Unique features of stimulus-based probabilistic reversal learning

Carl Harris[1#], Claudia Aguirre[2#], Saisriya Kolli[2], Kanak Das[2], Alicia Izquierdo[2-5*], Alireza Soltani[1*]

[1] Department of Psychological and Brain Science, Dartmouth College, Hanover, New Hampshire, United States 03755
[2] Department of Psychology, University of California-Los Angeles, Los Angeles, California, United States 90095
[3]The Brain Research Institute, University of California-Los Angeles, Los Angeles, California, United States 90095
[4]Integrative Center for Learning and Memory, University of California-Los Angeles, Los Angeles, California, United States 90095
[5]Integrative Center for Addictions, University of California-Los Angeles, Los Angeles, California, United States 90095

#co-first authors

*co-senior authors

Funding
This work was supported by R01 DA047870 (Soltani and Izquierdo). We acknowledge UCLA's Graduate Division Graduate Summer Research Mentorship and Graduate Research Mentorship programs (Aguirre), the NSF Graduate Research Fellowship Program (Aguirre), and the Whitcome Summer Research Fellowship (Kolli).

Disclosure
There is no conflict of interest or need for disclosure.

Corresponding author email: aizquie@psych.ucla.edu

Abstract

Reversal learning paradigms are widely-used assays of behavioral flexibility with their probabilistic versions being more amenable to studying integration of reward outcomes over time. Prior research suggests differences between initial learning and learning following reversals including higher learning rates, a greater need for inhibitory control, and more perseveration after reversals. However, it is not well-understood what aspects of stimulus-based reversal learning are unique to reversals, and whether and how differences between initial and post-reversal learning depend on reward probability. Here, we used a visual probabilistic discrimination and reversal learning paradigm during which male and female rats selected between a pair of stimuli associated with different reward probabilities. We compared various measures of accuracy, rewards collected, omissions, latencies, win-stay/lose-shift strategies, and indices of perseveration between two different reward probability schedules. We found that discrimination (pre-reversal) and reversal learning are behaviorally more unique than similar: longer choice latencies following incorrect trials, lesser win-stay and lose-shift strategies employed, and more perseveration in early reversal learning. Additionally, fit of choice behavior using reinforcement learning models revealed a lower sensitivity to the difference in subjective reward values (greater exploration) and higher learning rates for the reversal phase. Interestingly, a consistent reward probability group difference emerged with a richer environment associated with longer reward collection latencies than a leaner environment. We also replicated previous reports on sex differences in reversal learning. Future studies should systematically compare the neural correlates of fine-grained behavioral measures to reveal possible dissociations in how the circuitry is recruited in each phase.

Keywords*: reversal learning, probabilistic learning, win-stay, lose-shift, stimulus-reward, reinforcement learning, orbitofrontal cortex, DREADDs*

Introduction

A critical feature of goal-directed, instrumental behavior is the ability to discriminate stimuli that predict reward from those that do not, and further, to flexibly update the response to those stimuli if predictions become inaccurate. Discrimination learning paradigms in rodents often involve pairing of an action (e.g., lever pressing, digging in a bowl, nosepoking stimuli on a touchscreen, or displacing an object) with an outcome (e.g., a desirable food reward). In typical paradigms, two or more stimuli are presented concurrently and the subject learns about the features of the stimuli that bring about reward and those that do not (e.g., nosepoking S_A results in a better probability of reward than S_B; pressing left lever yields better payout than right lever; scent A is more rewarded than scent B) (Alvarez & Eichenbaum, 2002; Dalton, Wang, Phillips, & Floresco, 2016; Eichenbaum, Fagan, & Cohen, 1986; Izquierdo et al., 2013; Schoenbaum, Chiba, & Gallagher, 2000; Schoenbaum, Nugent, Saddoris, & Setlow, 2002). With training, subjects become increasingly proficient at discrimination, in line with the associative rules imposed by the experimenter. The stimulus-reward rules can be deterministic (e.g., S_A results in a sucrose pellet reward and S_B does not) or probabilistic (e.g., S_A results in a better probability of reward over S_B), with deterministic and probabilistic schedules of reinforcement producing marked dissociations in the neural circuity recruited (Averbeck & Costa, 2017; Costa, Dal Monte, Lucas, Murray, & Averbeck, 2016).

In reversal learning paradigms (Izquierdo, Brigman, Radke, Rudebeck, & Holmes, 2017; Izquierdo & Jentsch, 2012), after either reaching a discrimination learning criterion for accuracy (Brushfield, Luu, Callahan, & Gilbert, 2008; Izquierdo et al., 2013; Stolyarova et al., 2019), a number of consecutive correct responses (Dalton, Phillips, & Floresco, 2014; Dalton et al., 2016), or a fixed block length of trials (Farashahi et al., 2017; Soltani & Izquierdo, 2019), the stimulus-reward contingencies are reversed. At reversal, the trained response no longer results in a better probability of reward, though it usually remains the more frequent one because of initial discrimination training. Indeed, usually reversals are acquired more slowly than the original discrimination, and younger subjects are quicker to learn than older ones (Brushfield et al., 2008; Schoenbaum, Setlow, Saddoris, & Gallagher, 2006). Perhaps partly due to this

difference with original learning, reversal learning is considered unique in its requirement of flexibility, because it involves the subject inhibiting the prepotent response and, instead, responding to stimuli that were previously irrelevant. Other popular views are that discrimination and reversal learning phases occupy different task "spaces" (Wilson, Takahashi, Schoenbaum, & Niv, 2014) and that the phases differ in the likelihood that changes in contingencies will occur (Jang et al., 2015). Importantly, probabilistic learning and reversal paradigms, in particular, are more amenable to the application of reinforcement learning (RL) models that can estimate parameters for choice behavior based on the integration of previous rewarded and non-rewarded trials (Lee, Seo, & Jung, 2012). Despite various favored accounts, it is not well-understood how *behaviorally unique* reversal learning is compared to original (initial discrimination) learning. For example, latency measures can be used to dissociate attention, decision speed, and motivation via analyses of the time taken to initiate trials, choose a stimulus, and collect reward, respectively (Aguirre et al., 2020). Here, we explored if such detailed trial-by-trial measures of latencies and omissions differ across learning phases. Further, we probed if there are differences between reward probability schedules on these various measures by comparing animals discriminating between Better and Worse options, rewarded with a probability schedule of 0.90 vs. 0.30, compared to 0.70 vs. 0.30. We also analyzed win-stay and lose-shift strategies, and perseveration and repetition metrics in each learning phase. Finally, we investigated if RL models fit these learning phases differently. We studied this in both male and female animals.

We found a higher perseveration index, lower probability of staying following choice of the better stimulus, and a higher probability of staying following choice of the worse stimulus, unique to early reversal learning. Worse choice latencies (a proxy for decision speed) were also longer in early reversal learning. We also report less use of win-stay and lose-shift strategies in the reversal phase, compared to the initial discrimination phase. As for RL models, we found a lower sensitivity to the difference in subjective reward values and higher learning rates during reversal than initial discrimination. There were also sex differences, particularly in reversal learning, which replicate previous findings using this paradigm. Interestingly the only consistent probability group difference we observed across discrimination and reversal learning phases

was in motivation (i.e. a richer environment was associated with longer reward collection latencies than a leaner environment). Collectively, our fine-grained analyses suggest that trial-by-trial behavioral measures of latencies and strategies may be particularly sensitive metrics to pair with neural correlate data in reversal learning. These measures may also be revealing in uncovering the unique substrates of flexible learning.

Methods

Subjects

Subjects were N=25 adult male (n=13) and female (n=12) Long-Evans rats (Charles River Laboratories) aged > post-natal-day (PND) 60 at the start of testing. Rats arrived to the vivarium between PND 40-60. The rats included in this report served as controls for two different experiments: one was a cohort of water (H_2O)-only drinking (n=8 female and n=8 male) rats that served as controls in an ethanol study (see *2-Bottle Choice*) and the other was a cohort of rats (n=5 female and n=4 male) that experienced surgical procedures (see *Surgery*), serving as controls for a study targeting the orbitofrontal cortex (OFC) with DREADDs. Importantly, all rats were the same age (PND 140-155) when pretraining commenced, and further, all rats were part of experiments that ran in parallel, minimizing differences between cohorts.

Before any treatment, all rats underwent a 3-day acclimation period during which they were pair-housed and given food and H2O *ad libitum*, and remained in cages with no experimenter interference. Following this 3-day acclimation period, animals were handled for 10 min per animal for 5 consecutive days. During the handling period, the animals were also provided food and H2O *ad libitum*. After the handling period, animals were individually-housed under standard housing conditions (room temperature 22–24° C) with a standard 12 h light/dark cycle (lights on at 6am). Following either 2-bottle choice or surgery, rats were tested on probabilistic discrimination and reversal learning, as below. All procedures were conducted in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and the Chancellor's Animal Research Committee at the University of California, Los Angeles.

2-Bottle Choice

Home cages were modified to allow for the placement of two bottles for drinking, whereas standard housing allows for only one bottle. Rats (n=16; 8 male and 8 female) included in this analysis were singly-housed and given access to 2 H_2O bottles simultaneously (no ethanol) for 10 weeks, the same duration that experimental animals were provided the choice of ethanol vs. H_2O. Weight of bottles was measured three times per week to measure consumption amounts, compared to a control cage placed on the same rack to account for leakage. Rats were not monitored for weight during this time.

Surgery

Viral Constructs

Rats (n=9; 4 male and 5 female) included in the present comparison were singly-housed and allowed to express DREADDs or eGFP in OFC for 6 weeks, the same duration experimental animals that were treated with clozapine-N-oxide (CNO) were allowed to express DREADDs. An adeno-associated virus AAV8 driving the hM4Di-mCherry sequence under the CaMKIIa promoter was used to express DREADDs on OFC neurons (AAV8-CaMKIIa-hM4D(Gi)-mCherry, packaged by Addgene) (Addgene, viral prep #50477-AAV8). A virus lacking the hM4Di DREADD gene and only containing the green fluorescent tag eGFP (AAV8-CaMKIIa-EGFP, packaged by Addgene) was also infused into OFC in separate groups of animals as a null virus control. Altogether there were n=5 DREADD/vehicle, n=2 eGFP/vehicle, and n=2 eGFP/CNO included in this control group.

Surgical Procedure

Infusion of AAV virus containing DREADD or eGFP (n=9) in OFC was performed using aseptic stereotaxic techniques under isoflurane gas (1-5% in O_2) anesthesia prior to any behavioral testing experience. Before surgeries were completed, all animals were administered 5mg/kg s.c. carprofen (NADA #141–199, Pfizer, Inc., Drug Labeler Code: 000069) and 1cc saline. After being placed in the stereotaxic apparatus (David Kopf; model 306041), the scalp was incised and retracted. The skull was leveled to ensure

that bregma and lambda were in the same horizontal plane. Small burr holes were drilled in the skull above the infusion target. Virus was bilaterally infused at a rate of 0.02 µl per minute for a total volume of 0.2 µl per hemisphere into OFC (AP = +3.7; ML= ±2.0; DV = −4.6, relative to bregma). After each infusion, 10 min elapsed before the syringe was pulled up.

Food Restriction

Five days prior to any behavioral testing, rats were placed on food restriction with females on average maintained 12-14 grams/ day and males given 16-18 grams/ day of chow. Food restriction level remained unchanged throughout behavioral testing, provided animals completed testing sessions. Water remained freely available in the home cage. Animals were weighed every other day and monitored closely to not fall below 85% of their maximum, free-feeding weight.

Learning

Pretraining. Behavioral testing was conducted in operant conditioning chambers outfitted with an LCD touchscreen opposing the sugar pellet dispenser. All chamber equipment was controlled by customized ABET II TOUCH software.

The pretraining protocol, adapted from established procedures (Stolyarova & Izquierdo, 2017), consisted of a series of phases: Habituation, Initiation Touch to Center Training (ITCT), Immediate Reward Training (IMT), designed to train rats to nosepoke, initiate a trial, and select stimuli. During habituation, rats were required to eat five pellets out of the pellet dispenser inside the chambers within 15 min before exposure to any stimuli on the touchscreen. ITCT began with the display of white graphic stimuli on the black background of the touchscreen. During this stage, a trial could be terminated for one of two reasons: if a rat touched the displayed image, or if the image display time (40 s) ended, after which the stimulus disappeared, a black background was displayed, and a 10 s inter-trial interval (ITI) ensued. If the rat did not touch within 40 s this was scored as an *initiation omission*. IMT began in the same way as ITCT, but the disappearance of the white graphic stimulus was now paired with the onset of a target image immediately

to the left or right of the stimulus (i.e. forced-choice). During this stage, a trial could be terminated for one of three reasons. First, if a rat touched the center display (i.e. white graphic stimulus) and touched the image displayed on either side, after which there was a dispensation of one sucrose pellet and illumination of the tray-light. Second, if the rat failed to touch the center white graphic stimulus after the display time ended (40 s), after which the stimulus disappeared, a black background was displayed, and a 10 s ITI ensued, scored as an *initiation omission*. Third, if the image display time (60 s) ended, after which the stimulus disappeared, a black background was displayed, and a 10 s ITI ensued, scored as a *choice omission*. Rats could also fail to respond to the center stimulus within 40 s during this phase (i.e. initiation omission, as in the previous phase). For habituation pretraining, the criterion for advancement was collection of all 5 sucrose pellets. For ITCT, the criterion to the next stage was set to 60 rewards consumed in 45 min. The criterion for IMT was set to 60 rewards consumed in 45 min across two consecutive days.

Probabilistic Discrimination Learning. After completion of all pretraining schedules, rats were advanced to the discrimination (initial) phase of the PRL task, in which they would initiate a trial by touching the white graphic stimulus in the center screen (displayed for 40 s), and choose between two visual stimuli presented on the left and right side of the screen (displayed for 60 s) counterbalanced between trials, assigned as the Better or Worse options, rewarded with a sucrose pellet, with probability $p_R(B)= 0.90$ or 0.70 and $p_R(W)=0.30$, respectively. If a trial was not initiated within 40 s, it was scored as an initiation omission. If a stimulus was not chosen, it was scored as a choice omission, and a 10 s ITI ensued. If a trial was not rewarded, a 5 s time-out would follow, subsequently followed by a 10 s ITI. Finally, if a trial was rewarded, a 10 s ITI would follow after the reward was collected (**Figure 1**). The criterion was set to 60 or more rewards consumed and selection of the better option in 70% of the trials or higher during a 60 min session across two consecutive days. After reaching the criterion for the discrimination phase, the rats advanced to the reversal phase beginning on the next session. Notably, one animal in the 90-30 group and five animals in the 70-30 group did not meet discrimination criterion and were force-reversed after 25+ days.

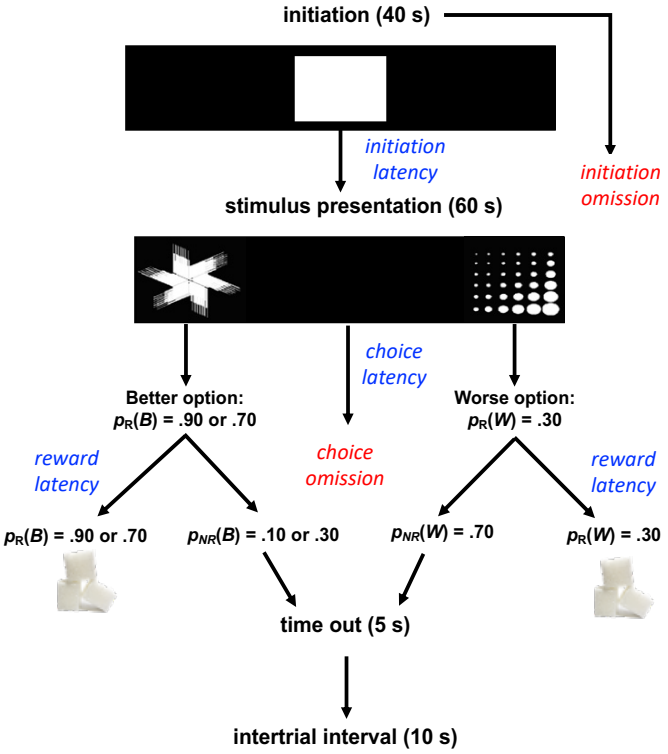

Figure 1. Task design. Schematic of probabilistic learning task. Rats initiated a trial by nosepoking the center stimulus (displayed for 40 s) and then selected between two visual stimuli pseudorandomly that were presented on either the left and right side of the screen (displayed for 60 s), assigned as the Better (B) and Worse (W) options. Correct nosepokes were rewarded with a sucrose pellet with probability $p_R(B)$=0.90 or 0.70 versus $p_R(W)$=0.30. If a trial was not rewarded [$p_{NR}(B)$ or $p_{NR}(W)$], a 5 s time-out would commence. If a stimulus was not chosen, it was considered a choice omission and a 10 s ITI would commence. Rats could also fail to initiate a trial, in which case, it was scored as an initiation omission. If a trial was rewarded, a 10 s ITI would follow reward collection. Other prominent measures collected on a trial-by-trial basis were trial initiation latency (time to nosepoke the center white square), choice latency (time to select between the two stimuli), and reward latency (time to collect reward in the pellet magazine).

Probabilistic Reversal Learning. After the discrimination phase, the rats advanced to the reversal phase during which rats were required to remap stimulus-reward contingencies and adapt to reversals in the reward probabilities. The stimuli associated with the p_R (B)=0.90 or 0.70 probability (i.e. better option), would now be associated with a $p_R(W)$=0.30 probability of being rewarded (i.e. worse option). Consistent with prior literature showing freely-behaving rodents exhibit slow learning on probabilistic reversals with visual stimuli (Aguirre et al., 2020), most animals from either cohort did not meet a 70% criterion before the termination of the study, so we limited our analyses to the most complete dataset in all animals.

Data Analyses and computational modeling

MATLAB (MathWorks, Natick, Massachusetts; Version R2019b) was used for all statistical analyses; MATLAB and Prism8 were used for figure preparation. Data were analyzed with a series of mixed-effects General Linear Models (GLM); omnibus analyses across discrimination and reversal phases in early learning (first 7 sessions), and then individual analyses within each phase separately. Learning data were analyzed with GLM (*fitglme* function; Statistics and Machine Learning Toolbox; MathWorks, Natick, Massachusetts; Version R2017a), with learning phase (discrimination vs reversal), probability group (90-30 vs 70-30), and sex (male vs female) as fixed factors, and individual rat as random factor. All Bonferroni post-hoc tests were corrected for number of comparisons. Statistical significance was noted when p-values were less than 0.05, and p-values between 0.05 and 0.07 were reported as a trend, or marginally significant. Major dependent variables include: probability correct, number of rewards (sucrose pellets earned), total initiation omissions (failure to initiate a trial), total choice omissions (failure to select a stimulus), and median latencies (to initiate a trial, to nosepoke the correct stimulus, to nosepoke the incorrect stimulus, and to collect reward). The latter we refer to as initiation-, correct-, incorrect-, and reward latencies, respectively.

Each trial was classified as a *win* if an animal received a sucrose pellet, and as a *loss* if no reward was delivered. Decisions were classified as better if the animal chose the more rewarding stimulus (stimulus with the larger probability of reward) and *Worse* if it chose the less rewarding stimulus. We classified decisions as *Stays* when a rat chose the same stimulus on the subsequent trial and as *Shifts* when it switched to the other alternative. From these first-order measures we were able to construct win-stay, the probability of choosing the same stimulus on the following trial after being rewarded, and lose-shift, the probability of choosing the alternative stimulus after not receiving a reward. These were further parsed into better or worse win-stay or lose-shift, depending on whether the win-stay/lose-shift followed selection of the better option or the worse option. Because we were primarily interested in the differences between the initial phases of discrimination and reversal learning, only the first seven sessions (i.e., early phase learning) for each animal were included in our analysis on response to reward feedback.

Metrics of repetition and perseveration. We also used two additional higher-order measures: repetition index and perseveration index. We calculated repetition index (Soltani, Noudoost, & Moore, 2013) as the difference between the actual probability of staying and the chance level of staying:

$$RI = P(stay) - (P(better) \times P(better) + P(worse) \times P(worse))$$

where $P(stay)$ is the actual probability of staying equal to the joint probability of choosing the option on two consecutive trials, $P(better)$ is the probability of choosing the more rewarded stimulus, and $P(worse)$ is the probability of choosing the less rewarded stimulus. We further parsed repetition index into RI_B and RI_W, which accounts for differing tendency to stay on better and worse options:

$$RI_B = P(stay, better) - (P(better) \times P(better))$$

$$RI_W = P(stay, worse) - (P(worse) \times P(worse))$$

Additionally, we introduced a perseverative index, analogous to the perseverative index defined in Brigman et al. (2008) and Brigman et al. (2010) as the number of same stimulus choices following a loss divided by the number of such "runs," or the ratio of first-presentation stimulus errors to consecutive errors.

We used these two measures, in addition to the probability of staying, to analyze repetitive behavior. As above, we only used the first seven sessions in calculating repetition measures.

Reinforcement learning model. We utilized two simple reinforcement learning models to capture animals' learning and choice behavior. In each model, the estimated subjective reward value of a choice was determined by prior choices and corresponding reward feedback. Specifically, subjective reward values were updated based on reward prediction error, the discrepancy between actual and expected reward. For each observed choice, we updated the subjective reward value of the corresponding option. Accordingly, if the more rewarded stimulus was chosen then $V_{choice} = V_{better}$, and if the less rewarded stimulus was selected $V_{choice} = V_{worse}$, where V was the option subjective reward value. We used the following learning rules to update V_{Choice}:

RL1: Model with a single learning rate. On a given trial t, the subjective reward value of the chosen stimulus is updated using the following function:

$$V_{Choice}(t+1) = V_{Choice}(t) + \alpha\big(r(t) - V_{Choice}(t)\big)$$

Where $V_{Choice}(t)$ is the subjective reward value on trial t, $r(t)$ is 1 (0) if trial t is rewarded (not rewarded) and α is the single learning rate.

RL2: Model with separate learning rates for rewarded and unrewarded trials. On a trial t, the following learning rule was used:

$$V_{Choice}(t+1) = \begin{cases} V_{Choice}(t) + \alpha_{Rew}\big(1 - V_{Choice}(t)\big), r(t) = 1 \\ V_{Choice}(t) - \alpha_{Unrew} V_{Choice}(t)), \quad r(t) = 0 \end{cases}$$

where α_{Rew} is the learning rate for rewarded trials and α_{Unrew} is the learning rate for unrewarded trials.

We then applied the following decision rule to determine the probability of selecting the better option on trial t:

$$P_{better}(t) = \frac{1}{1+exp\big(-\sigma(V_{better} - V_{worse})\big)},$$

where σ is the inverse temperature parameter or sensitivity to the difference in subjective reward values.

We fit each model's parameters to minimize the negative log likelihood of observed responses. The learning rate parameters (α, α_{Rew}, and α_{Unrew}) were restricted to values between 0.01 and 0.60, and σ was constrained to values greater than 0. Initial parameter values were initially selected from this range, then fit using MATLAB's *fmincon* function. For each set of trials, we performed 20 iterations with different initial, randomly selected parameter values to avoid local minima, and the best fit was selected from the iteration with the lowest negative log likelihood (*LL*). Additionally, we also calculated the Bayesian information criterion (BIC) for each fit, defined as

$$BIC = kln(\mathrm{n}) - 2(\mathrm{LL})$$

Where *k* is the number of parameters (2 in RL1, and 3 in RL2) and *n* is the number of trials.

To determine the relationship between learning rates and sensitivity to difference in subjective reward values, we calculated the Pearson correlation coefficient. For RL1, we calculated a single correlation

coefficient between α and σ. For RL2, we calculated two correlation coefficients: between α_{Rew} and σ and α_{Unrew} and σ.

Results

Comparisons of Sessions to Criterion, Omnibus measures

90-30 Reward Probability Control Groups. We first conducted statistical analyses to ensure the surgical control groups (s-groups) [i.e., DREADD/VEH, eGFP (CNO+VEH)] that constituted the 90-30 reward probability group did not differ significantly on omnibus measures: number of sessions to criterion (to 70% accuracy), probability of choosing the better option, and number of rewards collected during discrimination and reversal learning (i.e., first seven days of learning common to all rats). There was no effect of surgical control group (GLM: $\beta_{s\text{-}group}$ =4.17, p=0.46) or sex (GLM: β_{sex} =8.17, p=0.18) on sessions to reach criterion, probability correct (GLM: $\beta_{s\text{-}group}$ =0.004, p=0.96; GLM: β_{sex} =0.06, p=0.33), or number of rewards collected (GLM: $\beta_{s\text{-}group}$ =-6.39, p=0.78; GLM: β_{sex} =2.80, p=0.93) for discrimination learning. Similarly, for reversal learning we did not find any effect of surgical control group (GLM: $\beta_{s\text{-}group}$ =-0.17, p=0.96) or sex (GLM: β_{sex} =-2.67, p=0.48) on sessions to reach criterion, probability correct (GLM: $\beta_{s\text{-}group}$ =0.02, p=0.72; GLM: β_{sex} =0.10, p=0.22), or number of rewards collected (GLM: $\beta_{s\text{-}group}$ =-18.52, p=0.24; GLM: β_{sex} =-7.81, p=0.69). Given the lack of group differences between surgical control groups within the 90-30 reward probability cohort, we collapsed data across these groups for further analyses.

Overall measures. In total, the 70-30 probability group performed a total of 847 sessions, 417 during the discrimination phase and 430 during reversal (an average of 26.1 $\pm$ 3.1 during the discrimination phase and 26.9 $\pm$ 1.71 during the reversal phase). The 90-30 probability group performed 270 sessions, 118 during discrimination and 152 during reversal (an average of 13.1 $\pm$ 2.6 during discrimination and 16.9 $\pm$ 1.65 reversal), **Figure 2B** and **2D**. For discrimination learning, we found no effect of group (p=0.24), sex (p=0.21), or significant group*sex interaction (p=0.49) on sessions to reach criterion. The 90-30 reward probability group required an average of 11.56 $\pm$ 2.56 (M $\pm$ SEM) sessions while the 70-30 reward probability group required an average of 21.13 $\pm$ 3.26 sessions to reach a 70% criterion; females required

an average of 11.92 ± 2.47 sessions, while males required an average of 23.92 ± 3.59 sessions. For reversal learning, there was a significant effect of group (GLM: β_{group} =8.28, p=0.02), with the 90-30 reward probability group requiring more session to reach criterion on average (17.44 ± 1.68) than the 70-30 reward probability (23.50 ± 1.95). However, there was no effect of sex (p=0.20), with males (17.67 ± 1.97) and females (24.69 ± 1.80) requiring a comparable amount of sessions to reach criterion for reversal learning, and no significant group*sex interaction (p=0.41).

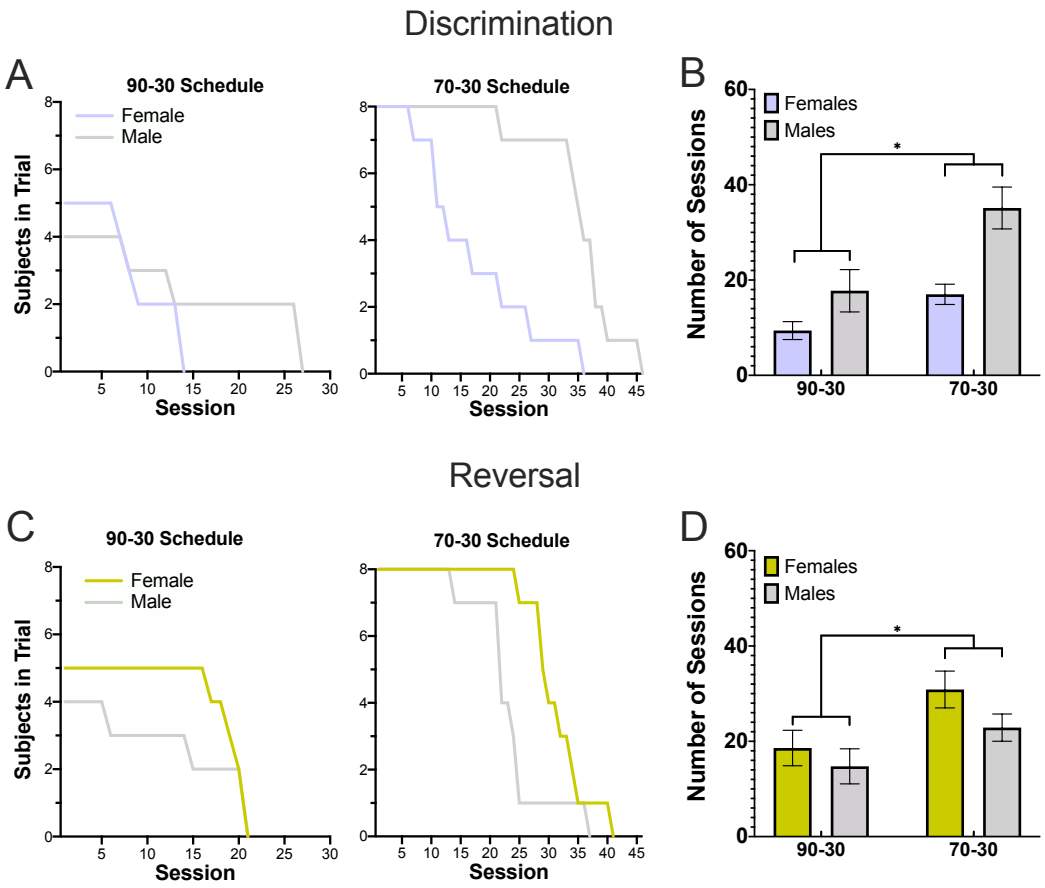

Figure 2. Greater number of completed sessions for the 70-30 reward probability schedule in both discrimination and reversal learning. (A-B) Plotted are the number of subjects per session (A) and the number of sessions to criterion (B) during the discrimination (pre-reversal) phase. The 70-30 reward probability group completes significantly more sessions during discrimination than the 90-30 group. **(C-D)** The same as A-B but during for reversal learning. The 70-30 reward probability group again completes significantly more sessions than the 90-30 group. Bars indicate ± SEM *p ≤0.05

Comparisons of Accuracy and Rewards Collected

Comparisons between early Discrimination & Reversal Learning. We first fitted GLMs that combined analysis of the first 7 days of learning across both phases (discrimination and reversal). For *probability correct,* we found that all rats exhibited learning by demonstrating an increase in choosing the better option across days (GLM: β_{day} =0.05, p=0.01). Overall animals chose the better option more in the discrimination phase (GLM: β_{phase} =-0.23, p=0.002) than the reversal phase. For the *number of rewards*, all rats increased their number of rewards collected across days for both learning phases (GLM: β_{day} =5.98, p=0.004). There were no difference between reward probability group (GLM: β_{group} =-2.20, p=0.86), sex (GLM: β_{sex} =29.13, p=0.09), or learning phase (GLM: β_{phase} =-0.64, p=0.96), nor were there any factor interactions on probability correct or number of rewards.

Comparisons within discrimination learning. We fitted GLMs to test the effect of reward probability group (90-30, 70-30), sex (female, male), sessions (referred to as day, below), and their 2-way and 3-way interactions on all the dependent measures of interest, for the discrimination phase separately. For *probability correct*, we found that all rats exhibited learning by demonstrating an increase in choosing the better option across days, (GLM: β_{day} =0.036, p<0.001), irrespective of reward probability group or sex. There was no effect of reward probability group (p=0.56), or sex (p=0.61), and no group*sex (p=0.27), group*day (p=0.48), sex*day (p=0.64), or group*sex*day (p=0.32) interaction on probability correct (**Figure 3A**). For the *number of rewards*, all rats exhibited a trend to increase rewards collected across days (GLM: β_{day} =2.095, p=0.067). There was no effect of reward probability group (p=0.68), or sex (p=0.64), and no group*sex (p=0.38), group*day (p=0.64), sex*day (p=0.27), or group*sex*day (p=0.24) interaction on number of rewards (**Figure 3B**).

Comparisons within Reversal Learning. As with the discrimination phase, we fitted separate GLMs to test the effect of reward probability group (90-30, 70-30), sex (female, male), sessions, and their 2-way and 3-way interactions on all the dependent measures of interest, for the reversal phase separately. For *probability correct*, we found that all rats exhibited learning by demonstrating an increase in choosing the better option across days (GLM: β_{day} =0.017, p<0.001). There was a significant effect of sex (GLM: β_{sex}

=0.076, p=0.01) with males improving faster than females, but no effect of probability group (p=0.90), and no group*sex (p=0.77) interaction. There were significant group*day (GLM: $\beta_{group*day}$ =-0.006, p=0.02) and sex*day (GLM: $\beta_{sex*day}$=-0.007, p<0.01) interactions, with post-hoc analyses revealing that the 90-30 group increasingly chose the better option at a faster rate than the 70-30 group, although this was only marginally significant (p=0.057), but there was no group*sex*day interaction (p=0.12) on reward probability correct (**Figure 3C**). For the *number of rewards*, unlike the trend in discrimination, we found no increase in rewards collected across days for reversal (p=0.32). Additionally, there was no significant effect of reward probability group (p=0.32), or sex (p=0.19), and no group*sex (p=0.47), group*day (p=0.80), sex*day (p=0.68), or group*sex*day (p=0.90) interaction on number of rewards, **Figure 3D**.

Summary. The omnibus comparison within each phase of learning revealed that all animals demonstrated learning by increasing their accuracy (i.e. probability of choosing the better option) and increasing the number of rewards collected across days, regardless of phase, reward probability group, or sex, during early discrimination and reversal learning. Separate analyses of discrimination and reversal learning yielded similar results, with the exception that when the phases were analyzed individually and entirely (not just early learning), we found that animals did not significantly increase their number of rewards collected during the reversal phase.

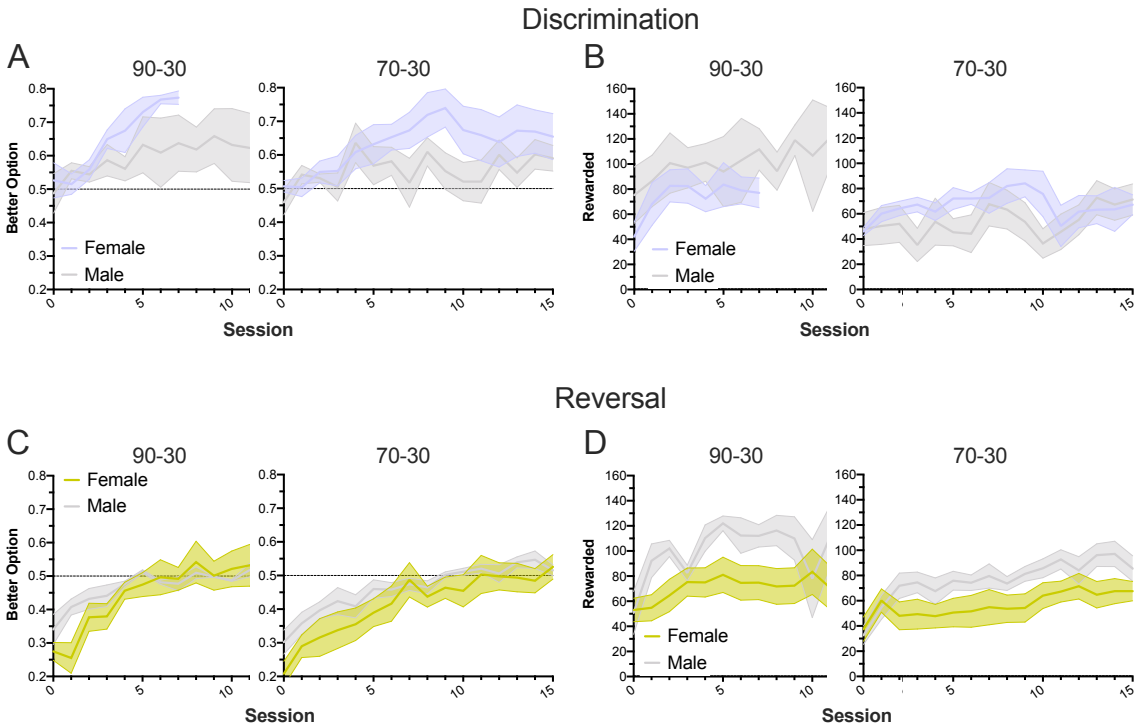

Figure 3. Both reward probability groups and both sexes increase selection of the better option over time, at a similar rate, with fewer rewards collected over time in the reversal phase. (A-B) Proportion of better option selections (A) and number of rewards in a session (B) in the discrimination (pre-reversal) phase, showing the first 10 and 15 trials. Both groups increase selection of the better option and receive more rewards per session over time, with no significant differences between groups or sexes. **(C-D)** Same as A-B, but in the reversal phase. Again, animals in both reward probability groups improve accuracy over time, with no differences by group or sex. However, fewer rewards are collected over time in the reversal phase.

Comparisons of Omissions and Latencies

Comparisons between early Discrimination & Reversal Learning. Similar to above, we fitted GLMs that combined analysis of the first 7 days of learning across both phases (discrimination and reversal). For total number of *initiation omissions*, there was a significant phase*group*sex (GLM: $\beta_{phase*group*sex}$ =-158.10, p=0.01) interaction, with post-hoc comparisons revealing 70-30 males exhibited more initiation omissions in the discrimination phase than the reversal phase (p=0.01). There was also a significant group*sex interaction on this measure (GLM: $\beta_{group*sex}$ =164.90, p=0.02), but post-hoc tests were not significant after accounting for the number of comparisons. There was no effect of phase

(p=0.30), reward probability group (p=0.53), or sex (p=0.26), and no significant phase*group (p=0.16), or phase*sex (p=0.97) interactions on this measure. For total number of *choice omissions*, there was no effect of phase (p=0.55), reward probability group (p=0.44), or sex (p=0.18), and no significant phase*group (p=0.23), phase*sex (p=0.23), group*sex (p=0.11), or phase*group*sex (p=0.22) interactions.

Next we analyzed median initiation latencies, correct and worse choice latencies, and reward collection latencies. For *initiation latencies*, we found a marginal group*sex interaction (GLM: $\beta_{group*sex}$ =3.35, p=0.059), but no effect phase (p=0.10), or reward probability group (p=0.77), and no significant phase*group (p=0.27), phase*sex (p=0.54), group*sex (p=0.06), or phase*group*sex (p=0.17) interactions. For *better choice latencies*, there was a significant phase*sex interaction (GLM: $\beta_{phase*sex}$ =-0.21, p=0.01), with males exhibiting longer better choice latencies in the discrimination phase compared to the reversal phase (p=0.01), but no effect of phase (p=0.65), reward probability group (p=0.48), or sex (p=0.28), and no significant phase*group (p=0.06), group*sex (p=0.71), or phase*group*sex (p=0.08) interactions. For *worse choice latencies*, there was a significant effect of phase (GLM: β_{phase} =0.10, p=0.003), with animals in the reversal phase exhibiting longer latencies than in the discrimination phase, but no effect of reward probability group (p=0.51), sex (p=0.45), or significant phase*group (p=0.26), phase*sex (p=0.07), group*sex (p=0.88), phase*group*sex (p=0.49) interactions. For *reward collection latencies*, we found a significant effect of reward probability group (GLM: β_{group} =-0.58, p=0.0001), with the 90-30 group taking longer to collect reward than the 70-30 group, but no effect of phase (p=0.48) or sex (p=0.41). There was a significant phase*sex interaction (GLM: $\beta_{phase*sex}$ =-0.23, p=0.02), with females exhibiting longer latencies than males in the reversal phase (p=0.01), and a phase*group*sex interaction (GLM: $\beta_{phase*group*sex}$ =0.29, p=0.02), with 90-30 males in the discrimination phase than 90-30 males in the reversal phase, but no significant phase*group (p=0.69) or group*sex (p=0.23) interaction.

Comparisons within Discrimination Learning. We also used GLMs to fit the total number of initiation and choice omissions for the discrimination phase separately. For total number of *initiation omissions*, there was a marginal group*sex interaction (GLM: $\beta_{group*sex}$ =462.88, p=0.058) with males in the

70-30 group exhibiting more initiation omissions than males in the 90-30 group (p=0.04) (**Figure 4A**). There was no effect of reward probability group (p=0.55), or sex (p=0.93). For total number of *choice omissions*, there was no effect of reward probability group (p=0.22), or sex (p=0.13), and no group*sex interaction (p=0.12).

We next analyzed median initiation latencies, correct and worse choice latencies, and reward collection latencies. For *initiation latencies*, there was a significant effect of sex (GLM: β_{sex} =-1.726, p=0.049**; Figure S2A**) with females taking longer to initiate trials, but no effect of reward probability group (p=0.39; **Figure S1A**) or group*sex interaction (p=0.09), **Figures 4B**. For *better choice latencies*, there was no effect of reward probability group (p=0.66; **Figure S1B**), sex (p=0.70; **Figure S2B**) or group*sex interaction (p=0.42; **Figures 4C)**. Similarly, for *worse choice latencies*, we did not find an effect of reward probability group (p=0.47; **Figure S1C**) or sex (p=0.77; **Figure S2C**), and no group*sex interaction (p=0.37). For *reward collection latencies*, there was a significant effect of reward probability group (GLM: β_{group} =-0.60, p<0.0001; **Figure S1D**) with the 90-30 group taking longer to collect reward than the 70-30 group (**Figure 4D**), but no effect of sex (p=0.27; **Figure S2D**) and no group*sex interaction (p=0.12).

Comparisons within Reversal Learning. We next analyzed total number of initiation and choice omissions for reversal learning. For total number of *initiation omissions*, we found an effect of sex (GLM: β_{sex} =-189.3, p=0.04) with females exhibiting more omissions than males, but no effect of reward probability group (p=0.12) or group*sex interaction (p=0.35) (**Figure 4E**). For total number of *choice omissions*, we found a group*sex interaction (GLM: $\beta_{group*sex}$ =-3.675, p=0.02) with 90-30 males making more omissions than 70-30 males (p=0.056), but no effect of reward probability group (p=0.30) or sex (p=0.28).

We then analyzed median initiation latencies, correct and worse choice latencies, and reward collection latencies for the reversal phase only. For *initiation latencies*, as for discrimination, we found no significant effect of sex (GLM: β_{sex} =-1.450, p<0.001; **Figure S2E**) with females taking longer to initiate trials, but no effect of reward probability group (p=0.81**; Figure S1E**) or group*sex interaction (p=0.96), **Figure 4F**. For *better choice latencies*, there was no significant effect of reward probability group (p=0.16;

Figure S1F), sex (p=0.11; **Figure S2F**) or significant group*sex interaction (p=0.31; **Figure 4G)**. For *worse choice latencies*, there was no effect of reward probability group (p=0.50; **Figure S1G**) or sex (p=0.19; **Figure S2G**), and no group*sex interaction (p=0.54). For *reward collection latencies*, there was a significant effect of reward probability group (GLM: β_{group} =-0.488, p<0.001; **Figure S1H**) with the 90-30 group taking longer to collect reward than the 70-30 group (**Figure 4H**), but no effect of sex (p=0.16; **Figure S2H**) and no group*sex interaction (p=0.12).

Summary. Worse choice latencies were longer in the reversal phase compared to the discrimination phase in early days of learning for all animals. Better choice latencies were longer in discrimination than reversal learning for males, but not for females in early learning. Both of these findings point to differences in decision speed between the two phases. Robust reward probability group differences emerged for reward collection latencies, with the 90-30 group (i.e. animals experiencing the richer reward schedule) exhibiting longer latencies than the 70-30 group, a pattern that was observed both during early learning, across both phases, as well as in discrimination and reversal phases when analyzed in their entirety. More subtle effects were also observed; females exhibited longer reward latencies during early reversal compared to males, whereas 90-30 males exhibited longer latencies during early discrimination learning than in early reversal learning. Males in the 70-30 group exhibited a greater number of initiation omissions in the discrimination phase than the reversal phase, however, this pattern was restricted to early learning and not seen throughout each phase. Conversely, sex differences in initiation latencies revealed that females in both probability groups exhibited longer initiation latencies than males, with this pattern observed overall in each phase, but not in early learning.

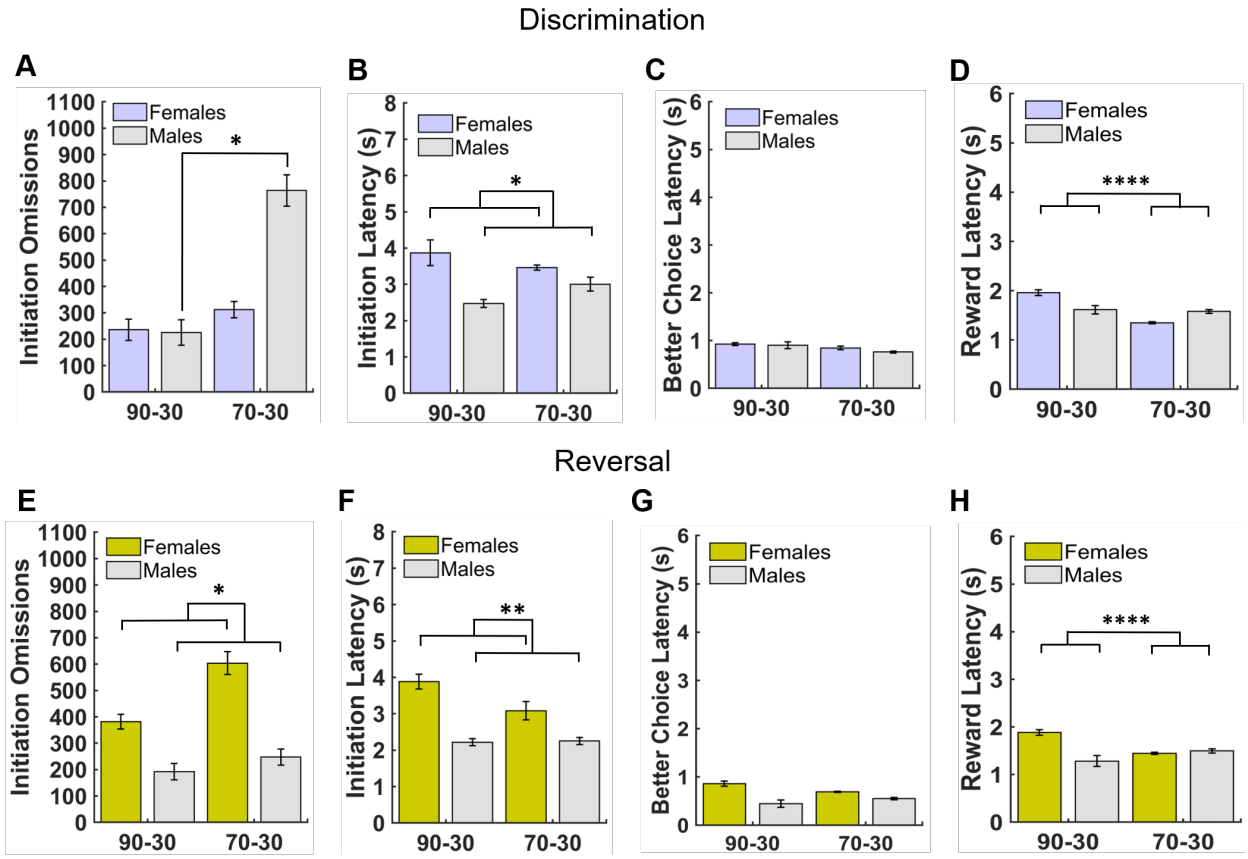

Figure 4. Similar patterns in Omissions and Latencies for reward probability groups during both discrimination and reversal learning. (A) There was a reward probability group difference in initiation omissions in the discrimination (pre-reversal) phase, with more omissions in 70-30 males. (B) There were significant sex differences in initiation latencies in discrimination, with females exhibiting longer initiation latencies than males. (C) There were no group or sex differences on better choice latencies in discrimination. (D) There were significant probability group differences in reward collection latencies in discrimination, with the 90-30 reward probability group exhibiting longer latencies than the 70-30 reward probability group. (E) There were sex differences in initiation omissions in reversal learning, with females exhibiting more initiation omissions than males. (F) There were sex differences in initiation latencies in reversal learning, with females exhibiting longer initiation latencies than males. (G) There were no group or sex differences on better choice latencies in the reversal phase. (H) There were significant probability group differences in reward collection latencies in the reversal phase, with the 90-30 reward probability group exhibiting longer latencies than the 70-30 reward probability group. Bars indicate ± SEM $^{\#}p=0.06$, $*p \leq 0.05$, $**p \leq 0.01$ $***p \leq 0.001$, $****p \leq 0.0001$

Comparisons of Response to Reward Feedback

Comparisons between early Discrimination & Reversal Learning. For win-stay, we found a marginally significant effect of phase on employing the win-stay strategy (GLM: β_{phase} =-0.13, p=0.052), with animals using this strategy more in the discrimination phase than in the reversal phase, a significant

effect of reward probability group (GLM: β_{group} =-0.07, p=0.04), with animals in the 90-30 reward probability group using this strategy more than those in the 70-30 reward probability group, but no effect of sex (p=0.12). There was also a significant phase*group interaction (GLM: $\beta_{phase*group}$ =0.11, p=0.04), with animals in the 90-30 reward probability group using this strategy more in the discrimination phase than the reversal phase (p=0.001), but no significant phase*sex (p=0.48), group*sex (p=0.55), or phase*group*sex (p=0.68) interactions. For lose-shift, there was a significant effect of phase (GLM: β_{phase} =-0.16, p=0.0001), with animals employing this strategy more in the discrimination learning phase than the reversal phase. There was a significant effect of reward probability group (GLM: β_{group} =-0.06, p=0.002), with the 90-30 group using the lose-shift strategy more than the 70-30 group. Finally, there was also an effect of sex (GLM: β_{sex} =0.04, p=0.04), with males employing this strategy more than females, but no significant phase*group (p=0.48), phase*sex (p=0.43), group*sex (p=0.21), phase*group*sex (p=0.52) interactions.

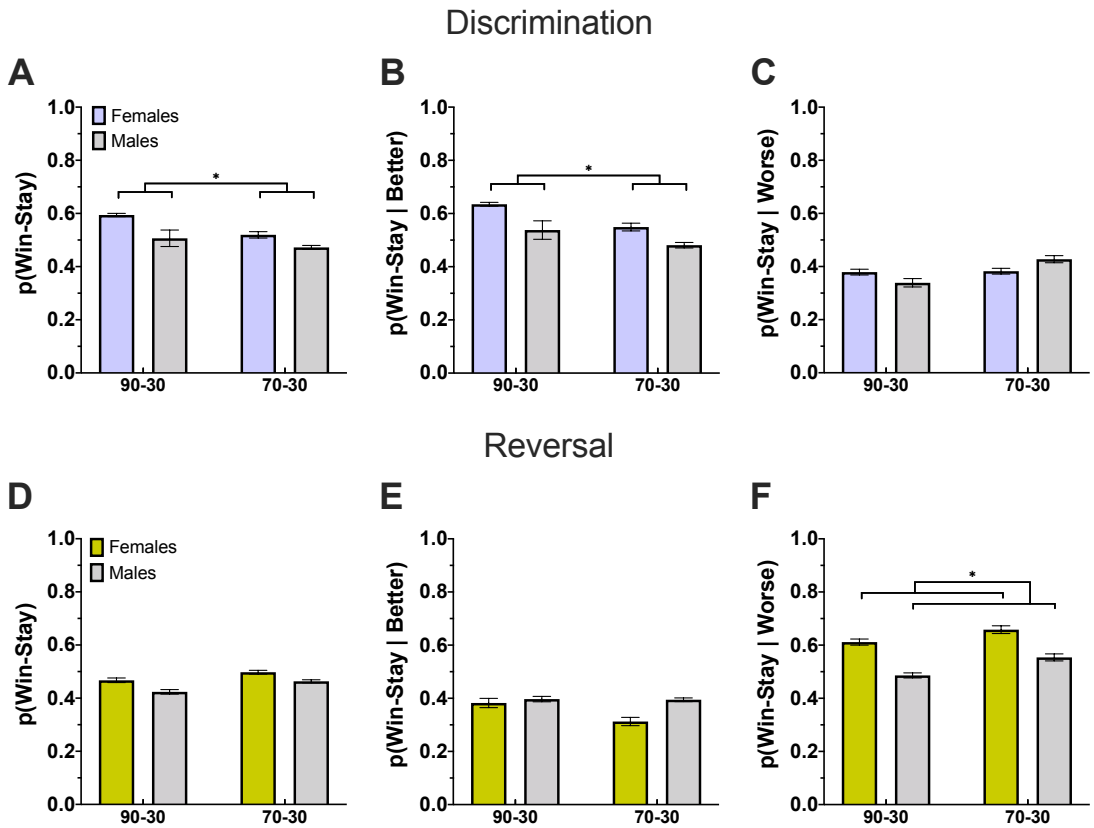

Figure 5. Greater overall win-stay and win-stay on the better option in the 90-30 group during discrimination. (**A-C**) Plotted are proportion of win-stay responses overall (A), after choosing the better option (B), and after choosing the worse option (C) during discrimination. Overall win-stay and win-stay on the better option is used more often in the 90-30 group than the 70-30 group. **(D-F)** The same as A-C but for reversal. We find no significant effects on overall win-stay and win-stay on the better option, but do find females are more likely to apply a win-stay strategy after choosing the worse option than males. Bars indicate $\pm$ *SEM* *p ≤0.05.

Comparisons within Discrimination Learning. We analyzed potential differences in win-stay/lose-shift strategies on stimulus responses employed by each reward probability group and by sex, in the discrimination phase separately. For win-stay, there was no effect of reward probability group (p=0.41), or sex (p=0.25), and no group*sex interaction (p=0.96) (**Figure 5A)**. When we considered win-stay *on the better option* (win-stay|better), we found a significant effect of reward probability group (GLM: β_{group} =-0.086, p=0.049), but no effect of sex (p=0.13) or significant group*sex (p=0.72) interaction (**Figure 5B**). When we analyzed win-stay *on the worse option* (win-stay|worse), we found no significant effect of reward probability group (p=0.92), sex (p=0.26), or group*sex interaction (p=0.15) (**Figure 5C**). However, for

lose-shift, we did find a significant effect of reward probability group (GLM: β_{group} =-0.087, p<0.001) with the 90-30 group using this strategy more than the 70-30 group, a significant effect of sex (GLM: β_{sex} =0.035, p=0.04) with the males using this strategy more than females, but no group*sex interaction (p=0.88), (**Figure 6A**). For lose-shift|better we found a significant effect of sex (GLM: β_{sex} =0.21, p=0.023), but no reward probability group effect (p=0.19) or significant group*sex interaction (p=0.09) (**Figure 6B**). When lose-shift|worse was considered, we found no significant effects of group (p=0.25), sex (p=0.85) or group*sex interaction (p=0.43) (**Figure 6C**).

Comparisons within Reversal Learning. We similarly analyzed win-stay/lose-shift strategies on stimulus responses employed during the reversal phase, separately. We found males were less likely to employ a win-stay strategy (GLM: β_{sex} =-0.069, p=0.055) **Figure 5D**, but no effect of reward probability group (p=0.52) or group*sex interaction (p=0.60). We found no significant group (p=0.21) or sex (p=0.71) differences, and no group*sex interaction (p=0.27) for win-stay|better (**Figure 5E**). We did, however, find a significant effect of sex (GLM: β_{sex} =-0.13, p=0.00025) on win-stay|worse, with males less likely to employ the strategy than females, but no reward probability group effect (p=0.31) or group*sex interaction (p=0.73; **Figure 5F**). For lose-shift, we found a significant effect of sex (GLM: β_{sex} =0.059, p=0.02) with the males using this strategy more than females, but no effect of reward probability group (p=0.99) or group*sex interaction (p=0.37; **Figure 6D**). Additionally, we found no significant effect reward probability group (p=0.44), sex (p=0.87) or group*sex interaction (p=0.46) on lose-shift|better (**Figure 6E**). We did observe a slightly higher lose-shift|worse value in males (GLM: β_{sex} =0.071, p=0.038), but no effect of group (p=0.16) or group * sex interaction (p=0.76) (**Figure 6F**).

Summary. Both win-stay and lose-shift strategies were employed more frequently during early discrimination learning compared to early reversal learning. Males employed the lose-shift strategy more than females in both phases overall, which was consistent with what was observed during early learning. The 90-30 reward probability group used both win-stay and lose-shift strategies more in early learning, regardless of phase. The 90-30 reward probability group also used the win-stay strategy more in early

discrimination than early reversal learning, but used the lose-shift strategy more than the 70-30 group overall during the discrimination phase, suggesting a possible change in strategy with experience.

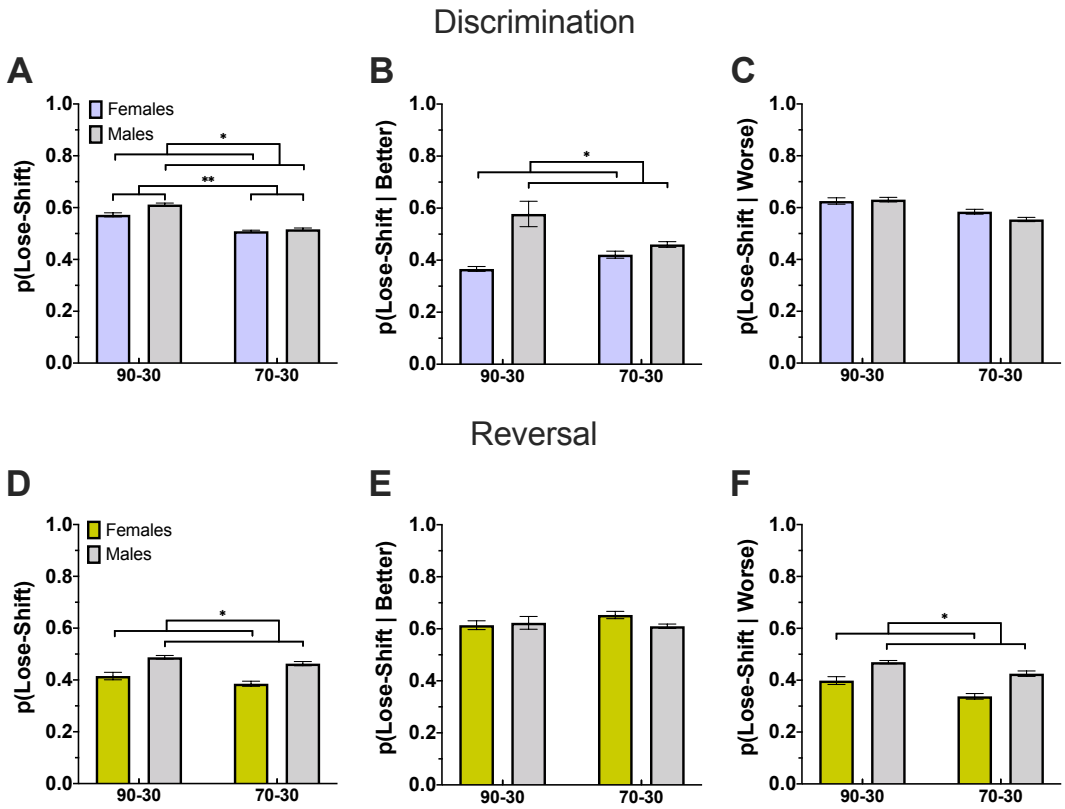

Figure 6. Greater overall lose-shift during discrimination than reversal. (A-C) Plotted is the proportion lose-shift overall (A), following selection of the better option (B), and following selection of the worse option (C) in the discrimination phase. The overall lose-shift proportion is higher in the 90-30 reward probability group than the 70-30 group, but there are no group differences when parsed by better or worse selections. Males are more likely to employ a lose-shift strategy than females overall and following selection of the better option. **(D-F)** Same as A-C, but in the reversal phase. Males are more likely to follow a lose-shift strategy overall and following selection of the worse outcome. Bars indicate $\pm\, SEM$ *p ≤0.05, **p ≤0.01

Comparisons of Repetition measures

Comparisons between early Discrimination & Reversal Learning. We next fitted GLMs to examine the effects of different factors on repetition in choice behavior. Although we did not find any significant effect of reversal on overall $p(Stay)$, we found a significantly lower $\mathrm{p(Stay \mid Better)}$ following reversal (GLM: β_{phase} =-0.25, p<0.0001) and a significantly higher $\mathrm{p}(Stay \mid Worse)$ (GLM: β_{phase} =0.23, p<0.0001),

in reversal compared to the discrimination phase, irrespective of group or sex. There was also an increased perseveration index in the reversal phase, relative to discrimination (GLM: β_{phase} =0.37, p<0.0001), regardless of reward probability group or sex. For *RI,* we find a significantly lower value in the 70-30 group (GLM: β_{group} =-0.024, p=0.028) and in males (GLM: β_{sex} =-0.062, p=0.015), but no significant effects of phase (p=0.41), group*phase (p=0.19), group*sex (p=0.084), phase*group (p=0.43), or phase*group*sex interaction (p=0.46). Similarly, for RI_B we observe lower values in the 70-30 group (GLM: β_{group} =-0.012, p=0.025) and in males (GLM: β_{sex} =-0.031, p=0.016), but no significant difference by phase (p=0.40), group*phase (p=0.19), group*sex (p=0.08), phase*sex (p=0.43), or group*phase*sex interaction(p=0.45). This pattern holds for RI_W as well, with a lower RI_W in the 70-30 group (GLM: β_{group} =-0.012, p=0.031) and males (GLM: β_{sex} =-0.031, p=0.014), and no significant effects of phase (p=0.41), phase*group (p=0.18), group*sex (p=0.087), phase*sex (p=0.43) or group*phase*sex interaction (p=0.46).

Comparisons within Discrimination Learning. Additionally, we compared the effects of reward probability group, sex, and their interactions on measures of repetition including, $p(Stay)$ and repetition index (both overall and following better and worse option selections), as well as perseveration index. For $p(Stay)$, we found no significant differences by group (p=0.17), sex (p=0.10), or group * sex interaction (p = 0.40), **Figure 7A**. There was a marginally significant effect of group on $\mathrm{p(Stay \mid Better)}$ (GLM: β_{group} =-0.077, p=0.067), but no significant effect of sex (p=0.11) or group*sex interaction (p=0.54), **Figure 7B**. There were no significant effects of group (p=0.38), sex (p=0.56), or group*sex interaction (p=0.26) on $\mathrm{p}(Stay \mid Worse)$, **Figure 7C**. For repetition index, we observe a lower value in the 70-30 reward probability group (GLM: β_{group} =-0.024, p=0.033), and in males (GLM: β_{sex} =-0.062, p=0.019) than the 90-30 group and females, respectively, but no group*sex interaction (p=0.091), **Figure S3A**. We then considered repetition index parsed by better, RI_B and worse, RI_W, stimulus selections. For RI_B, we found that the 70-30 reward probability group exhibited a significantly lower value (GLM: β_{group} =-0.012, p=0.030), as did males (GLM: β_{sex} =-0.031, p=0.020), but no group*sex interaction (p=0.088), **Figure S3B**. For RI_W, we similarly observed a significantly lower value in the 70-30 reward probability group (GLM:

β_{group} =-0.012, p=0.036) and males (GLM: β_{sex} =-0.031, p=0.018) than the 90-30 group and females, respectively, but no significant group*sex interaction (p=0.094), **Figure S3C**. We also considered perseveration index, and found no effect of reward probability group (p=0.42), sex (p=0.87), or significant group*sex interaction (p=0.50; **Figure 7D**).

Comparisons within Reversal Learning. Again, we compared $p(Stay)$, repetition index (and their variants), as well as perseveration index for the reversal phase alone. For $p(Stay)$, we found a significantly lower value in males (GLM: β_{sex} =-0.062, p=0.0072), but no effect of reward probability group (p=0.18), or group*sex interaction (p=0.99), **Figure 7E**. For $\mathrm{p(Stay \mid Better)}$ there were no significant effects of reward probability group (p=0.26), sex (p=0.75), or group*sex interaction (p=0.33), **Figure 7F**. For $\mathrm{p(Stay \mid Worse)}$ we observed a significant effect of sex (GLM: β_{sex} =-0.087, p=0.0070), with a lower value in males than females, but no effect of reward probability group (p=0.18) or group*sex interaction (p=0.93), **Figure 7G**. For overall repetition index, we found a marginally lower value in males (GLM: β_{sex} =-0.038, p=0.061), but no effect of group (p=0.97), or group * sex interaction (p = 0.32), **Figure S3D**. Similarly, we observe a marginally significant effect of sex on RI_B (GLM: β_{sex}=-0.018, p =0.062) (**Figure S3E**) and RI_W (GLM: β_{sex} =-0.019, p=0.059) (**Figure S3F**), but no significant effects of reward probability group (p=0.99) or group*sex interaction (p=0.31). Additionally, we found no effect of reward probability group (p=0.40) or group*sex interaction (p=0.42), but a marginally significant effect of sex (GLM: β_{sex} =-0.13, p=0.064) on perseveration index, **Figure 7H**.

Summary. We found more perseveration, lower $\mathrm{p(Stay \mid Better)}$, and higher $\mathrm{p(Stay \mid Worse)}$, during early reversal learning, compared to the discrimination phase. This indicates animals continued to perform according to contingencies learned in discrimination. During the discrimination phase, we observed less repetition in the 70-30 reward probability group, as evidenced by the lower $p(Stay)$ values. Even when accounting for difference in reward schedule using the RI measure, we still observed less repetition (lower RI, particularly RI_W) in the 70-30 reward probability group and in males, suggesting probability group and sex have an effect on repetitive behavior that are not due to differences in the

propensity to be rewarded. In the reversal phase, we observe higher variability in males, as evidenced by both the concomitant higher $p(Stay)$ and lower RI, particularly on the worse option. Interestingly, we observe no significant group differences in repetition measures during overall reversal, suggesting that reward schedule may have a greater effect on initial learning.

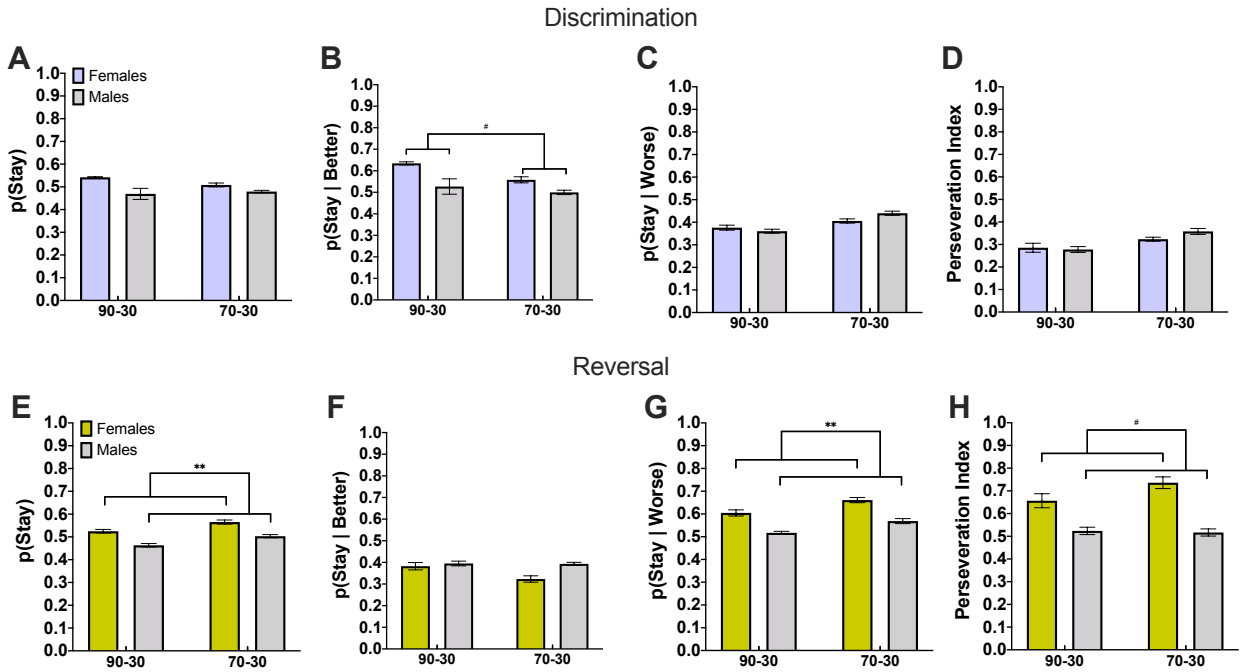

Figure 7. Greater perseveration, lower $\mathbf{p(Stay \mid Better)}$, and higher $\mathbf{p(Stay \mid Worse)}$ during reversal. (A-D) Plotted is the reward probability of staying overall (A), after selecting the better option (B), and after selecting the worse option (C), as well as the perseveration index (D) during the discrimination phase. During the discrimination phase, the only sex- or reward probability group-dependent difference we observe is a higher probability of staying after selecting the correct option for the 90-30 reward probability group. **(E-H)** Same as A-D, but for reversal. Greater perseveration in the reversal phase. Overall, males are less likely to stay than females. This is primarily the result of a lower propensity to stay after selecting the worse option (as within-group sex differences are close to zero for $p(Stay|Better)$). Males also have a lower perseveration index than females in the reversal phase. Bars indicate $\pm\, SEM$ $^{\#}$p=0.06, **p ≤0.01

Comparisons of Estimated Parameters based on Fit of Choice Behavior with RL models

Comparisons between Discrimination & Reversal Learning. To gain more insight into learning and decision making, we next compared estimated model parameters (learning rate (α) and sensitivity to difference in subjective reward values, σ) based on two RL models across groups in each of the two phases. Comparison of estimated parameters based on RL1 revealed that the average learning rate was significantly

higher during reversal than discrimination phase (difference in α =0.12; two sample t-test: t(48)=3.74, p<0.0001) and at the same time, sensitivity to subjective reward values was significantly lower (difference in σ=-1.42; two sample t-test: t(48)=-7.10, p<0.00001), which corresponds to enhanced exploration. Additionally, to better explore the relationship between learning rate and sensitivity to difference in reward values, we calculated the correlation coefficients between parameters across groups. We found a significant negative correlation between learning rate and sensitivity to difference in reward values during reversal (r=-0.68, p=0.0018) but not discrimination learning (r=-0.38, p=0.061).

We found consistent results using RL2 that involved separate learning rates for rewarded and unrewarded trials. More specifically, the rewarded learning rate, α_{Rew}, increased significantly during reversal than discrimination phase (difference in α_{Rew} =0.13; two sample t-test: t(48)=3.38, p=0.0014), as did the unrewarded learning parameter, α_{unrew} (difference in α_{Unrew} =0.093; two sample t-test: estimate: t(48)=3.24, p=0.0022), whereas the sensitivity to difference in subjective reward values decreased (difference in σ =-1.42; two sample t-test: t(df)=-6.67, p<0.0001). To ensure that the observed simultaneous increase in learning rates and decrease in sensitivity was not due to model fitting procedure, we tested the correlation between learning rates and sensitivity to difference in reward values. We found no significant correlation between α_{Rew} and sensitivity to difference in reward values during the discrimination (r=-0.25, p=0.22) or reversal phase (r=-0.26, p=0.22), or between α_{Unrew} and sensitivity to difference in reward values in the discrimination phase (r=-0.36, p=0.072). However, there was a significant correlation between α_{Unrew} and sensitivity in the reversal phase (r=-0.61, p=0.0012). Together, these results suggest that observed simultaneous increase in learning rates and decrease in sensitivity in the reversal phase was not due to our fitting procedure and instead, happened concurrently due to independent neural mechanisms.

Comparisons within Discrimination Learning. Using estimated parameters based on RL1, we found that the 90-30 reward probability group had a learning rate $\alpha = 0.029 \pm 0.019$, sensitivity to difference in reward values $\sigma = 1.50 \pm 0.20$ and fit of $BIC = -2141 \pm 604$ (**Figure 8A**). The 70-30 reward probability group had a learning rate $\alpha = 0.028 \pm 0.018$, sensitivity to difference in reward values

$\sigma = 1.81 \pm 0.27$ and fit of $BIC = -4061 \pm 616$. We found no significant differences between reward probability groups in terms of α (two sample t-test: *t* (23)= -0.025, p=0.98) or σ (two sample t-test; *t* (23)= -0.81; p=0.42).

Using RL2, we found that for the 90-30 reward probability group $\alpha_{Rew} = 0.046 \pm 0.036$, $\alpha_{Unrew} = 0.022 \pm 0.012$, sensitivity to difference in reward values $\sigma = 1.50 \pm 0.20$ with fit of $BIC = -2134 \pm 604$ (**Figure 8B**). The 70-30 reward probability group had $\alpha_{Rew} = 0.034 \pm 0.017$, $\alpha_{Unrew} = 0.025 \pm 0.014$, and sensitivity to difference in reward values $\sigma = 1.93 \pm 0.28$ with fit of $BIC = -4053 \pm 616$. Similarly, using this model we did not find any significant difference between the reward probability groups for α_{Rew} (two sample t-test; *t*(23) = 0.27; p = 0.78), α_{Unrew} (two sample t-test; *t*(23)= -0.22; p=0.82), or σ (two sample t-test; *t*(23) = 1.06; p=0.30).

Comparisons within Reversal Learning. As in the discrimination phase, we again fit choice behavior for each subject in the reversal phase using the two RL models. Using RL1, (**Figure 8C**) we found that the 90-30 reward probability group had a learning rate $\alpha = 0.17 \pm 0.041$, sensitivity to difference in reward values $\sigma = 0.18 \pm 0.010$ and fit of $BIC = -3127 \pm 319$. The 70-30 reward probability group had a learning rate $\alpha = 0.10 \pm 0.028$, sensitivity to difference in reward values $\sigma = 0.33 \pm 0.10$ and fit of $BIC = -5202 \pm 391$. We found no significant difference between the α (two sample t-test; *t*(23)= -1.38; p=0.18) or σ (two sample t-test; *t*(23) = 0.97; p=0.34) parameters between groups.

Using RL2 (**Figure 8D**), we found that for the 90-30 group $\alpha_{Rew} = 0.22 \pm 0.066$, $\alpha_{Unrew} = 0.11 \pm 0.038$, sensitivity to difference in reward values $\sigma = 0.18 \pm 0.10$ with fit of $BIC = -3119 \pm 391$. The 70-30 group had $\alpha_{Rew} = 0.19 \pm 0.046$, $\alpha_{Unrew} = 0.096 \pm 0.031$, sensitivity to difference in reward values $\sigma = 0.38 \pm 0.11$ with a fit of $BIC = -5194 \pm 390$. Similarly, in this model we found no significant difference between the reward probability groups' α_{Rew} (two sample t-test; *t*(23) = -0.41; p=0.68), α_{Unrew} (two sample t-test; *t*(23)= -0.23; p = 0.82), or σ (two sample t-test; *t*(23) = 1.25; p = 0.23) parameters.

Summary. In analyses across discrimination and reversal phases, we found that sensitivity to difference in reward values was lower during reversal than discrimination, and learning rate(s) were higher

(the single learning rate for RL1, and both learning rates for RL2) during reversal relative to the discrimination phase. These results indicate that reversal caused faster learning and more exploration at the same time. However, within each phase of learning, the estimated model parameters were not significantly different between groups in either phase. We also found no sex-dependent differences. This suggests the slight increased probability of reward corresponding to the better option in the 90-30 reward probability group, as compared to the 70-30 group, was not large enough to induce significantly different learning or alternatively, this effect could not be captured by our RL models.

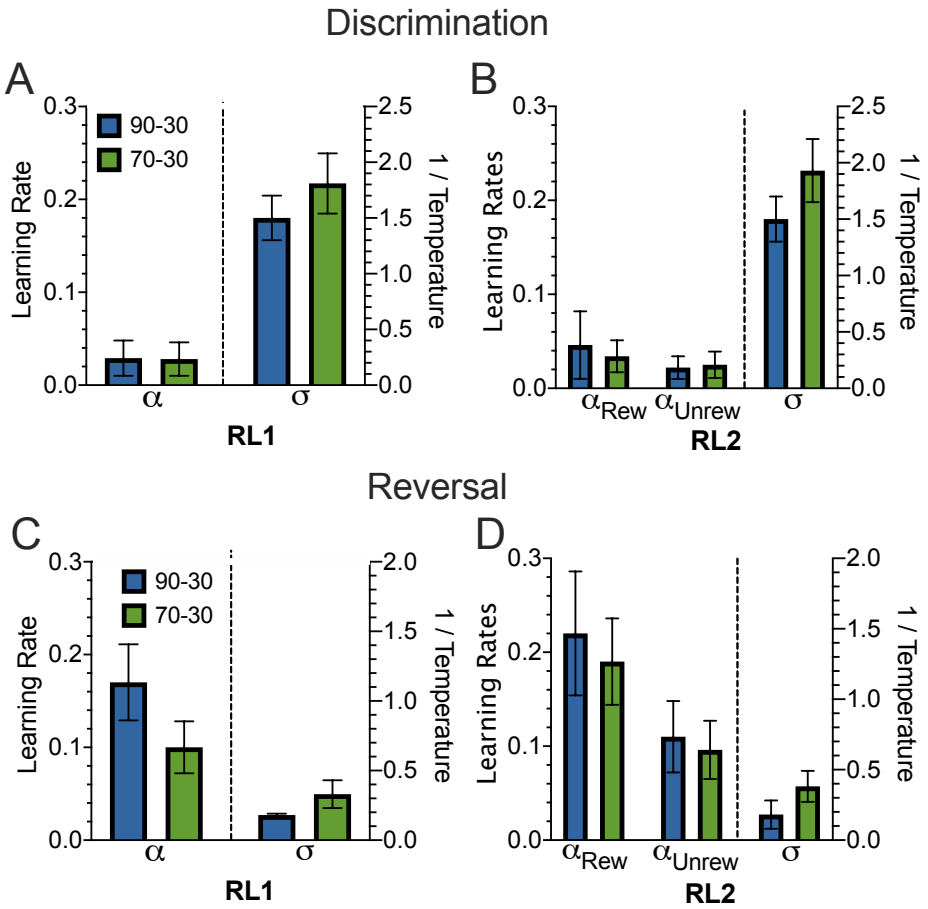

Figure 8. Higher learning rate and lower sensitivity to difference in subjective reward values in reversal compared to discrimination learning. (A-B) Learning parameters and sensitivity to difference in reward values for the single-learning rate (A) and double-learning rate (B) models fitted to choice behavior in the discrimination (pre-reversal) phase. We find no significant difference in parameter values between reward probability groups. **(C-D)** Same as A-B, but for choices after the reversal. Again, we find no significant difference between reward probability groups. However, we do find significantly higher learning rates (for both the single and two learning rate models), as well as significantly lower sensitivity to difference in reward values parameters, following reversal.

Discussion

Here, we used a stimulus-based probabilistic discrimination and reversal paradigm with different probabilities of reward (i.e., 90/30, 70/30) to test learning and performance on several measures (i.e., probability of choosing the better option, initiation and choice omissions and latencies, perseveration and repetition measures, and win-stay/lose-shift strategies). We also examined whether fit of choice data using RL models would reveal differences between the two learning phases, and tested for potential differences in learning rates (single, separate) and explore/exploit behavior. We found higher learning rates in the reversal than discrimination phase, which mirror recent reports (described in more detail below). Animals also exhibited decreased sensitivity to the difference in subjective reward values of the two options during the reversal learning phase compared to discrimination, indicative of greater exploration. Additionally, consistent with the long-standing idea of reversal learning as a measure of inhibitory control we found increased perseveration in early reversal compared to early discrimination learning. Further, we found that differences in reward collection latencies depended on the degree to which the environment was profitable (90% vs 70% reward). Finally, the only pronounced sex differences were uncovered for initiation latencies across both learning phases, and a greater number of initiation omissions during reversal learning, suggesting greater attentional demands in females relative to males, which replicate effects of a recent study by our group (Aguirre et al., 2020). We elaborate on these findings below.

Perseveration

All animals exhibited more perseveration during early reversal learning than early discrimination learning. This finding supports those by Verharen, den Ouden, Adan, and Vanderschuren (2020) in which they simulated data of thousands of probabilistic reversal learning sessions using a Q-learning model consisting of separate learning rates for learning from positive (i.e. rewarded) and negative (i.e. nonrewarded) feedback, a beta parameter, and a stickiness parameter indicative of perseveration. They found that a greater number of reversals occurred when the stickiness parameter value was high (i.e., greater perseveration), but only when both learning rates were also high (Verharen et al., 2020).

Interestingly, we found reward probability group differences for perseveration, with the 90-30 group exhibiting greater perseveration than the 70-30 group, but only during discrimination learning. This finding implies that more consistent reward feedback (i.e. higher reward probability associated with the better option) in the 90-30 group, promoted perseverative responding in early learning. Furthermore, males demonstrated greater exploratory behavior (i.e. lower perseveration) during discrimination learning, in line with previous research showing greater impulsivity in male rats (Lukkes, Thompson, Freund, & Andersen, 2016; Palanza, Gioiosa, & Parmigiani, 2001).

Animals will perseverate on stimulus choice during *early* reversal learning, yet the lack of perseveration across reversal learning in its entirety demonstrates that animals eventually alternate and switch according to the new reward contingencies over time. Greater perseveration during early reversal is consistent with the long-standing idea that reversal learning is a measure of inhibitory control, such that the inability to disengage from previously rewarding behavior after a change in contingency may be reflective of compulsive and even impulsive response tendencies, commonly associated with drug dependence (Izquierdo & Jentsch, 2012). Additionally, there is evidence that inflexible responding in reversal learning may be genetically related to impulsivity (Crews & Boettiger, 2009; Fineberg et al., 2010; Franken, van Strien, Nijs, & Muris, 2008; Groman, James, & Jentsch, 2009; Groman & Jentsch, 2013; Jentsch et al., 2014). Given our findings, the classic idea of stimulus-based reversal learning as a measure of inhibitory control, which has recently been overshadowed by other conceptualizations, should be revisited. It is particularly interesting that we show 'stickiness' in the selection of visual stimuli, aside from spatial- or response-based perseveration, which we control for here by pseudorandomly presenting the stimuli on the left- vs. right- sides of the screen.

Learning rates

We found that both (single and separate learning rates) RL models yielded higher learning rates in reversal than discrimination learning. The general trend of increased learning rates following reversal and decreased sensitivity to the difference in subjective reward values is consistent with existing literature (Costa, Tran, Turchi, & Averbeck, 2015; Massi, Donahue, & Lee, 2018). The increased learning rate

suggests a more rapid integration of reward feedback, while the lower sensitivity to difference in reward values indicates rats choose the higher-valued option less consistently, corresponding to greater exploration. These changes may reflect an increase in environment volatility following the reversal as well as more exploration to identify the better choice option. Bathellier, Tee, Hrovat, and Rumpel (2013) accounted for a similar change in learning rate in mice by assuming slower initial learning during discrimination due to weaker initial synaptic weights, and much faster learning during reversal due to the same synapses being activated, but at a state where synaptic weights are stronger than they were in discrimination. Future modeling studies are needed to explain how a single reversal can induce both higher learning rates and decreased sensitivity to subjective values.

Win-stay, lose-shift and reward collection latencies

Animals exhibited less consistent response to reward feedback during the reversal compared to the discrimination phase, indicative of noisier behavior (more exploration), weaker learning, or both. This observation dovetails the observation that animals more often used reward-dependent choice strategies (i.e. win-stay, lose-shift) in the early discrimination compared to the reversal phase.

Probability group differences across both learning phases were found for reward collection latencies, with the 90-30 group exhibiting longer reward collection times than the 70-30 group, suggesting the 90-30 group may experience attenuations in motivation. Latency to collect reward is commonly used as a measure of motivation, whereas stimulus response latency is often used as a measure of processing or decision-making speed in the Five-Choice Serial Reaction Time Task (5CSRTT) (Amitai & Markou, 2010; Asinof & Paine, 2014; Bari, Dalley, & Robbins, 2008; Bushnell & Strupp, 2009; Chudasama et al., 2003; Remmelink, Chau, Smit, Verhage, & Loos, 2017; Robbins, 2002; Robinson et al., 2009). One interpretation of this finding is that animals may have been more motivated to confirm whether they had received a reward when there was more uncertainty involved (i.e., when the probability for the better option was lower), and conversely, were more confident in their decision when there was already a high reward probability associated with choosing the better option in the 90-30 group. Another related explanation is that animals exert more effort and display more vigor when in leaner and more

uncertain reward environments (Amsel, 1967; McNamara, Fawcett, & Houston, 2013), which in our experiments was directly tied to the differences in reward probabilities associated with the better option. Because there was greater reward uncertainty associated with the better option for the 70-30 group, rats may have exerted more effort to retrieve rewards compared to the 90-30 group. Interestingly, although animals generally increased their number of rewards collected across days during discrimination learning, this was actually not the case for reversal learning where the number of rewards collected over time was relatively flat. Despite the reward collection latency differences, we found no difference in the number of cached rewards between the probability groups, so these differences are likely not due to satiety *per se*, but instead, an effect of an experience with overall reward rate over time.

Sex differences

Females in both probability groups exhibited longer initiation latencies than males, which remained a consistent finding across learning phases. Females also exhibited more initiation omissions than males but this was only observed during the reversal learning phase. Measures of attention in rodents vary, but can include number of omissions (i.e., failure to make a response before the end of the trial), as well as percentage of correct responses (i.e., accuracy), usually on the 5CSRTT as described above. Our data supports the interpretation that females exhibit greater attentional demands (and consequently poorer performance) than males in the reversal phase. Prior work utilized initiation omissions as measures of attentional deficits, and we found similar sex-specific effects for both initiation omissions and latencies. These effects were limited to specific phases, with females exhibiting more initiation omission during the discrimination learning phase, and longer initiation latencies during late reversal learning (Aguirre et al., 2020).

Conclusion

We conclude that discrimination and reversal learning are behaviorally more unique than they are similar: longer choice latencies following incorrect trials, fewer rewards collected over time, lesser win-stay and lose-shift strategies employed, and more perseveration in early reversal learning. Additionally, fitting choice behavior with reinforcement learning models, we found a lower sensitivity to the difference

in subjective reward values (greater exploration) and higher learning rate for the reversal compared to original discrimination learning. Interestingly, a consistent reward probability group difference emerged with a richer environment associated with longer reward collection latencies than a leaner environment. We also replicate previous reports on sex differences in reversal learning. Future studies should probe the neural correlates of these fine-grained behavioral measures, as these have been under-utilized and may reveal marked dissociations in how the circuitry is recruited in each phase.

Supplemental Figures

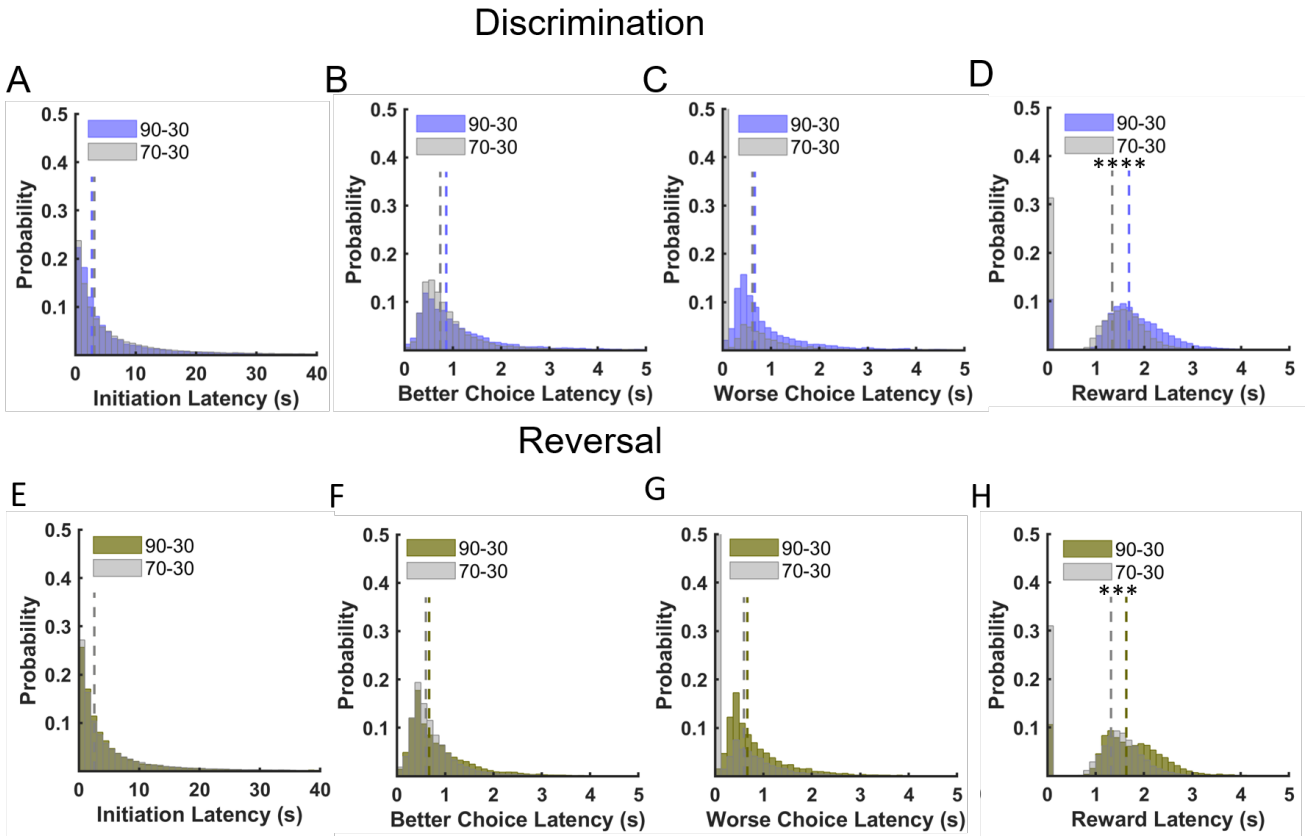

Figure S1. 90-30 probability group was consistently slower to collect reward than 70-30 probability group. **(A)** There were no group difference in initiation latencies for discrimination learning. **(B)** No significant group difference for better choice latencies in discrimination learning. **(C)** No significant group difference in worse choice latencies in discrimination learning. **(D)** The 90-30 group exhibited longer reward collection latencies than the 70-30 group in discrimination learning. **(E)** No significant group difference in initiation latencies in reversal learning. **(F)** No significant group difference in better choice latencies in reversal learning. **(G)** No significant group differences in worse choice latencies in reversal learning. **(H)** The 90-30 group exhibited longer reward collection latencies than the 70-30 group in reversal learning. Dashed lines in histograms of latencies represent group medians. ***p ≤0.001, ****p≤0.0001

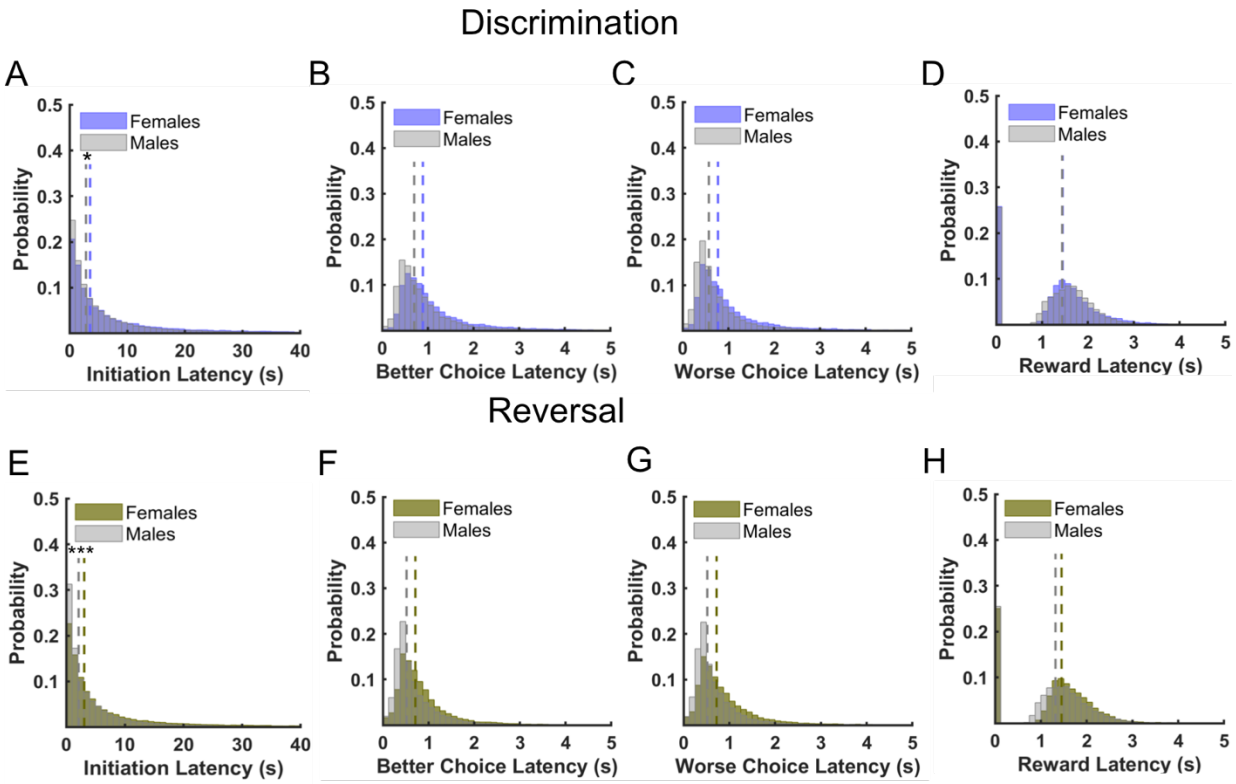

Figure S2. Females exhibited longer initiation latencies compared to males in discrimination and reversal phases, whereas males exhibited longer 'better choice' latencies compared to females only in the reversal phase. (**A**) Females exhibited longer initiation latencies than males in discrimination learning. **(B)** No significant sex difference on better choice latencies in discrimination learning. **(C)** No sex difference in worse choice latencies in discrimination learning. **(D)** No sex difference in reward collection latencies in discrimination learning**. (E)** Females exhibited longer initiation latencies than males in discrimination learning. **(F)** Males exhibited longer better choice latencies than females in reversal learning**. (G)** No sex differences in worse choice latencies in reversal learning. **(H)** No sex difference in reward collection latencies in reversal learning. Dashed lines in histograms of latencies represent group medians. *$p \leq 0.05$,***$p \leq 0.001$

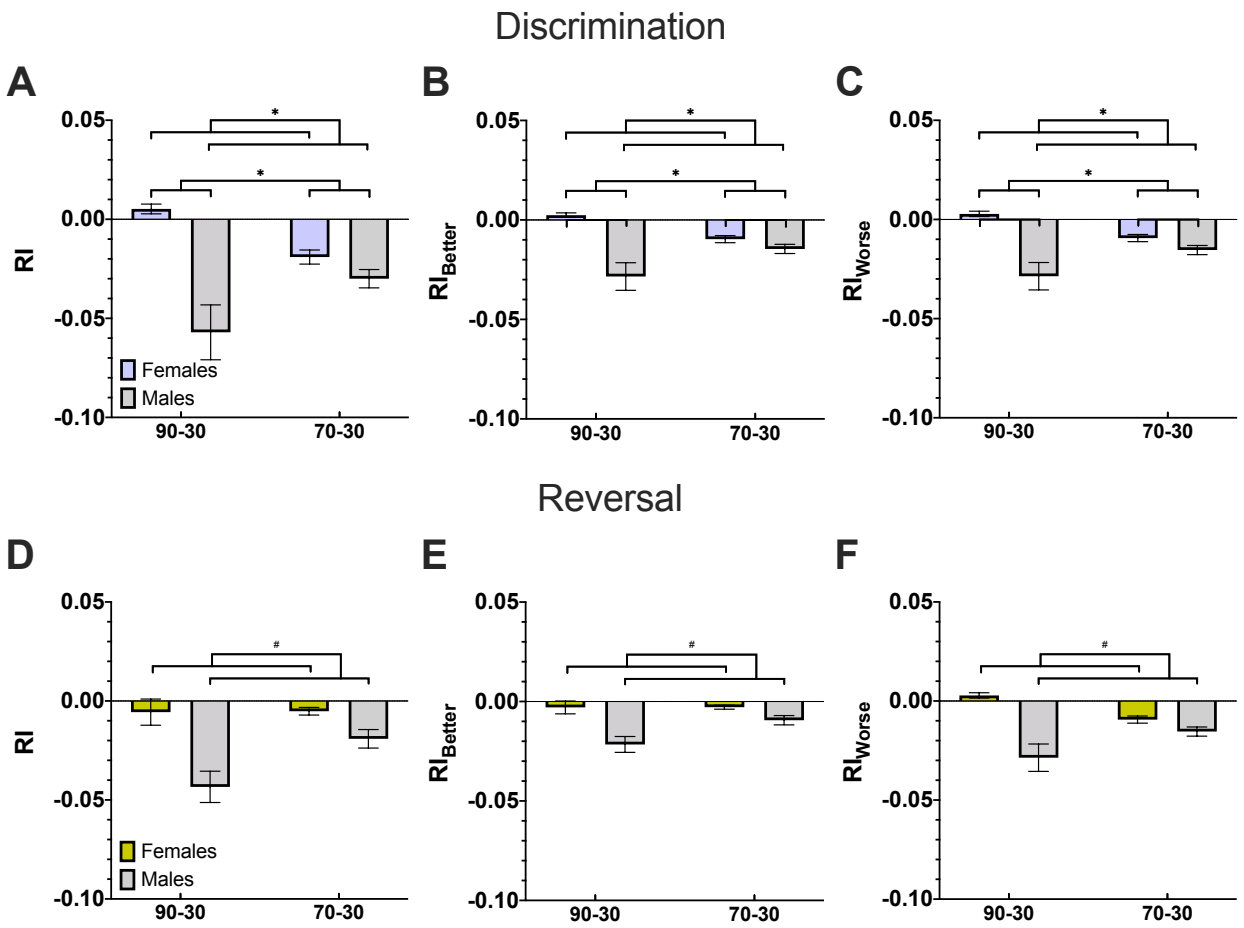

Figure S3. Lower repetition index measures for males as compared to females in both phases, and for the 70-30 group as compared to the 90-30 group in the discrimination phase. (A-C) Overall repetition index (A), repetition index for the better option (B), and repetition index for the worse option (C) in the discrimination phase. We find the 70-30 group has a significantly lower RI and RI_B, and marginally significantly lower RI_W. Additionally, males have significantly lower values for all three measures. **(D-F)** Same as A-C, but in the reversal phase. We find marginally significant lower RI, RI_B, and RI_W in males than females. Bars indicate $\pm\ SEM$ $^{\#}$p=0.06, *p ≤0.05

Acknowledgements
We thank Alexandra Stolyarova for programming the probability schedules and trial-by-trial outputs in the Lafayette Instrument chambers.

References

Aguirre, C. G., Stolyarova, A., Das, K., Kolli, S., Marty, V., Ray, L., . . . Izquierdo, A. (2020). Sex-dependent effects of chronic intermittent voluntary alcohol consumption on attentional, not motivational, measures during probabilistic learning and reversal. *PLoS One, 15*(6), e0234729. doi:10.1371/journal.pone.0234729

Alvarez, P., & Eichenbaum, H. (2002). Representations of odors in the rat orbitofrontal cortex change during and after learning. *Behav Neurosci, 116*(3), 421-433. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12049323

Amitai, N., & Markou, A. (2010). Disruption of performance in the five-choice serial reaction time task induced by administration of N-methyl-D-aspartate receptor antagonists: relevance to cognitive dysfunction in schizophrenia. *Biol Psychiatry, 68*(1), 5-16. doi:10.1016/j.biopsych.2010.03.004

Amsel, A. (Ed.) (1967). *Partial reinforcement effects in vigor and persistence: Advances in frustration theory derived from a variety of within-subjects experiments.* (Vol. 1). New York: Academic Press.

Asinof, S. K., & Paine, T. A. (2014). The 5-choice serial reaction time task: a task of attention and impulse control for rodents. *J Vis Exp*(90), e51574. doi:10.3791/51574

Averbeck, B. B., & Costa, V. D. (2017). Motivational neural circuits underlying reinforcement learning. *Nat Neurosci, 20*(4), 505-512. doi:10.1038/nn.4506

Bari, A., Dalley, J. W., & Robbins, T. W. (2008). The application of the 5-choice serial reaction time task for the assessment of visual attentional processes and impulse control in rats. *Nat Protoc, 3*(5), 759-767. doi:10.1038/nprot.2008.41

Bathellier, B., Tee, S. P., Hrovat, C., & Rumpel, S. (2013). A multiplicative reinforcement learning model capturing learning dynamics and interindividual variability in mice. *Proc Natl Acad Sci U S A, 110*(49), 19950-19955. doi:10.1073/pnas.1312125110

Brigman, J. L., Feyder, M., Saksida, L. M., Bussey, T. J., Mishina, M., & Holmes, A. (2008). Impaired discrimination learning in mice lacking the NMDA receptor NR2A subunit. *Learn Mem, 15*(2), 50-54. doi:10.1101/lm.777308

Brigman, J. L., Mathur, P., Harvey-White, J., Izquierdo, A., Saksida, L. M., Bussey, T. J., . . . Holmes, A. (2010). Pharmacological or genetic inactivation of the serotonin transporter improves reversal learning in mice. *Cereb Cortex, 20*(8), 1955-1963. doi:10.1093/cercor/bhp266

Brushfield, A. M., Luu, T. T., Callahan, B. D., & Gilbert, P. E. (2008). A comparison of discrimination and reversal learning for olfactory and visual stimuli in aged rats. *Behav Neurosci, 122*(1), 54-62. doi:10.1037/0735-7044.122.1.54

Bushnell, P. J., & Strupp, B. J. (2009). Assessing Attention in Rodents. In nd & J. J. Buccafusco (Eds.), *Methods of Behavior Analysis in Neuroscience*. Boca Raton (FL).

Chudasama, Y., Passetti, F., Rhodes, S. E., Lopian, D., Desai, A., & Robbins, T. W. (2003). Dissociable aspects of performance on the 5-choice serial reaction time task following lesions of the dorsal anterior cingulate, infralimbic and orbitofrontal cortex in the rat: differential effects on selectivity, impulsivity and compulsivity. *Behav Brain Res, 146*(1-2), 105-119. doi:10.1016/j.bbr.2003.09.020

Costa, V. D., Dal Monte, O., Lucas, D. R., Murray, E. A., & Averbeck, B. B. (2016). Amygdala and Ventral Striatum Make Distinct Contributions to Reinforcement Learning. *Neuron, 92*(2), 505-517. doi:10.1016/j.neuron.2016.09.025

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal learning and dopamine: a bayesian perspective. *J Neurosci, 35*(6), 2407-2416. doi:10.1523/JNEUROSCI.1989-14.2015

Crews, F. T., & Boettiger, C. A. (2009). Impulsivity, frontal lobes and risk for addiction. *Pharmacol Biochem Behav, 93*(3), 237-247. doi:10.1016/j.pbb.2009.04.018

Dalton, G. L., Phillips, A. G., & Floresco, S. B. (2014). Preferential involvement by nucleus accumbens shell in mediating probabilistic learning and reversal shifts. *J Neurosci, 34*(13), 4618-4626. doi:10.1523/JNEUROSCI.5058-13.2014

Dalton, G. L., Wang, N. Y., Phillips, A. G., & Floresco, S. B. (2016). Multifaceted Contributions by Different Regions of the Orbitofrontal and Medial Prefrontal Cortex to Probabilistic Reversal Learning. *J Neurosci, 36*(6), 1996-2006. doi:10.1523/JNEUROSCI.3366-15.2016

Eichenbaum, H., Fagan, A., & Cohen, N. J. (1986). Normal olfactory discrimination learning set and facilitation of reversal learning after medial-temporal damage in rats: implications for an account of preserved learning abilities in amnesia. *J Neurosci, 6*(7), 1876-1884. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/3734866

Farashahi, S., Donahue, C. H., Khorsand, P., Seo, H., Lee, D., & Soltani, A. (2017). Metaplasticity as a Neural Substrate for Adaptive Learning and Choice under Uncertainty. *Neuron, 94*(2), 401-414 e406. doi:10.1016/j.neuron.2017.03.044

Fineberg, N. A., Potenza, M. N., Chamberlain, S. R., Berlin, H. A., Menzies, L., Bechara, A., . . . Hollander, E. (2010). Probing compulsive and impulsive behaviors, from animal models to endophenotypes: a narrative review. *Neuropsychopharmacology, 35*(3), 591-604. doi:10.1038/npp.2009.185

Franken, I. H., van Strien, J. W., Nijs, I., & Muris, P. (2008). Impulsivity is associated with behavioral decision-making deficits. *Psychiatry Res, 158*(2), 155-163. doi:10.1016/j.psychres.2007.06.002

Groman, S. M., James, A. S., & Jentsch, J. D. (2009). Poor response inhibition: at the nexus between substance abuse and attention deficit/hyperactivity disorder. *Neurosci Biobehav Rev, 33*(5), 690-698. doi:10.1016/j.neubiorev.2008.08.008

Groman, S. M., & Jentsch, J. D. (2013). Identifying the molecular basis of inhibitory control deficits in addictions: neuroimaging in non-human primates. *Curr Opin Neurobiol, 23*(4), 625-631. doi:10.1016/j.conb.2013.03.001

Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience, 345*, 12-26. doi:10.1016/j.neuroscience.2016.03.021

Izquierdo, A., Darling, C., Manos, N., Pozos, H., Kim, C., Ostrander, S., . . . Rudebeck, P. H. (2013). Basolateral amygdala lesions facilitate reward choices after negative feedback in rats. *J Neurosci, 33*(9), 4105-4109. doi:10.1523/JNEUROSCI.4942-12.2013

Izquierdo, A., & Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology (Berl), 219*(2), 607-620. doi:10.1007/s00213-011-2579-7

Jang, A. I., Costa, V. D., Rudebeck, P. H., Chudasama, Y., Murray, E. A., & Averbeck, B. B. (2015). The Role of Frontal Cortical and Medial-Temporal Lobe Brain Areas in Learning a Bayesian Prior Belief on Reversals. *J Neurosci, 35*(33), 11751-11760. doi:10.1523/JNEUROSCI.1594-15.2015

Jentsch, J. D., Ashenhurst, J. R., Cervantes, M. C., Groman, S. M., James, A. S., & Pennington, Z. T. (2014). Dissecting impulsivity and its relationships to drug addictions. *Ann N Y Acad Sci, 1327*, 1-26. doi:10.1111/nyas.12388

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annu Rev Neurosci, 35*, 287-308. doi:10.1146/annurev-neuro-062111-150512

Lukkes, J. L., Thompson, B. S., Freund, N., & Andersen, S. L. (2016). The developmental inter-relationships between activity, novelty preferences, and delay discounting in male and female rats. *Dev Psychobiol, 58*(2), 231-242. doi:10.1002/dev.21368

Massi, B., Donahue, C. H., & Lee, D. (2018). Volatility Facilitates Value Updating in the Prefrontal Cortex. *Neuron, 99*(3), 598-608 e594. doi:10.1016/j.neuron.2018.06.033

McNamara, J. M., Fawcett, T. W., & Houston, A. I. (2013). An adaptive response to uncertainty generates positive and negative contrast effects. *Science, 340*(6136), 1084-1086. doi:10.1126/science.1230599

Palanza, P., Gioiosa, L., & Parmigiani, S. (2001). Social stress in mice: gender differences and effects of estrous cycle and social dominance. *Physiol Behav, 73*(3), 411-420. doi:10.1016/s0031-9384(01)00494-2

Remmelink, E., Chau, U., Smit, A. B., Verhage, M., & Loos, M. (2017). A one-week 5-choice serial reaction time task to measure impulsivity and attention in adult and adolescent mice. *Sci Rep, 7*, 42519. doi:10.1038/srep42519

Robbins, T. W. (2002). The 5-choice serial reaction time task: behavioural pharmacology and functional neurochemistry. *Psychopharmacology (Berl), 163*(3-4), 362-380. doi:10.1007/s00213-002-1154-7

Robinson, E. S., Eagle, D. M., Economidou, D., Theobald, D. E., Mar, A. C., Murphy, E. R., . . . Dalley, J. W. (2009). Behavioural characterisation of high impulsivity on the 5-choice serial reaction time task: specific deficits in 'waiting' versus 'stopping'. *Behav Brain Res, 196*(2), 310-316. doi:10.1016/j.bbr.2008.09.021

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (2000). Changes in functional connectivity in orbitofrontal cortex and basolateral amygdala during learning and reversal training. *J Neurosci, 20*(13), 5179-5189. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10864975

Schoenbaum, G., Nugent, S. L., Saddoris, M. P., & Setlow, B. (2002). Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *Neuroreport, 13*(6), 885-890. doi:10.1097/00001756-200205070-00030

Schoenbaum, G., Setlow, B., Saddoris, M. P., & Gallagher, M. (2006). Encoding changes in orbitofrontal cortex in reversal-impaired aged rats. *J Neurophysiol, 95*(3), 1509-1517. doi:10.1152/jn.01052.2005

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nat Rev Neurosci, 20*(10), 635-644. doi:10.1038/s41583-019-0180-y

Soltani, A., Noudoost, B., & Moore, T. (2013). Dissociable dopaminergic control of saccadic target selection and its implications for reward modulation. *Proc Natl Acad Sci U S A, 110*(9), 3579-3584. doi:10.1073/pnas.1221236110

Stolyarova, A., & Izquierdo, A. (2017). Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning under uncertainty. *Elife, 6*. doi:10.7554/eLife.27483

Stolyarova, A., Rakhshan, M., Hart, E. E., O'Dell, T. J., Peters, M. A. K., Lau, H., . . . Izquierdo, A. (2019). Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. *Nat Commun, 10*(1), 4704. doi:10.1038/s41467-019-12725-1

Verharen, J. P. H., den Ouden, H. E. M., Adan, R. A. H., & Vanderschuren, L. (2020). Modulation of value-based decision making behavior by subregions of the rat prefrontal cortex. *Psychopharmacology (Berl), 237*(5), 1267-1280. doi:10.1007/s00213-020-05454-7

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron, 81*(2), 267-279. doi:10.1016/j.neuron.2013.11.005