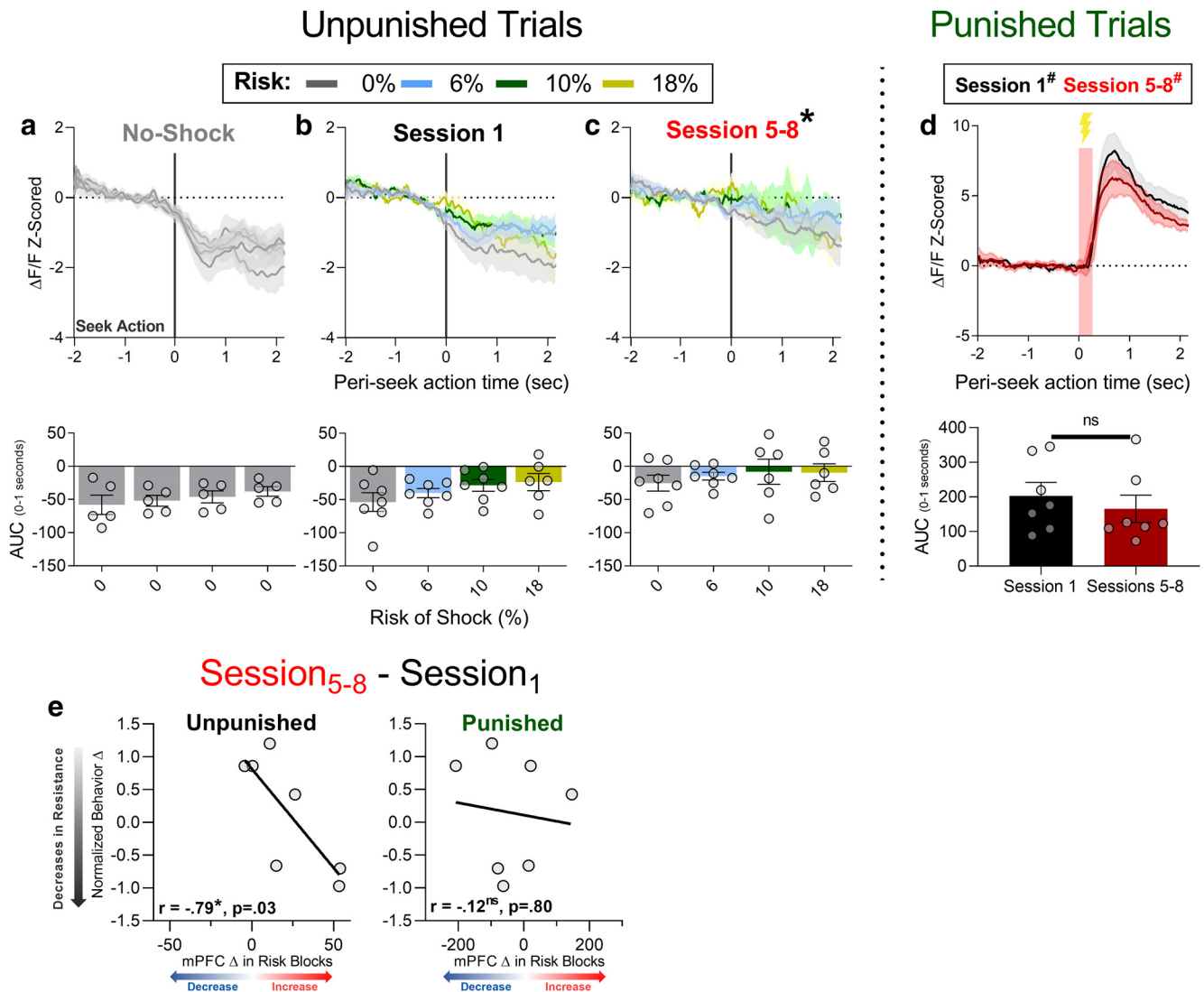
Figure 4. Changes in mPFC neural calcium activity following seek action execution before and after probabilistic punishment task learning. *a–c*, Seek actions that were unpunished resulted in decreases in calcium activity at the time of action execution both before introduction of probabilistic punishment (no-shock, gray) and in the Session 1 (black) before probabilistic punishment was learned. These decreases were attenuated following stabilization (learning) of the probabilistic punishment task in Session 5–8 (red). *d*, Footshock (red bar) after a seek action produced ro-bust increases in calcium activity in the mPFC in Session 1 and Session 5–8 that were significantly greater than the 0% risk block. Bottom, mPFC responses to punishment did not change after learning of the probabilistic punishment task. *e*, Individual differences in performance between Session 5–8 and Session 1 revealed that greater decreases in punishment resistance (shaded black arrow) were associated with greater increases in mPFC activity after risky, unpunished seek actions, but not after punished seek actions. Black line indicates the linear trendline for the data. *p < 0.05, Session 5–8 versus Session 1 or no-shock. #p < 0.05 versus 0% risk of shock. n = 5–7. ns = not significant. Gray circles represent individual data points.

Finally, we examined whether learning-related changes in mPFC calcium activity during seek action generalize to execution of safe actions or to reward delivery. The advantage of the seek-take task structure is that the take response has the same mechanics but carries no punishment risk. Unlike seek trials, mPFC activity in unpunished trials did not change in Session 1 compared with Session 5–8 during take action or reward delivery (effect of session or risk or interaction: unpunished take action F values < 1, p values > 0.48; reward unpunished F values < 1.2, p values > 0.33; Fig. 5a–c). In punished trials, the take response appeared to show a more robust decrease in calcium activity than in unpunished trials. However, this result was not signifi-cant (effect of risk block: $F_{(1,6)} = 4.7$, $p = 0.073$) and should be interpreted with caution as it may be confounded by the large multisecond increases in calcium activity seen from footshock administration (Fig. 5d, top). Nevertheless, significant effects of session (i.e., learning) were not observed for the take action

(paired t test: $t_{(6)} = 1.2$, $p = 0.26$) nor the reward retrieval (paired t test: $t_{(6)} = 1.5$, $p = 0.17$; Fig. 5d, bottom) in punished trials.

Behavioral and pharmacological manipulations of probabilistic punishment task behavior

Value of reward or punishment may change even after action-punishment contingencies are learned. Thus, animals must appropriately adapt their behavior to such changes. To assess whether the current task is sensitive to shifting reward or punish-ment contingencies, we did three additional behavioral experi-ments to manipulate reinforcement and punishment values after task learning.

To decrease the reinforcing value of the food reward, subjects were given 22 h of *ad libitum* access to standard chow in the home cage before the task. This manipulation decreased punish-ment resistance when there was 30%–60% risk of shock (risk
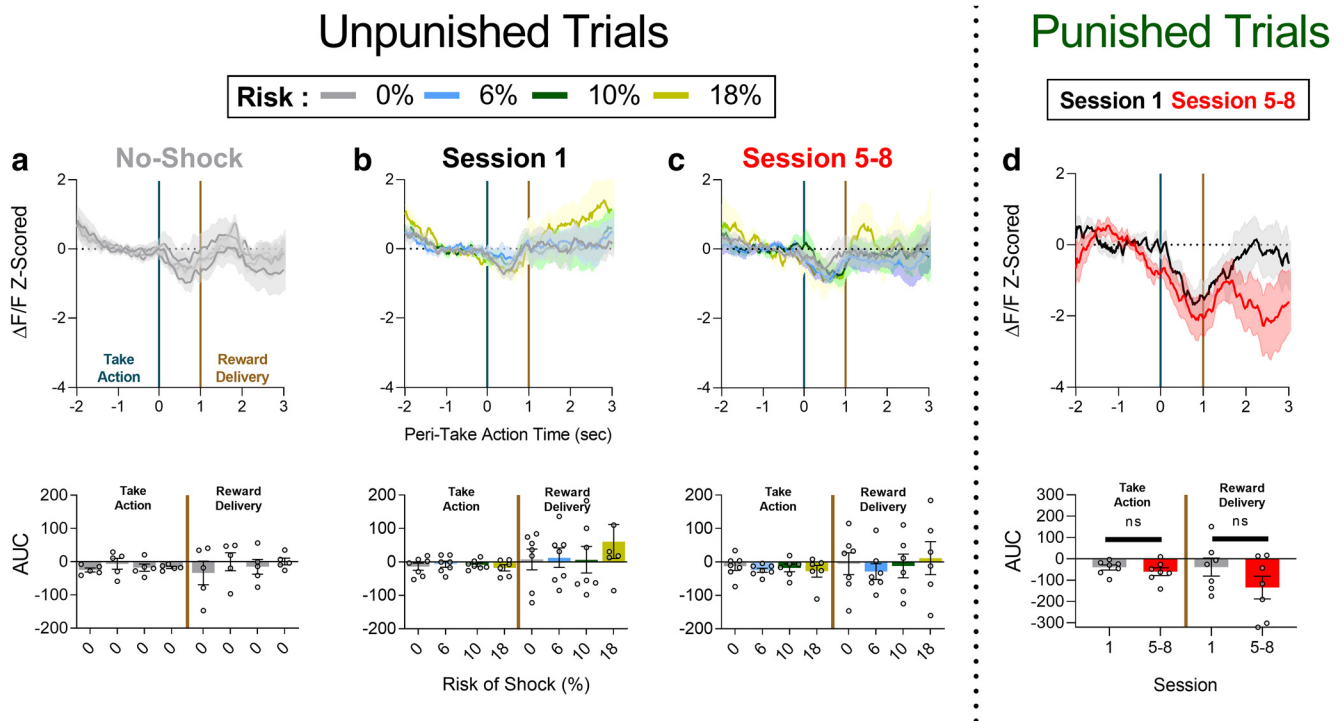
Figure 5. Neural calcium activity in the mPFC for the take action (blue vertical line) and reward delivery (brown vertical line). *a–c*, Take actions and reward delivery produced little effect on mPFC calcium activity and were not influenced by task learning. *d*, Punished trials were not significantly different between Session 1 and Session 5–8 for the take action nor for reward delivery. $n = 5$–7. ns = not significant. Gray circles represent individual data points.

block × satiety interaction: $F_{(2.9,32.4)} = 3.02$, $p = 0.045$; *post hoc p* values < 0.046; Fig. 6a) but not during the 0%-18% risk blocks (*post hoc p* values > 0.24). Satiation also increased latency to complete the seek action (risk × satiety interaction: $F_{(1,10)} = 24.04$, $p < 0.01$; Fig. 6b), an effect seen more profoundly in risk blocks (*post hoc p* < 0.01) but also in the 0% risk block (*post hoc p* = 0.045).

To assess whether the task was sensitive to the value of punishment, and to help produce comparable behavioral levels across subjects, we adjusted the shock intensity until levels of action suppression were similar between subjects. Stable behavior was acquired after 3–12 d of adjustment of shock intensity in ~0.05 mA increments. Subjects reliably responded to manipulation in the intensity of punishment, which overall decreased punishment resistance (titration × risk block interaction: $F_{(2.2,24.4)} = 22.8$, $p < 0.001$; Fig. 6c). *Post hoc* analyses revealed that, overall, task completion decreased after shock adjustment in 30%-60% risk blocks ($p < 0.01$). Seek response latency during risk blocks also increased after titration of shock intensity (titration × risk interaction: $F_{(1,11)} = 32.2$, $p < 0.01$; Fig. 6d), but no effect of titration was observed on seek latency in the 0% risk block (*post hoc p* = 0.21).

To assess whether subjects could flexibly adapt to the omission of punishment, they underwent extinction sessions in which the probabilistic footshock was no longer presented after the seek action. Extinction of shock risk increased task completion in risk blocks (extinction day × risk block interaction: $F_{(3.1,33.6)} = 24.95$, $p < 0.001$; Fig. 6e). This was apparent for blocks with previous shock risk of ≥10% for the first and second extinction sessions. Increases in trial completion were observed in the second extinction day compared with the first extinction day (main effect of session: $F_{(1.47,16.1)} = 134.5$, p values > 0.005). Extinction of probabilistic punishment also resulted in decreases in seek latency (extinction day × risk block interaction: $F_{(1.5,16.4)} = 15.46$, $p < 0.001$; Fig. 6f). As early as the first extinction sessions, seek latency in risk blocks decreased nonsignificantly from 66 to 26 s (*post hoc p* = 0.078 vs shock). However, a continued significant decrease in seek latency to 4 s was observed in Extinction 2 (*post hoc p* > 0.001, Extinction 2 vs shock). No changes were seen in the 0% risk block for seek latency between the shock risk sessions and either of the extinction sessions (all p values > 0.11).

Finally, we asked whether the behavioral responses to probabilistic punishment have relevance to anxiety states by testing the impact of the anxiolytic drug and GABA$_a$ receptor-positive allosteric modulator diazepam (1 and 2 mg/kg) on the probabilistic punishment task. These low doses of diazepam were chosen so that motor behavior would not be impacted to the degree that animals could not complete the task; 1.0 mg/kg diazepam was first tested when all subjects were given the same 0.25 mA shock intensity and did not change trial completion or seek action latency (main effect of treatment: $F_{(1,11)} = 0.14$, $p = 0.71$, $F_{(1,11)} = 0.23$, $p = 0.64$; Fig. 7a,b). However, it was possible that a ceiling effect precluded detection of significant changes for many of the subjects. We therefore tested diazepam after action suppression was titrated using shock intensity (as shown in Fig. 6c). To control for the ascending dose order or additive effects, we also analyzed an additional saline injection session that was at least 48 h after the last diazepam test. Diazepam produced increases in trial completion under probabilistic punishment (main effect of treatment: $F_{(2.1,22.93)} = 6.9$, $p < 0.01$) for both doses (*post hoc p* values < 0.022; Fig. 7c). These anticonflict effects of diazepam complicated interpretation of seek action latency changes, as subjects were completing trials at higher risk blocks than at baseline. Consequently, we assessed seek latency up to the 10% risk block as this was the risk block where all subjects were completing more

than two trials (i.e., did not meet exclusion criteria) on the first saline day. Diazepam attenuated seek latency increases (risk × treatment interaction: $F_{(1.57,17.3)} = 7.76$, $p = 0.02$; Fig. 7d) at the 6% risk block (*post hoc* p values < 0.04) and nonsignificantly at the 10% risk block for 2 mg/kg diazepam (*post hoc* p values = 0.07 $_{1 mg/kg}$ and 0.051 $_{2 mg/kg}$). These effects were not observed at higher risk blocks, although there were increased amounts of variability (data not shown). Importantly, these low doses of diazepam had no effect on locomotor reactivity to the shock with comparable activity levels seen on the saline day (mean$_{Diazepam}$ ± SEM: 2.3 ± 0.24, mean$_{Saline}$ ± SEM: 2.13 ± 0.23; paired t test: $t_{(10)} = 0.52$, $p = 0.62$).

Sex as a biological variable in probabilistic punishment resistance

The work above was done in both male and female rats. After the completion of data collection, without *a priori* hypothesis, we analyzed behavioral data with sex as a factor. While the aim of this study was not to study sex differences, the constructs of anxiety and impulsive reward seeking relevant to this task show stark sex differences in prevalence. Overall, the learning pattern of the task was similar between sexes, with stabilization of both male and female behavior after ~4 or 5 sessions as seen in Figure 8a (main effect of session: $F_{(3.27,84.8)} = 9.98$, $p < 0.001$) which did not interact with sex (sex × session interaction: $F_{(11,285)} = 1.67$, $p = 0.08$).

Once behavior was stable, however, females displayed increased sensitivity to probabilistic punishment compared with males with greater blockwise decreases in trial completion for females compared with males (sex × risk block × session interaction: $F_{(5,130)} = 7.2$, $p < 0.001$; Fig. 8b). This difference was only present when the risk of shock was ≥10% (*post hoc* p values < 0.03) and not during 0% or 6% risk blocks (*post hoc* p values > 0.07). Interestingly, the "sensitive" subgroup observed in Figure 2a exclusively included female subjects, whereas moderate and resistant subgroups contained both males and females albeit in different proportions (Fig. 8c).

Other task behaviors were also significantly different between males and females. While latencies to complete the punished seek action increased during risk blocks compared with no-shock conditions, females showed heightened increases in seek latency during risk blocks in probabilistic punishment sessions (sex × risk block × session interaction: $F_{(1,26)} = 5.7$, $p = 0.02$, *post hoc* $p < 0.01$; Fig. 8d). No differences were observed for seek action
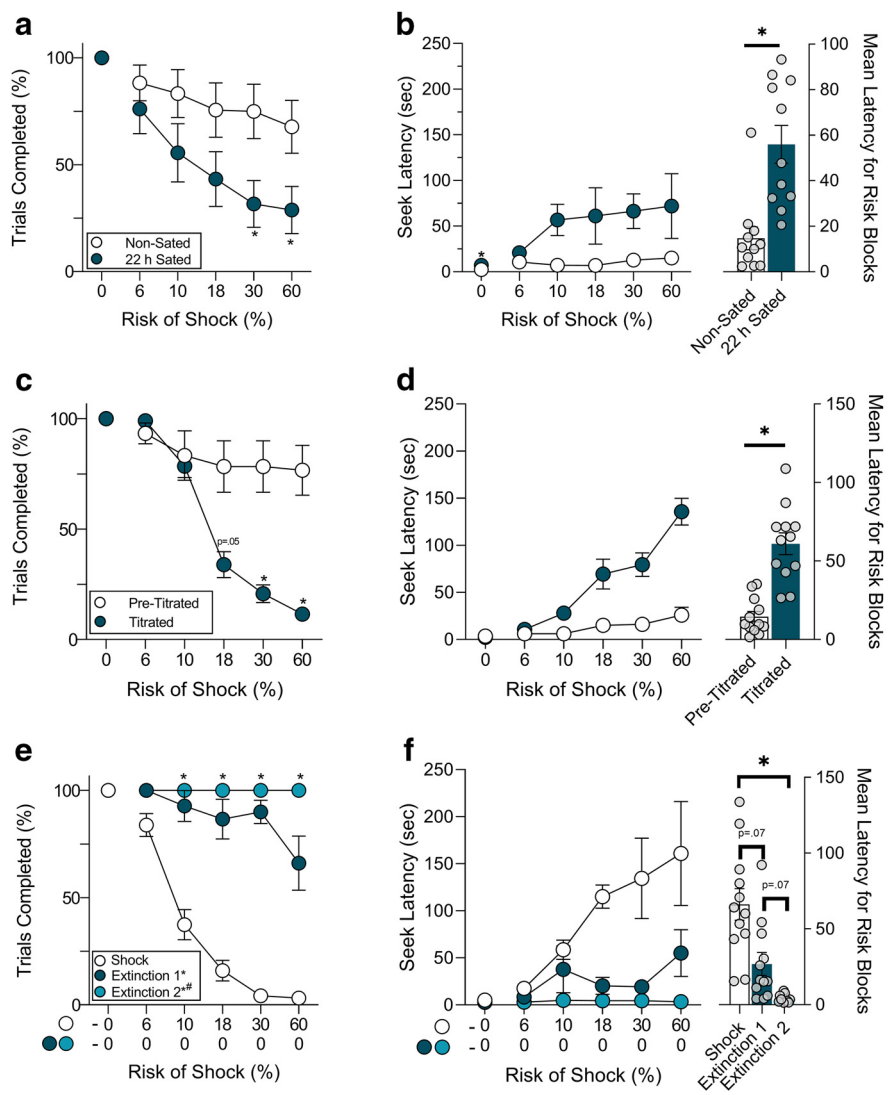

Figure 6. Behavioral effects of manipulation of reinforcer and punisher value. *a*, Decreasing reward value with 22 h of home cage access to *ad libitum* food before a session (blue) decreased trial completion under probabilistic punishment and (*b*) potentiated increases in latency to complete the punished seek action compared with the previous session (white). One subject (female) was excluded from seek latency analysis in *b* due to inability to complete more than one trial in any of the risk blocks. *c, d*, Adjusting the subjective value of the footshock punishment by changing shock intensity (blue) resulted in changes in trial completion under probabilistic punishment and seek action latency. *e*, Extinguishing the risk of punishment by omission of the footshock presentation resulted in an increase in trial completion and (*f*) decreases in seek action latency over the two extinction sessions (blue/light blue). *p < 0.05 versus white symbols. #p < 0.05, Extinction 1 versus Extinction 2. n = 12, except where otherwise noted. Gray circles represent individual data points.

latencies at the 0% risk block when no shocks were given (*post hoc* $p = 0.99$). Females were slower to complete the take action compared with males (main effect of sex: $F_{(1,26)} = 7.03$, $p = 0.014$). While these differences appeared to depend on receiving punishment (sex × trial type interaction: $F_{(2,52)} = 3.34$, $p = 0.043$; Fig. 8e), *post hoc* testing indicated no significant differences between males and females in the 0% risk block, and between unpunished or punished trials in risk blocks (*post hoc* p values > 0.088). Both males and females showed similar latencies to retrieve the food reward, suggesting comparable motivation to acquire the reward (main effect of sex or sex × risk block × session interaction: all F values < 1.1, p values > 0.29; Fig. 8f). To more directly assess underlying reward motivation differences, a cohort was also tested with a PR task following extinction of probabilistic punishment. Males and females displayed
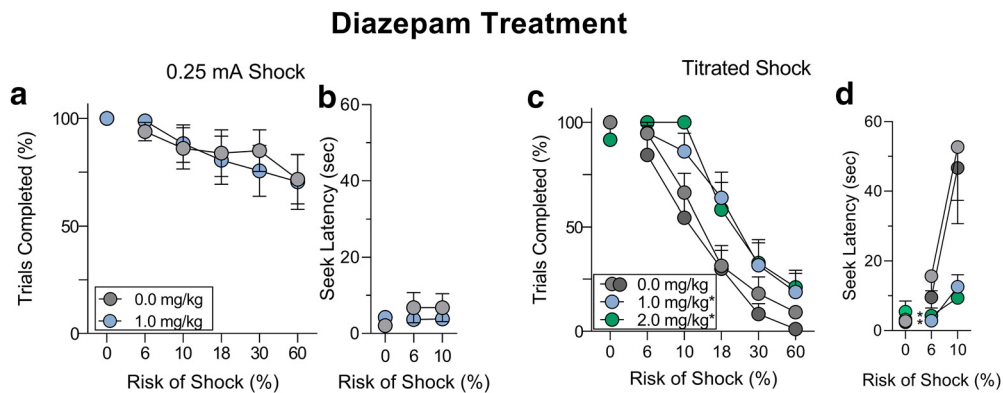
Diazepam Treatment

Figure 7. Effects of systemic diazepam treatment. ***a***, ***b***, Under conditions where the shock intensity was fixed at 0.25 mA for all subjects, no effects of 1.0 mg/kg diazepam (light blue) were observed on trial completion nor seek action latency compared with saline (light gray). ***c***, After behavior was titrated such that all subjects showed sensitivity to probabilistic punishment, 1.0 and 2.0 mg/kg diazepam increased resistance to punishment and (***d***) attenuated increases in seek action latency at lower risks. Performance returned to saline levels after diazepam testing (dark gray). $*p < 0.05$, 1.0 and 2.0 versus 0.0 mg/kg. $n = 12$.

comparable motivation, as measured through PR breakpoint, to obtain the food reward (unpaired t test: $t_{(10)} = 0.72$, $p = 0.48$; Fig. 8g).

To determine whether these effects were due to differences in body size, in one cohort we tested performance after adjusting shock intensity for body weight (1 mA/kg). For males it was observed that trial completion decreased when the risk of shock was $\geq 30\%$ (effect of risk block: $F_{(1.6,11)} = 6.45$, p values < 0.04). Females, however, showed significant decreases at $\geq 10\%$ shock risks (effect of risk block: $F_{(1.6,11.4)} = 11.6$, p values < 0.001). Importantly, body weight-adjusted shock intensity had no effect on the amount of trial completion under probabilistic punishment for either sex (effect of shock intensity: F values < 0.1, p values > 0.7; Fig. 8h), suggesting that body weight is not a critical factor in punishment resistance. This was further supported by a second cohort exposed to a shock threshold procedure where no differences between sexes were seen in shock intensity required to elicit a flinch response (unpaired t test: $t_{(10)} = 0.71$, $p = 0.49$; Fig. 8i). These data suggest that differences in punishment resistance were not due to general sensory differences between males and females. Finally, both sexes showed similar activity in response to the shock and during ITI periods, suggesting similar reactivity to the shock despite body weight differences (unpaired t tests: t values < 2.06, p values > 0.05; Fig. 8j).

While estrous cycle was not systematically investigated in the present study, analysis of female subject data during the 5 consecutive stability sessions (the length of the estrous cycle) did not reveal any consistent fluctuations on a day-to-day basis (data not shown). To assess whether overall trial completion or seek latency was differentially affected by sex in other tasks manipulations, such as shock extinction, satiety tests, and diazepam treatment, we performed additional two-way ANOVAs using sex and manipulation or treatment as factors. No significant sex × treatment or sex × manipulation interactions were observed for overall trial completion (F values < 2.1, p values > 0.12) or mean seek latency in risk blocks (F values < 1, p values > 0.46).

Integration of probabilistic punishment in males and females after titrating shock intensity

Both males and females showed changes in task behavior when we adjusted shock intensity to produce similar overall levels of trial completion (Fig. 9a). These procedures produced nearly identical probabilistic punishment resistance and seek latency increases between sexes (effect of sex or sex × risk block

interaction: F values < 1.5, p values > 0.26; Fig. 9b,c). However, the intensity of shock needed to achieve these comparable behavioral results was significantly higher in males compared with females (unpaired t test: $t_{(10)} = 3.47$, $p < 0.01$; Fig. 9b, inset). To better understand whether males and females integrate risk of punishment into reward-guided actions differently, we modeled action suppression by risk of punishment using a 4PLR similar to those used to assess cost-benefit decision-making (Friedman et al., 2017). We fit a sigmoid to individual data from titrated shock trials, when presumably the subjective suppressive effects of the shock were equal, and revealed three distinct phases: a high action phase, a transition phase, and a low action phase (Fig. 9d–f). Effects of probabilistic punishment on action suppression (trial completion) were well predicted by the model (R^2 = 0.64-0.97), and comparison of small sample size-corrected Akaike information criteria values between the 4PLR model and a linear regression revealed the 4PLR was the preferred model (paired t test: $t_{(11)} = 2.4$, $p = 0.03$). The use of the 4PLR model allowed us to assess whether the integration of punishment risk into behavioral actions differed between sexes. This was achieved by fitting a straight line to the transition from high to low action (i.e., the linear portion of the sigmoid; Fig. 9d). The similar slope steepness (unpaired t test: $t_{(10)} = 0.30$, $p = 0.76$; Fig. 9g) revealed that males and females demonstrated comparable patterns of integrating punishment risk when transitioning from high to low action.

Discussion

Actions we execute to obtain a reward are often associated with the probability of harm occurring. Learning about this probability allows for appropriate computation of risk and guides future action by weighting that risk against the value of obtaining a reward. Impaired learning of this probability may be critical for the pathogenesis of anxiety or reckless and impulsive behavior. To investigate this mode of probability learning, we developed a seek-take instrumental task where the seek action was associated with varying probability of punishment while the take action remained safe and was followed by reward delivery. Animals learned to adapt to probabilistic punishment and exhibited a stable but individualized pattern for inhibiting reward-guided actions as a function of punishment probability. Recording of neural activity in the mPFC during the task revealed that risky action encoding in this region is involved in learning of punishment probability. In particular, the behavioral measure of
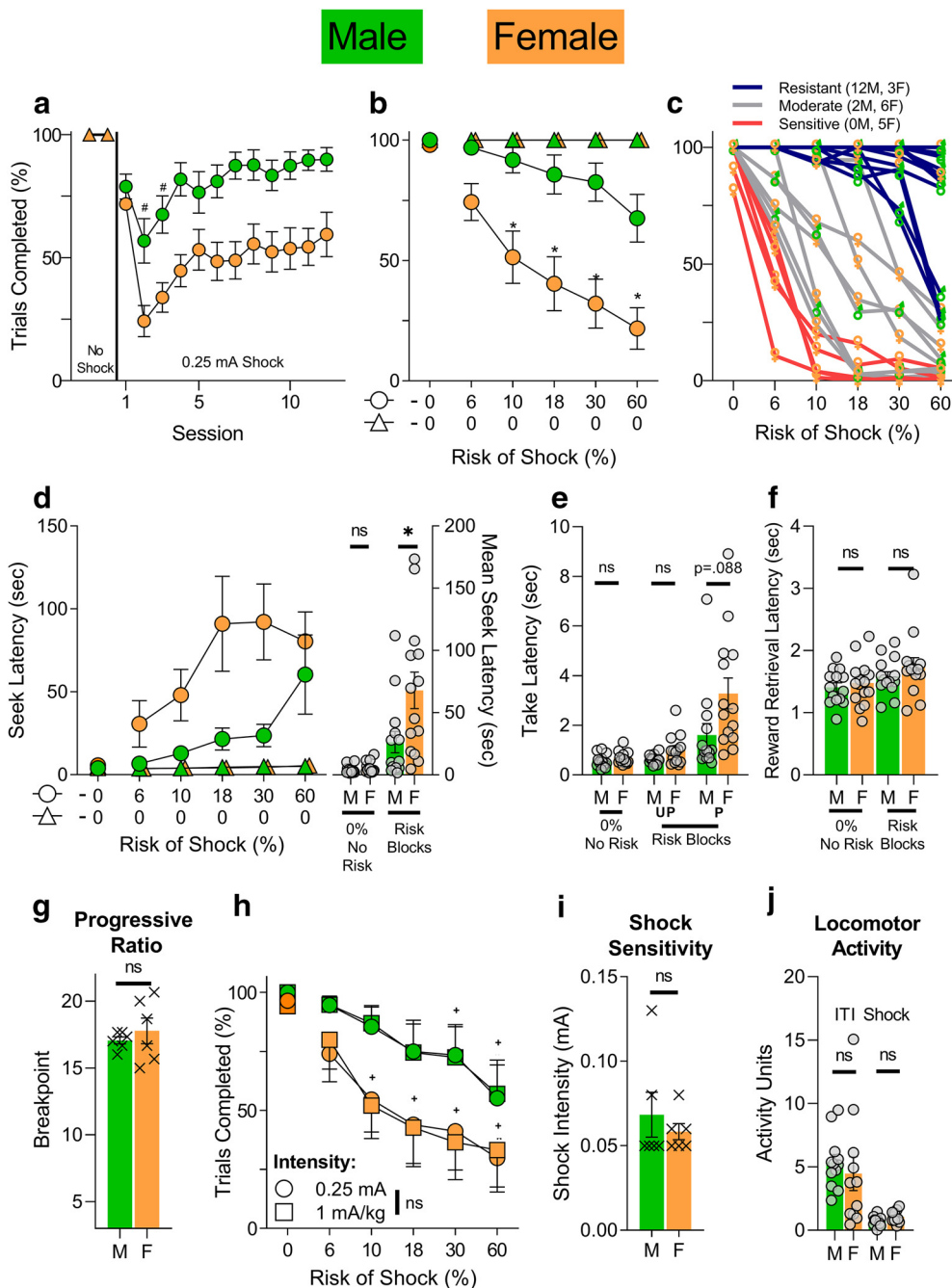
Figure 8. Assessment of sex differences in the probabilistic punishment task. Green represents male (M) data. Orange represents female (F) data. *a*, Patterns of task acquisition were similar between males and females, both of which showed stable behavior after about Sessions 4 and 5. *b*, Generally, males were less sensitive to punishment compared with females; that is, males completed more trials than females after task behavior stabilized (circles), although they did not differ when no footshock risk was present (triangles). *c*, Individual data demonstrating punishment-resistant (blue lines) and moderate (gray lines) subcategorized males were observed, but only females were observed to be resistant, moderate, and sensitive (red lines) to punishment. *d*, Females showed increased latencies to complete the seek action compared with males when the risk of shock was present. *e*, Take action latency increased in females compared with males, although this effect was not uniquely attributed to risk nor if the take action was preceded by a punished (P) or unpunished outcome (UP). *f*, Latency to retrieve the reward did not differ between sexes. *g*, Motivation for sucrose pellets was similar as assessed through a PR task. *h*, Titrating shock intensity based on body weight failed to exert a significant effect in males or females. *i*, Flinch sensitivity to the shock did not differ between sexes. *j*, Sex differences did not appear to be related to general locomotor activity nor shock reactivity during the task. *p < 0.05 versus female. #p < 0.05 versus Session 1. +p < 0.05 versus 0% risk block. n = 12-28. Gray symbols and x's represent individual data points. ns = not significant.

learning was associated with changes in phasic mPFC neuronal activity during risky seek actions but not during the safe take actions or reward delivery. The task was further characterized by establishing that sex is a critical biological variable and that inhibition of behavior as a function of punishment probability is sensitive to manipulations in reinforcer and punisher value, and anxiolytic treatment.

Punishment probability learning during reward-guided actions

Our task provides a tool to measure punishment probability learning. During the first session where animals were exposed to the risk of punishment during the seek action, their behavior remained unchanged until the risk increased to 18%. But in subsequent sessions, animals adjusted their behavior earlier. A
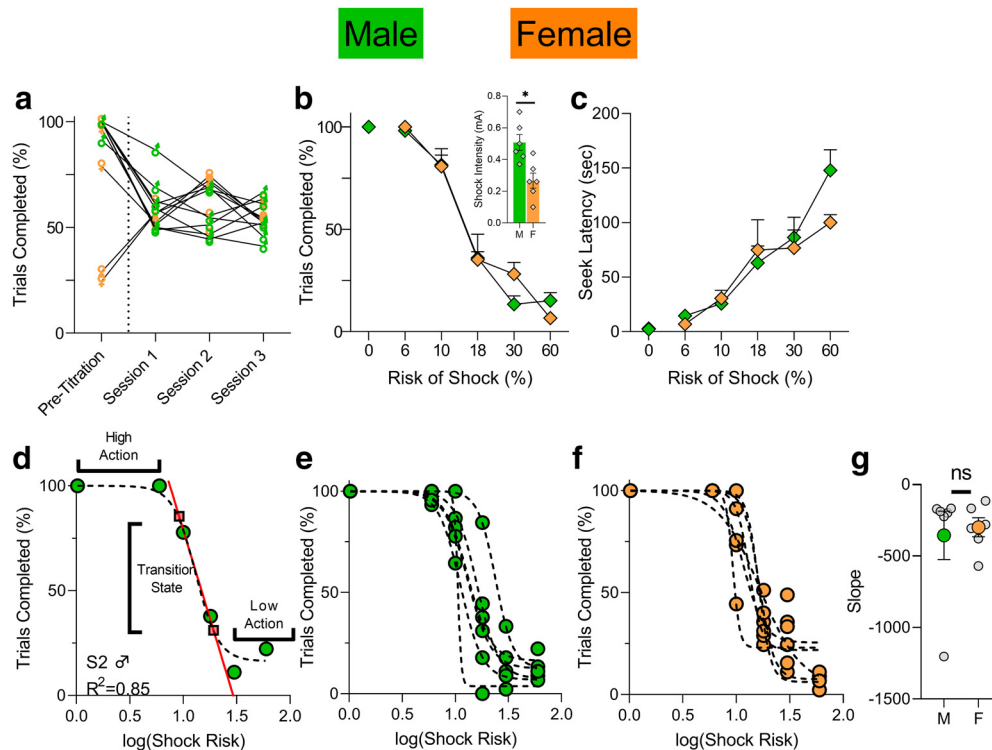
Figure 9. Titration of behavior using shock intensity and modeling for males (green) and females (orange). *a*, Altering shock intensities produced increases or decreases in task completion to result in near 50% suppression of task completion. *b*, *c*, These manipulations produced near-identical changes in trial completion and seek action latency in males and females. *b*, Inset, Higher shock intensity for males was required to produce comparable male and female behavior. *d*, A representative subject's data showing high, low, and transition states and the fit of the 4PLR model (dashed line). The linear portion of the sigmoid was determined by fitting a linear line (red line) to the bend points of the sigmoid (red squares). *e*, *f*, Individual risk effect curves revealed that behavior was well captured using a 4PLR to assess high action, low action, and transition states for behavior during risk of punishment for both males and females. *g*, The steepness of the transition state (i.e., the linear portion of the sigmoid) was not different between males (M) and females (F). Gray symbols and panels *d*, *e*, *f* represent individual data points. $n = 6$/sex. $*p < 0.05$. ns = not significant.

critical aspect of this learning process was that a robust change in behavior was only seen for the risky seek actions, and not for risk-free take actions or reward retrieval actions. This supports the notion that changes in behavior as training progressed were not due to reduced motivation or general motor effects, and were due to punishment probability learning.

The stability of performance after learning allowed us to examine the effect of several manipulations on performance. This led to the following key observations on how reward-guided actions are impacted by punishment probability. First, although behavior stabilized, there was a high level of individual variability, in particular with respect to when subjects stopped responding. Some subjects displayed complete resistance to the risk of punishment, whereas others were more sensitive. Open field behavior, a traditional method of assessing anxiety, was not associated with these individual differences. This indicates that individual differences observed in our task are not due to inherent trait anxiety but relate to learning and expression of punishment probability. Behavioral differences were also absent in the no-shock trials, indicating that motivation to work for reward, in the absence of punishment risk, was not a factor in differences to risk of punishment. Thus, the present task provides a valuable behavioral tool for future investigation of individual differences in the emergence of phenotypes related to anxiety and impulsivity.

Second, after learning, behavior was flexibly influenced by changes in the value of the reinforcer or punisher. The sensitivity of behavior in the current task to value manipulation is consistent with human behavior, providing a valid clinical model for assessing physiological or maladaptive reward and punishment valuation or risk depreciation (Bechara et al., 2002; Hartley and Phelps, 2012).

Third, the increases in the latency of the seek action may provide a novel model for the anxious apprehension state commonly associated with some anxiety disorders. Anticipation of, and adaptation to, potential harm are fundamental features of anxiety (Grillon et al., 2009). Consistent with this notion, the anxiolytic diazepam reduced the impact of punishment risk on seek action execution and latency. In the context of anxiety, another interesting and clinically relevant observation was that, when the risk of shock was removed after learning, seek action latency remained elevated until the second extinction session. The sustained anxiety-like behavior despite extinction of punishment may provide a useful model for assessing normal or pathologic coping with changes in punishment risk over time.

mPFC and learning of punishment probability

Localized lesions and manipulations of neuronal activity have demonstrated that learning of action-outcome associations involves the mPFC (Balleine and Dickinson, 1998; Ostlund and Balleine, 2005). Electrophysiological recordings during instrumental learning show that this learning is expressed at a dynamic level throughout the PFC by emergence of a phasic response during action execution (Sturman and Moghaddam, 2011; del Arco et al., 2017). Moreover, while the adaptive response of individual neurons is both inhibitory and excitatory, the net population response following action execution is largely inhibitory (Mulder et al., 2003; Homayoun and Moghaddam, 2009). After learning,

phasic response of mPFC neurons during action execution is flexible and changes with learning of new rules about outcome contingencies (Simon et al., 2015; del Arco et al., 2017). Given these studies, and that mPFC is implicated in fear conditioning and other models of punishment representation (Corcoran and Quirk, 2007; Park and Moghaddam, 2017a,b), we had hypothesized that learning of punishment probability is, in part, represented in mPFC. Fiber photometry was used to assess changes in population activity because it allows for evaluation of mPFC encoding of shock during learning.

The inhibitory response during peri-seek action periods of the no-shock blocks was consistent with previous studies that have recorded unit activity during action execution in instrumental goal-directed tasks (Simon et al., 2015; Hong et al., 2019), suggesting that our output measure reflects phasic neuronal activity. Alternatively, these inhibitory responses may represent disengagement of prefrontal cortical regions when motor actions become automatic or well learned (Wu et al., 2004; Sturman and Moghaddam, 2011; Kupferschmidt et al., 2017). We observed a significant reduction in this peri-seek action phasic response as punishment risk was learned. Importantly, this change in neuronal activity, similar to the change in behavior, was selective to the seek action. Responses to events that were not associated with risk (i.e., take action and reward delivery) did not significantly change with learning, strengthening the notion that mPFC ensembles are selectively encoding learning of punishment risk. This is a novel learning role for the PFC and consistent with its established role in mediating punishment-related decision-making after learning (Friedman et al., 2015; Orsini et al., 2018).

Further studies are needed to establish the neuronal basis of the reduced inhibitory response in calcium activity seen during probabilistic punishment learning. One possible mechanism is changes in the recruitment of inhibitory interneurons, which then influence the activity of the excitatory pyramidal cells. Another possibility is changes in the recruitment of neuromodulators, such as dopamine and norepinephrine, which generally inhibit the firing rate of spontaneously active neurons. Dopamine and norepinephrine projections to the mPFC are sensitive to stress- and anxiety-provoking contexts (Deutch et al., 1990; Pezze and Feldon, 2004; Morilak et al., 2005). While these modulators generally do not produce overt excitatory or inhibitory responses on their target cells, they may influence ongoing responses.

The excitatory response to shock was consistent with previous studies showing that PFC responds strongly to stressors by increasing glutamate release (Moghaddam, 1993). While it has been proposed that mPFC may adapt and desensitize its response to known stressors (McKlveen et al., 2015), we did not observe an overt reduction in phasic response to the footshock, suggesting that any PFC-mediated learning of probabilistic punishment in this task may be unrelated to adaptation to pain perception. It is, however, important to consider that population-level activity measured in fiber photometry may arise through a variety of processes. For example, while no change in population-level response may indicate a stable response of a brain region, it may also reflect a bidirectional change in both excitatory and inhibitory responses. Consequently, future studies with cell-specific and functional manipulations will advance our understanding of punishment learning.

Sex as a biological variable in probabilistic punishment resistance

Male and female rats learned the probabilistic punishment task at the same rate, but after learning, sex differences indicated greater risky action apprehension and sensitivity to punishment in females. This effect was not related to motivation to obtain reward, body size differences, or basic shock reactivity as adjusting shock for body weight failed to alter punishment resistance. These findings are consistent with, and complement, the emerging data involving sex-related differences in risk taking during reward-seeking behavior (Van den Bos et al., 2013; Orsini and Setlow, 2017; Becker and Charthoff, 2018).

The sex differences in seek action latency, however, dissipated when shock intensity was individually adjusted to produce comparable levels of overall trial completion. This suggested that, if the subjective value of the punishment is normalized, there is no sex difference in transition from resisting punishment to inhibiting behavioral responding. This concept was verified using a four-parameter sigmoid model, where we observed that the transition from high to low action states had a similar steepness in both sexes. The sigmoidal pattern revealed in this model is similar to that reported in choice-based decision-making tasks (Friedman et al., 2017).

Given the sexual dimorphisms seen in symptoms of psychiatric illnesses, including impulsivity and anxiety, our overall observation on sex differences in punishment-resistant behavior highlights the importance of using biological sex as a variable to inform our understanding of the neuronal basis of reward-motivated actions.

In conclusion, the present study provides a powerful behavioral model for determining biological and environmental factors that influence the resistance of reward-seeking behavior to probabilistic punishment. The learning of this form of probabilistic contingencies appears to involve changes in the mPFC as this region flexibly adapted its response to risky actions as contingencies were learned. Our data further emphasizes the importance of studying the impact of sex as a biological variable.

References

Azrin NH, Holz WC, Hake DF (1963) Fixed-ratio punishment. J Exp Anal Behav 6:141–148.

Babayan BM, Uchida N, Gershman SJ (2018) Belief state representation in the dopamine system. Nat Commun 9:1891.

Baeg EH, Kim YB, Jang J, Kim HT, Mook-Jung I, Jung MW (2001) Fast spiking and regular spiking neural correlates of fear conditioning in the medial prefrontal cortex of the rat. Cereb Cortex 11:441–451.

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407–419.

Bechara A, Dolan S, Hindes A (2002) Decision-making and addiction: II. Myopia for the future or hypersensitivity to reward? Neuropsychologia 40:1690–1705.

Becker JB, Chartoff E (2018) Sex differences in neural mechanisms mediating reward and addiction. Neuropsychopharmacology 44:166–183.

Chen BT, Yau HJ, Hatch C, Kusumoto-Yoshida I, Cho SL, Hopf FW, Bonci A (2013) Rescuing cocaine-induced prefrontal cortex hypoactivity prevents compulsive cocaine seeking. Nature 496:359–362.

Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499:295–300.

Cooper SE, Goings SP, Kim JY, Wood RI (2014) Testosterone enhances risk tolerance without altering motor impulsivity in male rats. Psychoneuroendocrinology 40:201–212.

Corcoran KA, Quirk GJ (2007) Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears. J Neurosci 27:840–844.

Del Arco A, Park J, Wood J, Kim Y, Moghaddam B (2017) Adaptive encoding of outcome prediction by prefrontal cortex ensembles supports behavioral flexibility. J Neurosci 37:8363–8373.

Deutch AY, Clark WA, Roth RH (1990) Prefrontal cortical dopamine depletion enhances the responsiveness of mesolimbic dopamine neurons to stress. Brain Res 521:311–315.

Ersche KD, Gillan CM, Jones PS, Williams GB, Ward LH, Luijten M, de Wit S, Sahakian BJ, Bullmore ET, Robbins TW (2016) Carrots and sticks fail to change behavior in cocaine addiction. Science 352:1468–1471.

Friedman A, Homma D, Gibb LG, Amemori KI, Rubin SJ, Hood AS, Riad MH, Graybiel AM (2015) A corticostriatal path targeting striosomes controls decision-making under conflict. Cell 161:1320–1333.

Friedman A, Homma D, Bloem B, Gibb LG, Amemori KI, Hu D, Delcasso S, Truong TF, Yang J, Hood AS, Mikofalvy KA, Beck DW, Nguyen N, Nelson ED, Toro Arana SE, Vorder Bruegge RH, Goosens KA, Graybiel AM (2017) Chronic stress alters striosome-circuit dynamics, leading to aberrant decision-making. Cell 171:1191–1205.e28.

Gabriel DB, Freels TG, Setlow B, Simon NW (2019) Risky decision-making is associated with impulsive action and sensitivity to first-time nicotine exposure. Behav Brain Res 359:579–588.

Grillon C, Pine DS, Lissek S, Rabin S, Bonne O, Vythilingam M (2009) Increased anxiety during anticipation of unpredictable aversive stimuli in posttraumatic stress disorder but not in generalized anxiety disorder. Biol Psychiatry 66:47–53.

Hartley CA, Phelps EA (2012) Anxiety and decision-making. Biol Psychiatry 72:113–118.

Homayoun H, Moghaddam B (2009) Differential representation of Pavlovian-instrumental transfer by prefrontal cortex subregions and striatum. Eur J Neurosci 29:1461–1476.

Hong DD, Huang WQ, Ji AA, Yang SS, Xu H, Sun KY, Cao A, Gao WJ, Zhou N, Yu P (2019) Neurons in rat orbitofrontal cortex and medial prefrontal cortex exhibit distinct responses in reward and strategy-update in a risk-based decision-making task. Metab Brain Dis 34:417–429.

Jean-Richard-Dit-Bressel P, Killcross S, McNally GP (2018) Behavioral and neurobiological mechanisms of punishment: implications for psychiatric disorders. Neuropsychopharmacology 43:1639–1650.

Kupferschmidt DA, Juczewski K, Cui G, Johnson KA, Lovinger DM (2017) Parallel, but dissociable, processing in discrete corticospinal inputs encodes skill learning. Neuron 96:476–489.e5.

Lopes G, Bonacchi N, Frazão J, Neto JP, Atallah BV, Soares S, Moreira L, Matias S, Itskov PM, Correia PA, Medina RE, Calcaterra L, Dreosti E, Paton JJ, Kampff AR (2015) Bonsai: an event-based framework for processing and controlling data streams. Front Neuroinform 9:7.

Matias S, Lottem E, Dugue GP, Mainen ZF (2017) Activity patterns of serotonin neurons underlying cognitive flexibility. 6:e20552. Elife

Mulder AB, Norquist RE, Orgut O, Pennartz C (2003) Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. Behav Brain Res 146:77–88.

McEwen BS, Morrison JH (2013) The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. Neuron 79:16–29.

McKlveen JM, Myers B, Herman JP (2015) The medial prefrontal cortex: coordinator of autonomic, neuroendocrine and behavioural responses to stress. J Neuroendocrinol 27:446–456.

Menegas W, Akiti K, Amo R, Uchida N, Watabe-Uchida M (2018) Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. Nat Neurosci 21:1421–1430.

Moghaddam B (1993) Stress preferentially increases extraneuronal levels of excitatory amino acids in the prefrontal cortex: comparison to hippocampus and basal ganglia. J Neurochem 60:1650–1657.

Morilak DA, Barrera G, Echevarria DJ, Garcia AS, Hernandez A, Ma S, Petre CO (2005) Role of brain norepinephrine in the behavioral response to stress. Prog Neuropsychopharmacol Biol Psychiatry 29:1214–1224.

Orsini CA, Setlow B (2017) Sex differences in animal models of decision making. J Neurosci Res 95:260–269.

Orsini CA, Willis ML, Gilbert RJ, Bizon JL, Setlow B (2016) Sex differences in a rat model of risky decision making. Behav Neurosci 130:50–61.

Orsini CA, Heshmati SC, Garman TS, Wall SC, Bizon JL, Setlow B (2018) Contributions of medial prefrontal cortex to decision making involving risk of punishment. Neuropharmacology 139:205–216.

Ostlund SB, Balleine BW (2005) Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. J Neurosci 25:7763–7770.

Park J, Moghaddam B (2017a) Impact of anxiety on prefrontal cortex encoding of cognitive flexibility. J Neurosci 345:193–202.

Park J, Moghaddam B (2017b) Risk of punishment influences discrete and coordinated encoding of reward-guided actions by prefrontal cortex and VTA neurons. Elife 6:e30056.

Pascoli V, Terrier J, Hiver A, Luscher C (2015) Sufficiency of mesolimbic dopamine neuron stimulation for the progression to addiction. Neuron 88:1054–1066.

Paxinos G, Watson C (1998) The rat brain in stereotaxic coordinates, Ed 4. San Diego: Academic.

Pelloux Y, Everitt BJ, Dickinson A (2007) Compulsive drug seeking by rats under punishment: effects of drug taking history. Psychopharmacology (Berl) 194:127–137.

Pelloux Y, Murray JE, Everitt BJ (2015) Differential vulnerability to the punishment of cocaine related behaviours: effects of locus of punishment, cocaine taking history and alternative reinforcer availability. Psychopharmacology (Berl) 232:125–134.

Pezze MA, Feldon J (2004) Mesolimbic dopaminergic pathways in fear conditioning. Prog Neurobiol 74:301–320.

Richardson NR, Roberts DC (1996) Progressive ratio schedules in drug self-administration studies in rats: a method to evaluate reinforcing efficacy. J Neurosci Methods 66:1–11.

Sebaugh J, McCray P (2003) Defining the linear portion of a sigmoid-shaped curve: bend points. Pharmaceut Statist 2:167–174.

Simon NW, Gilbert RJ, Mayse JD, Bizon JL, Setlow B (2009) Balancing risk and reward: a rat model of risky decision making. Neuropsychopharmacology 34:2208–2217.

Simon NW, Wood J, Moghaddam B (2015) Action-outcome relationships are represented differently by medial prefrontal and orbitofrontal cortex neurons during action execution. J Neurophysiol 114:3374–3385.

Soares S, Atallah BV, Paton JJ (2016) Midbrain dopamine neurons control judgment of time. Science 354:1273–1277.

Söderpalm B, Engel JA (1988) Biphasic effects of clonidine on conflict behavior: involvement of different alpha-adrenoceptors. Pharmacol Biochem Behav 30:471–477.

St Onge JR, Floresco SB (2010) Prefrontal cortical contribution to risk-based decision making. Cereb Cortex 20:1816–1828.

Sturman DA, Moghaddam B (2011) Reduced neuronal inhibition and coordination of adolescent prefrontal cortex during motivated behavior. J Neurosci 31:1471–1478.

van den Bos R, Homberg J, de Visser L (2013) A critical review of sex differences in decision-making tasks: focus on the Iowa Gambling Task. Behav Brain Res 238:95–108.

Vanderschuren LJ, Minnaard AM, Smeets JA, Lesscher HM (2017) Punishment models of addictive behavior. Curr Opin Behav Sci 13:77–84.

Verharen JP, van den Heuvel MW, Luijendijk M, Vanderschuren L, Adan RA (2019) Corticolimbic mechanisms of behavioral inhibition under threat of punishment. J Neurosci 39:4353–4364.

Volkow ND, Fowler JS, Wang GJ (2003) The addicted human brain: insights from imaging studies. J Clin Invest 111:1444–1451.

Wu T, Kansaku K, Hallett M (2004) How self-initiated memorized movements become automatic: a functional MRI study. J Neurophysiol 91:1690–1698.