



Published in final edited form as:

*Neurobiol Learn Mem.* 2016 May ; 131: 101–108. doi:10.1016/j.nlm.2016.03.016.

## Strain commonalities and differences in response-outcome decision making in mice

Kelsey S. Zimmermann<sup>1,2,3</sup>, Chia-Chun Hsu<sup>2,3</sup>, and Shannon L. Gourley<sup>1,2,3</sup>

<sup>1</sup> Departments of Pediatrics and Psychiatry, Emory University School of Medicine

<sup>2</sup> Graduate Program in Neuroscience, Emory University

<sup>3</sup> Yerkes National Primate Research Center, Emory University

### Abstract

The ability to select between actions that are more vs. less likely to be reinforced is necessary for survival and navigation of a changing environment. A task termed “response-outcome contingency degradation” can be used in the laboratory to determine whether rodents behave according to such goal-directed response strategies. In one iteration of this task, rodents are trained to perform two food-reinforced behaviors, then the predictive relationship between one instrumental response and the associated outcome is modified by providing the reinforcer associated with that response non-contingently. During a subsequent probe test, animals can select between the two trained responses. Preferential engagement of the behavior most likely to be reinforced is considered goal-directed, while non-selective responding is considered a failure in response-outcome conditioning, or “habitual.” This test has largely been used with rats, and less so with mice. Here we compiled data collected from several cohorts of mice tested in our lab between 2012-2015. Mice were bred on either a C57BL/6 or predominantly BALB/c strain background. We report that both strains of mice can use information acquired as a result of instrumental contingency degradation training to select amongst multiple response options the response most likely to be reinforced. Mice differ, however, during the training sessions when the familiar response-outcome contingency is being violated. BALB/c mice readily generate perseverative or habit-like response strategies when the only available response is unlikely to be reinforced, while C57BL/6 mice more readily inhibit responding. These findings provide evidence of strain differences in response strategies when an anticipated reinforcer is unlikely to be delivered.

### Keywords

prelimbic; infralimbic; orbitofrontal; striatum; reward; outcome

---

**Contact:** Shannon L. Gourley, PhD, Department of Pediatrics, Yerkes National Primate Research Center, Emory University, 954 Gatewood Dr. NE, Atlanta GA 30329, 404-727-2482, shannon.l.gourley@emory.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Introduction

The ability to recognize the relationships between actions and their outcomes, and to select a particular behavior based on how likely it is to be reinforced, is an important tool that allows organisms to engage in flexible, goal-directed decision making. Accordingly, mice, rats, and humans can select actions based on the likelihood that they will be reinforced (Balleine and O'Doherty, 2010). Habits, by contrast, are stimulus-elicited and can develop through, for example, repetition. Habits can be advantageous, freeing cognitive resources to attend to other events and environmental stimuli during the execution of a familiar behavior, but the expression of stimulus-elicited habits can also be maladaptive, for instance in the case of habitual drug seeking in addiction or inflexible, habit-like rumination in depression.

One way that we and others have dissociated goal-directed actions from habits in experimental rodents is through the use of a modified form of classical response-outcome contingency degradation tasks (Dickinson, 1980; Hammond, 1980), depicted in figure 1. First, mice are trained to generate two nose poke responses reinforced with food. Following this training, the response-outcome relationship associated with one of the responses is “degraded” by providing the corresponding reinforcer non-contingently. “Goal-directed” decision making is inferred by preferential engagement of the behavior that is likely to be reinforced, relative to the behavior that is unlikely to be reinforced, during a subsequent probe test. By contrast, mice that have developed habits engage both responses equally, even though one is no longer likely to be reinforced.

Over the course of several experiments, we have observed that healthy mice bred on C57BL/6 and BALB/c strain backgrounds can preferentially engage a response that is highly predictive of reinforcement during a probe test following response-outcome contingency degradation (Gourley et al., 2012; Gourley et al., 2013a; Swanson et al., 2013; Swanson et al., 2015; Zimmermann et al., 2015). Only mice bred on a C57BL/6 background, however, consistently inhibit responding *during* our typical 25-minute contingency degradation training session, relative to a session in which responding is reinforced. This contrast can be appreciated, for example, by comparing ‘floxed’ *Bdnf* mice bred on a mixed, but predominantly C57BL/6, background in Gourley et al., 2012, 2013b with those bred on a mixed, but predominantly BALB/c, background in Hinton et al., 2014.

Here we report side-by-side comparisons of behavioral sensitivity (or insensitivity) to response-outcome contingency degradation training in C57BL/6 and BALB/c mice tested in our laboratory between 2012-2015. These tests followed 1-2 weeks of nose poke training using schedules of reinforcement that would be expected to promote goal-directed action selection, as opposed to stimulus-response habits. We report that responding during a brief instrumental contingency degradation training session readily decreases, relative to a session in which responding is reinforced, in mice bred on a C57BL/6 background, but not in mice bred on a predominantly BALB/c background. In BALB/c mice, responding is maintained, or even increased, during an initial experience with the instrumental contingency degradation procedure. Nonetheless, all mice are capable of using information acquired as a result of this “degradation training” to subsequently select amongst multiple responses the response that is most predictive of reinforcement. These findings suggest that BALB/c mice

are more susceptible to generating habit-like or perseverative-like response patterns when the only available response is unlikely to be reinforced, even while consolidating new information regarding response-outcome contingencies to guide subsequent action selection strategies. These strain differences may be capitalized upon in studies regarding response inhibition and the expression and suppression of reward-seeking behavior.

## Methods

### Subjects

Mice were wild type mice bred on a C57BL/6 background or commercially-available “floxed” *Bdnf* mice bred on a predominantly BALB/c background (Jackson Labs). In “floxed” mice, the introduction of Cre Recombinase decreases *Bdnf* expression. In the absence of Cre Recombinase (as in this report), mice are intact and express normal levels of *Bdnf*. Both sexes were tested. Mice were 8 weeks of age, maintained on a 12-hour light cycle (0700 on), and provided food and water *ad libitum* except during instrumental conditioning. At this point, body weights were reduced to 90-95% of baseline, depending on the experiment, to motivate food-reinforced instrumental responding. Procedures were Emory University IACUC-approved.

### Instrumental response training

Mice were trained to nose poke for food reinforcement (20 mg grain-based pellets; Bioserv) using standard illuminated Med-Associates conditioning chambers with 2 nose poke apertures located on opposite sides of the chamber walls. Mice were trained using a fixed ratio 1 (FR1) schedule of reinforcement; 30 pellets were available for responding on each aperture, resulting in 60 pellets/session and equivalent experience with the response-outcome contingencies associated with both responses. Seven or 14 daily training sessions were conducted as indicated, during which all mice acquired the responses. In the final experiment depicted in figure 2, mice were shifted to a random interval 30-second schedule of reinforcement (RI-30) during the final 2 training sessions, as noted graphically.

Sessions ended when mice acquired 60 pellets or at 70 minutes (figure 2a,d), 135 minutes (figure 2g), or 25 minutes (figure 3).

### Response-outcome contingency degradation and progressive ratio testing

After response training, one nose poke aperture was occluded, leaving one recess available. During this session, reinforcers were delivered into the magazine, independent of animals' responding, for 25 min at a rate that was yoked to each animal's individual reinforcement rate from the previous session (Gourley et al., 2012; Barker et al., 2013; Gourley et al., 2013a; Gourley et al., 2013b; Swanson et al., 2013; Barker et al., 2014; Hinton et al., 2014; Gross et al., 2015; Swanson et al., 2015; Zimmermann et al., 2015). In another session, only the opposite aperture was available, and responding was reinforced using a variable ratio 2 (VR2) schedule of reinforcement for 25 min (Baldwin et al., 2002). Thus, one response became significantly more predictive of reinforcer delivery than the other (see Hinton et al., 2014 and Butkovich et al., 2015 for response-pellet delivery probability

distributions). The order of these sessions and the location of the “degraded” aperture within the chamber were counter-balanced.

The following day, both apertures were available simultaneously for 10 min; responding during this probe test was nonreinforced.

In one experiment, a second identical probe test was conducted the next day, and responding during both tests is reported. In these same mice, responding was then reinstated using a single 25-min training session and a VR2 schedule of reinforcement. The following day, mice were tested using a progressive ratio schedule of reinforcement. In this case, the response:reinforcement requirement progressive increased by 4 with each reinforcer delivered (i.e., 1,5,9,13,17...responses for a single food pellet). The highest response:reinforcement ratio the animal achieves is considered the “break point ratio.” Sessions ended when the mice had not responded on the active aperture (the aperture associated with the nondegraded contingency during instrumental contingency degradation testing) for 5 min, or at 3 hours.

### Cortical viral vector infusions

This report represents, in part, a re-analysis of data collected from control mice in several experiments conducted in our lab between 2012-2015. As such, the BALB/c mice represented in figure 2 had received prefrontal cortex-targeted infusions of lentiviruses expressing Green Fluorescence Protein (Emory Viral Vector Core). In these cases, mice were anaesthetized with ketamine/dexdomitor or ketamine/xylazine and placed in a digitized stereotaxic frame (Stoelting). The scalp was incised, skin retracted, bregma and lambda identified, the head leveled, and coordinates located. Viral vectors were infused in a volume of 0.5  $\mu$ l at AP+2.0, DV-2.5, ML $\pm$ 0.1 or AP+2.6, DV-2.8, ML $\pm$ 1.2, and the skin was sutured following infusion. Mice recovered for at least 2 weeks prior to food restriction and testing. Importantly, the mice in figure 3 are intact.

### Statistics

Response rates during training were compared by ANOVA (“to be degraded” vs. “to be non-degraded”) with repeated measures. Response rates during the instrumental contingency degradation training sessions and during the probe tests were compared by paired t-tests or two-way ANOVA as appropriate. Break point ratios were compared by t-test.

### Results

Figure 1 outlines the behavioral assay used here. Mice were trained to respond on two nose poke recesses for food reinforcement. Following response acquisition, mice were exposed to a single session of instrumental contingency degradation training. In this case, one of the recesses was occluded, and the predictive relationship between responding on the remaining aperture and the associated outcome was weakened by providing reinforcers non-contingently. In this case, pellets are delivered (by chance) within 2 seconds following only ~2.5-7% of all responses (Hinton et al., 2014; Butkovich et al., 2015). In another session, only the opposite aperture was available, and responding remained reinforced according to a VR2 schedule of reinforcement. Thus, the response-outcome contingencies associated with

both responses changed, and one response became much less likely to be reinforced. Next, mice were given access to both apertures during a brief probe test conducted in extinction.

Figure 2 depicts response acquisition curves and response rates during and after the violation of familiar instrumental contingencies in several cohorts of C57BL/6 and BALB/c mice tested in our lab between 2012-2015. Statistics associated with this figure are reported in table 1. C57BL/6 mice readily acquired the nose poke responses (fig.2a). Mice subsequently inhibited responding during a session when food pellets associated with the one available response were delivered non-contingently, relative to a test session when responding on the other aperture was reinforced (fig.2b). During a probe test, these mice also preferentially engaged the response most closely linked with reinforcement, evidence of knowledge of the action-outcome contingencies (fig.2c). This was true of both female (fig.2a-c) and male (fig. 2d-f) mice.

BALB/c mice were tested according to the same conditions. In this case, mice readily acquired the instrumental responses (fig.2g), but they failed to modify response rates when the food pellets associated with one of the responses were provided non-contingently, relative to a test session when responding on the other aperture was reinforced (fig.2h). Despite this apparent insensitivity to instrumental contingency degradation, these mice nonetheless preferentially generated the response most likely to be reinforced when both response operandi were available simultaneously the following day (fig.2i).

These findings demonstrate that mice bred on a C57BL/6 background readily inhibit responding when a familiar behavior fails to produce the expected outcome. Under the same conditions, BALB/c mice show no adjustment of response rates. Despite these differences, all mice can use information acquired as a result of this training session to choose amongst multiple behavioral response options and preferentially engage the response that is most closely coupled with reinforcement during a subsequent probe test.

One might ask: Are BALB/c mice capable of detecting the change in response-outcome contingency during our brief instrumental contingency degradation procedure? To address this question, we conducted a new experiment using naïve, intact male C57BL/6 and BALB/c mice tested simultaneously. We prolonged the initial nose poke training using shorter training sessions over the course of 2 weeks and again, a fixed ratio 1 schedule of reinforcement, which would be expected to bias decision making towards goal-directed response strategies (Dickinson et al., 1983). Response rates did not differ between groups (fig.3a) [effect of response and interaction  $F_{3, 1}$ ; main effect of session  $F_{(13,169)}=12.9$ ,  $p<0.001$ ].

Following 14 days of training, we then again decreased the likelihood that one response would be reinforced. Again, BALB/c mice failed to inhibit responding, relative to a session when responding was reinforced. And with this more protracted nose poke training history, responding was in fact energized when the only available response became unlikely to be reinforced (fig.3b) [strain  $\times$  response interaction  $F_{(1,13)}=7.2$ ,  $p=0.02$ ]. This pattern suggests that BALB/c mice can indeed detect the change in response-outcome relationship, and that the strains differ in their reaction to this expectancy violation, with C57BL/6 mice

withholding responding and BALB/c energizing responding, reminiscent of an extinction burst. Consistent with this perspective, when we analyzed response rates in 5-min bins, the greatest difference was in the first 5 min (fig.3c) [interaction  $F_{(4,52)}=6$ ,  $p<0.001$ ]. Additionally of note, magazine entries differed, with C57BL/6 mice generating consistent entry rates, while magazine entries were greatly increased in the BALB/c mice when the response-outcome relationship was most violated (fig.3d) [interaction  $F_{(1,26)}=4.7$ ,  $p=0.04$ ].

Despite these differences, mice of both strains preferentially generated the response most likely to be reinforced during a subsequent probe test when both response operandi were available (fig.3e) [main effect of response  $F_{(1,13)}=15$ ,  $p=0.002$ ]. Importantly, this response pattern was not obviously attributable to the recency of the contingency degradation experience (*i.e.*, whether the contingency degradation procedure occurred 1 or 2 days prior to the probe test) (fig.3f).

Notably, a main effect of strain was also detected during the initial probe test, with BALB/c mice generating significantly higher response rates overall (fig.3e) [main effect  $F_{(1,13)}=12$ ,  $p=0.004$ ]. To determine whether this effect was persistent, we conducted a second, identical probe test the following day. A main effect of response was again detected, indicating that mice still preferentially generated the response that was most likely to be reinforced in a goal-directed fashion (fig.3g) [ $F_{(1,13)}=4.7$ ,  $p<0.05$ ]. We identified no strain effect, however, indicating that heightened response rates in BALB/c mice in fig.3e were transient [ $F<1$ ].

As indicated above, one factor that could contribute to heightened response rates in BALB/c mice is a greater propensity to generate an “extinction burst” when a reinforcer is not delivered; this could be associated with greater “motivation” or value assigned to the food reinforcer, as has been previously suggested (see Discussion). We thus tested these same mice using a classical progressive ratio schedule of reinforcement wherein the response requirement increases with each sequential reinforcer delivered. Under these conditions, BALB/c mice generated a substantially higher number of total responses and achieved higher break point ratios [ $t_{14}=3.7$ ,  $p=0.007$ ;  $t_{14}=4.1$ ,  $p=0.002$ , respectively]. Break points are shown in fig.3h.

Finally, we replicated the instrumental contingency degradation procedure in intact female BALB/c mice. Response rates associated with the “to be degraded” and “to be non-degraded” contingencies did not differ during response acquisition (fig.3i) [interaction  $F<1$ ; main effect  $F_{(1,16)}=3.9$ ,  $p=0.07$ ]. During a contingency degradation training period, however, response rates were again energized when the only available response was unlikely to be reinforced, relative to when responding was explicitly reinforced (fig.3j) [ $t_8=-2.98$ ,  $p=0.02$ ]. When multiple response options were available, however, mice preferentially engaged the response most likely to be reinforced in a goal-directed manner (fig.3k) [ $t_8=2.6$ ,  $p=0.03$ ].

Three mice were excluded from the experiments reported in fig.3 due to response rates lying  $>2$  standard deviations above the mean or a persistent response bias during training (1 C57BL/6 mouse and 2 BALB/c mice).



## Discussion

Here we compile data from several experiments conducted in our laboratory between 2012-2015 to compare behavioral sensitivity to instrumental contingency degradation between mice bred on two strain backgrounds commonly used in neuroscience research. We trained mice to nose poke for food reinforcers using protocols and schedules of reinforcement that would be expected to promote behavioral sensitivity to changes in familiar response-outcome associative contingencies. Then, we blocked one nose poke recess and provided the food reinforcers associated with the other independently of the animals' behavior. C57BL/6 mice inhibited responding during this 25-minute training session, relative to a session when responding was reinforced. By contrast, BALB/c mice failed to inhibit, or even energized, responding during this period. Nonetheless, when multiple response options were available in a subsequent probe test, all mice preferentially engaged the response most likely to be reinforced in a goal-directed manner. In other words, even though BALB/c mice did not readily inhibit responding *during* instrumental contingency degradation training, they nonetheless used information acquired as a result of this training to later express goal-directed action selection strategies.

A dissociation between response patterns generated during training from those generated during a subsequent probe test may seem paradoxical, but it is not dissimilar to a phenomenon sometimes observed in conditioned fear extinction experiments: Specifically, freezing during an initial extinction training session in fear-conditioned mice may not differ from freezing rates generated during training, despite the absence of the aversive unconditioned stimulus (*e.g.*, foot shock). However, freezing then decreases during a second or third fear extinction training session, providing evidence that the mouse utilized the information learned previously to guide behavioral response strategies (*e.g.*, to freeze *vs.* to explore) (for instance, see: (Andero et al., 2011; Gafford et al., 2012)). Similarly, BALB/c mice here clearly acquired new outcome-based information as a result of contingency degradation training, even if they initially failed to inhibit responding relative to a session when responding was reinforced. By contrast, C57BL/6 mice readily inhibited responding when the only available response strategy was unlikely to be rewarded, as shown in figure 2 and previously demonstrated in other reports (Wiltgen et al., 2007; Gourley et al., 2012; Gourley et al., 2013b; Butkovich et al., 2015; Parnaudeau et al., 2015; Swanson et al., 2015).

In a final set of experiments, we prolonged the initial nose poke training period. Then, the contingency associated with one of the nose poke responses was again violated. Response rates in extensively-trained BALB/c mice were energized compared to a session when a different response was reinforced. This finding suggests that the BALB/c mice recognized the change in response-outcome contingency. BALB/c mice also generated higher magazine entry rates during this session. Additionally, mice of both strains consistently increased response rates when they transitioned from a fixed ratio 1 schedule of reinforcement during nose poke training to a variable ratio 2 schedule during the “non-degraded” session, further evidence of the recognition of response-outcome contingency. Together, our findings thus indicate that BALB/c mice can display behavioral sensitivity to response-outcome contingencies, even though they generate perseverative-like (figure 3) or habit-like (figure 2)

response strategies when their only available instrumental response is unlikely to be reinforced.

### Neurobiological differences between the C57BL/6 and BALB/c mouse strains

What might account for behavioral differences between the C57BL/6 and BALB/c mouse strains? The BALB/c strain is characterized by a highly anxious phenotype compared to mice bred on a C57BL/6 background and is considered “stress-susceptible” (Anisman et al., 2001; Mehta and Schmauss, 2011; Savignac et al., 2011). Additionally, BALB/c mice can respond at higher rates than C57BL/6 mice for food and liquid reinforcers (Deroche et al., 1997; Johnson et al., 2009; Johnson et al., 2010). Notably however, when responding for ethanol (Elmer et al., 1987) or cocaine (Deroche et al., 1997), C57BL/6 mice respond at significantly higher rates than the BALB/c strain. This suggests that food is uniquely reinforcing for BALB/c mice, particularly under the conditions of pronounced food restriction — *e.g.*, such that body weights during testing fall to 85% of initial body weight (Deroche et al., 1997; Johnson et al., 2009; Johnson et al., 2010) and/or when the reinforcer is highly palatable (sucrose pellets or sweet liquid) (Hutsell and Newland, 2013). Variability in stress and anxiety systems, along with the ostensibly higher value placed on food reinforcement by BALB/c mice (see also figure 3h), could dramatically affect food-reinforced instrumental decision making.

The dopaminergic system in BALB/c mice is also demonstrably different than that of the C57BL/6 strain. For example, expression and activity levels of midbrain tyrosine hydroxylase are significantly higher in BALB/c mice (Vadasz et al., 2007), and the density of somatodendritic dopamine D2 autoreceptors is lower in the substantia nigra pars compacta (SNc) (Kanes et al., 1993), suggesting diminished capacity for negative feedback on dopamine release. Notably, the SNc shapes appetitive conditioning through phasic dopaminergic bursts in response to unexpected presentation of rewards or predictive stimuli during task acquisition, and suppression of activity when an expected reward is withheld (Schultz, 1998; Brown et al., 1999; Tan and Bullock, 2008). Therefore, during instrumental contingency degradation training, when an expected reinforcer is not delivered upon completion of a response, a BALB/c midbrain dopaminergic system that has a comparative resistance to inhibition (relative to the C57BL/6 midbrain) may not immediately demonstrate a suppression of activity, which could lead to slower recognition of prediction error and delayed inhibition of the response that is unlikely to be reinforced.

The medial prefrontal cortical (mPFC) dopaminergic system is implicated in goal-directed action selection — for instance, dopamine D2 receptor activation or D1 inhibition in the infralimbic prefrontal cortex (IL) promotes flexible, goal-directed responding even after habits have otherwise formed (Barker et al., 2013), and infusions of dopamine itself into the IL also induce goal-directed response strategies at the expense of stimulus-response habits (Hitchcott et al., 2007). D2 receptor density in C57BL/6 mice is significantly higher in the lateral caudate putamen, nucleus accumbens, and SNc, while D1 receptor density at these sites trends higher in BALB/c mice (Kanes et al., 1993). If the expression patterns observed in the striatum and the midbrain are maintained in cortical regions, BALB/c mice may have



increased expression of D1 receptors and decreased expression of D2 receptors in the IL compared to C57BL/6 mice, making them more prone to habit-based response strategies.

### What about extinction?

Multiple studies have compared the performance of mice bred on these two strain backgrounds in other appetitive conditioning tasks, including response extinction. This is relevant because BALB/c mice trained for two weeks prior to the instrumental contingency degradation procedure generated higher response rates than C57BL/6 mice during the “degradation training” session and in the initial probe test conducted in extinction (figure 3c,e). These exaggerated response rates were transient and thus reminiscent of an “extinction burst.” This behavior, also referred to as “frustrative nonreward,” refers to instrumental responding that is acutely energized by the absence of an anticipated reinforcer (Amsel, 1958; Cooper, 1987; Gray, 1988; Lerman and Iwata, 1995; Corr, 2002).

Interestingly, some prior reports have identified no pronounced differences between C57BL/6 and BALB/c mice in the extinction of responding for food reinforcers (Lederle et al., 2011; Malkki et al., 2011). These studies report between-sessions response extinction, however, leaving open the possibility that these analyses failed to capture within-session strain differences, which occurred within 5 min and quickly normalized here (figure 3c). It is also important to note that instrumental contingency degradation involves the unpredictable delivery of food, introducing an element of ambiguity that is presumably present to a lesser degree in classical extinction conditioning, in which the reinforcer is withheld entirely. This could influence responding in a strain-dependent fashion. In support of this perspective, both BALB/c and C57BL/6 mice can inhibit responding according to fixed ratio schedules of reinforcement that progressively increase from 15 to 45, 90, 180, 360, and ultimately 590 over the course of several days (Hutsell and Newland, 2013), while the rapid and un-signaled modification in response-outcome contingencies here resulted in considerable strain differences. We suggest that the uncertainty engendered by instrumental contingency degradation or even the constantly changing response requirements of a within-session progressive ratio response schedule, rather than simply differences in extinction conditioning, stimulate responding in BALB/c mice and contribute to strain differences. Interestingly, perseverative nose poking in the BALB/c mice could be considered similar to the “near-miss” effect described in problematic gambling, in which the perception of an action *almost* being reinforced potentiates activity in reward-related circuits and encourages prolonged task engagement (Cocker and Winstanley, 2015).

### Conclusions

We report evidence that the behavioral response to a violation of familiar instrumental contingencies is – at least acutely – different between strains of mice commonly used in behavioral neuroscience research, and that C57BL/6 mice readily inhibit responding when an instrumental contingency is violated, while BALB/c mice are apparently indifferent or even energize responding. Nonetheless, mice bred on both C57BL/6 and BALB/c strain backgrounds can use information acquired as a result of an instrumental contingency degradation procedure to guide subsequent decision making in a goal-directed manner.

Establishing phenotypic differences between commonly used mouse strains may be important to comprehensively interpreting data and possibly to elucidating mechanisms of disease. Reduced promoter region methylation of *htr3a*, the serotonin receptor 3A (5-HT3A) gene, in outbred mice was recently associated with insensitivity to the degradation of an ethanol-reinforced response-outcome contingency (Barker et al., 2014). This finding led investigators to the discovery that a 5-HT3 antagonist interferes with ethanol seeking, but that effective dosing parameters are dependent on each individual's endophenotype. This discovery could inform the development of individualized treatment approaches for alcohol abuse. The experiments discussed here contribute to a growing body of literature dedicated to the task of defining distinct behavioral, anatomical, molecular, and genetic differences between mouse strains that are broadly used in biomedical research, with the ultimate goal of capitalizing on biological variability between strains.

## Acknowledgements

This work was supported by Children's Healthcare of Atlanta, the Brain and Behavior Research Foundation when Dr. Gourley was the Foundation's Katherine Deschner Family Investigator and PHS MH101477. The Yerkes National Primate Research Center is supported by the Office of Research Infrastructure Programs/OD P51OD011132. The Emory Viral Vector Core is supported by an NINDS Core Facilities grant, P30NS055077. We thank Ms. Amanda Allen for her assistance.

## References

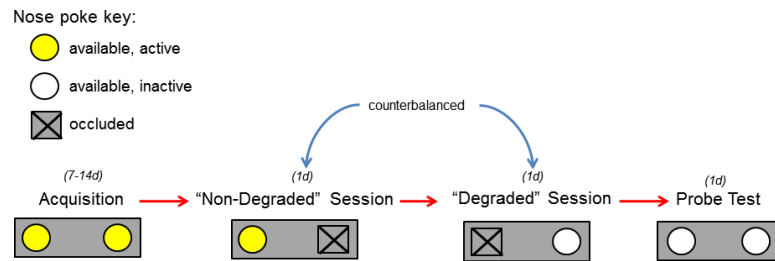
- Amsel A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol Bull.* 1958; 55:102–119. [PubMed: 13527595]
- Andero R, Heldt SA, Ye K, Liu X, Armario A, Ressler KJ. Effect of 7,8-dihydroxyflavone, a small-molecule TrkB agonist, on emotional learning. *Am J Psychiatry.* 2011; 168:163–172. [PubMed: 21123312]
- Anisman H, Hayley S, Kelly O, Borowski T, Merali Z. Psychogenic, neurogenic, and systemic stressor effects on plasma corticosterone and behavior: mouse strain-dependent outcomes. *Behav Neurosci.* 2001; 115:443–454. [PubMed: 11345969]
- Baldwin AE, Sadeghian K, Kelley AE. Appetitive instrumental learning requires coincident activation of NMDA and dopamine D1 receptors within the medial prefrontal cortex. *J Neurosci.* 2002; 22:1063–1071. [PubMed: 11826135]
- Barker JM, Torregrossa MM, Taylor JR. Bidirectional modulation of infralimbic dopamine D1 and D2 receptor activity regulates flexible reward seeking. *Front Neurosci.* 2013; 7:126. [PubMed: 23882177]
- Barker JM, Zhang H, Villafane JJ, Wang TL, Torregrossa MM, Taylor JR. Epigenetic and pharmacological regulation of 5HT3 receptors controls compulsive ethanol seeking in mice. *Eur J Neurosci.* 2014; 39:999–1008. [PubMed: 24772465]
- Brown J, Bullock D, Grossberg S. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *J Neurosci.* 1999; 19:10502–10511. [PubMed: 10575046]
- Butkovich LM, Depoy LM, Allen AG, Shapiro LP, Swanson AM, Gourley SL. Adolescent-onset GABAA  $\alpha 1$  silencing regulates reward-related decision making. *Eur J Neurosci.* 2015; 42:2114–2121. [PubMed: 26096050]
- Cocker PJ, Winstanley CA. Irrational beliefs, biases and gambling: exploring the role of animal models in elucidating vulnerabilities for the development of pathological gambling. *Behav Brain Res.* 2015; 279:259–273. [PubMed: 25446745]
- Cooper, JOH.; T.E.; Heward, WL. *Applied Behavior Analysis.* Macmillan; New York: 1987.
- Corr PJ. J. A. Gray's reinforcement sensitivity theory: tests of the joint subsystems hypothesis of anxiety and impulsivity. *Personality and Individual Differences.* 2002; 33:511–532.

- Deroche V, Caine SB, Heyser CJ, Polis I, Koob GF, Gold LH. Differences in the liability to self-administer intravenous cocaine between C57BL/6 x SJL and BALB/cByJ mice. *Pharmacol Biochem Behav.* 1997; 57:429–440. [PubMed: 9218267]
- Dickinson, A. *Contemporary Animal Learning Theory.* Cambridge University Press; Cambridge: 1980.
- Dickinson A, Nicholas D, Adams CD. The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology.* 1983; 35:35–51.
- Elmer GI, Meisch RA, George FR. Differential concentration-response curves for oral ethanol self-administration in C57BL/6J and BALB/cJ mice. *Alcohol.* 1987; 4:63–68. [PubMed: 3828066]
- Gafford GM, Guo JD, Flandreau EI, Hazra R, Rainnie DG, Ressler KJ. Cell-type specific deletion of GABA(A)alpha1 in corticotropin-releasing factor-containing neurons enhances anxiety and disrupts fear extinction. *Proc Natl Acad Sci U S A.* 2012; 109:16330–16335. [PubMed: 22992651]
- Gourley SL, Olevska A, Gordon J, Taylor JR. Cytoskeletal determinants of stimulus-response habits. *J Neurosci.* 2013a; 33:11811–11816. [PubMed: 23864670]
- Gourley SL, Olevska A, Zimmermann KS, Ressler KJ, Dileone RJ, Taylor JR. The orbitofrontal cortex regulates outcome-based decision-making via the lateral striatum. *Eur J Neurosci.* 2013b; 38:2382–2388. [PubMed: 23651226]
- Gourley SL, Swanson AM, Jacobs AM, Howell JL, Mo M, Dileone RJ, Koleske AJ, Taylor JR. Action control is mediated by prefrontal BDNF and glucocorticoid receptor binding. *Proc Natl Acad Sci U S A.* 2012; 109:20714–20719. [PubMed: 23185000]
- Gray JA. Behavioural and neural-system analyses of the actions of anxiolytic drugs. *Pharmacol Biochem Behav.* 1988; 29:767–769. [PubMed: 2901114]
- Gross C, Raj N, Molinaro G, Allen AG, Whyte AJ, Gibson JR, Huber KM, Gourley SL, Bassell GJ. Selective role of the catalytic PI3K subunit p110beta in impaired higher order cognition in fragile X syndrome. *Cell Rep.* 2015; 11:681–688. [PubMed: 25921527]
- Hammond LJ. The effect of contingency upon the appetitive conditioning of free-operant behavior. *J Exp Anal Behav.* 1980; 34:297–304. [PubMed: 16812191]
- Hinton EA, Wheeler MG, Gourley SL. Early-life cocaine interferes with BDNF-mediated behavioral plasticity. *Learn Mem.* 2014; 21:253–257. [PubMed: 24737916]
- Hitchcott PK, Quinn JJ, Taylor JR. Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. *Cereb Cortex.* 2007; 17:2820–2827. [PubMed: 17322558]
- Hutsell BA, Newland MC. A quantitative analysis of the effects of qualitatively different reinforcers on fixed ratio responding in inbred strains of mice. *Neurobiol Learn Mem.* 2013; 101:85–93. [PubMed: 23357283]
- Johnson JE, Pesek EF, Newland MC. High-rate operant behavior in two mouse strains: a response-bout analysis. *Behav Processes.* 2009; 81:309–315. [PubMed: 19429225]
- Johnson JM, Bailey JM, Johnson JE, Newland MC. Performance of BALB/c and C57BL/6 mice under an incremental repeated acquisition of behavioral chains procedure. *Behav Processes.* 2010; 84:705–714. [PubMed: 20412839]
- Kanes SJ, Hitzemann BA, Hitzemann RJ. On the relationship between D2 receptor density and neuroleptic-induced catalepsy among eight inbred strains of mice. *J Pharmacol Exp Ther.* 1993; 267:538–547. [PubMed: 7901398]
- Lederle L, Weber S, Wright T, Feyder M, Brigman JL, Crombag HS, Saksida LM, Bussey TJ, Holmes A. Reward-related behavioral paradigms for addiction research in the mouse: performance of common inbred strains. *PLoS One.* 2011; 6:e15536. [PubMed: 21249214]
- Lerman DC, Iwata BA. Prevalence of the extinction burst and its attenuation during treatment. *J Appl Behav Anal.* 1995; 28:93–94. [PubMed: 16795857]
- Malkki HA, Donga LA, De Groot SE, Neuro BMPC, Battaglia FP, Pennartz CM. Towards mouse models of perseveration: a heritable component in extinction of operant behavior in fourteen standard and recombinant inbred mouse lines. *Neurobiol Learn Mem.* 2011; 96:280–287. [PubMed: 21624482]
- Mehta M, Schmauss C. Strain-specific cognitive deficits in adult mice exposed to early life stress. *Behav Neurosci.* 2011; 125:29–36. [PubMed: 21319884]

- Parnaudeau S, Taylor K, Bolkan SS, Ward RD, Balsam PD, Kellendonk C. Mediodorsal thalamus hypofunction impairs flexible goal-directed behavior. *Biol Psychiatry*. 2015; 77:445–453. [PubMed: 24813335]
- Savignac HM, Finger BC, Pizzo RC, O'leary OF, Dinan TG, Cryan JF. Increased sensitivity to the effects of chronic social defeat stress in an innately anxious mouse strain. *Neuroscience*. 2011; 192:524–536. [PubMed: 21635938]
- Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol*. 1998; 80:1–27. [PubMed: 9658025]
- Swanson AM, Allen AG, Shapiro LP, Gourley SL. GABAA $\alpha$ 1-mediated plasticity in the orbitofrontal cortex regulates context-dependent action selection. *Neuropsychopharmacology*. 2015; 40:1027–1036. [PubMed: 25348603]
- Swanson AM, Shapiro LP, Whyte AJ, Gourley SL. Glucocorticoid receptor regulation of action selection and prefrontal cortical dendritic spines. *Commun Integr Biol*. 2013; 6:e26068. [PubMed: 24563705]
- Tan CO, Bullock D. A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. *J Neurosci*. 2008; 28:10062–10074. [PubMed: 18829964]
- Vadasz C, Smiley JF, Figarsky K, Saito M, Toth R, Gyetvai BM, Oros M, Kovacs KK, Mohan P, Wang R. Mesencephalic dopamine neuron number and tyrosine hydroxylase content: Genetic control and candidate genes. *Neuroscience*. 2007; 149:561–572. [PubMed: 17920205]
- Wiltgen BJ, Law M, Ostlund S, Mayford M, Balleine BW. The influence of Pavlovian cues on instrumental performance is mediated by CaMKII activity in the striatum. *Eur J Neurosci*. 2007; 25:2491–2497. [PubMed: 17445244]
- Zimmermann KS, Yamin JA, Rainnie DG, Ressler KJ, Gourley SL. Connections of the Mouse Orbitofrontal Cortex and Regulation of Goal-Directed Action Selection by Brain-Derived Neurotrophic Factor. *Biol Psychiatry*. 2015

**Highlights**

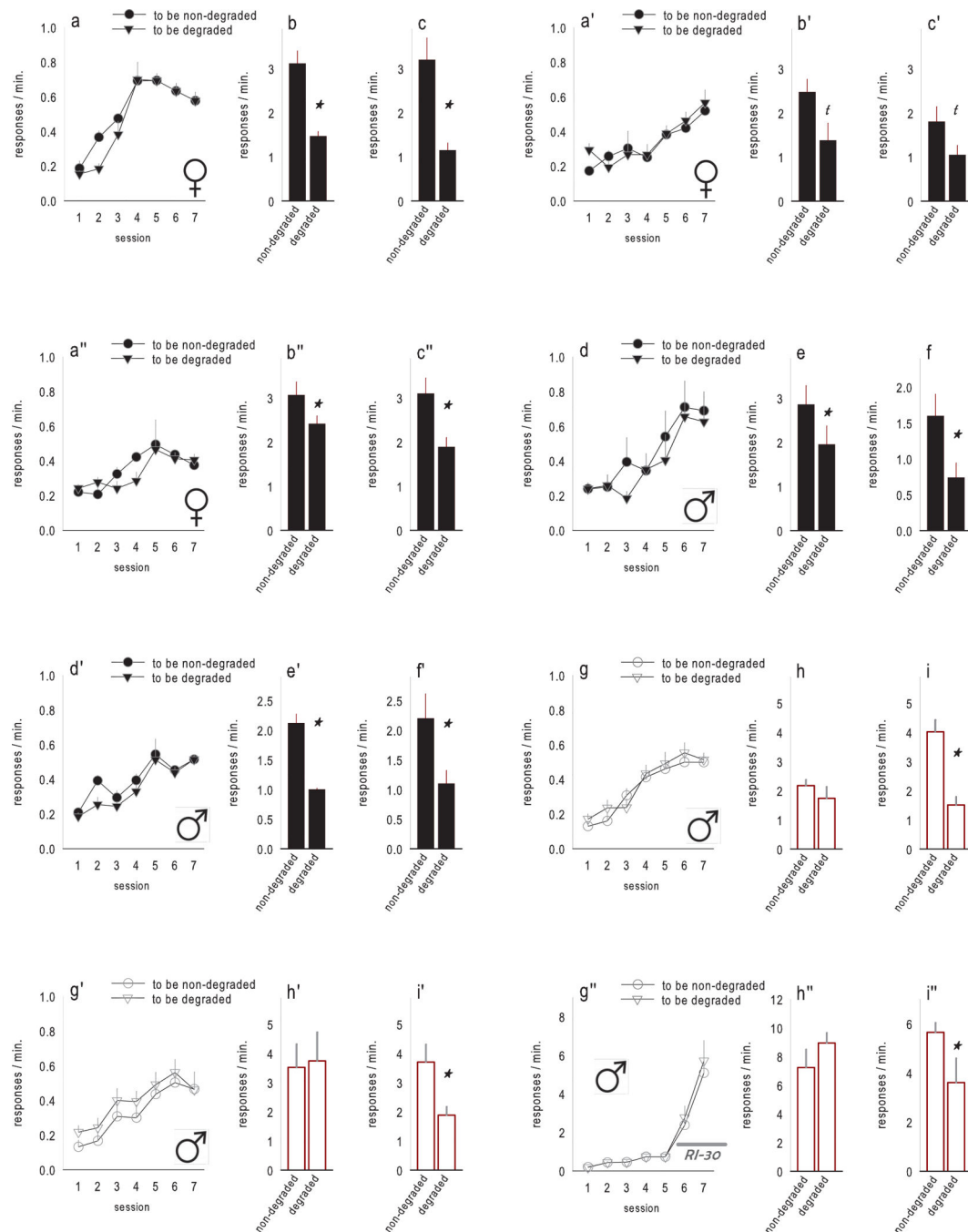
- The C57BL/6 and BALB/c mouse strains are commonly used in neuroscience research.
- Mice of both strains can select between 2 actions the one most likely to be rewarded.
- Strains differ when choice is restricted, and familiar responses are unreinforced.
- BALB/c mice favor perseverative-like responding, while B6 mice inhibit responding.



**Figure 1. Task schematic**

Mice are trained for 7 or 14 days, during which responding on 2 available nose poke recesses is reinforced with food. Following instrumental response acquisition, one of the recesses is occluded, while the other remains available and responding is reinforced. In a separate session, the opposite aperture is available but inactive; instead, reinforcer pellets are delivered independent of responding, thus “degrading” the action-outcome contingency associated with that response. Next, mice are allowed access to both apertures during a brief probe test conducted in extinction. Goal-directed action selection is characterized by preferential engagement of the response that is most likely to be reinforced. “d” refers to days.

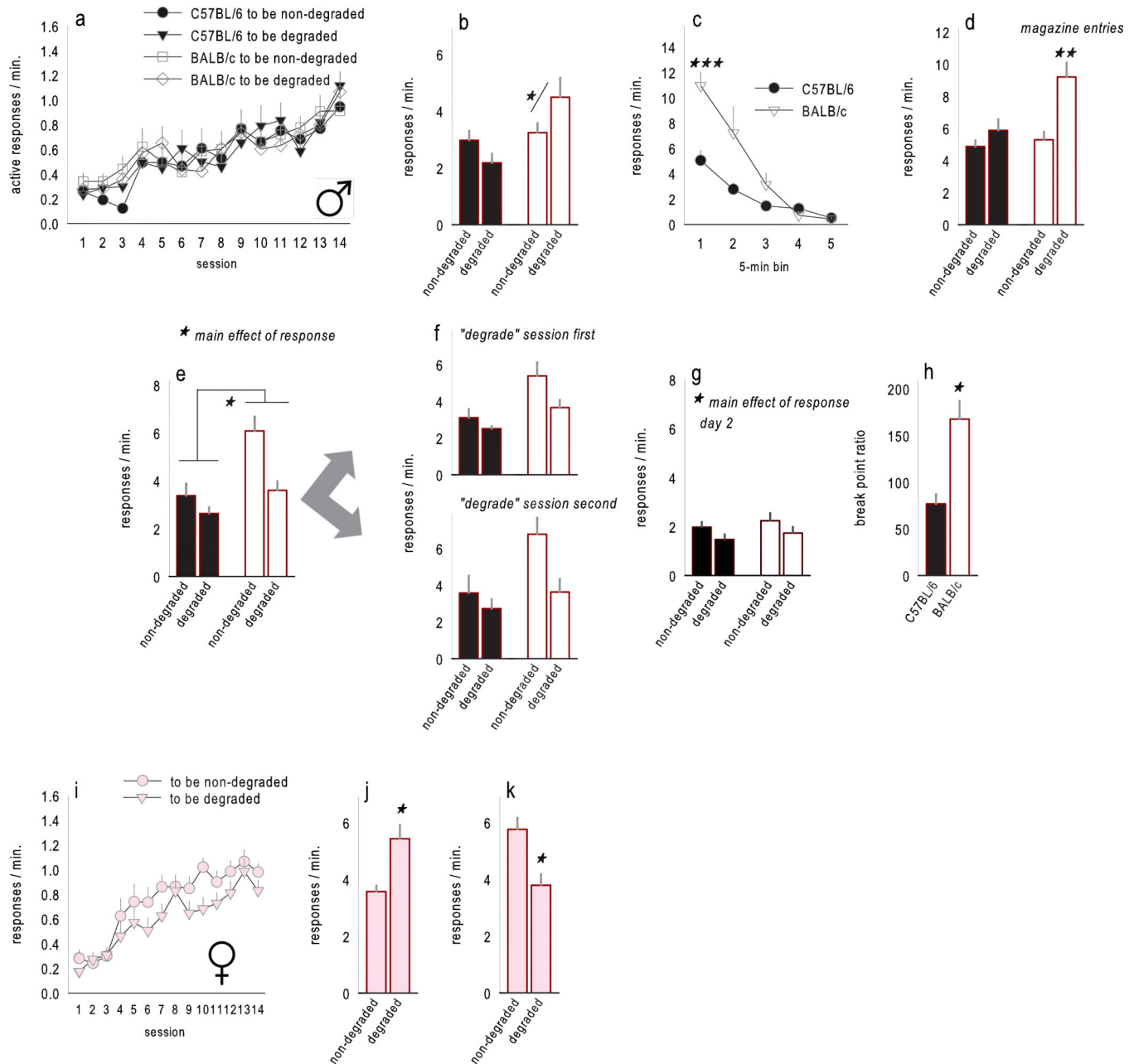




**Figure 2. Mice bred on a BALB/c, but not C57BL/6, strain background appear acutely insensitive to response-outcome contingencies**

(a) Food-reinforced response rates during the acquisition of two distinct nose poke responses in three different cohorts (a-c, a-c', and a-c'') of female C57BL/6 mice are shown. (b) Mice inhibited responding during a session when the food pellet associated with the one available response was delivered non-contingently ("degraded"), relative to a test session when the available response was reinforced ("non-degraded"). Only one nose poke recess was available during each of these sessions (see again, fig.1). (c) During a probe test when both recesses were available simultaneously, mice preferentially generated the response most

likely to be reinforced. (d) Two separate cohorts of male mice (d-f, d-f') bred on a C57BL/6 background generated the same response profiles, developing food-reinforced instrumental responses, and then (e) decreasing responding when responding was unlikely to generate a reinforcer. (f) During a probe test, mice preferentially engaged the response most likely to be reinforced. (g) Three cohorts of BALB/c mice (g-i, g-i', gi'') also acquired the instrumental responses. (h) Unlike C57BL/6 mice, BALB/c mice did not inhibit responding during instrumental contingency degradation training. (i) Nonetheless, when presented with two response options, BALB/c mice preferentially generated the response most likely to be reinforced. "RI-30" refers to an RI-30-second schedule of reinforcement. Symbols and bars represent means+SEMs, \* $p<0.05$ . "t" refers to  $p=0.06-0.08$ .



**Figure 3. Mice bred on a BALB/c, but not C57BL/6, background generate high rates of responding when a familiar response-outcome contingency is weakened**

(a) In a side-by-side comparison, C57BL/6 and BALB/c mice could acquire two distinct instrumental responses, with no response preferences or strain differences. (b) Following two weeks of training using a fixed ratio 1 schedule of reinforcement, the likelihood that one response would be reinforced was decreased ("degraded"). During this training session, BALB/c mice responded robustly, at even higher rates than when a different response was reinforced ("non-degraded"). (c) When we analyzed response rates in 5-min bins, the most significant difference was identifiable within the first 5 min, after which the strains did not differ significantly. (d) Magazine entries were also elevated in BALB/c mice during the

session when the available response was no longer likely to be reinforced (“degraded”). (e) Nonetheless, in a subsequent probe test, all mice preferentially engaged the response most likely to be reinforced. Additionally, BALB/c mice generated higher response rates overall. (f) This could not be attributed to the recency of contingency degradation training (1 or 2 sessions prior to the probe test). (g) In a second probe test, all mice again preferentially generated the response most likely to be reinforced. (h) In a progressive ratio test, BALB/c mice generated higher break point ratios. (i) As with male BALB/c mice, female BALB/c mice can also acquire two distinct instrumental responses, with no response preferences. (j) Following two weeks of training, the likelihood that one response would be reinforced was reduced. During this training session, BALB/c mice responded robustly, at higher rates than when responding was reinforced. (k) Nonetheless, when both response recesses were simultaneously available, mice preferentially engaged the response most likely to be reinforced. Symbols and bars represent means+SEMs,  $*p<0.05$  as indicated,  $**p<0.002$  vs. all other groups,  $***p<0.001$  vs. C57BL/6.

**Table 1**  
**Statistical reporting corresponding to figure 2**

Gray highlighting annotates experiments using C57BL/6 mice, while the white background refers to studies with BALB/c mice.

Figure	Response acquisition (interaction; main effect of response)	Training (paired t-test)	Probe test (paired t-test)
Fig.2a-c	$F_{(6,60)}=0.94, p=0.48; F_{(1,10)}=1.0, p=0.33$	$t_5=8.9, p<0.001$	$t_5=5.52, p=0.003$
Fig.2a-c'	$F_{(6,72)}=0.6, p=0.73; F_{(1,12)}=0.17, p=0.69$	$t_6=2.1, p=0.08$	$t_6=2.29, p=0.06$
Fig.2a-c''	$F_{(6,120)}=0.68, p=0.67; F_{(1,20)}=0.28, p=0.6$	$t_{10}=2.35, p=0.04$	$t_{10}=5.29, p<0.001$
Fig.2d-f	$F_{(6,48)}=0.55, p=0.77; F_{(1,8)}=0.28, p=0.61$	$t_4=3.8, p=0.02$	$t_4=3.15, p=0.04$
Fig.2d-f'	$F_{(6,60)}=0.46, p=0.84; F_{(1,10)}=3.13, p=0.1$	$t_5=7.73, p<0.001$	$t_5=3.83, p=0.01$
Fig.2g-i	$F_{(6,132)}=0.48, p=0.82; F_{(1,22)}=0.39, p=0.54$	$t_{11}=1.1, p=0.3$	$t_{11}=6.12, p<0.001$
Fig.2g-i'	$F_{(6,96)}=0.15, p=0.99; F_{(1,16)}=0.9, p=0.36$	$t_8=-0.31, p=0.76$	$t_8=4.1, p=0.003$
Fig.2g-i''	$F_{(6,36)}=0.18, p=0.98; F_{(1,6)}=0.21, p=0.66$	$t_3=-1.19, p=0.32$	$t_3=3.38, p=0.04$