Effects of Acute Stress on Goal-Directed and Habitual Decision-Making in Female Rats

EFFECTS OF ACUTE STRESS ON GOAL-DIRECTED AND HABITUAL DECISION-MAKING IN FEMALE RATS

A Thesis Presented

by

Russell Jacobs Dougherty

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Arts
Specializing in Psychology

August, 2024

Defense Date: May 10, 2024
Thesis Examination Committee:

Donna J. Toufexis, Ph.D., Advisor
David C. Jangraw, Ph.D., Chairperson
Eric A. Thrailkill, Ph.D.
Travis P. Todd, Ph.D.
Holger Hoock, DPhil, Dean of the Graduate College

ABSTRACT

Decision-making in dynamic environments involves a tradeoff between deliberate and habitual modes of behavior. Optimal decision-making often requires a balance between these two systems, but following stress, organisms tend to shift away from deliberation and towards habitual control. Although this has been demonstrated across a variety of tasks involving multiple memory systems, few studies have examined how acutely stressful events may differentially affect the learning and performance of an instrumental response. It is also unknown how acute stress affects the integration of goal-directed actions and habits in hierarchical planning of instrumental behavior. Here, we examined whether acute stress experienced before acquisition training (Experiment 1), or prior to the test of expression (Experiment 2) would influence goal-directed and habitual control of an instrumental response in female rats using an outcome devaluation paradigm. Experiment 3 assessed the effect of acute stress on the integration of goal-direction and habit by asking whether stress facilitates the execution of habitual sequences of actions in a serial decision task.

The results of Experiment 1 indicated that acute stress experienced before acquisition promotes habit learning, while the results of Experiment 2 indicated that habitual control of responding was present when stress was experienced just prior to the test. These data suggest that acute stress may generally facilitate habitual control over instrumental responding in free-operant tasks when response feedback is withheld. In Experiment 3, we found that on a strategic level, rats exhibited goal-directed planning of habitual action sequences under both stressed and non-stressed conditions. However, stress caused distinct effects on the structure of goal-directed planning and the performance of habitual action sequences, decreasing the flexibility and precision with which behavior was executed. These results indicate that for serial decision tasks which incentivize active monitoring of response feedback, the overall structure of decision-making remains intact following stress despite modest degradation in action planning and sequence execution. Collectively, these findings may shed light on the cognitive mechanisms by which stress contributes to impairments in decision-making, as occurs in numerous psychological conditions.

CITATIONS

Material from this thesis has been published in the following form:

Dougherty, R., Thrailkill, E. A., Mohammed, Z., VonDoepp, S., Hilton-Vanosdall, E., Charette, S., Van Horn, S., Quirk, A., Kraus, A., & Toufexis, D. J.. (2024). Acute stress facilitates habitual behavior in female rats. *Physiology & Behavior*, *275*, 114456. https://doi.org/10.1016/j.physbeh.2024.114456

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Donna Toufexis, for her support and advice throughout my research and writing process. She not only encouraged me to join the accelerated master's program, but also provided me with the freedom to pursue my own research ideas and to use novel methods that aligned with my academic interests. I would also like to give a huge thanks to Dr. Eric Thrailkill for his critical support in all three experiments for this thesis, but especially in Experiment 3. It would not have been possible without his guidance, programming skill, and constant willingness to troubleshoot as we developed the new task.

I would like to thank all the research assistants in the Toufexis lab who helped run these experiments: Zaidan Mohammed, Sarah VonDoepp, Ella Hilton-Vanosdall, Sam Charette, Sarah Van Horn, Adrianna Quirk, Adina Kraus, Caroline Genest, and Auny Kussad.

I would also like to thank Dr. Tim Stickle for his guidance and support in statistical analysis for Experiment 3. Multilevel logistic modeling was new territory for me, and his discussions provided me with the confidence that I was on the right track when I was unsure.

Lastly, I would like to thank my committee members, the Biobehavioral Cluster, and the Department of Psychological Science for their involvement throughout the thesis process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

Chapter 2

Chapter 3

CHAPTER 1: LITERATURE REVIEW

Stress changes voluntary behavior and decision making in various ways. Under acute stress, organisms tend to shift from deliberate planning of behavior towards faster and less cognitively effortful habits (Braun & Hauber, 2013; Meier et al., 2022; Schwabe et al., 2011; Schwabe & Wolf, 2009, 2010). This can be adaptive insofar as it allows us to offload well-rehearsed routines to an automatic form of control, but it can also become harmful when we relegate tasks that are best left to goal-directed planning to those of habit. Goal-directed behaviors are motivated by knowledge of action-outcome relationships and the outcome's current value, and as such they reflect an organism's present beliefs and desires. Habits, which emerge with extensive training, are supported by reinforcement history and performed without evaluation of their consequences (C. D. Adams, 1982; C. D. Adams & Dickinson, 1981; Balleine & Dezfouli, 2019). This deliberation makes goal-directed actions more flexible to changes in circumstance but also more costly both in terms of time and cognitive resources, while habits can be rapidly deployed at the expense of immediate sensitivity to changes in the values of choices (Smith & Graybiel, 2016). Deficits in goal-directed control caused or exacerbated by stress are thought to underly psychological conditions such as obsessive-compulsive disorder (OCD), substance abuse, post-traumatic stress disorder (PTSD), and depression, among others (T. G. Adams et al., 2018; Goodman et al., 2012; Griffiths et al., 2014; Hogarth, 2020; Vaghi et al., 2017; Voon et al., 2015; Yoshida et al., 2021), making them an important target for study.

1.1. Assays of goal-direction and habit

Assays of instrumental behavior have classically relied on outcome devaluation or degrading the experienced contingency between the behavior and reinforcement to parse out goal-directed and habitual processes (Dickinson, 1985; Balleine & O'Doherty, 2010). Outcome devaluation is typically achieved through selectively satiating the subject to the outcome (often a food reinforcer), or by pairing the reinforcer with an aversive stimulus to abolish its value. Following the devaluation phase, subjects are then tested in extinction conditions – where outcome delivery is omitted – so that the motivation of the behavior can be assessed independent of immediate reinforcement effects. Response rates in the devalued and non-devalued groups can then be compared to measure the sensitivity of the behavior to devaluation (Adams & Dickinson, 1981; reviewed in Watson & de Wit, 2018). If the behavior is goal-directed, it is expected that those in the devalued group would remain sensitive to the subjective quality of the outcome and make significantly less responses than those in the non-devalued group. Conversely, if the behavior is performed out of habit, no difference in response rates would be expected between those in the devalued and non-devalued groups.

1.2. Neural systems for goal-direction and habit

Research has identified two distinct memory systems responsible for coordinating goal-directed actions and habits. The goal-directed system is implemented in a cortico-striatal network involving the dorsomedial striatum (DMS), prelimbic cortex (PL) and orbitofrontal cortex (OFC), with inactivation of these areas attenuating goal-direction (Balleine & O'Doherty, 2010; Corbit & Balleine, 2003; Gremel & Costa, 2013; Shipman et al., 2018; Yin et al., 2005). Evidence from rodent studies also points to a role of the dorsal HC (dHC) in encoding causal relationships between actions and outcomes (Corbit et al., 2002; Corbit & Balleine, 2000; Miller et al., 2017), as well as for the basolateral amygdala (BLA) in encoding features of outcome incentive value (Balleine et al., 2003; Blundell et al., 2001).

On the other hand, habitual control is instantiated within a cortico-striatal circuit involving the DLS and infralimbic cortex (IL), with phasic neuronal firing in these areas coordinating the bounds of learned sequences of responses (Coutureau & Killcross, 2003; Killcross & Coutureau, 2003; Smith & Graybiel, 2016; Yin et al., 2004). Notably, artificial inactivation of components of either the goal-directed or habitual system can cause a switch in control, such that behavior becomes in-line with that of the opposite system. For example, inactivation of the DMS via blockade of the glutamatergic NMDA receptors in a goal-directed animal can cause it to behave habitually (Yin et al., 2005), and likewise lesioning the DLS of an animal can prevent it from forming a habit (Yin et al., 2004). Artificially manipulating these areas is not the only way to cause such a switch, as compounds released in the stress response have a similar effect.

1.3. Modulation of goal-directed and habitual systems by stress

In chronically stressed rats, brain areas involved in habitual control such as the DLS and nucleus accumbens core show hypertrophy and increased dendritic branching, while those in the goal-directed system, such as the DMS and mPFC, show atrophy and signs of neuronal damage (Dias-Ferreira et al., 2009; Taylor et al., 2014). These rats also exhibit habitual behavior at a level of training where goal-direction would otherwise be expected. A similar phenomenon can be seen in rats trained to perform maze navigation tasks, where the injection of anxiogenic drugs (a pharmacological mimicry of stress) into the BLA selectively impairs consolidation in a hippocampal-dependent place task while enhancing performance in a DLS-dependent response learning task (Wingard & Packard, 2008). Stress produces similar effects in humans, biasing decision-making towards procedural strategies dependent upon the sensorimotor striatum.

For instance, probabilistic classification learning (PCL) tasks can be solved through HC-based single-cue declarative strategies or striatum-based multi-cue learning (Knowlton et al., 1994, 1996; Poldrack et al., 2001). In a 2012 experiment, Schwabe & Wolf found pre-training stress to bias subjects towards the use of striatum-based procedural strategies, and the extent to which subjects relied upon procedural strategies was positively correlated with pre- and post-test salivary cortisol levels. Furthermore, subjects in the stressed group who exhibited increased hemodynamic activity in the HC during this test performed worse in the task, suggesting that the use of striatum-based procedural strategies conveys an adaptive benefit under stress. In a subsequent experiment by this group, the experimenters found that administration of the mineralocorticoid receptor (MR) antagonist spironolactone blocked the stress-induced shift towards striatal-based procedural strategy use, providing further evidence that hormones released in the stress response are involved in this memory system tradeoff (Schwabe et al., 2013).

Collectively, these studies suggest a stress-induced tradeoff between cognitive and habit memory systems that includes not only a switch between the control within the striatum (DMS vs. DLS) but that extends to other brain regions involved in deliberation (dHC, OFC) and possibly within those responsible for procedural automation (DLS).

Whether stress does so through independently altering goal-directed processes, systems for habitual responding, or their integration, is yet unclear.

1.4. Stress and reinforcement learning

Tasks that involve serially dependent decisions can be used to assess the integration of goal-directed and habitual processes in decision-making. Multi-stage decision tasks present subjects with repeated trials in which a reward can be earned through performing the correct sequence of choices. These choices can be selected through different systems, such as on the basis of their experienced reinforcement history or on prospective evaluation of their expected values (Daw et al., 2005; Dolan & Dayan, 2013). These two strategies are formalized in reinforcement learning (RL) frameworks as model-free and model-based control, respectively (Sutton & Barto, 1998). The neural substrates of these two systems are shared with those involved in instrumental, spatial, and PCL tasks, and react similarly to stress (Dolan & Dayan, 2013; Wirz et al., 2018).

Under stress, the usage of model-based strategies dependent upon areas of the prefrontal cortex (PFC), striatum, and the dHC is inversely correlated with cortisol response, supporting that the neuroendocrine stress response may degrade the computation of goal-directed planning (Daw et al., 2011; Miller et al., 2017; Otto, Raio, et al., 2013; Radenbach et al., 2015; see also van Timmeren et al., 2023). Model-free control processes in the striatum, on the other hand, are left intact or in some conditions enhanced following stress (Hollerman & Schultz, 1998; Otto, Raio, et al., 2013; Park et al., 2017; Schultz & Dickinson, 2000). As occurs in navigational learning, PCL, and other instrumental conditioning tasks, the results of these studies suggest that stress can lead to myopic decision-making by impairing deliberate planning in favor of responses taken based on retrospective learned values.

1.5. Aims of the present thesis

Despite these advances, much remains unknown about the effects of stress on aspects of learning, memory, and decision making in instrumental behavior. Rodent studies have tested the effects of acute stress on retrieval of instrumental learning (Braun & Hauber, 2013), but the effects of acute stress on instrumental acquisition in rodents have yet to be studied, to our knowledge. Thus, whether the effects acute stress experienced prior to learning a response differ from those when experienced directly prior to expression of the behavior remains unknown. Human studies of stress prior to instrumental acquisition have found acute stress to prompt habitual behavior (Schwabe et al., 2010; Schwabe & Wolf, 2009) or have failed to demonstrate goal-directed behavior in control groups (Smeets et al., 2023). Studying the effects of stress both prior to acquisition and retrieval in a well-controlled rodent experiment may provide clarification of these findings and open the door to studies on a brain circuit level.

Increasingly, research in instrumental behavior has advanced the idea that goal-directed and habitual systems operate hierarchically and may both be involved in goal selection, planning, and response performance (Ballard et al., 2024; Balleine & Dezfouli, 2019; Favila et al., 2024; Ferguson et al., 2024; Frölich et al., 2023; Schreiner et al., 2020). These theories suggest that rather than directly competing for control over behavior, these systems may organize behavior through collaborative integration of goal-directed and habitual processes. Conceptualizing instrumental behavior in this way affords simultaneous investigation of how aspects of goal-direction and habit may change under different physiological conditions. This approach has yet to be applied to understanding how stress may affect the planning and execution of instrumental behaviors, and accordingly little is known about how stress may affect hierarchical control over actions and habits.

Filling these gaps is the central aim of this thesis. In Experiments 1 and 2, we assess the impact of acute stress before training or testing of an instrumental response on its status as goal-directed or habitual. In Experiment 3, we explore the effects of acute stress on hierarchical planning of instrumental action sequences in a two-stage decision-making task. With these experiments we hope to increase understanding of how stress alters the fundamental components of instrumental learning and decision-making in dynamic environments.

CHAPTER 2: ACUTE STRESS FACILITATES HABITUAL BEHAVIOR IN FEMALE RATS

Dougherty, R., Thrailkill, E. A., Mohammed, Z., VonDoepp, S., Hilton-Vanosdall, E., Charette, S., Van Horn, S., Quirk, A., Kraus, A., & Toufexis, D. J. (2024). Acute stress facilitates habitual behavior in female rats. *Physiology & Behavior*, *275*, 114456. https://doi.org/10.1016/j.physbeh.2024.114456

Abstract

Instrumental behavior can reflect the influence of goal-directed and habitual systems. Contemporary research suggests that stress may facilitate control by the habitual system under conditions where the behavior would otherwise reflect control by the goal-directed system. However, it is unclear how stress modulates the influence of these systems on instrumental responding to achieve this effect, particularly in females. Here, we examine whether a mild psychogenic stressor experienced before acquisition training (Experiment 1), or prior to the test of expression (Experiment 2) would influence goal-directed and habitual control of instrumental responding in female rats. In both experiments, rats acquired an instrumental nose-poke response for a sucrose reward. This was followed by a reinforcer devaluation phase in which half the rats in Stressed and Non-Stressed conditions received pairings of the sucrose pellet with illness induced by lithium chloride until they rejected the pellet when offered. The remaining rats received a control treatment consisting of pellets and illness on separate days (Unpaired). Control by goal-directed and habitual systems was evaluated in a subsequent nonreinforced test of nose poking. The results of Experiment 1 indicated that the Non-Stressed Paired group reduced nose-poking compared to the Unpaired controls, identifying the response as goal directed, whereas the Stressed Paired and Unpaired groups made a similar number of nose pokes identifying the response as habitual despite a similar amount of training. Results from Experiment 2 indicated habitual control of nose-poke responding was present when stress was experienced just prior to the test. Collectively, these data suggest that stress may facilitate habitual control by altering the relative influence of goal-directed and habitual processes underpinning instrumental behavior. These results may be clinically relevant for understanding the contributions of stress to dysregulated instrumental behavior in compulsive pathologies.

2.1. Introduction

Stress can cause perseveration of established behavioral responses and hinder the ability to adapt behavior to changing environments (Schwabe & Wolf, 2010; Otto et al., 2013; Meier et al., 2022). This influence on behavioral flexibility may play a role in maladaptive voluntary behaviors that characterize diverse psychiatric conditions, including substance use disorder, obsessive-compulsive disorder, and the enduring fear experienced in post-traumatic stress disorder (Schwabe et al., 2011; Voon et al., 2015; Goodman et al., 2012). Stress also has an important influence on instrumental behavior in the laboratory (Schwabe & Wolf, 2009; reviewed in Wirz et al., 2018). Instrumental responding can be goal-directed in the sense that it is sensitive to the current value of the reinforcing outcome. In contrast, responses can be insensitive to the current value of the outcome and therefore habitual (Adams & Dickinson, 1981; Adams, 1982; Dickinson, 1985).

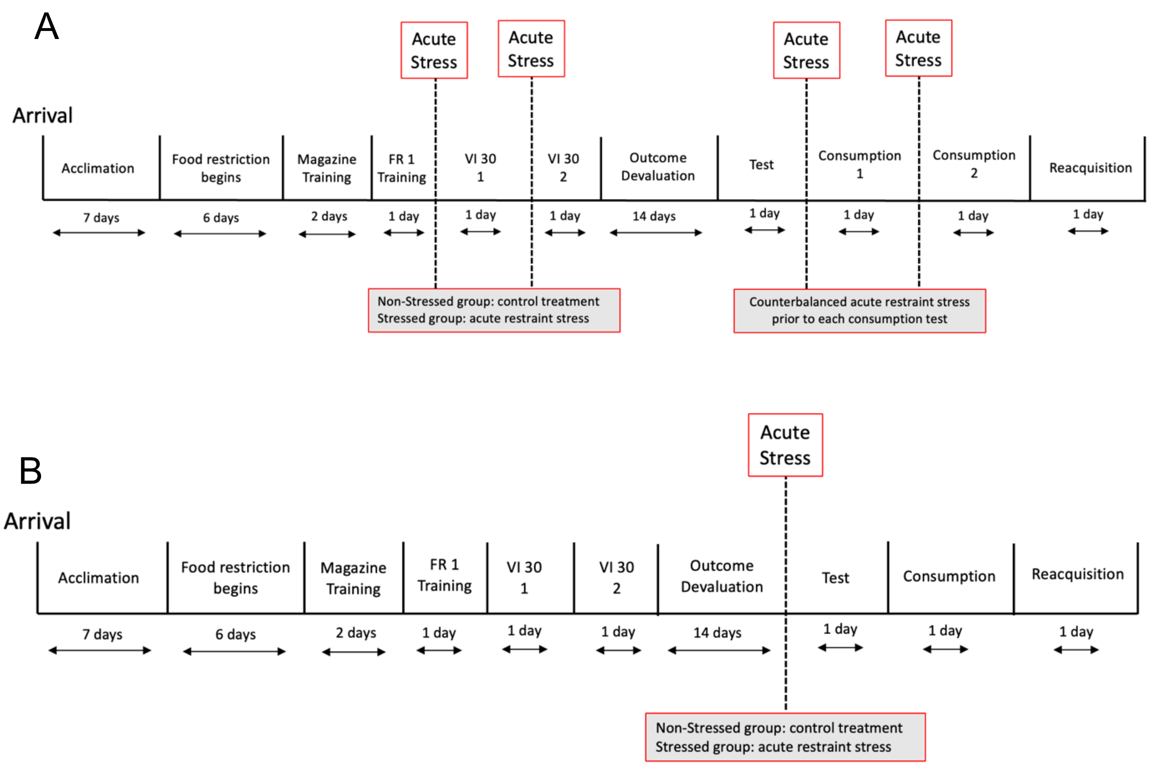
In the laboratory, goal-direction and habit can be identified with the outcome devaluation method. In a typical procedure, a rat, for instance, first acquires an instrumental response (e.g., a nose poke) paired with the contingent presentation of a reinforcing outcome. In a second phase, the response is not available to be performed and the outcome is devalued with taste aversion learning or specific satiety (Adams & Dickinson, 1981; reviewed in Watson & de Wit, 2018). The following test phase consists of the opportunity to make the response without the outcome. If the behavior is goal-directed, rats that received the devaluation treatment will make fewer responses in comparison to a control group that received identical training but did not experience a change in the outcome value. In contrast, following devaluation, if responding does not differ from controls, then the response is not sensitive to changed value and meets criteria for habit.

Stress seems to facilitate control by the habit system in both humans and rodents (Schwabe & Wolf, 2009; Dias-Ferreira et al., 2009). However, while most human studies have used acute stressors (Schwabe & Wolf, 2009; Schwabe & Wolf, 2010; Meier et al., 2022), those conducted in rodents have primarily used chronic stress paradigms (Dias-Ferreira et al., 2009; Taylor et al., 2014). The extent to which the timing of the acute stressor (i.e., before response acquisition or before testing) modulates its influence on habit, also remains unclear. Further, while stress facilitated habit control in male rats (Braun & Hauber, 2013), little is known about how stress influences control by goal-directed and habitual processes in female rats. This is especially relevant given women's higher risk for stress-related disorders (Bangasser & Valentino, 2014), and the sexually dimorphic influences of stress on other learning processes (Conrad et al., 2004; Wood & Shors, 1998; McDowell et al., 2015).

The present experiments extended our recent work with female rats (Schoenberg et al., 2019; Schoenberg et al., 2022) to examine whether a mild acute stressor prior to acquisition training (Experiment 1) or before testing (Experiment 2) enhances habitual behavior when goal-directed behavior would otherwise be expected (diagrammed in Figure 1). Based on studies of spatial learning in rodents and reward-based decision-making in humans (for review see Packard & Goodman, 2012; Schwabe & Wolf, 2011), we hypothesized that acute stress would provoke habitual responding in each experiment.

2.2. Materials and methods

Figure 1. *Experimental Timeline*



Note. A) Experimental timeline for Experiment 1: Stress before acquisition. B) Experimental timeline for Experiment 2: Stress before test.

2.2.1. Subjects

Experiments 1 and 2 were conducted successively and each included a cohort of 36 Female Long Evans rats (Charles River, Quebec; 75–90 days old at the time of arrival). Sample size was based on a prior experiment in this laboratory that found sensitivity of nose-poke responding to outcome devaluation under training conditions identical to those described below (Schoenberg et al., 2019). Rats were pair-housed with unlimited access to water in a climate-controlled colony room maintained at 23 °C with a 12-hour light/dark cycle (7:00 – 19:00 light-on period). Rats were handled by an experimenter each day. Experimental sessions were conducted each day at approximately 08:30. Rats were maintained at 85% of their free-feeding weights and were given supplementary feeding approximately 4 hr after each session. During experimental phases that included acute restraint stress (see below), we conducted vaginal cytology to confirm that normal estrus cycling continued following the stressor. Vaginal lavage was performed on these days approximately 4 hr after each experimental session, prior to supplementary feeding. All procedures were approved by the Institutional Animal Care and Use Committee at the University of Vermont.

2.2.2. Apparatus

The instrumental training equipment consisted of 6 standard rat operant chambers (model ENV-007; Med Associates, St. Albans, VT) housed within individual sound-attenuating cabinets. In the center of the right-facing chamber wall was a food cup (5 cm x 5 cm) into which a pellet dispenser (model ENV-203-45; Med Associates) could deliver 45-mg sucrose pellets (Bio-Serv). Entries to the food cup were recorded by an infrared beam. To the right of the food cup was a nose-poke response manipulandum (model ENV-114; Med Associates). The nose poke measured 2.5 cm in diameter and 2 cm deep and was located 6.3 cm (to center) above the chamber floor near the side panel that functioned as the chamber door. Entries into the nose-poke were recorded by an infrared beam. Experimental events were controlled and data were recorded by a computer located in an adjacent room.

The apparatus for acute restraint stress consisted of a cylindrical restraining device 9 cm x 15 cm (D x H) placed in a brightly lit room upon the center of a countertop.

2.2.3. Procedure

2.2.3.1. Experiment 1

The timeline of Experiment 1 is diagrammed in Figure 1a. Prior to the experiment, rats were assigned to an operant chamber in which they received all training and testing procedures. A house light located on the rear wall of the chamber signaled the beginning and end of each session.

2.2.3.1.1. Magazine training

On the first two days, all rats received 30-min sessions in which one sucrose pellet was delivered into the food cup every 60 s, on average (variable-time 60 s schedule; VT 60 s). Pellet delivery was signaled by the sound of the pellet striking the food cup; no other stimuli were presented. The nose-poke manipulandum was removed during magazine training.

2.2.3.1.2. Nose-poke training

The nose-poke device was installed in each chamber and rats received two days of free-operant nose-poke training. Each nose-poke response resulted in the delivery of a sucrose pellet (a fixed-ratio 1 schedule; FR 1). Sessions terminated after the delivery of 25 pellets or 1 hr, whichever occurred first. No other stimuli were presented during these sessions.

2.2.3.1.3. Acute stress

Rats were randomly assigned to Stressed and Non-Stressed groups. Prior to the next 2 sessions, rats in the Stressed group received 60 min of acute restraint stress in a separate brightly lit laboratory room directly before each session. Acute restraint stress has previously been demonstrated to induce robust physiological and behavioral stress responses in rats (Reis et al., 2011). Concurrently, the Non-Stressed group spent 60 min in a standard cage in a darkened room. For all rats, a variable-interval (VI) schedule arranged a sucrose pellet to become available contingent on a nose-poke response after 30 s, on average (VI 30 s). Sessions terminated after 40 reinforcers were delivered. The overall amount of instrumental training (130 reinforced responses over 4 sessions) was based on previous studies in the laboratory that found evidence of sensitivity to outcome devaluation, and hence goal-direction, following the identical amount of training with female rats (Schoenberg et al., 2019).

2.2.3.1.4. Outcome devaluation

On the following day, nose-poke manipulanda were removed from the operant chambers and the rats were randomly assigned to have the sucrose pellet devalued via pairings with an injection of lithium chloride (LiCl; Paired group) or to a non-devalued control group that received equivalent exposure to sucrose pellets and LiCl on different days (Unpaired group). Cage mates always received opposite devaluation treatments. The outcome devaluation procedure consisted of 2-day cycles (cf. Thrailkill & Bouton, 2015). On odd days (e.g., day 1, 3, 5, 7, 9, 11, 13), all rats received a 10 ml/kg intraperitoneal (i.p.) injection of 0.15 M LiCl immediately following the session. On even days, all rats received an i.p. injection of the same volume of 0.9% physiological saline. On Day 1, the Paired group received 40 sucrose pellets according to VT 30 s, the Unpaired group was placed in the chamber but did not receive pellets. On Day 2, the Unpaired group received 40 sucrose pellets on VT 30 s and Paired group did not receive pellets. This 2-day cycle repeated for the remainder of the devaluation phase. The number of pellets eaten was recorded following each session by counting any pellets left in the food cup or chamber floor. The number of pellets delivered was adjusted each cycle to match the mean number of pellets consumed by the Paired group in the preceding cycle. Cycles repeated until the number of pellets consumed by rats in the Paired group reached zero.

2.2.3.1.5. Test

On the day following the final devaluation cycle, the nose-poke device was reinstalled in the operant chambers and all rats received a 12-min test. Rats were placed in the chamber and responses were recorded but had no programmed consequences.

2.2.3.1.6. Consumption test

On days subsequent to the test day, rats received pellet consumption tests to confirm the total devaluation of the outcome. Rats were placed in the operant chamber and 10 pellets were delivered according to VT 30 s. The nose-poke devices were again physically removed from the chamber and pellet consumption was recorded. All rats received one consumption test session with 60-min acute stress immediately beforehand and one without (order counterbalanced) to assess whether the presence or absence of stress would influence the expression of the aversion to the pellets in the Paired group as indexed by free pellet consumption. Acute stress procedures were identical to those performed in acquisition.

2.2.3.1.7. Reacquisition

The next day, nose-poke manipulanda were again reinstalled in the chambers and rats received a 30-min session in which nose-poking could earn pellets according to VI 30 s.

2.2.3.2. Experiment 2

The timeline of Experiment 2 is diagrammed in Figure 1b.

2.2.3.2.1. Training

Magazine and nose-poke training procedures were identical to Experiment 1. All rats received VI-30 s training on Days 3 and 4 similar to the Non-Stressed group in Experiment 1 except that stress was not administered before the session.

2.2.3.2.2. Outcome devaluation

Following instrumental training, half the rats were randomly assigned to Paired and Unpaired groups and received the same outcome devaluation procedure as described in Experiment 1.

2.2.3.2.3. Test

On the day following the final devaluation cycle, rats were randomly assigned to Stressed and Non-Stressed groups. The Stressed group received 60 min of acute restraint stress in a separate brightly lit laboratory room immediately before the test session. The Non-Stressed group spent 60 min in a standard cage in a darkened room. The test session was the same as described in Experiment 1. Rats were placed in the chamber and allowed to make the nose poke response for 12 min. As before, responses were recorded but had no programmed consequences.

2.2.3.2.4. Consumption and reacquisition

On the following days, rats received the same pellet consumption and 30-min reacquisition test sessions as described in Experiment 1 with the exception that consumption testing was carried out only once and was not preceded by stress.

2.2.4. Statistical Analysis

Response rates (nose pokes per minute) during the VI 30-s training sessions were analyzed using analysis of variance (ANOVA). Stress (Stressed or Non-Stressed) and Devaluation (Paired or Unpaired) were entered as between-subject factors, and Session was entered as a within-subject factor. Counts of pellets consumed by Stressed Paired and Non-Stressed Paired groups in each outcome devaluation cycle were analyzed in a Stress by Devaluation by Session ANOVA.

Nose-poke and food-cup response rates from the test were analyzed as a proportion of baseline (nose-poke or food-cup entry rates at test divided by respective response rates in the final session VI 30-s training session) with planned comparison tests of the Devaluation groups in each Stress condition (Bouton et al., 2021; A. J. D. Nelson & Killcross, 2013; A. Nelson & Killcross, 2006; Schoenberg et al., 2022). The planned comparisons of devaluation groups for each stress condition were of *a priori* interest (cf. Trask et al., 2020). Pellets eaten in the consumption test and proportion of baseline response rates in the reacquisition test were analyzed with a Stress by Devaluation ANOVA.

Response rate microstructure during the final nose-poke training session and the test were analyzed by parsing nonreinforced nose-poke responses not immediately followed by checking the food cup (discrete/independent responses) from nose-pokes that were followed by a food cup entry within the subsequent 2.5-s period (nose-poke to food-cup sequences; cf. Halbout et al., 2019). These two types of response rate are illustrated in Figure 3A. Because rats earned all possible rewards during the final training session, every animal exhibited approximately 40 reinforced action sequences and thus, only nonreinforced responses were analyzed. Independent nose-poke rates and nose-poke to food-cup sequences were calculated for the final VI 30-s training session and test session. Each rate measure from the test was analyzed as a proportion of baseline (i.e., divided by the respective rate from the final training session). In each analysis, separate Stress by Devaluation ANOVAs and planned comparisons of Devaluation groups compared independent nose-poke rates and nose-poke to food-cup sequences.

For all analyses, response rate outliers were operationalized as observations two standard deviations above or below the group mean (Z +/- 2; Field, 2007). Exclusion criteria for Paired condition animals was consumption of an average of 4 or more pellets across the two consumption sessions (Experiment 1) or in the single consumption session (Experiment 2). Sphericity of the data was assessed using Mauchly's test ($p < 0.05$), Greenhouse-Geisser ($\varepsilon > 0.75$) or Huynh-Feldt ($\varepsilon < 0.75$) to determine if corrections of degrees of freedom were necessary. Effect sizes are reported when relevant to study hypotheses. Alpha was set at $p < 0.05$ for all tests. Bayesian independent samples t-tests were performed to quantify the evidence in favor of the null hypothesis when relevant (Thrailkill & Bouton, 2015; de Wit et al., 2018). Bayes factors were computed in JASP using a noninformative default Cauchy prior (JASP Team, 2023).

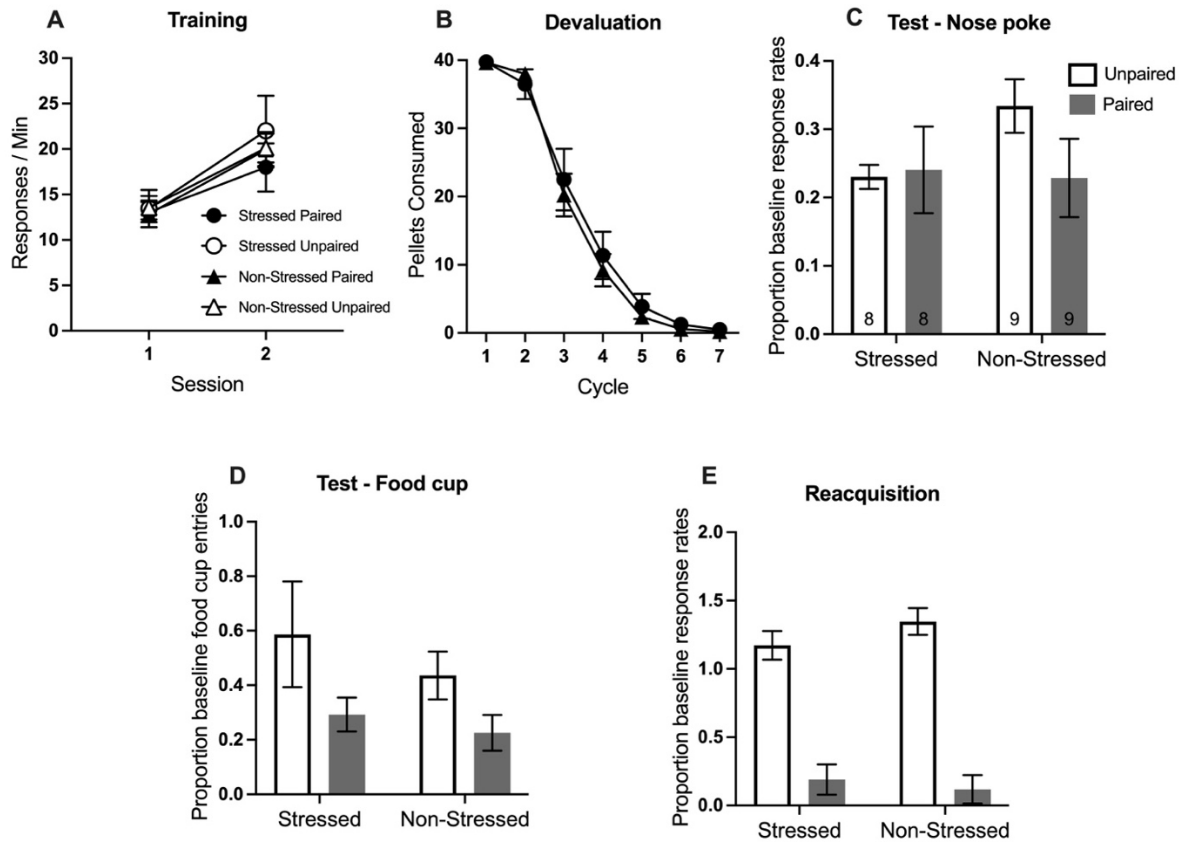
2.3. Results

2.3.1. Experiment 1

2.3.1.1. Vaginal cytology

Our analysis of vaginal cytology found no evidence to suggest that acute stress disrupted the estrous cycle.

Figure 2. *Results of Experiment 1*



Note. A) Mean nose-poke rates (responses per minute) over the two sessions of training on a VI 30-s schedule. B) Mean number of sucrose pellets consumed by Paired condition animals during each cycle of the outcome devaluation phase. C) Mean nose-poke rates during the test as a proportion of nose-poke rate in the final VI 30-s session (proportion of baseline). D) Mean food-cup entry rates during the test as a proportion of baseline. E) Mean nose-pokes per minute during the reacquisition test as a proportion of baseline. Error bars denote standard error of the mean.

2.3.1.2. Nose-poke training

As shown in Figure 2A, rats increased nose-poke response rates across the two training sessions and pre-session acute stress had no apparent effect on response rates in the Stressed group. This result was supported by a repeated measures ANOVA with a significant main effect of session, $F(1, 30) = 50.59$, $p < .001$, $\eta_p^2 = .63$. Effects involving stress or devaluation did not approach significance, Fs < 1.

2.3.1.3. Outcome devaluation

Pellets consumed by Paired groups in each cycle of the devaluation phase are shown in Figure 2B. Unpaired groups consumed all pellets (not shown). Paired groups reached the criterion of zero consumption of pellets by the seventh devaluation cycle. The ANOVA indicated no main effect of stress group or stress group by session interaction, *F*s < 1, which suggests that stress treatment did not significantly influence the outcome devaluation process.

2.3.1.4. Test

The results of the nose-poke test are shown in Figure 2C**.** Two animals were identified as outliers and excluded from all analyses based on their performance in this test: One each from Group Stressed Unpaired, $Z = 2.24$, and Group Stressed Paired, $Z = 2.46$. Consistent with our hypothesis, Non-Stressed Paired rats exhibited lower proportion baseline response rates ($M = 0.23$, $SD = 0.17$) than their Unpaired counterparts ($M = 0.33$, $SD = 0.12$). In contrast, Stressed rats exhibited similar nose-poke rates across Paired and Unpaired conditions (respectively: $M = 0.24$, $SD = 0.18$; $M = 0.23$, $SD = 0.05$). Due to positive skew in this data, we applied a logarithm (base 10) transformation to normalize the distribution. In the Non-Stressed group, Bonferroni-corrected planned comparisons found significantly lower proportion of baseline response rates in Group Paired, $F(1, 30) = 4.73$, $p = .04$, $\eta_p^2 = .14$. In contrast, Stressed Paired and Stressed Unpaired groups did not differ statistically, $F(1, 30) = 0.26$, $p = .61$, $\eta_p^2 = .01$. Further, a Bayesian independent samples t-test performed on the non-transformed data found anecdotal evidence for the null hypothesis of no difference between the Stressed Paired and Unpaired groups, $BF_{01} = 2.32$.

Food cup entry rates during the test are shown in Figure 2D. These data were also positively skewed and again logarithm (base 10) transformed. Like the pattern observed for the nose-poke response, the planned pairwise comparisons with Bonferroni corrections confirmed a significant devaluation effect on food-cup entry rates in the Non-Stressed group, $F(1, 30) = 5.26$, $p = .03$, $\eta_p^2 = .15$. The effect was not significant in the Stressed group, $p = .13$.

2.3.1.5. Consumption and reacquisition tests

Pellet consumption tests confirmed the success of the outcome devaluation procedure and that stress did not influence sucrose pellet valuation. In each pellet consumption test following stress and no stress, Stress and Non-Stressed Paired animals consumed close to no pellets, while those in the Unpaired condition consumed almost all pellets (see Table 1). No animals met exclusion criteria for pellet consumption. Results of the reacquisition test are shown in Figure 2E. Unpaired groups reacquired nose-poke responding to circa-baseline levels while Paired groups made few responses. The analysis found a significant devaluation effect, $F(1, 30) = 118.32$, $p < .001$, $\eta_p^2 = .80$. Effects involving stress did not approach significance, Fs < 1.

Table 1
Results of Experiment 1 Consumption Tests

Group	Unstressed Consumption	Stressed Consumption
Stressed Unpaired	10.00 (0.00)	9.75 (0.71)
Stressed Paired	0.88 (1.13)	0.38 (0.74)
Non-Stressed Unpaired	10.00 (0.00)	10.00 (0.00)
Non-Stressed Paired	0.11 (0.33)	0.56 (1.33)

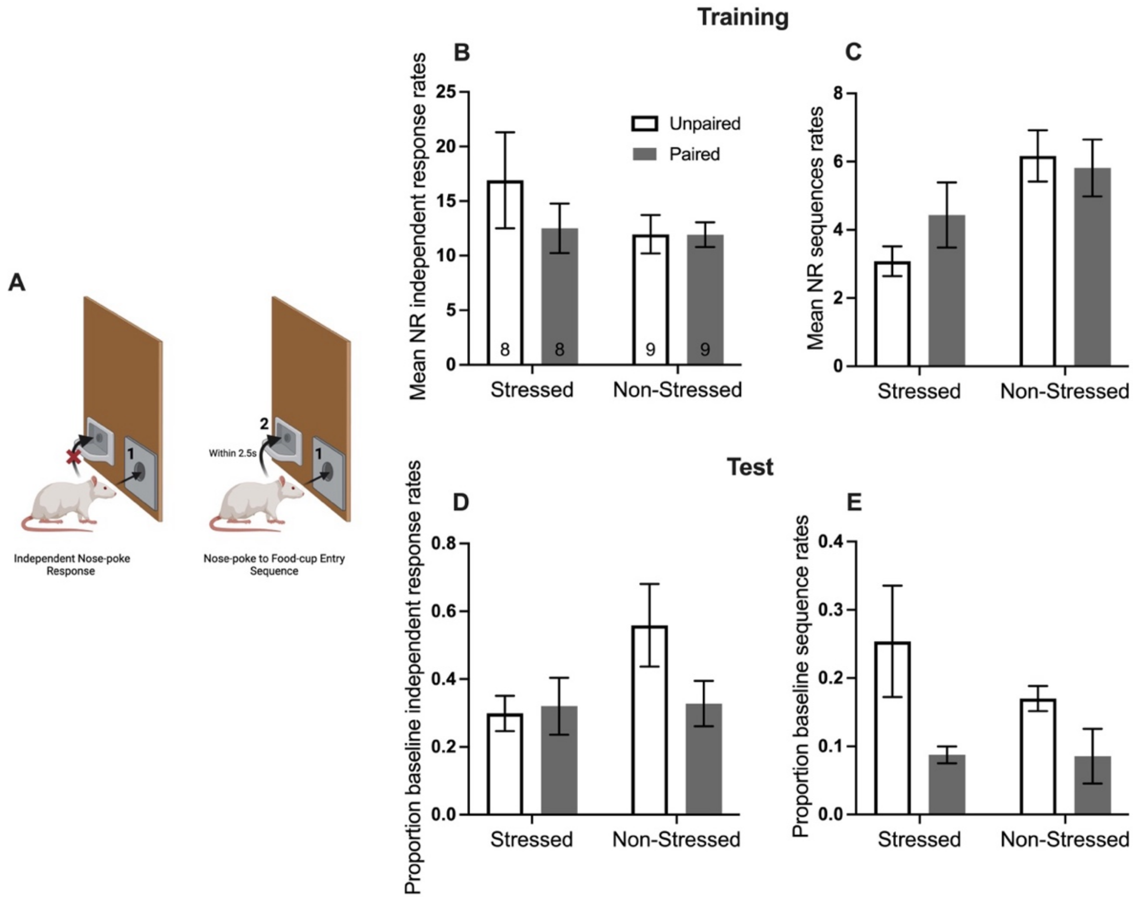
Note: Mean (standard deviation) pellet consumption during the unstressed and stressed consumption tests by stress and devaluation group.

2.3.1.6. Response rate microstructure

Figure 3B shows the nonreinforced independent nose-pokes per minute during the final session of VI 30-s training. A stress by devaluation ANOVA found no significant effects of stress, devaluation, or interaction, largest $F(1, 30) = 1.15$. Figure 3C shows nonreinforced nose-poke to food-cup sequence rates during the final training session. The same ANOVA found a significant effect of stress, $F(1, 30) = 8.29$, $p = .01$, $\eta_p^2 = .22$, and no effects involving devaluation, largest $F(1, 30) = 1.21$.

Proportion of baseline independent nose-poke rates during the test are shown in Figure 3D. This data was positively skewed and logarithm (base 10) transformed prior to analysis. Bonferroni corrected pairwise comparisons of Paired and Unpaired groups did not approach significance in the Stressed condition, $F < 1$. The same comparison of Non-Stressed Paired and Unpaired groups approached significance, $F(1, 30) = 3.53$, $p = .07$, $\eta_p^2 = .11$. Proportion of baseline nose-poke to food-cup sequence rates during the test are shown in Figure 3E. These data were also positively skewed and logarithm transformed. Planned comparisons with Bonferroni corrections found evidence of higher proportion baseline nose-poke to food-cup sequence rates in the Unpaired condition for Stressed, $F(1, 30) = 4.88$, $p = .04$, $\eta_p^2 = .14$, and Non-Stressed groups , $F(1, 30) = 11.14$, $p = .002$, $\eta_p^2 = .27$.

Figure 3. *Results of Experiment 1 response sequence analyses*



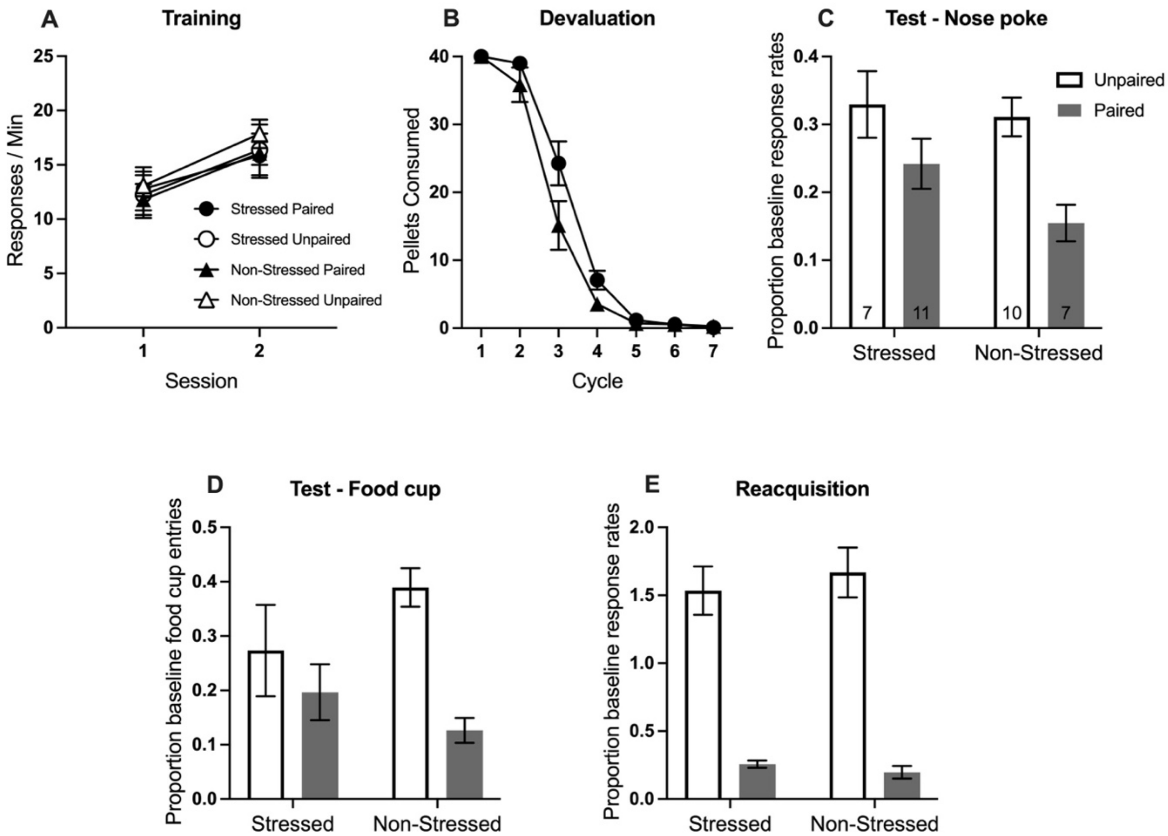
Note. A) Graphic depicting independent nose-poke responses and nose-poke to food-cup entry sequences. B) Mean nonreinforced (NR) independent nose-poke responses per minute during the final VI 30-s training session. C) Mean NR nose-poke to food-cup entry sequences per minute during the final VI 30-s training session. D) Mean independent nose-poke responses per minute during the test as a proportion of the independent response rates from the final VI 30-s training session (proportion of baseline). E) Mean nose-poke to food-cup entry sequences as a proportion of mean sequence rate in the final VI 30-s training session (proportion of baseline). Error bars denote standard error of the mean.

2.3.2. Experiment 2

2.3.2.1. Vaginal cytology

As in Experiment 1, we found no evidence of disruption to normal estrous cycling following acute restraint stress.

Figure 4. *Results of Experiment 2*



Note. A) Mean nose-poke rates (responses per minute) in the two VI 30-s training sessions. B) Mean number of sucrose pellets consumed by Paired condition animals during each cycle of the outcome devaluation phase. C) Mean nose-poke rates during the test as a proportion of nose-poke rate in the final VI 30-s training session (proportion of baseline). D) Mean food-cup entries per minute during the test as a proportion food-cup rates in the final VI 30-s training session. E) Mean nose-pokes per minute during the reacquisition test as a proportion of baseline. Error bars denote standard error of the mean.

2.3.2.2. Nose-poke training

Figure 4A shows nose-poke response rate in the first and second VI 30-s training sessions. There was a significant effect of session, $F(1, 31) = 65.93$, $p < .001$, $\eta_p^2 = .68$, and no effect of stress, devaluation, or interaction, Fs < 1.

2.3.2.3. Outcome devaluation

Pellets consumed by Paired groups in each cycle of the devaluation phase are shown in Figure 4B. Unpaired groups consumed all pellets (not shown). The analysis found a significant effect of stress, $F(1, 16) = 5.50$, $p = .03$, $\eta_p^2 = .26$, and no interaction with session, $F(1.61, 25.70) = 2.60$, $p = .10$, $\eta_p^2 = .14$. All rats met the criterion of complete cessation of pellet consumption by the end of devaluation.

2.3.2.4. Test

The Experiment 2 test results are shown in Figure 4C. One rat from Group Non-Stressed Unpaired was identified as an outlier and excluded from all analyses ($Z = 2.62$). Stressed Paired rats had slightly lower proportion baseline response rates than Stressed Unpaired rats (respectively: $M = 0.24$, $SD = 0.12$; $M = 0.33$, $SD = 0.13$). This difference was more pronounced in the Non-Stressed group (Paired: $M = 0.15$, $SD = 0.07$; Unpaired: $M = 0.31$, $SD = 0.09$). Planned comparisons with Bonferroni corrections indicated that Paired and Unpaired groups did not differ significantly in the Stressed condition, $F(1, 31) = 2.86$, $p = .10$, $\eta_p^2 = .08$. A Bayesian independent samples t-test found anecdotal evidence in favor of the null hypothesis of no devaluation effect in the Stressed group, $BF_{01} = 1.20$. Paired and Unpaired groups significantly differed in the Non-Stressed condition, $F(1, 31) = 8.81$, $p = .01$, $\eta_p^2 = .22$.

Proportion of baseline food cup entry rates during the test are shown in Figure 4D. In Group Non-Stressed, proportion of baseline food cup entry rates were lower in Paired compared to Unpaired animals. Food cup rates appeared to not differ in Group Stressed. Planned pairwise comparisons with Bonferroni corrections found that Paired and Unpaired groups differed in the Non-Stressed condition, $F(1, 31) = 12.22$, $p = .001$, $\eta_p^2 = .28$, and not in Group Stressed, $F(1, 31) = 1.08$, $p = .31$, $\eta_p^2 = .03$.

2.3.2.5. Consumption and reacquisition tests

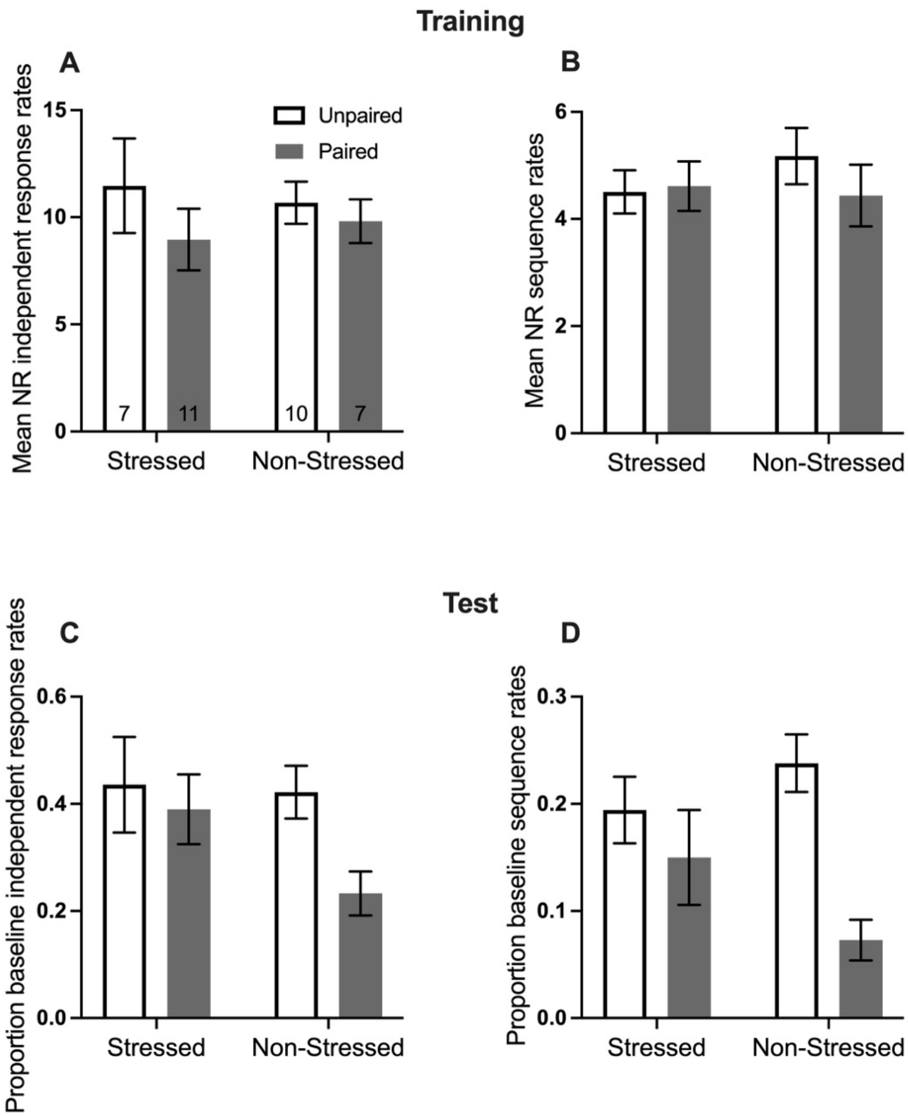
The results of the consumption test confirmed the success of the outcome devaluation procedure. Paired groups rejected nearly all pellets ($M = 0.83$, $SD = 1.42$). Unpaired groups consumed all pellets offered ($M = 10$, $SD = 0$). In the reacquisition test, stress condition did not influence the observation that Paired groups clearly failed to reacquire the nose-poking for the sucrose pellet while the Unpaired reacquired nose-poking to circa-baseline levels (see Figure 4E). There was a significant Devaluation effect, $F(1, 31) = 108.80$, $p < .001$, $\eta_p^2 = .78$, and no effects involving stress group approached significance, F's < 1.

2.3.2.6. Response rate microstructure

Mean nonreinforced independent nose-poke rates in the final training session are shown in Figure 5A. These rates did not differ between groups at this point in the experiment, F's < 1. As shown in Figure 5B, there were also no differences in nonreinforced nose-poke to food-cup entry sequence rates, F's < 1.

Figure 5C shows the proportion of baseline independent nose-poke rates per minute during the test. Independent nose pokes were lower in Paired compared to Unpaired Non-Stressed groups and supported by a Bonferroni-corrected planned comparison, $F(1, 31) = 4.18$, $p = .049$, $\eta_p^2 = .12$. These rates did not differ significantly in the Stressed group, $F < 1$. The proportion of baseline nose-poke to food-cup entry sequence rate, shown in Figure 5D, was lower in the Paired group in the Non-Stressed condition, $F(1, 31) = 10.55$, $p = .003$, $\eta_p^2 = .25$. Devaluation did not influence sequence rates in the Stressed group, $F < 1$.

Figure 5. *Results of Experiment 2 response sequence analyses*



Note. A) Mean NR independent nose-poke responses per minute during in the final VI 30-s training session. B) Mean NR nose-poke to food-cup entry sequences per minute during the final VI 30-s training session. C) Mean independent nose-poke responses per minute during the test as a proportion of baseline independent response rates from the final VI 30-s training session. D) Mean nose-poke to food-cup entry sequences as a proportion of baseline sequence rates from the final VI 30-s training session. Error bars denote standard error of the mean.

2.4. Discussion

Acute psychogenic stress resulted in habitual behavior, as evidenced by insensitivity to outcome devaluation, at a training level that otherwise maintained goal-directed behavior. In two experiments utilizing the outcome devaluation method and intact female rats, we found that acute restraint stress prior to training, and stress directly before the test both rendered a nose-poke response insensitive to outcome devaluation. These findings suggest two ways in which acute stress can alter the flexibility of decision-making. Experiment 1 suggests that acute stress facilitated habit learning, whereas Experiment 2 suggests that stress experienced after learning and just before testing may bias instrumental responding in female rats toward being expressed as habit.

Studies with human participants indicate that acute stress may shift control of instrumental behavior from goal-directed to habitual (Schwabe et al., 2012; Schwabe & Wolf, 2009, 2010; see also Buabang et al., 2022; Smeets et al., 2023). Nonhumans studies have found evidence of habit following chronic stress (Dias-Ferreira et al., 2009; Taylor et al., 2014) and acute stress (Braun & Hauber, 2013) in males. To our knowledge the present study is the first to examine facilitation of habit by acute stress in female rats.

Although across-experiment comparisons should be interpreted with caution, it does appear that male and female rats may differ in sensitivity to the severity of the stress event. The present study found that 60 min of restraint stress facilitated habitual responding in female rats. In contrast, Braun & Hauber (2013) found that males rats remained goal-directed after 60 min of restraint stress and habit was observed only after multiple stress exposures in male rats. Further work including direct comparisons of female and male rats under the same conditions is needed to assess a potential sex difference.

Other work on sex differences suggests that female rats exhibit similar or greater corticosterone responses to restraint stress than males (Aloisi et al., 1994; Babb et al., 2013; Conrad et al., 2004), as well as different effects of restraint stress on spatial (Conrad et al., 2004) and Pavlovian memory (reviewed in Shors et al., 2000). These differences, along with ovarian hormone effects in females, may mediate the differential effects of stress upon these behavior systems in males and females. Indeed, studies of ovariectomized female rats suggest that the behavioral effects of acute stress can be reversed in the absence of ovarian hormones, and that corticosterone is not required to produce stress-induced learning deficits in females (Wood et al., 2001; Wood & Shors, 1998). Similarly, recent evidence from this laboratory suggests that ovarian hormones are required for habit formation in female rats (unpublished data), and cyclic estrogen and progesterone are both required to facilitate this process (Schoenberg et al., 2022). Further research is needed to ascertain whether females are more prone to expressing habit following stress compared to males and whether such changes are mediated by gonadal hormones or adrenocortical reactivity.

Prior work has shown that, like other interoceptive states, stress may acquire a modulatory function over instrumental food-seeking behaviors (Miles et al., 2018; Schepers & Bouton, 2019). That is, nose-poking to earn the sucrose pellet while stressed could allow rats to encode the stress state as modulating the relationship between nose-poking and pellets. We assessed this potential influence in Experiment 1. Here, rats in the Stressed group received VI 30-s training for the reward in the presence of stress (note all rats received initial response training under identical nonstressed conditions). Devaluation occurred under nonstressed conditions for all rats. Given prior exposure to sucrose pellets in a stressed state, it was possible for Stressed Paired rats to have learned that stress signaled nondevalued sucrose pellets. We assessed such a possibility by testing the consumption of sucrose pellets under stressed and non-stressed conditions. The results clearly suggest that stress did not influence the value of sucrose in the consumption test. Therefore, insensitivity to devaluation observed in the Stressed groups is not complicated the possibility of the value of the sucrose pellet being modulated by stress.

While such comparisons should be made with caution, it is potentially notable that different overall levels of responding were observed in the Stressed groups in Experiments 1 and 2. In Experiment 1, responding in the Paired and Unpaired Stressed groups was similar to the Paired Non-stressed group (Figure 2C). Some work has suggested that the behavioral output of the animals during the test is a composite of both goal-directed and habitual components (Dickinson et al., 1995; Thrailkill & Bouton, 2015). In this way, it is possible that stress before acquisition may have resulted in insensitivity to outcome devaluation by reducing responding in both groups. This interpretation is speculative given the lack of influence of stress on instrumental responding during the training sessions (Figure 2A); the two groups were virtually identical. However, it remains possible that the effect of stress on instrumental control occurred during a post-training consolidation period or interfered with retrieval at the test. These results contrast with the those in Experiment 2, where stress before the test did not appear to produce an overall reduction in the Stressed groups. Altogether, these comparisons suggest directions that could clarify the possibly diverse influences of stress on instrumental behavior when experienced before training versus before testing.

To further explore these possibilities, we compared the execution of discrete nose-poke responses and nose-poke-food-cup entry sequences (Balleine & Dezfouli, 2019; Dezfouli & Balleine, 2013; Smith & Graybiel, 2016). Discrete reward-seeking responses not immediately followed by checking the food cup may be more adaptive under ambiguous reinforcement circumstances than seeking-retrieval sequences which are more resource costly (Halbout et al., 2019). On the other hand, in a reward rich environment (as in training) rats follow their reward-seeking response with an attempt to retrieve the reward more often. Previous research suggests that outcome devaluation can reduce performance of discrete seeking responses but spare seeking-retrieval sequences (Halbout et al., 2019). That is, following devaluation treatment, discrete actions may remain goal-directed while action sequences may be more apt to become habits (Watson et al., 2022). Discrete reward-seeking responses and seeking-retrieval sequences may also be dissociable at the neural level (Halbout et al., 2019; Ostlund et al., 2009).

Following this logic, in Experiment 1, if acute stress before training diminished goal-directed behavior, then Paired and Unpaired groups should not differ in their discrete nose-poking rates. Likewise, if stress had a weaker influence on the habitual component of responding, then nose-poke-food-cup sequence rate should remain sensitive to outcome devaluation. This is indeed what we found. When stressed prior to training, only the discrete nose-poke responses were insensitive to devaluation (Figure 3D-E). Interestingly, this was not the case when acute stress occurred before the test of expression in Experiment 2. Here, both discrete nose-poke and nose-poke-food-cup sequences were insensitive to the devaluation treatment (Figure 5C-D). Taken together, these results may suggest that stress before training impaired expression of goal-direction, whereas stress before the test enhanced habitual control. Further studies are needed to investigate the diverse influence of stress on aspects of instrumental behavior, particularly with female rats.

To summarize, two experiments with minimally-trained female rats found that acute stress administered either before instrumental conditioning or before the test resulted in the expression of habit in instrumental behavior, albeit by distinct mechanisms. These results may have important implications for understanding how stress may lead to behavioral inflexibility in disorders associated with compulsive responding to stimuli. This knowledge could help us engineer interventions for reinstating goal-directed control, or instituting productive habitual control, and reducing problematic behaviors in these pathologies.

References

Adams, C. D. (1982). Variations in the Sensitivity of Instrumental Responding to Reinforcer Devaluation. *The Quarterly Journal of Experimental Psychology Section B*, 34(2b), 77–98. https://doi.org/10.1080/14640748208400878

Adams, C. D., & Dickinson, A. (1981). Instrumental Responding following Reinforcer Devaluation. *The Quarterly Journal of Experimental Psychology Section B*, 33(2b), 109–121. https://doi.org/10.1080/14640748108400816

Aloisi, A. M., Steenbergen, H. L., van de Poll, N. E., & Farabollini, F. (1994). Sex-dependent effects of restraint on nociception and pituitary-adrenal hormones in the rat. *Physiology & Behavior*, *55*(5), 789–793. https://doi.org/10.1016/0031-9384(94)90061-2

Babb, J. A., Masini, C. V., Day, H. E. W., & Campeau, S. (2013). Stressor-specific effects of sex on HPA axis hormones and activation of stress-related neurocircuitry. *Stress (Amsterdam, Netherlands)*, *16*(6), 664–677. https://doi.org/10.3109/10253890.2013.840282

Balleine, B. W., & Dezfouli, A. (2019). Hierarchical Action Control: Adaptive Collaboration Between Actions and Habits. *Frontiers in Psychology*, 10, 2735. https://doi.org/10.3389/fpsyg.2019.02735

Bangasser, D. A., & Valentino, R. J. (2014). Sex differences in stress-related psychiatric disorders: Neurobiological perspectives. *Frontiers in Neuroendocrinology*, 35(3), 303–319. https://doi.org/10.1016/j.yfrne.2014.03.008

Bouton, M. E., Allan, S. M., Tavakkoli, A., Steinfeld, M. R., & Thrailkill, E. A. (2021). Effect of context on the instrumental reinforcer devaluation effect produced by taste aversion learning. *Journal of Experimental Psychology. Animal Learning and Cognition*, 47(4), 476–489. https://doi.org/10.1037/xan0000295

Braun, S., & Hauber, W. (2013). Acute stressor effects on goal-directed action in rats. *Learning & Memory*, 20(12), 700–709. https://doi.org/10.1101/lm.032987.113

Buabang, E. K., Boddez, Y., Wolf, O. T., & Moors, A. (2022). The role of goal-directed and habitual processes in food consumption under stress after outcome devaluation with taste aversion. *Behavioral Neuroscience*. https://doi.org/10.1037/bne0000439

Conrad, C. D., Jackson, J. L., Wieczorek, L., Baran, S. E., Harman, J. S., Wright, R. L., & Korol, D. L. (2004). Acute stress impairs spatial memory in male but not female rats: Influence of estrous cycle. *Pharmacology Biochemistry and Behavior*, 78(3), 569–579. https://doi.org/10.1016/j.pbb.2004.04.025

de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., Evans, M., Mirza, H., & Gillan, C. M. (2018). Shifting the Balance Between Goals and Habits: Five Failures in Experimental Habit Induction. *Journal of Experimental Psychology. General*, *147*(7), 1043–1065. https://doi.org/10.1037/xge0000402

Dezfouli, A., & Balleine, B. W. (2013). Actions, Action Sequences and Habits: Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized. *PLOS Computational Biology*, 9(12), e1003364. https://doi.org/10.1371/journal.pcbi.1003364

Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., Costa, R. M., & Sousa, N. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science* (New York, N.Y.), 325(5940), 621–625. https://doi.org/10.1126/science.1171203

Dickinson, A. (1985). Actions and habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 308(1135), 67–78. https://doi.org/10.1098/rstb.1985.0010

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995). Motivational control after extended instrumental training. *Animal Learning & Behavior*, 23(2), 197–206. https://doi.org/10.3758/BF03199935

Field, A.P., 2007. Discovering Statistics Using SPSS: (And Sex, Drugs and Rock'n'roll), 2. ed. SAGE Publ. reprinted.

Goodman, J., Leong, K.-C., & Packard, M. G. (2012). Emotional modulation of multiple memory systems: Implications for the neurobiology of post-traumatic stress disorder. *Reviews in the Neurosciences*, 23(5–6). https://doi.org/10.1515/revneuro-2012-0049

Halbout, B., Marshall, A. T., Azimi, A., Liljeholm, M., Mahler, S. V., Wassum, K. M., & Ostlund, S. B. (2019). Mesolimbic dopamine projections mediate cue-motivated reward seeking but not reward retrieval in rats. *ELife*, 8, e43551. https://doi.org/10.7554/eLife.43551

JASP Team (2023). JASP (Version 0.17.3)[Computer software].

McDowell, A. L., Fransen, K. M. H., Elliott, K. S., Elghouche, A., Kostylev, P. V., O'Dea, P. K., & Garraghty, P. E. (2015). Sex Differences and the Impact of Chronic Stress and Recovery on Instrumental Learning. *Neuroscience Journal*, 2015, 1–12. https://doi.org/10.1155/2015/697659

Meier, J. K., Staresina, B. P., & Schwabe, L. (2022). Stress diminishes outcome but enhances response representations during instrumental learning. *ELife*, 11, e67517. https://doi.org/10.7554/eLife.67517

Miles, O. W., Thrailkill, E. A., Linden, A. K., May, V., Bouton, M. E., & Hammack, S. E. (2018). Pituitary Adenylate Cyclase-Activating Peptide in the Bed Nucleus of the Stria Terminalis Mediates Stress-Induced Reinstatement of Cocaine Seeking in Rats. *Neuropsychopharmacology*, *43*(5), Article 5. https://doi.org/10.1038/npp.2017.135

Nelson, A. J. D., & Killcross, S. (2013). Accelerated habit formation following amphetamine exposure is reversed by D1, but enhanced by D2, receptor antagonists. *Frontiers in Neuroscience*, 7, 76. https://doi.org/10.3389/fnins.2013.00076

Nelson, A., & Killcross, S. (2006). Amphetamine Exposure Enhances Habit Formation. *Journal of Neuroscience*, 26(14), 3805–3812. https://doi.org/10.1523/JNEUROSCI.4305-05.2006

Ostlund, S. B., Winterbauer, N. E., & Balleine, B. W. (2009). Evidence of Action Sequence Chunking in Goal-Directed Instrumental Conditioning and Its Dependence on the Dorsomedial Prefrontal Cortex. *Journal of Neuroscience*, *29*(25), 8280–8287. https://doi.org/10.1523/JNEUROSCI.1176-09.2009

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, 110(52), 20941–20946. https://doi.org/10.1073/pnas.1312011110

Packard, M. G., & Goodman, J. (2012). Emotional arousal and multiple memory systems in the mammalian brain. *Frontiers in Behavioral Neuroscience*, 6. https://doi.org/10.3389/fnbeh.2012.00014

Reis, D. G., Scopinho, A. A., Guimarães, F. S., Corrêa, F. M. A., & Resstel, L. B. M. (2011). Behavioral and Autonomic Responses to Acute Restraint Stress Are Segregated within the Lateral Septal Area of Rats. *PLoS ONE*, 6(8), e23171. https://doi.org/10.1371/journal.pone.0023171

Schepers, S. T., & Bouton, M. E. (2019). Stress as a context: Stress causes relapse of inhibited food seeking if it has been associated with prior food seeking. *Appetite*, 132, 131–138. https://doi.org/10.1016/j.appet.2018.10.016

Schoenberg, H. L., Bremer, G. P., Carasi-Schwartz, F., VonDoepp, S., Arntsen, C., Anacker, A. M. J., & Toufexis, D. J. (2022). Cyclic estrogen and progesterone during instrumental acquisition contributes to habit formation in female rats. *Hormones and Behavior*, 142, 105172. https://doi.org/10.1016/j.yhbeh.2022.105172

Schoenberg, H. L., Sola, E. X., Seyller, E., Kelberman, M., & Toufexis, D. J. (2019). Female rats express habitual behavior earlier in operant training than males. *Behavioral Neuroscience*, 133(1), 110–120. https://doi.org/10.1037/bne0000282

Shors, T. J., Beylin, A. V., Wood, G. E., & Gould, E. (2000). The modulation of Pavlovian memory. *Behavioural Brain Research*, *110*(1), 39–52. https://doi.org/10.1016/S0166-4328(99)00183-7

Schwabe, L., & Wolf, O. T. (2009). Stress Prompts Habit Behavior in Humans. *The Journal of Neuroscience*, 29(22), 7191–7198. https://doi.org/10.1523/JNEUROSCI.0979-09.2009

Schwabe, L., & Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology*, 35(7), 977–986. https://doi.org/10.1016/j.psyneuen.2009.12.010

Schwabe, L., & Wolf, O. T. (2011). Stress-induced modulation of instrumental behavior: From goal-directed to habitual control of action. *Behavioural Brain Research*, *219*(2), 321–328. https://doi.org/10.1016/j.bbr.2010.12.038

Schwabe, L., Dickinson, A., & Wolf, O. T. (2011). Stress, habits, and drug addiction: A psychoneuroendocrinological perspective. *Experimental and Clinical Psychopharmacology*, 19(1), 53–63. https://doi.org/10.1037/a0022212

Schwabe, L., Tegenthoff, M., Höffken, O., & Wolf, O. T. (2012). Simultaneous Glucocorticoid and Noradrenergic Activity Disrupts the Neural Basis of Goal-Directed Action in the Human Brain. *Journal of Neuroscience*, 32(30), 10146–10155. https://doi.org/10.1523/JNEUROSCI.1304-12.2012

Smeets, T., Ashton, S. M., Roelands, S. J. A. A., & Quaedflieg, C. W. E. M. (2023). Does stress consistently favor habits over goal-directed behaviors? Data from two preregistered exact replication studies. *Neurobiology of Stress*, 23, 100528. https://doi.org/10.1016/j.ynstr.2023.100528

Smith, K. S., & Graybiel, A. M. (2016). Habit formation. *Dialogues in clinical neuroscience*, *18*(1), 33–43. https://doi.org/10.31887/DCNS.2016.18.1/ksmith

Taylor, S. B., Anglin, J. M., Paode, P. R., Riggert, A. G., Olive, M. F., & Conrad, C. D. (2014). Chronic stress may facilitate the recruitment of habit- and addiction-related neurocircuitries through neuronal restructuring of the striatum. *Neuroscience*, 280, 231–242. https://doi.org/10.1016/j.neuroscience.2014.09.029

Thrailkill, E. A., & Bouton, M. E. (2015). Contextual control of instrumental actions and habits. *Journal of Experimental Psychology. Animal Learning and Cognition*, 41(1), 69–80. https://doi.org/10.1037/xan0000045

Trask, S., Shipman, M. L., Green, J. T., & Bouton, M. E. (2020). Some factors that restore goal-direction to a habitual behavior. *Neurobiology of Learning and Memory*, 169, 107161. https://doi.org/10.1016/j.nlm.2020.107161

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., Schreiber, L. R. N., Gillan, C., Fineberg, N. A., Sahakian, B. J., Robbins, T. W., Harrison, N. A., Wood, J., Daw, N. D., Dayan, P., Grant, J. E., & Bullmore, E. T. (2015). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, 20(3), Article 3. https://doi.org/10.1038/mp.2014.44

Watson, P., & de Wit, S. (2018). Current limits of experimental research into habits and future directions. *Current Opinion in Behavioral Sciences*, 20, 33–39. https://doi.org/10.1016/j.cobeha.2017.09.012

Watson, P., O'Callaghan, C., Perkes, I., Bradfield, L., & Turner, K. (2022). Making habits measurable beyond what they are not: A focus on associative dual-process models. *Neuroscience & Biobehavioral Reviews*, *142*, 104869. https://doi.org/10.1016/j.neubiorev.2022.104869

Wirz, L., Bogdanov, M., & Schwabe, L. (2018). Habits under stress: Mechanistic insights across different types of learning. *Current Opinion in Behavioral Sciences*, 20, 9–16. https://doi.org/10.1016/j.cobeha.2017.08.009

Wood, G. E., & Shors, T. J. (1998). Stress facilitates classical conditioning in males, but impairs classical conditioning in females through activational effects of ovarian hormones. *Proceedings of the National Academy of Sciences of the United States of America*, 95(7), 4066–4071.

Wood, G. E., Beylin, A. V., & Shors, T. J. (2001). The contribution of adrenal and reproductive hormones to the opposing effects of stress on trace conditioning males versus females. *Behavioral Neuroscience*, 115(1), 175–187. https://doi.org/10.1037/0735-7044.115.1.175

CHAPTER 3: EFFECTS OF STRESS ON HIERACHICAL PLANNING OF SEQUENTIAL ACTIONS

Dougherty, R., Thrailkill, E.A., Van Horn, S., Kussad, A., & Toufexis, D.J. Effects of acute stress on hierarchical planning of sequential actions. Manuscript in preparation for submission.

Abstract

Stressful events typically enhance habitual forms of behavior across various types of decision-making. Research into this phenomenon has largely assessed the role of stress in the tradeoff between goal-directed and habitual systems, or their putative computational analogs: model-based and model-free reinforcement learning. However, recent work has advanced the idea that goal-direction and habit may exist in a hierarchy, where both contribute to the selection and performance of behavior. In this model, habits are characterized by sequences of responses executed as a functional unit and are selected by a higher-level planner. This account of habits may help explain the role of stress in promoting habitual action sequences in decision-making tasks and in exacerbating behavioral stereotypies. Our aim in the present study was to assess whether stress prompted the usage of habitual action sequences over the selection of discrete actions in a serial decision task. We trained 16 female Long Evans rats in a two-stage binary choice task that could be solved using either chunked sequences of actions or step-by-step discrete choices. Rats then underwent two probe tests, one following 60 minutes of restraint stress and one under control conditions, to identify the extent to which stress affected higher-level planning of behavior and the usage of habitual action sequences. We found that under both stressed and control conditions, rats exhibited goal-directed planning of habitual sequences of actions. However, following stress, rats showed a greater tendency to repeat just-performed sequences of actions, indicating that stress may induce modest exploitative biases in action planning. Stress also resulted in longer latencies between stage 1 and stage 2 responses in the task – marginally degrading action sequence integrity despite conserving their selection. Taken together, these findings suggest that acute stress does not disrupt the overall macrostructure of behavior in two-stage decision-making; although, it does alter the microstructure of goal-directed and habitual control themselves. Further, these results imply that the extent to which stress occludes goal-directed control over behavior may depend on the incentive structure and attentional demands of the decision environment.

3.1. Introduction

Following stress or extensive practice, the behavior of organisms tends to become invariant and stereotyped, with actions following each other in quick succession. This invariant behavior is characteristic of habits, and exists in a variety of behavior systems including grooming behaviors (Aldridge et al., 2004; Kalueff et al., 2016), maze navigation strategies (Smith & Graybiel, 2016), the structure of instrumental behavior chains (Dezfouli & Balleine, 2012, 2013; Frölich et al., 2023; Jin & Costa, 2010), and ritual behaviors such as those in obsessive compulsive disorder (OCD) and autism spectrum disorder (ASD) (Banca et al., 2023; Graybiel, 2008; Ridley, 1994). Dual process theories of instrumental behavior suggest a tension between this habitual system, driven by stimulus-responses associations, and a goal-directed system governed by action-outcome associations which is sensitive to changes in contingency and outcome value (C. D. Adams & Dickinson, 1981; Dickinson & Balleine, 1994). In instrumental conditioning experiments, stress has been shown to bias control towards the habitual system (Dias-Ferreira et al., 2009; Schwabe & Wolf, 2009, 2010); however, the facilitation of habits and stereotyped behaviors by stress appears to extend across multiple behavioral tasks and memory systems. In addition to promoting habits at the expense of goal-directed actions in instrumental conditioning, stress also prompts the usage of habits at the expense of cognitive memory in other dual-solution tasks (reviewed in Packard et al., 2021; Wirz et al., 2018). Habitual behavior in these tasks is often defined by invariant sequences of actions, or action chunking (Balleine & Dezfouli, 2019).

Action sequences are concatenated series of responses which become organized into a functional unit with increasing practice (Balleine & Dezfouli, 2019; Dezfouli et al., 2014; Smith & Graybiel, 2013, 2016). These chunked representations of behavior are exhibited across various tasks including those classically thought to contain habitual patterns of action, such as motor skill learning and operant paradigms (Dezfouli & Balleine, 2013, 2019; Halbout et al., 2019; Keele, 1968; Nissen & Bullemer, 1987; Pew, 1966; van Elzelingen et al., 2022), but may also extend to other types of memory for planning with chunked representation structures (e.g., Gobet & Simon, 1998; Huys et al., 2015). By concatenating together actions that are often executed in a sequence, a course of action can be evaluated based on the expected value of the sequence as a whole (e.g., R1, R2, R3) (Dezfouli & Balleine, 2012). This reduces time spent on deliberation over the outcomes of each action within the sequence, allowing for faster and smoother performance of the macro-behavior. However, action chunking also comes at the expense of insensitivity to changes in the values of outcomes within the sequence of actions, as planning occurs only at the level of the outcome of the entire sequence. This property – the automatic elicitation of the next response within a sequence – is termed open-loop control, as opposed to evaluation of the outcome of each constituent response (closed-loop) (Dezfouli & Balleine, 2012).

The role of habitual action sequences in voluntary behavior has been demonstrated in maze navigation (Barnes et al., 2005; Graybiel, 1998, 2008; Thorn et al., 2010), operant outcome devaluation paradigms (Dezfouli et al., 2014), and most clearly, in multi-stage decision tasks (Daw et al., 2011; Dezfouli & Balleine, 2013, 2019). Action sequences can be identified in these tasks through the offline revaluation of actions within the sequence. Sequences form as subjects are trained to make binary choices in two distinct task stages, with the choice taken in the second stage leading to reward. The task is arranged such that after the subject executes a response in stage 1 (R_1), it is then transitioned to a stage 2 state (denoted by a discriminative stimulus) where it must then take the opposite action (R_2) to earn a reward. However, on any given trial, in only one stage 2 state a reward can be earned – which shifts across trials with some small probability. The subject must then track the stage 2 state being rewarded and take the correct stage 1 action as to bring them to that state.

Optimal behavior in stage 1 of the task thus entails an understanding of the two-stage nature of the task, wherein each stage is encoded as a distinct event with an opposing appropriate response, as well as active tracking of trial-to-trial reinforcement. Tracking trial-to-trial reinforcement to guide behavior requires consistent sensitivity to action-outcome contingency - staying on the same stage 1 action following trials in which a reward is earned, and switching stage 1 actions when reward is omitted. Sensitivity to action-outcome contingency is a key criterion of goal-directed behavior (Dickinson & Balleine, 1994), and as such, selection of stage 1 actions according to preceding reinforcement suggests goal-directed planning of choices.

There are several possible solutions to selecting actions in stage 2 of the task. First, because the stage 2 state is signaled by a discriminative stimulus (e.g., a light cue), actions can be selected on the basis of their association with reward in the presence of the cue, which remains consistent throughout the task. These responses could represent discrete actions selected separately from those of stage 1. Alternatively, since response contingencies between stage 1 and 2 actions are arranged in a deterministic manner (i.e.,

R_1R_2 -> reward), subjects could concatenate responses together to form a predictably rewarded sequence, and indeed, this has been demonstrated in humans and in rodents (Dezfouli et al., 2014; Dezfouli & Balleine, 2013, 2019). This concatenation of responses into a unit allows for goal-directed planning of behavior at the level of the sequence, rather than at the level of its constituent responses.

Action sequences can be identified in the two-stage task through several indices. First is the characteristic reduction in time to perform the sequence that occurs with habit formation (Balleine & Dezfouli, 2019; Smith & Graybiel, 2013). Second is the addition of probe sessions, which include rare trials wherein stage 1 actions could bring the subject to the opposite stage 2 state, occurring on the order of 20% of trials (Dezfouli & Balleine, 2019). This creates a situation where to earn reward, the subject must inhibit performance of the previously learned action sequence and instead deliberately take the action signaled by the discriminative stimulus in that stage to earn reward. With this addition, a tension is established between step-by-step deliberation and execution of the sequence, where optimality is achieved through a combination of both strategies (Dezfouli & Balleine, 2012). Habitual action sequences have been shown to interfere with goal-directed choice under such circumstances (Frölich et al., 2023).

Although stress has been shown time and time again to prompt a switch to from deliberation to habitual behavior, it has yet to be shown that it does so by means of accentuating performance of the action sequences that may underly these habits. Other studies have examined stress in the context of maze learning performance (Packard et al., 2021), instrumental tasks (Dias-Ferreira et al., 2009; Schwabe & Wolf, 2009, 2010), and even two-stage tasks (Cremer et al., 2021; Otto, Gershman, et al., 2013; Otto, Raio, et al., 2013; Radenbach et al., 2015; Raio et al., 2020), but none have examined the extent to which stress affects action sequence behavior.

These previous studies conducted on stress in a two-stage task have evaluated its role in the tradeoff between model-based reinforcement learning (RL), which selects actions according to expected values estimated from a learned model of the environment, and model-free RL, which selects actions based on stored values acquired through previous experience (Dolan & Dayan, 2013). These systems have been proposed as analogs to goal-direction and habit respectively (Daw et al., 2005), but as noted by others in the field (Balleine & Dezfouli, 2019; Dezfouli et al., 2014; Miller et al., 2019), the notion of habit as model-free control fails to capture several key aspects of habitual behavior. Most notably, model-free RL cannot account for insensitivity to changes in action-outcome contingency when reward is omitted in contingency degradation, as it predicts that the error signal induced by the omission of reward would result in rapid policy updates and reduced responding (Dezfouli & Balleine, 2012). These pitfalls are avoided by instead conceptualizing habits as sequences of actions proceeding under open-loop control, subsumed under the governance of a higher-level model-based planner.

The goal of this experiment was to directly test the extent to which stress facilitates the usage of habitual action sequences during serial decision-making. To accomplish this, we trained a cohort of 16 female Long Evans rats in a two-stage decision-making task adopted from Dezfouli & Balleine (2019), with a within-subjects evaluation of the effects of 60 minutes of restraint stress on probe test performance. Each rat underwent one probe test following restraint stress, and one test in its absence. We hypothesized that following the stressor, rats would exhibit goal-directed control of habitual action sequences during the task, while under non-stressed conditions they would use a mixture of discrete actions and habitual action sequences.

3.2. Materials and methods

3.2.1. Subjects

16 Female Long Evans rats (Charles River, Quebec; 75–90 days old at the time of arrival) were used in this experiment. This sample size is consistent with other studies using two-stage decision-making task in rodents (Dezfouli & Balleine, 2019; Miller et al., 2017, 2022). Rats were pair-housed with unlimited access to water in a climate-controlled colony room maintained at 23°C with a 12-hour light/dark cycle (7:00 – 19:00 light-on period), and were handled by an experimenter each day. Experimental sessions were conducted each day at approximately 08:30. Rats were maintained at 85% of their free-feeding weights and given supplementary feeding approximately 4 hr after each session. All procedures have been approved by the Institutional Animal Care and Use Committee at the University of Vermont.

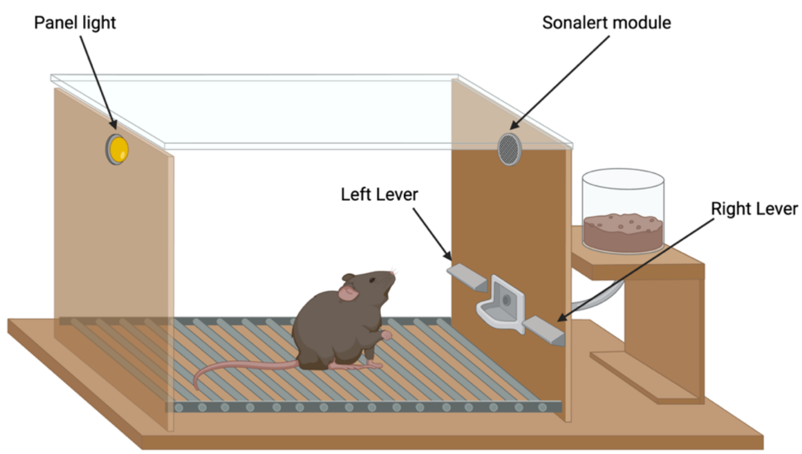
3.2.2. Apparatus

The instrumental training equipment consisted of 8 standard rat operant chambers (model ENV-008-VP; Med Associates, St. Albans, VT) housed within individual sound-attenuating cabinets. Operant chamber configurations are diagrammed in Figure 6. In the center of the right-facing chamber wall is a food cup (5 cm x 5 cm) into which a pellet dispenser (model ENV-203-45; Med Associates) can deliver 45-mg sucrose pellets (Bio-Serv). Entries to the food cup were recorded by an infrared beam. Two retractable lever manipulanda were positioned adjacent to the food cup, one to its right and one to its left

(Med Associates model: ENV-112CM). One panel light was installed on the left-facing chamber wall, centered, and directly below the chamber ceiling. Each chamber was illuminated by a single 150 lumen red LED bulb mounted to the roof of the sound-attenuating cabinet. Ventilation fans provided background noise. The auditory tone was delivered through a sonalert module mounted in the center of the right-facing chamber wall directly below the ceiling. Experimental events were controlled, and data was recorded by a computer located in an adjacent room.

The apparatus for acute restraint stress was a cylindrical restraining device 9 cm x 15 cm (D x H) placed in a brightly lit room upon the center of a countertop.

Figure 6. *Apparatus*



Note. Operant chamber configuration used throughout the experiment.

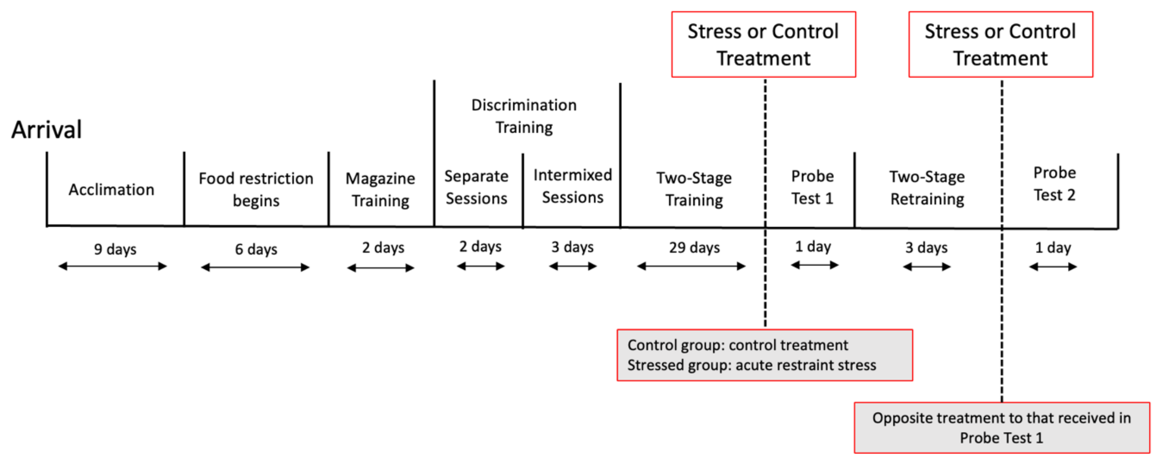
3.2.3. Procedure

The timeline of this experiment is diagrammed in Figure 7. Prior to the experiment, rats were assigned to an operant chamber in which they received all training and testing procedures. All sessions began two minutes after rats were loaded into their chambers. For all phases other than magazine training, session onset was signaled by lever insertion into the chamber, and session termination was signaled by lever retraction.

3.2.3.1. Magazine training

On the first two days, all rats received sessions in which one sucrose pellet was delivered into the food cup every 60 s, on average (variable-time 60 s schedule; VT 60 s), terminating after the delivery of 30 pellets. Pellet delivery was signaled by the sound of the pellet striking the food cup; no other stimuli were presented. Lever manipulanda were not extended during these sessions and the panel light was off.

Figure 7. *Experimental timeline*



Note. Experimental timeline for Experiment 3.

3.2.3.2. Discrimination training

On the next day, all rats received two lever training sessions in which the left lever or the right lever was available alone and lever-presses were reinforced on a free operant fixed-ratio 1 schedule (FR 1; i.e., every response is reinforced). Once 20 reinforcers were earned on this schedule, the session entered a discriminated operant training phase where lever presses were reinforced on an FR 1 schedule during presentations of one of two light stimuli (steady activation of the panel light, S1, or blinking activation of the panel light, S2). In this portion of the session the response was reinforced during the stimulus but not during the intertrial intervals (ITIs), which varied around an average 35-s length. Only the targeted lever was extended into the chamber in these sessions. Each trial terminated when the lever was pressed or 60 seconds had elapsed, thereby turning off the stimulus and entering the ITI. One session was administered for each stimulus (e.g., first session: S1: "R"+; second session: S2: "L"+) and these sessions were separated by 60 minutes. The pairings of stimuli and levers were counterbalanced across rats, as were the order of sessions. Sessions ended after 30 discriminated operant trials were complete or after 60 minutes, whichever occurred first.

The following day, all rats received two more sessions identical to those described above except that the sessions no longer began with a free operant phase; instead, consisting solely of the discriminated operant training component. Again, these sessions ended after 30 discriminated operant trials were complete or after 60 minutes, whichever occurred first.

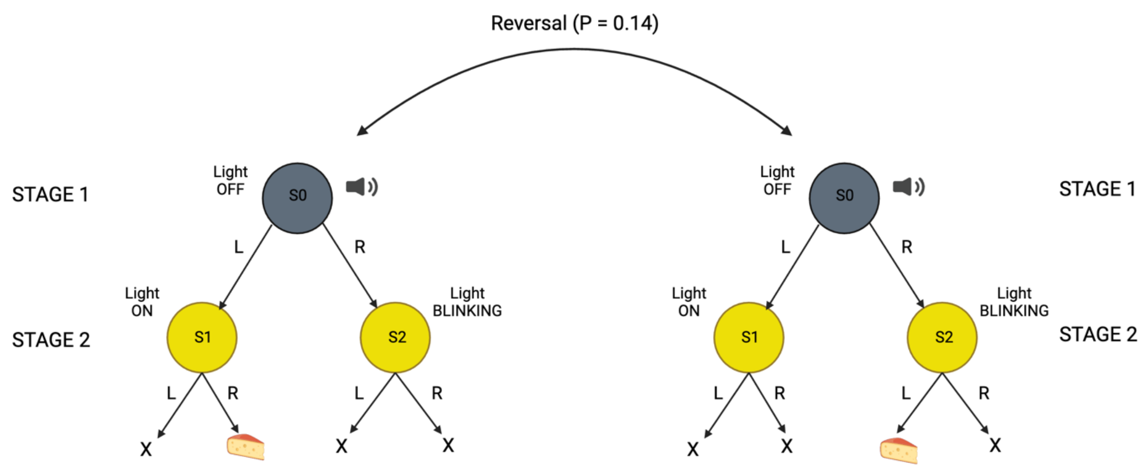
On the following three days, rats received three sessions containing intermixed trials (e.g., S1: "R"+ and S2: "L"+) with both levers continuously extended into the chamber. Each trial terminated (turning off the stimulus) when the correct lever was pressed, or 10-s had elapsed. Again, the response was reinforced during the stimulus but not during the variable 35-s ITI. One session was given per day, terminating after 30 reinforcers were earned with a maximum time limit of 60 minutes.

3.2.3.3. Two-stage task training

Next, rats were trained on the two-stage task diagrammed in Figure 8. The first stage of the task was signaled by an auditory tone, and the panel light was turned off. This configuration of stimuli was denoted as state 0 (S0). In this stage, rats could make a left lever press (“L”) or a right lever press (“R”) to terminate stage 1 of the task, immediately turning off the auditory tone and shifting the task to its second stage. Each stage 1 action (“L” or “R”) took the rat to a different second stage state (i.e., “L” goes to state 1 (S1); “R” goes to state 2 (S2)). The two stage 2 states – S1 and S2 – were signaled by steady activation of the panel light (S1) or blinking activation of the panel light at 5 Hz (S2). In these stage 2 states, to earn a sucrose pellet reward the rat would need to take the opposite action it took in took in stage 1. For instance, if the rat took “L” in stage 1, thereby taking it to S1 in stage 2, it must then take “R” to earn a sucrose pellet reward. If this same rat were to take “R” in stage 1, thereby taking it to S2 in stage 2, it would then need to take “L” to earn a reward. After an action is made in stage 2 the trial was terminated and a new trial was initiated, beginning again from stage 1. All animals underwent one session per day, terminating after 60 reinforcers were earned or 60 minutes was reached, whichever occurred first. Mapping of stage 1 action transitions to stage 2 states was counterbalanced across rats.

On any given trial only one stage 2 state delivered reward if the correct stage 2 action was taken. This reward-possible state reversed with a probability of 0.14 after four reinforcers were earned. Put another way, only one sequence of actions (i.e., “S0-L” to “S1-R”, or “S0-R” to “S2-L”) could lead to reward on a trial. Because the reward-possible state was likely to remain the same from trial to trial, optimal behavior entailed repeating the same stage 1 action on the next trial to transition to the previously rewarded state. However, if no reward was earned in the previous trial, the optimal choice was to switch stage 1 actions in order to transition to the other stage 2 state, as it was now likely that this state was reward-possible. Therefore, to maintain an optimal rate of reinforcement in this task rats must remain sensitive to these “reversals” and adjust their stage 1 choice by “staying” on the same stage 1 action or “switching” to the other in light of these events in a goal-directed manner.

Figure 8. *Two-stage task*



Note. Two-stage task design. “L” and “R” indicate left lever press and right lever press actions. “X” indicates no reward received; cheese symbol indicates reward. “S0”, “S1”, and “S2” indicate state 0, state 1, and state 2. Text referring to light (e.g., “Light OFF”) describes the status of the panel light located in the chamber. The sound symbol indicates that the auditory tone is activated. “Reversal (P = 0.14)” describes the property that on any given trial, only one stage 2 state (i.e., S1 or S2) could be rewarded, and this reward-possible state switches with a probability of 0.14 following four reinforced trials.

We expected that it would take multiple training sessions before rats acquired the correct state-space of the task, and would likely first act according to a simplified state-space representation where S1 and S2 had not yet been encoded as outcomes of actions taken in S0 (as seen in Dezfouli & Balleine, 2019). This would be demonstrated by a tendency to repeat the action most proximal to reinforcement (i.e., the action taken in stage 2) at the commencement of the next trial – thereby switching stage 1 actions from the last trial. In contrast, optimal behavior could be guided by a correct state-space representation where S0, S1, and S2 are encoded as different states signaled by their environmental cues. Using this state-space representation the rat could select actions according to state properties and take the stage 1 action that led to the reward-possible stage 2 state on the previous trial. Therefore, if rats learn to stay on the previously rewarded stage 1 action, this could suggest that their behavior is guided by the complete state-space of the task and that they are actively tracking prior reinforcement trial-to-trial to guide goal-directed action selection.

Training occurred in three blocks across which task parameters were successively adjusted to facilitate acquisition. Acquisition was measured in terms of the log odds of staying on the same stage 1 action following reward (log odds > 0 indicates greater likelihood of staying on the same stage 1 action). In Block 1 (sessions 1-7), the intertrial interval was 1s and the probability of a reversal was set to 0.50. In Block 2 (sessions 8-11), the intertrial interval was increased to 3s and the probability of reversal was again set at 0.50. Finally, in block 3 (sessions 12-29), the intertrial interval was eliminated and the probability of reversal was set to 0.14 following four reinforced trials. This final stipulation meant that the rats necessarily gained experience with both lever press sequences because a reversal could not occur until they had successfully completed the correct sequence and were reinforced for doing so. These training sessions continued until the mean log odds for staying on the same stage 1 action after earning reward was greater than zero for at least three training sessions.

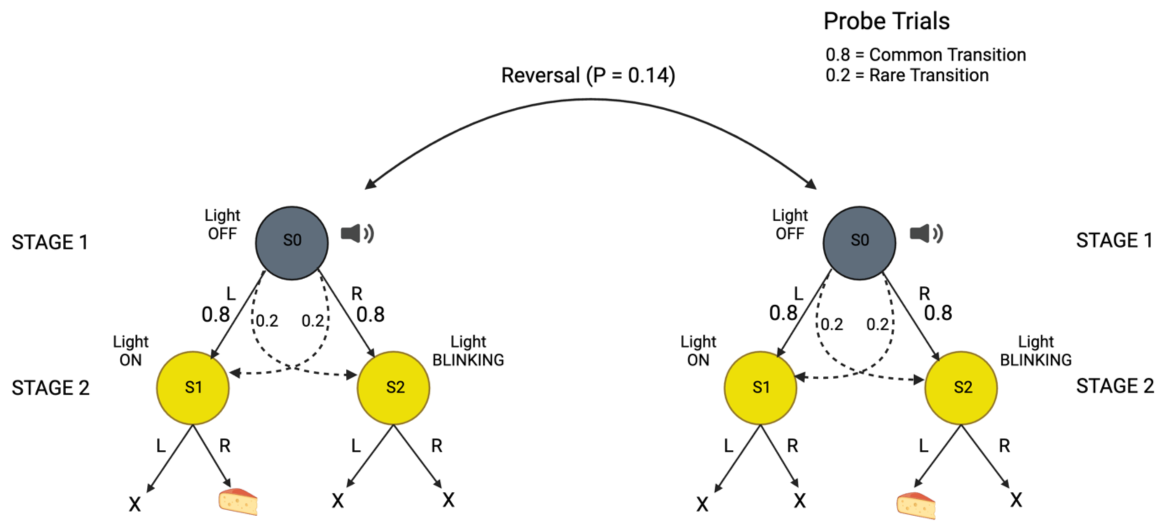
3.2.3.4. Probe tests and acute stress

An additional possibility was that rats could approximate optimal stay/switch behavior by learning to chunk together actions that earned reward into a sequence to be executed as a functional unit. For example, if “L” in stage 1 and “R” in stage 2 were rewarded, these actions could be concatenated into “L->R” to achieve the same effect. This sequence of actions can therefore be selected based on the computed value of the sequence – effectively bypassing the need to evaluate constituent actions. The use of such a strategy would not be apparent from behavior in the two-stage task alone, apart from decreased latency between stage 1 and stage 2 choices due to procedural automaticity (Dezfouli & Balleine, 2019). Thus, probe tests were required to parse the usage of action sequences and deliberate choice at stage 2 of the task.

Because action sequences are executed as a unit, they are insensitive to changes in action-outcome relationships within the sequence. This makes action sequences habitual in the sense that their constituent responses are automatically elicited by the immediately preceding behavior without evaluation of their consequences. The probe test shifted S0 to S1 and S2 transitions from being deterministic (e.g., S0-“L”->S1 with $P = 1.00$) to probabilistic (e.g., S0-“L”->S1 with $P = 0.80$; S0-“L”->S2 with $P = 0.20$) to expose actions taken in stage 2 that were inconsistent with state cues in this stage (diagrammed in Figure 9). Common transitions, which were the same stage 1 to stage 2 state transitions as those of two-stage training, occurred with a probability of 0.80. In rare transitions, the stage 1 action that led to S1 or S2 in training instead lead to the opposite state, occurring with a probability of 0.20. If step-by-step deliberation was engaged during these rare trials, the rat should have adjusted its choice in stage 2 according to the state discriminative stimulus (the light cue) and took the correct stage 2 action. Otherwise, if choices were guided by a habitual action sequence rats would fail to adjust their stage 2 action on rare trials because the second action in the sequence was not individually evaluated. Likewise, deliberation would be suggested by longer reaction times in stage 2 while evaluation takes place, while action sequences would be performed more quickly. The action-outcome mappings for stage 2 states do not change, meaning that if a rat earned reward for "R" in S1 and for "L" in S2 in training this remains true in probe tests as well. These contingencies were always deterministic.

Following the two-stage training phase all rats received two probe tests, one following 60 minutes of restraint stress and the other under control (non-stressed) conditions similar to those of training. The restraint stress took place immediately before the test session in an adjacent brightly lit laboratory room, to which the rats had no prior exposure. For the control treatment, rats were handled for a similar duration but left in their home cages in their colony room. All rats received both treatments and the ordering of stressed and non-stressed test sessions was counterbalanced, with cage mates always receiving the same treatment. These probe test sessions were identical to the two-stage training sessions in all facets other than the newly imposed state transition structure described above. Probe tests were separated by three sessions of two-stage retraining (same methods as prior two-stage training), one session per day, to wash out effects of exposure to the first test session.

Figure 9. *Probe test*



Note. Probe test structure. “L” and “R” indicate left lever press and right lever press actions. “X” indicates no reward received; cheese symbol indicates reward. “S0”, “S1”, and “S2” indicate state 0, state 1, and state 2. Text referring to light (e.g., “Light OFF”) describes the status of the panel light located in the chamber. The sound symbol indicates that the auditory tone is activated. “Reversal (P = 0.14)” describes the property that on any given trial, only one stage 2 state (i.e., S1 or S2) could be rewarded, and this reward-possible state switches with a probability of 0.14 following four reinforced trials. “0.8” and “0.2” indicate transition probabilities for stage 1 (i.e., S0) to stage 2 (i.e., S1 or S2) transition.

3.2.5. Statistical analyses

3.2.5.1. Two-stage task training

Stage 1 action data from two-stage task training was analyzed using a generalized linear mixed model (GLMM) analysis with logistic regression. Stage 1 actions (stay [1] or switch [0]) from each training block were regressed on reward earned on the previous trial (reward [1] or no reward [0]), session, and a reward X session interaction term. Intercept was included as a random effect varying across rats. Stage 2 reaction times (RTs) were regressed on session in a GLMM analysis, again including intercept as a random effect varying across rats.

3.2.5.2. Probe tests

Trial reward rates were calculated as rewards earned divided by the total number of trials and were compared across stressed and non-stressed conditions in a paired samples *t*-test. Trial attempt rates were calculated as trials attempted per minute and were again compared across stress conditions in a paired samples *t*-test.

Stage 1 action data from the probe tests were analyzed in a GLMM with logistic regression. Stage 1 actions were regressed on reward earned on the previous trial, stress (stressed [1] or non-stressed [0]), and a reward X stress interaction term. GLMM with logistic regression was also used for the analysis of stage 2 actions during the probe tests. Stage 1 actions, reward earned on the previous trial, stress, and interaction terms for stage 1 action X reward and stage 1 action X reward X stress were entered as predictors, and the outcome variable was repetition of the same stage 2 action as on the previous trial (stay or switch). Only those trials in which the stage 2 state differed from that of the previous trial were entered into this analysis. For both GLMM analyses of stage 1 and stage 2 action data, intercept and slopes for each predictor were included as random effects varying across rats.

Discrimination accuracy during stage 2 of the task was analyzed for common and rare trials separately by comparing the percentage of correct discriminations across stressed and non-stressed conditions in a paired samples *t*-test. Median stage 1 and stage 2 RTs per rat were also compared across stressed and non-stressed conditions in paired samples *t*-tests. The relationship between stage 2 RTs and discrimination accuracy was assessed for all trials and rare trials specifically using two identical GLMMs with logistic regression. Discrimination (correct [1] or incorrect [0]) was regressed on stage 2 RT, stress, and a stress X stage 2 RT interaction term, allowing intercept to vary randomly across rats. Follow-up GLMM analyses were conducted for each stress condition including random effects of intercept and RT.

The relationship between double-lever presses (i.e., pressing the same lever in stage 2 as was just pressed during stage 1) and discrimination accuracy was assessed for all trials using a GLMM with logistic regression. Double-presses (double-press [1] or non-double-press [0]) were regressed on stage 2 RT, stress, and a stress X stage 2 RT interaction term, including random intercepts and slopes for each predictor varying across rats.

3.2.5.3. Two-stage task retraining

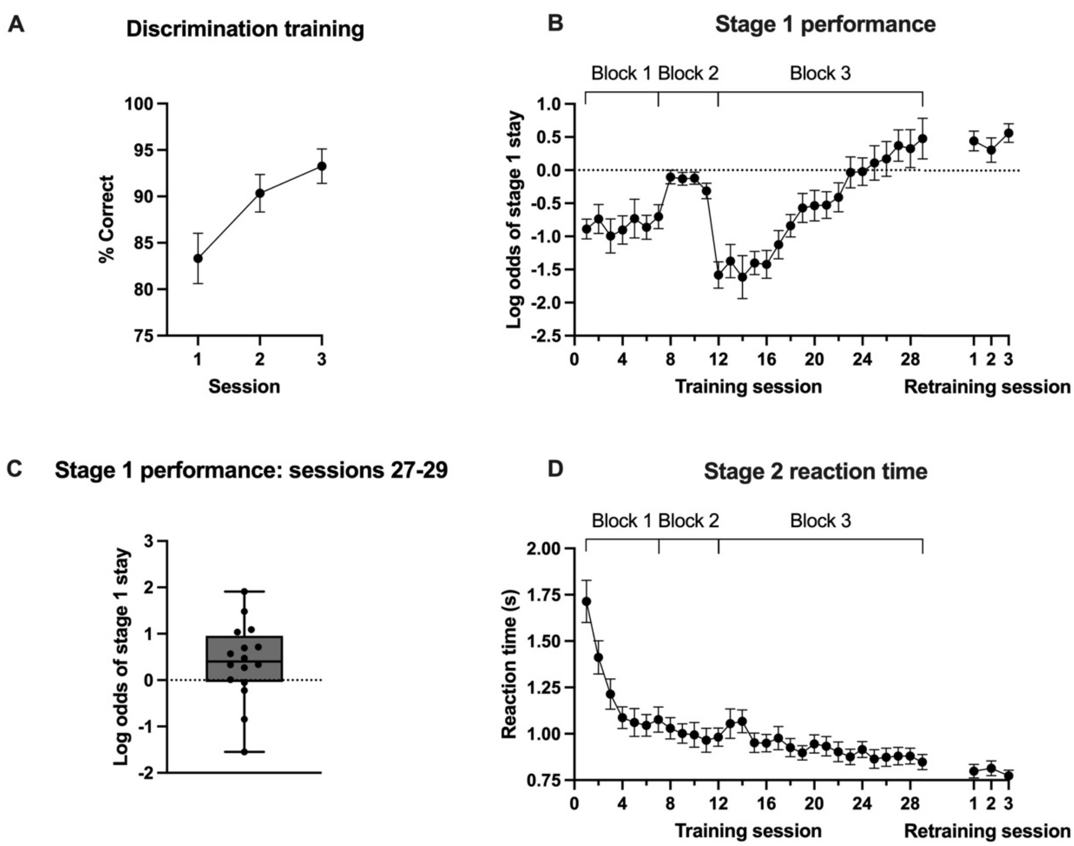
Stage 1 action data from two-stage task retraining was analyzed in a GLMM with logistic regression. Stage 1 actions were regressed on reward earned on the previous trial. Intercept and reward were included as random effects varying across rats.

3.2.5.4. General

Statistical analyses were conducted in SPSS (IBM, version 29) and Excel (Microsoft Corporation, 2024). For all GLMMs, tests of fixed effects and coefficients were performed with robust estimation of the parameter estimates covariance matrices. RT data was logarithm (base 10) transformed prior to all GLMM analyses to correct for violations of normality. 95% confidence intervals are reported. Alpha was set at $p < 0.05$ for all tests.

3.3. Results

Figure 10. *Training results*



Note. A) Mean discrimination accuracy (percent correct) from intermixed discrimination training. B) Mean logarithm of the odds ratio (log odds) of staying on the same stage 1 action after being rewarded on the preceding trial. Log odds > 0 indicates greater likelihood of staying than switching after earning reward. Log odds = 0 indicates equal preference for both actions. C) Log odds of staying on the same stage 1 action following reward averaged across the final three sessions (27-29) of training. Dots indicate individual data points. D) Reaction time in stage 2 of the task. Block 1: ITI = 1s, P(reversal) = 0.50; Block 2: ITI = 3s, P(reversal) = 0.50; Block 3: ITI = 0s, P(reversal) = 0.14 following 4 O's. Error bars indicate standard error of the mean (SEM).

3.3.1. Intermixed discrimination training

Results of discrimination training are shown in Figure 10A. Rats increased discrimination accuracy across sessions and mean accuracy reached levels greater than 90% by the final day of discrimination training (Day 3: *M* = 93.27%, *SD* = 7.44).

3.3.2. Two-stage task training

The results of two-stage task training are shown in Figure 10. In Block 1, reward earned on previous trials was a significant negative predictor of stage 1 action (β = -0.85, t[17,967] = -4.84, p < .001, CI [-1.19, -0.50]), indicating a greater likelihood of switching stage 1 actions following reward. Session was a significant positive predictor of staying on the same stage 1 action (β = 0.07, t[17,967] = 4.98, p < .001, CI [0.05, 0.10]), meaning that the tendency of rats to repeat the same stage 1 action trial-by-trial increased by session. The reward X session interaction, however, was not significant (β = 0.01, p = 0.64), suggesting that rats were not becoming increasingly likely to repeat stage 1 actions following reward across sessions.

In Block 2, after the ITI was increased from 1s to 3s, the effect of reward on staying on the same stage 1 action increased, however it was not significant (β = 0.43, p = 0.34). Effects of session and the reward X session interaction were also nonsignificant (respectively: β = 0.04, p = 0.13; β = -0.06, p = 0.18). This change in task parameters and the resulting behavioral effects suggested that during the ITI, rats were repeating the just-performed sequence of actions – thereby causing a mistiming of their sequence of responses with the commencement of the next trial. In effect, this created a mirage of positive change in the coefficient for reward, but the failure to increase the effect of reward across sessions indicated that rats were not becoming increasingly likely to repeat stage 1 actions following reward. Instead, they may have been randomly allocating responses to opposing sequences of lever presses (i.e., “L->R” and “R->L”) – effectively matching responses with their relative rates of reinforcement (Herrnstein, 1961). Under the condition where reversals occurred with a 0.50 probability (as in Blocks 1 and 2), this set up a situation in which each sequence was reinforced on approximately half of its attempts. This behavior was identified during Block 2, and the program was revised to update the reversal probability from 0.50 to 0.14.

In Block 3, the ITI was reduced to 0s to prevent any responses from being made between trials and the probability of reversal was corrected to 0.14. Here, the effect of previous reward was significant (β = -3.21, t[49,386] = -11.05, p < .001, CI [-3.78, -2.64]) but negative, reflecting the fact that for the majority of training the mean log odds of staying on the same stage 1 action remained below zero. The effect of session was also significant and negative (β = -.04, t[49,386] = -4.36, p < .001, CI [-0.06, -0.02]), indicating that rats became less perseverative as training continued and they increasingly attempted both stage 1 responses. Importantly, the reward X session interaction was significant and positive (β = 0.13, t[49,386] = 7.23, p < .001, CI [0.09, 0.16]), meaning that the probability of repeating the same stage 1 action following reward on the previous trial increased across sessions. This trend is shown clearly in Figure 10B. Ultimately, 12/16 (75%) rats finished training with log odds ratios greater than zero when averaged across the final three training sessions (Mdn = 0.40, IQR = -0.04, 0.95; See Figure 10C).

This heterogeneity in training endpoints may be due to a failure of some rats to represent the complete state-space of the task, instead treating the task as if composed of a single state and consistently repeating the action most proximal to reward. This could be indicated by the most negative log odds ratios (Rat 5: $M(\beta)$ = -1.55, Rat 11: $M(\beta)$ = -0.84), which suggest a tendency to switch stage 1 actions from the preceding trial and repeat the action just rewarded in stage 2. Alternatively, log odds ratios closer to zero (Rat 16: $M(\beta)$ = -0.22, Rat 10: $M(\beta)$ = -0.05, Rat 14: $M(\beta)$ = 0.01) could suggest a failure to track reinforcement history during the task and use it to guide goal-directed selection of action sequences, as these rats were likely selecting both sequences of actions at similar rates irrespective of preceding reinforcement.

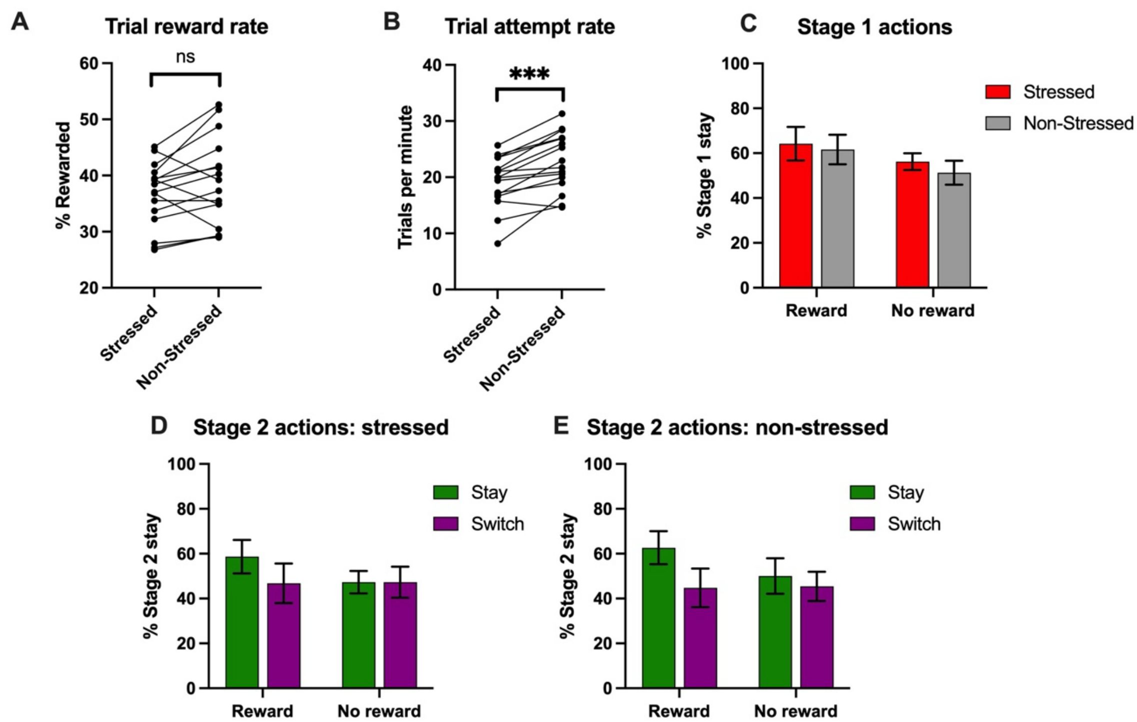
Stage 2 RTs decreased across training sessions, reflecting increased automaticity in the task. These results are shown in Figure 10D. Results indicated that across all blocks, session was significantly associated with reduced RTs (β = -2.67, t[76,345] = -3.18, p < .001, CI [-4.41, -1.03]). The variance in random intercept was also significant (Z = 2.74, p = .01), suggesting that rats exhibited substantial variety in RTs.

3.3.3. Probe tests

Results of the probe tests are shown in Figures 11 and 12. Rats achieved similar levels of reward in the task across stressed and non-stressed conditions. Reward rates were slightly lower when stressed (M = 36.60%, SD = 5.74) compared to non-stressed (M = 38.81%, SD = 7.76); although this difference was not significant (t[15] = -1.91, p = 0.08; see Figure 11A). When stressed, rats also attempted significantly less trials per minute (M = 19.16, SD = 4.57) than under non-stressed control conditions (M = 22.80, SD = 5.08; t[15] = -5.63, p < .001; see Figure 11B), possibly reflecting a stress-induced motivational deficit or greater time taken to perform actions within trials. These possibilities are discussed in further depth shortly.

Stage 1 action data is shown in Figure 11C. Across stressed and non-stressed conditions, reward earned on the preceding trial was a significant positive predictor of staying on the same stage 1 action (β = 0.42, t[5,253] = 2.58, p = .01, CI [0.10, 0.73]). This indicates that actions taken in stage 1 of the task during the probe tests were likely guided by online tracking of trial-to-trial reinforcement in a goal-directed manner. Stress was also significantly associated with repetition of stage 1 actions (β = 0.19, t[5,253] = 2.81, p = .005, CI [0.06, 0.32]), indicating that independent of previous reward, rats became marginally more perseverative when stressed – repeating the same stage 1 action more often. The reward X stress interaction was not significant (β = -.07, p = 0.62), suggesting that reward earned on preceding trials guided behavior to a similar extent in stress and in its absence. The variance in the random effect of reward was significant (Z =

Figure 11. *Probe test performance*



Note. A) Trial reward rate expressed as percent of trials rewarded. B) Trial attempt rate expressed as trials attempted per minute. C) Percentage of trials where the stage 1 action from the preceding trial was repeated as a function of whether the preceding trial was rewarded (reward/no reward), and whether the test occurred under stressed or non-stressed conditions. D) (stressed) and E) (non-stressed): Percentage of trials where the stage 2 action from the preceding trial was repeated as function of whether the preceding trial was rewarded (reward/no reward) and whether the stage 1 action from the preceding trial was repeated (stay/switch). Dots represent individual data points. ns = nonsignificant; *** = p < .001. Error bars indicate 95% confidence intervals.

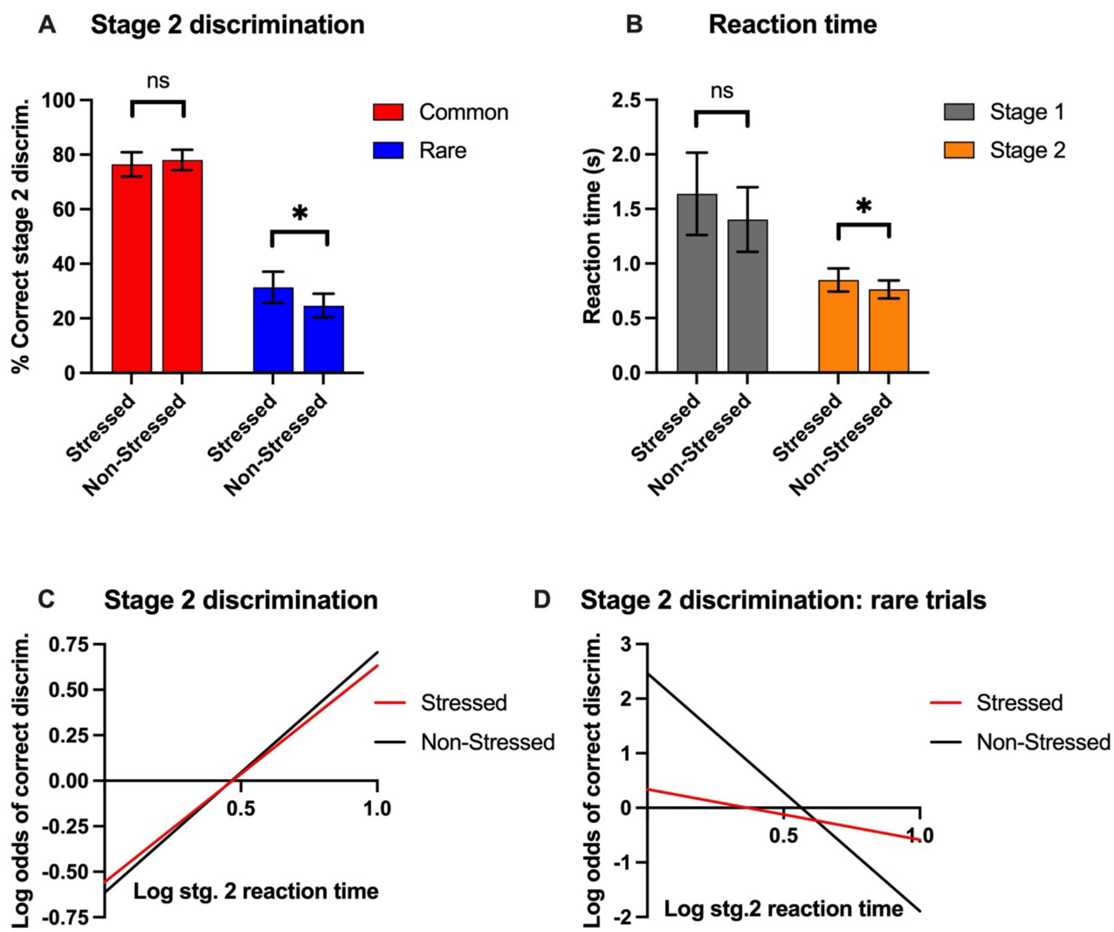
3.00, $p = .003$), in-line with the heterogeneity observed in training endpoints.

Stage 2 action data is shown in Figures 11D and 11E. Effects of staying on the same stage 1 action, reward earned on the preceding trial, and stress were not significant (all t's < 1). However, the stage 1 action X reward interaction was significant ($\beta = 0.60$, $t[2,505] = 2.19$, $p = .03$, CI [0.06, 1.13]), meaning that across stress conditions, after earning reward on the preceding trial and staying on the same stage 1 action rats also tended to repeat stage 2 actions – thereby repeating the previously rewarded action sequence despite having entered a different stage 2 state. This interaction did not significantly differ between stressed and non-stressed conditions ($\beta = -.18$, $p = 0.29$), suggesting that rats used habitual actions sequences to a similar extent irrespective of stress.

Stage 2 discrimination accuracy data is shown in Figure 12A. Mean discrimination accuracies for common transition trials were similar under stressed (M = 76.44%, SD = 8.37) and non-stressed conditions (M = 78.04%, SD = 7.02). These accuracies did not significantly differ ($t[15] = -0.85$, $p = 0.41$). For rare trials, discrimination accuracies were slightly higher when stressed (M = 31.41%, SD = 10.79, CI [25.66, 37.16]) compared to non-stressed (M = 24.66%, SD = 8.21, CI [20.28, 29.03]), which was significant ($t[15] = 2.21$, $p = 0.043$). Confidence intervals for both stressed and non-stressed conditions on rare trials were entirely below 50%, indicating that rats consistently made incorrect discriminations at an above-chance level.

Stage 1 and 2 RT data are shown in Figure 12B. Although, mean stage 1 RTs were longer under stressed conditions (M = 1.64s, SD = 0.71) than non-stressed (M = 1.40s, SD = 0.56), the difference was not significant ($t[15] = 1.77$, $p = 0.10$). Stage 2 RTs were significantly longer when stressed (M = 0.85s, SD = 0.20) compared to non-stressed (M = 0.76s, SD = 0.16; $t[15] = 2.39$, $p = 0.03$), suggesting a weaker coupling of stage 1 and stage 2 responses. Longer RTs following stress may have also contributed to the decrease in trials attempted per minute, as trials took more time to complete on average.

Figure 12. *Probe test discrimination accuracies and reaction times*



Note. A) Stage 2 discrimination accuracy (percent correct) during stressed and non-stressed probe tests as a function of whether the trial was common or rare. B) RTs (s) during stressed and non-stressed probe tests for stage 1 and stage 2 of the task. C) Log odds of making a correct stage 2 discrimination as a function of log transformed stage 2 RT. D) Log odds of making a correct stage 2 discrimination as a function of log transformed stage 2 RT for rare trials only. Lines depict stress condition. ns = not significant, * = $p < .05$. Error bars indicate 95% confidence intervals.

Under both stressed and non-stressed conditions, greater stage 2 RT was associated with higher likelihood of making a correct stage 2 discrimination ($\beta = 0.66$, $t[5,253] = 2.64$, $p = .01$, CI [0.17, 1.15]). Effects of stress and the stress X stage 2 RT interaction were not significant (respectively: $\beta = -.02$, $p = 0.97$; $\beta = -.03$, $p = 0.92$), suggesting that overall discrimination accuracy and the effect of RT on discrimination accuracy did not differ between stress conditions. Regression lines for each stress condition are shown in Figure 12C.

The positive association between stage 2 RT and likelihood of making a correct discrimination was likely due to the fact that for trials in which rats immediately pressed the same lever in stage 2 as they had just pressed in stage 1 (i.e., a double-press), RTs were significantly shorter ($\beta = -0.92$, $t[5,253] = -1.96$, $p = .05$, CI [-1.847, -0.001]). Effects of stress ($\beta = 0.16$, $p = 0.86$) and stress X stage 2 RT ($\beta = -0.01$, $p = 0.98$) were nonsignificant, indicating that the amount of double-presses did not differ between stress conditions, nor did the association between RT and double-presses. Together, this suggests that the relationship between increased stage 2 RT and discrimination accuracy is likely not due to increased deliberation. Rather, it is likely due a decrease in erroneous double lever presses with increasing RT, as double-presses terminated the trial faster than crossing the chamber to press the opposite lever during a correct action sequence.

This relationship was reversed for rare transition trials, where across stress conditions greater stage 2 RTs were associated with a lower likelihood of making a correct stage 2 discrimination ($\beta = -1.84$, $t[1,076] = -2.19$, $p = .03$, CI [-3.49, -0.19]). Effects of stress and stress X stage 2 RT were nonsignificant (respectively: $\beta = -2.07$, $p = 0.16$; $\beta = 1.28$, $p = 0.12$). Separate GLMMs for each stress group indicated that when stressed, the negative association between stage 2 RT and discrimination accuracy was not significant (β = -0.59, p = 0.26) compared to a much stronger significant negative association under non-stressed conditions (β = -1.89, t[502] = -2.22, p = .03, CI [-3.57, -0.22]). These regression lines are shown in Figure 12D. Here, the finding that shorter RTs were associated with greater likelihood of a correct discrimination could be due to the fact that on rare trials, double-presses (which were erroneous on common trials) were more often the correct response to make and tend to take less time than opposing responses.

3.3.4. Two-stage task retraining

Two-stage task retraining data is shown in Figure 10. Results of the GLMM analysis indicated that following reward earned on the previous trial, rats were significantly more likely to stay on the same stage 1 action (β = 0.41, t[6,060] = 3.02, p = .003, CI [0.14, 0.67]). This supports the idea that most rats had successfully acquired the task and were selecting stage 1 actions based on the tracking of prior reinforcement. However, there was still significant variance in the random effect of reward (Z = 2.33, p = 0.02), reproducing the heterogeneity in performance observed during training and testing.

3.4. Discussion

These findings suggest that under both stressed and non-stressed conditions, female rats used goal-directed planning over habitual action sequences to solve the two-stage decision-making task. In-line with our hypothesis about behavior under stress, we found that action planning was guided by online tracking of trial-to-trial reinforcement and that stage 1 and stage 2 responses were executed as a functional unit without evaluation at intermediate steps. Contrary to our hypothesis about behavior under non-stressed conditions, rats did not deliberate at the level of individual actions during the test and instead executed responses again as a habitual sequence of actions. These results suggest that for responses executed in a habitual sequence, acute stress does not occlude goal-directed control over the sequence, nor does it prompt the deconstruction of the sequence into closed-loop control.

Goal-directed planning of action sequences was evidenced by the sensitivity of stage 1 actions in the probe tests to reversals in action-outcome contingency on preceding trials. This sensitivity is supported by the significant effect of reward earned on preceding trials on repeating the same stage 1 action, indicating that following reward rats were more likely to repeat stage 1 actions, and likewise switch stage 1 actions when reward was omitted. Such behavior requires representing task stages as distinct events with opposing appropriate responses, as well as monitoring of trial-to-trial reinforcement to guide choice to stay on the same stage 1 response only when rewarded on the preceding trial.

Habitual execution of action sequences was reflected by the insensitivity of stage 2 actions to rare transitions during the probe tests. This insensitivity is supported by two main findings: 1) the significant association between repeating the stage 1 action made on the preceding trial after earning reward and following it with the same stage 2 action, despite encountering a different stage 2 state. This suggests that if rats earned reward on the preceding trial and repeated the same stage 1 action, they were likely to then repeat the same stage 2 action – reiterating the just-performed action sequence despite entering a different stage 2 state. And 2) stage 2 discrimination accuracies were markedly lower for rare trials compared to common trials, indicating a persistent failure to withhold the second action in the sequence and adjust responding according to the discriminative stimulus of the different state. These findings were consistent across stressed and non-stressed tests, suggesting that at the strategic level rats implemented similar modes of behavior in the task.

However, stress did have several effects on the nature of goal-directed planning of action sequences and their execution. First, holding constant the effects of previous reward, stress caused a marginal increase in the tendency of rats to repeat stage 1 actions. Although stage 1 actions were guided by active monitoring of reward, when stressed, rats exhibited a modest preference towards reenacting the just-performed sequence of responses. Thus, acute stress may bias goal-directed planning in this task towards more inflexible persistence on a course of actions, or framed differently, an aversion from shifting to alternative sequences of actions. This inflexibility may have contributed to the subtly lower reward rates observed during the stressed test.

At the level of action sequence execution, rats showed increased latency between responses within the sequence and made more correct discriminations on rare trials when stressed. However, because rats consistently made the incorrect discrimination on rare trials at a level greater than chance, this suggests that stress likely did not induce a shift to deliberate choice. More likely, it increased errors in action sequence execution as the timing of stage 1 and stage 2 actions became less tightly coupled. Taken together, these findings indicate that acute stress does not alter the overall macrostructure of planning in two-stage decision tasks; instead, it seems to affect the microstructure of goal-directed planning and habitual action sequence execution themselves.

The lack of deliberation at stage 2 of the task observed under non-stressed conditions could be attributable to a variety of factors. Although our task structure and training methodology was similar to Dezfouli & Balleine (2019), who found evidence of step-by-step deliberation in male rats, the present experiment differed in several notable ways. First, although all rats successfully acquired the stage 2 discrimination during intermixed discrimination training, the training was structured such that rats had the opportunity to attempt multiple responses within a 10s window if their first response was incorrect. This could have encouraged lever-switching behavior which may have generalized to two-stage task training, thus encouraging the formation of action sequences composed of opposing lever press responses. Indeed, research suggests that action sequences learned in one stage of a task can transfer to subsequent task stages and serve as the building blocks for future behaviors (Ferguson et al., 2024).

Secondly, parameters imposed during training Blocks 1 and 2 differed from the 0.14 reversal probability used throughout Dezfouli & Balleine's (2019) main experiment. In these blocks, the 0.50 reversal probability for the reward-possible stage 2 state allowed rats to experience reinforcement for actions taken in both possible stage 2 states, but also provided a situation in which each sequence of opposing lever presses was reinforced at an equivalent rate. This lack of predictability did not incentivize active monitoring of trial-to-trial reinforcement or deliberate choice, as emitting either lever press sequence was associated with equal probability of reinforcement. As a result, there was no impetus to acquire a correct representation of the task state-space in these blocks. There was however an incentive to acquire opposing lever press response sequences ("L->R" and "R->L"), as performance of these action sequences was the most valid means of earning reinforcement. Because the state transition structure remained the same in Block 3 as in Blocks 1 and 2, these sequences were again reinforced and could be carried over as the fundamental units of action to be selected by a goal-directed planner. This explanation is supported by the goal-directed planning of habitual action sequences revealed during the probe tests. Thus, future studies that confirm first attempt success rates during discrimination training and utilize a reversal probability of 0.14 throughout training may be more likely to find evidence of discrete actions and closed-loop control in probe tests.

The sex of the rats may have also played a role in the usage of habitual action sequences instead of step-by-step deliberation. Prior work by our laboratory has found that female rats shift from goal-directed to habitual control over instrumental responding with less training than males (Schoenberg et al., 2019); although, the extent to which this sex difference extends to instrumental sequences of actions has yet to be tested. Given the pronounced sex difference found for single instrumental responses, it seems quite plausible that females may also be more innately predisposed to concatenate responses into a functional action sequence. This too represents an important direction for future study.

These findings suggest several implications regarding the nature of goal-directed and habitual decision-making following stress. Like recent studies, our results challenge the notion that stress causes a general switch from goal-directed to habitual modes of behavior (Buabang et al., 2022; Smeets et al., 2023). The present findings indicate that rats can retain goal-directed planning following stress when the constraints of the task encourage it, as in when active monitoring of feedback is properly incentivized. It does appear however, that stress may cause a modest reduction in spontaneous exploration during the task as rats preferred to stick to one sequence of actions. Similar stress-induced biases in explore/exploit behavior have been shown previously in both rats and humans (Harms, 2017; Lenow et al., 2017; Matisz et al., 2021). As suggested by these authors, stressful events – which are inherently aversive – may promote negative appraisals of the environment that generalize to the decision-making setting. Because environmental appraisals strongly predict exploration of new actions and negative appraisals can lead to exploitative decision biases (Constantino & Daw, 2015), this may explain the increased tendency to repeat stage 1 actions following stress.

Further, although stress has been shown to enhance certain types of stereotyped behavior in other settings (Kalueff et al., 2016; MacLennan & Maier, 1983; Mosher et al., 2017; Ridley, 1994), acute stress does not appear to do so for the performance of instrumental action sequences. Stress instead resulted in increased latency between actions within a sequence and greater response variance on rare trials. These results may imply that for well-learned sequences of instrumental actions, acute stress marginally degrades action sequence integrity despite conserving their selection. Whether stress may promote a switch from deliberation over single instrumental actions to performing them in a chunked sequence remains an open question; conditions of the current experiment fostered the usage of chunked action sequences under both stressed and non-stressed conditions.

Considering widespread findings of acute stress enhancing habitual responding in free operant tasks (Braun & Hauber, 2013; Dias-Ferreira et al., 2009; Dougherty et al., 2024), it is notable that we did not find enhanced aspects of habitual behavior in the present experiment. One possible explanation for this disparity is the differing attentional demands determined by the predictability of each task (Pearce & Hall, 1980; Thrailkill et al., 2018). In the two-stage task, the reward associated with selecting an entire course of actions remains unpredictable throughout the task, requiring constant attention to maintain reinforcement. Although acute stress promotes heightened vigilance, which could cause more cautious responding as rats become alert to other sources of threat in the environment (Blanchard et al., 1990; van Marle et al., 2009), the dynamic nature of the tasks necessitates consistent monitoring and goal-directed control. On the other hand, the relationship between responses in stage 1 and 2 of the task is highly predictable and well-learned by the probe tests, thus requiring low attention. The predictability of this aspect of the task could afford rats the opportunity to allocate attention to monitoring the potentially threatening environment while experiencing relatively meager cost in terms of reinforcement. In this way, generalized vigilance could lead to decrements in response timing and disrupt action sequence execution.

Highly predictable relationships between response and reward are common to free operant tasks as well (Bouton, 2021), where the conditioning context serves as a cue indicating high probability of reinforcement. In these tasks, because responses always reliably predict the receipt of rewards, attentional demands remain low throughout. This allows the animal to maintain defensive monitoring of the environment at little cost to expected reinforcement by emitting responses at a high rate. In so doing, their sensitivity to outcome devaluation would be reduced as processing of the behavior becomes markedly lower than under non-stressed conditions, thus enhancing habitual control over the response. These differences suggest an important interaction between acute stress and task structure in determining the expression of goal-directed and habitual behavior.

Collectively, these findings indicate that rats can retain goal-directed planning of sequential actions following stress despite exhibiting several distinct changes in planning structure and action sequence execution. Rats became less exploratory in action selection and showed weaker coupling of actions within a sequence under stressed conditions. This knowledge could be relevant to understanding the relationship between stress and biases in goal-directed selection of courses of behavior, which is likely important as many of our most common behaviors are composed of a series of related actions which we select as a unitary set (Dezfouli & Balleine, 2012; Smith & Graybiel, 2016). Finding methods of attenuating these biases could help the development of interventions for disorders involving persistent stress-induced biases in decision-making.

References

Adams, C. D., & Dickinson, A. (1981). Instrumental Responding following Reinforcer Devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *33*(2b), 109–121. https://doi.org/10.1080/14640748108400816

Aldridge, J. W., Berridge, K. C., & Rosen, A. R. (2004). Basal ganglia neural mechanisms of natural movement sequences. *Canadian Journal of Physiology and Pharmacology*, *82*(8–9), 732–739. https://doi.org/10.1139/y04-061

Balleine, B. W., & Dezfouli, A. (2019). Hierarchical Action Control: Adaptive Collaboration Between Actions and Habits. *Frontiers in Psychology*, *10*, 2735. https://doi.org/10.3389/fpsyg.2019.02735

Banca, P., Ruiz, M. H., Gonzalez-Zalba, M. F., Biria, M., Marzuki, A. A., Piercy, T., Sule, A., Fineberg, N. A., & Robbins, T. W. (2023). Action-sequence learning, habits and automaticity in obsessive-compulsive disorder. *eLife*, *12*. https://doi.org/10.7554/eLife.87346.2

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., & Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, *437*(7062), Article 7062. https://doi.org/10.1038/nature04053

Blanchard, R. J., Blanchard, D. C., Rodgers, J., & Weiss, S. M. (1990). The characterization and modelling of antipredator defensive behavior. *Neuroscience & Biobehavioral Reviews*, *14*(4), 463–472. https://doi.org/10.1016/S0149-7634(05)80069-7

Bouton, M. E. (2021). Context, attention, and the switch between habit and goal-direction in behavior. *Learning & Behavior*, *49*(4), 349–362. https://doi.org/10.3758/s13420-021-00488-z

Braun, S., & Hauber, W. (2013). Acute stressor effects on goal-directed action in rats. *Learning & Memory*, *20*(12), 700–709. https://doi.org/10.1101/lm.032987.113

Buabang, E. K., Boddez, Y., Wolf, O. T., & Moors, A. (2022). The role of goal-directed and habitual processes in food consumption under stress after outcome devaluation with taste aversion. *Behavioral Neuroscience*. https://doi.org/10.1037/bne0000439

Constantino, S., & Daw, N. D. (2015). Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective & Behavioral Neuroscience*, *15*(4), 837–853. https://doi.org/10.3758/s13415-015-0350-y

Cremer, A., Kalbe, F., Gläscher, J., & Schwabe, L. (2021). Stress reduces both model-based and model-free neural computations during flexible learning. *NeuroImage*, *229*, 117747. https://doi.org/10.1016/j.neuroimage.2021.117747

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Dezfouli, A., & Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *The European Journal of Neuroscience*, *35*(7), 1036–1051. https://doi.org/10.1111/j.1460-9568.2012.08050.x

Dezfouli, A., & Balleine, B. W. (2013). Actions, Action Sequences and Habits: Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized. *PLOS Computational Biology*, *9*(12), e1003364. https://doi.org/10.1371/journal.pcbi.1003364

Dezfouli, A., & Balleine, B. W. (2019). Learning the structure of the world: The adaptive nature of state-space and action representations in multi-stage decision-making. *PLOS Computational Biology*, *15*(9), e1007334. https://doi.org/10.1371/journal.pcbi.1007334

Dezfouli, A., Lingawi, N. W., & Balleine, B. W. (2014). Habits as action sequences: Hierarchical action control and changes in outcome value. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130482. https://doi.org/10.1098/rstb.2013.0482

Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., Costa, R. M., & Sousa, N. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science (New York, N.Y.)*, *325*(5940), 621–625. https://doi.org/10.1126/science.1171203

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1–18. https://doi.org/10.3758/BF03199951

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Dougherty, R., Thrailkill, E. A., Mohammed, Z., VonDoepp, S., Hilton-Vanosdall, E., Charette, S., Van Horn, S., Quirk, A., Kraus, A., & Toufexis, D. J. (2024). Acute stress facilitates habitual behavior in female rats. *Physiology & Behavior*, *275*, 114456. https://doi.org/10.1016/j.physbeh.2024.114456

Ferguson, L. A., Matamales, M., Nolan, C., Balleine, B. W., & Bertran-Gonzalez, J. (2024). Adaptation of sequential action benefits from timing variability related to lateral basal ganglia circuitry. *iScience*, *27*(3), 109274. https://doi.org/10.1016/j.isci.2024.109274

Frölich, S., Esmeyer, M., Endrass, T., Smolka, M. N., & Kiebel, S. J. (2023). Interaction between habits as action sequences and goal-directed behavior under time pressure. *Frontiers in Neuroscience*, *16*. https://www.frontiersin.org/articles/10.3389/fnins.2022.996957

Gobet, F., & Simon, H. A. (1998). Expert Chess Memory: Revisiting the Chunking Hypothesis. *Memory*, *6*(3), 225–255. https://doi.org/10.1080/741942359

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, *70*(1–2), 119–136. https://doi.org/10.1006/nlme.1998.3843

Graybiel, A. M. (2008). Habits, Rituals, and the Evaluative Brain. *Annual Review of Neuroscience*, *31*(1), 359–387. https://doi.org/10.1146/annurev.neuro.29.051605.112851

Halbout, B., Marshall, A. T., Azimi, A., Liljeholm, M., Mahler, S. V., Wassum, K. M., & Ostlund, S. B. (2019). Mesolimbic dopamine projections mediate cue-motivated reward seeking but not reward retrieval in rats. *eLife*, *8*, e43551. https://doi.org/10.7554/eLife.43551

Harms, M. B. (2017). Stress and Exploitative Decision-Making. *Journal of Neuroscience*, *37*(42), 10035–10037. https://doi.org/10.1523/JNEUROSCI.2169-17.2017

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement,. *Journal of the Experimental Analysis of Behavior*, *4*(3), 267–272. https://doi.org/10.1901/jeab.1961.4-267

Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., Dayan, P., & Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3098–3103. https://doi.org/10.1073/pnas.1414219112

Jin, X., & Costa, R. M. (2010). Start/Stop Signals Emerge in Nigrostriatal Circuits during Sequence Learning. *Nature*, *466*(7305), 457–462. https://doi.org/10.1038/nature09263

Kalueff, A. V., Stewart, A. M., Song, C., Berridge, K. C., Graybiel, A. M., & Fentress, J. C. (2016). Neurobiology of rodent self-grooming and its value for translational neuroscience. *Nature Reviews. Neuroscience*, *17*(1), 45–59. https://doi.org/10.1038/nrn.2015.8

Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, *70*(6, Pt.1), 387–403. https://doi.org/10.1037/h0026739

Lenow, J. K., Constantino, S. M., Daw, N. D., & Phelps, E. A. (2017). Chronic and Acute Stress Promote Overexploitation in Serial Decision Making. *The Journal of*

Neuroscience, *37*(23), 5681–5689. https://doi.org/10.1523/JNEUROSCI.3618-16.2017

MacLennan, A. J., & Maier, S. F. (1983). Coping and the Stress-Induced Potentiation of Stimulant Stereotypy in the Rat. *Science*, *219*(4588), 1091–1093. https://doi.org/10.1126/science.6681679

Matisz, C. E., Badenhorst, C. A., & Gruber, A. J. (2021). Chronic unpredictable stress shifts rat behavior from exploration to exploitation. *Stress*, *24*(5), 635–644. https://doi.org/10.1080/10253890.2021.1947235

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*, *20*(9), Article 9. https://doi.org/10.1038/nn.4613

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2022). Value representations in the rodent orbitofrontal cortex drive learning, not choice. *eLife*, *11*, e64575. https://doi.org/10.7554/eLife.64575

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without Values. *Psychological Review*, *126*(2), 292–311. https://doi.org/10.1037/rev0000120

Mosher, L. J., Godar, S. C., Nelson, M., Fowler, S. C., Pinna, G., & Bortolato, M. (2017). Allopregnanolone mediates the exacerbation of Tourette-like responses by acute stress in mouse models. *Scientific Reports*, *7*(1), 3348. https://doi.org/10.1038/s41598-017-03649-1

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The Curse of Planning: Dissecting multiple reinforcement learning systems by taxing the central executive. *Psychological Science*, *24*(5), 10.1177/0956797612463080. https://doi.org/10.1177/0956797612463080

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. https://doi.org/10.1073/pnas.1312011110

Packard, M. G., Gadberry, T., & Goodman, J. (2021). Neural systems and the emotion-memory link. *Neurobiology of Learning and Memory*, *185*, 107503. https://doi.org/10.1016/j.nlm.2021.107503

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Pew, R. W. (1966). Acquisition of hierarchical control over the temporal organization of a skill. *Journal of Experimental Psychology*, *71*(5), 764–771. https://doi.org/10.1037/h0023100

Radenbach, C., Reiter, A. M. F., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.-J., Deserno, L., & Schlagenhauf, F. (2015). The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology*, *53*, 268–280. https://doi.org/10.1016/j.psyneuen.2014.12.017

Raio, C. M., Konova, A. B., & Otto, A. R. (2020). Trait impulsivity and acute stress interact to influence choice and decision speed during multi-stage decision-making. *Scientific Reports*, *10*(1), 7754. https://doi.org/10.1038/s41598-020-64540-0

Ridley, R. M. (1994). The psychology of perseverative and stereotyped behaviour. *Progress in Neurobiology*, *44*(2), 221–231. https://doi.org/10.1016/0301-0082(94)90039-6

Schoenberg, H. L., Sola, E. X., Seyller, E., Kelberman, M., & Toufexis, D. J. (2019). Female rats express habitual behavior earlier in operant training than males. *Behavioral Neuroscience*, *133*(1), 110–120. https://doi.org/10.1037/bne0000282

Schwabe, L., & Wolf, O. T. (2009). Stress Prompts Habit Behavior in Humans. *The Journal of Neuroscience*, *29*(22), 7191–7198. https://doi.org/10.1523/JNEUROSCI.0979-09.2009

Schwabe, L., & Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology*, *35*(7), 977–986. https://doi.org/10.1016/j.psyneuen.2009.12.010

Smeets, T., Ashton, S. M., Roelands, S. J. A. A., & Quaedflieg, C. W. E. M. (2023). Does stress consistently favor habits over goal-directed behaviors? Data from two preregistered exact replication studies. *Neurobiology of Stress*, *23*, 100528. https://doi.org/10.1016/j.ynstr.2023.100528

Smith, K. S., & Graybiel, A. M. (2013). A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics. *Neuron*, *79*(2), 361–374. https://doi.org/10.1016/j.neuron.2013.05.038

Smith, K. S., & Graybiel, A. M. (2016). Habit formation. *Dialogues in Clinical Neuroscience*, *18*(1), 33–43.

Thorn, C. A., Atallah, H., Howe, M., & Graybiel, A. M. (2010). Differential Dynamics of Activity Changes in Dorsolateral and Dorsomedial Striatal Loops During Learning. *Neuron*, *66*(5), 781–795. https://doi.org/10.1016/j.neuron.2010.04.036

Thrailkill, E. A., Trask, S., Vidal, P., Alcalá, J. A., & Bouton, M. E. (2018). Stimulus Control of Actions and Habits: A Role for Reinforcer Predictability and Attention in the Development of Habitual Behavior. *Journal of Experimental Psychology. Animal Learning and Cognition*, *44*(4), 370–384. https://doi.org/10.1037/xan0000188

van Elzelingen, W., Warnaar, P., Matos, J., Bastet, W., Jonkman, R., Smulders, D., Goedhoop, J., Denys, D., Arbab, T., & Willuhn, I. (2022). Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. *Current Biology*, *32*(5), 1163-1174.e6. https://doi.org/10.1016/j.cub.2021.12.027

van Marle, H. J. F., Hermans, E. J., Qin, S., & Fernández, G. (2009). From Specificity to Sensitivity: How Acute Stress Affects Amygdala Processing of Biologically Salient Stimuli. *Biological Psychiatry*, *66*(7), 649–655. https://doi.org/10.1016/j.biopsych.2009.05.014

Wirz, L., Bogdanov, M., & Schwabe, L. (2018). Habits under stress: Mechanistic insights across different types of learning. *Current Opinion in Behavioral Sciences*, *20*, 9–16. https://doi.org/10.1016/j.cobeha.2017.08.009

Comprehensive Bibliography

Adams, C. D. (1982). Variations in the Sensitivity of Instrumental Responding to Reinforcer Devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *34*(2b), 77–98. https://doi.org/10.1080/14640748208400878

Adams, C. D., & Dickinson, A. (1981). Instrumental Responding following Reinforcer Devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *33*(2b), 109–121. https://doi.org/10.1080/14640748108400816

Adams, T. G., Kelmendi, B., Brake, C. A., Gruner, P., Badour, C. L., & Pittenger, C. (2018). The Role of Stress in the Pathogenesis and Maintenance of Obsessive-Compulsive Disorder. *Chronic Stress*, *2*, 2470547018758043. https://doi.org/10.1177/2470547018758043

Aldridge, J. W., Berridge, K. C., & Rosen, A. R. (2004). Basal ganglia neural mechanisms of natural movement sequences. *Canadian Journal of Physiology and Pharmacology*, *82*(8–9), 732–739. https://doi.org/10.1139/y04-061

Aloisi, A. M., Steenbergen, H. L., van de Poll, N. E., & Farabollini, F. (1994). Sex-dependent effects of restraint on nociception and pituitary-adrenal hormones in the rat. *Physiology & Behavior*, *55*(5), 789–793. https://doi.org/10.1016/0031-9384(94)90061-2

Babb, J. A., Masini, C. V., Day, H. E. W., & Campeau, S. (2013). Stressor-specific effects of sex on HPA axis hormones and activation of stress-related neurocircuitry. *Stress (Amsterdam, Netherlands)*, *16*(6), 664–677. https://doi.org/10.3109/10253890.2013.840282

Ballard, I. C., Waskom, M., Nix, K. C., & D'Esposito, M. (2024). Reward Reinforcement Creates Enduring Facilitation of Goal-directed Behavior. *Journal of Cognitive Neuroscience*, 1–16. https://doi.org/10.1162/jocn_a_02150

Balleine, B. W., & Dezfouli, A. (2019). Hierarchical Action Control: Adaptive Collaboration Between Actions and Habits. *Frontiers in Psychology*, *10*, 2735. https://doi.org/10.3389/fpsyg.2019.02735

Balleine, B. W., Killcross, A. S., & Dickinson, A. (2003). The Effect of Lesions of the Basolateral Amygdala on Instrumental Conditioning. *The Journal of Neuroscience*, *23*(2), 666–675. https://doi.org/10.1523/JNEUROSCI.23-02-00666.2003

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *35*(1), 48–69. https://doi.org/10.1038/npp.2009.131

Banca, P., Ruiz, M. H., Gonzalez-Zalba, M. F., Biria, M., Marzuki, A. A., Piercy, T., Sule, A., Fineberg, N. A., & Robbins, T. W. (2023). Action-sequence learning, habits and automaticity in obsessive-compulsive disorder. *eLife*, *12*. https://doi.org/10.7554/eLife.87346.2

Bangasser, D. A., & Valentino, R. J. (2014). Sex differences in stress-related psychiatric disorders: Neurobiological perspectives. *Frontiers in Neuroendocrinology*, 35(3), 303–319. https://doi.org/10.1016/j.yfrne.2014.03.008

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., & Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, *437*(7062), Article 7062. https://doi.org/10.1038/nature04053

Blanchard, R. J., Blanchard, D. C., Rodgers, J., & Weiss, S. M. (1990). The characterization and modelling of antipredator defensive behavior. *Neuroscience & Biobehavioral Reviews*, *14*(4), 463–472. https://doi.org/10.1016/S0149-7634(05)80069-7

Blundell, P., Hall, G., & Killcross, S. (2001). Lesions of the basolateral amygdala disrupt selective aspects of reinforcer representation in rats. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *21*(22), 9018–9026. https://doi.org/10.1523/JNEUROSCI.21-22-09018.2001

Bouton, M. E. (2021). Context, attention, and the switch between habit and goal-direction in behavior. *Learning & Behavior*, *49*(4), 349–362. https://doi.org/10.3758/s13420-021-00488-z

Bouton, M. E., Allan, S. M., Tavakkoli, A., Steinfeld, M. R., & Thrailkill, E. A. (2021). Effect of context on the instrumental reinforcer devaluation effect produced by taste aversion learning. *Journal of Experimental Psychology. Animal Learning and Cognition*, *47*(4), 476–489. https://doi.org/10.1037/xan0000295

Braun, S., & Hauber, W. (2013). Acute stressor effects on goal-directed action in rats. *Learning & Memory*, *20*(12), 700–709. https://doi.org/10.1101/lm.032987.113

Buabang, E. K., Boddez, Y., Wolf, O. T., & Moors, A. (2022). The role of goal-directed and habitual processes in food consumption under stress after outcome devaluation with taste aversion. *Behavioral Neuroscience*. https://doi.org/10.1037/bne0000439

Conrad, C. D., Jackson, J. L., Wieczorek, L., Baran, S. E., Harman, J. S., Wright, R. L., & Korol, D. L. (2004). Acute stress impairs spatial memory in male but not female rats: Influence of estrous cycle. *Pharmacology Biochemistry and Behavior*, *78*(3), 569–579. https://doi.org/10.1016/j.pbb.2004.04.025

Constantino, S., & Daw, N. D. (2015). Learning the opportunity cost of time in a patch-foraging task. *Cognitive, Affective & Behavioral Neuroscience*, *15*(4), 837–853. https://doi.org/10.3758/s13415-015-0350-y

Corbit, L. H., & Balleine, B. W. (2000). The Role of the Hippocampus in Instrumental Conditioning. *Journal of Neuroscience*, *20*(11), 4233–4239. https://doi.org/10.1523/JNEUROSCI.20-11-04233.2000

Corbit, L. H., & Balleine, B. W. (2003). The role of prelimbic cortex in instrumental conditioning. *Behavioural Brain Research*, *146*(1), 145–157. https://doi.org/10.1016/j.bbr.2003.09.023

Corbit, L. H., Ostlund, S. B., & Balleine, B. W. (2002). Sensitivity to Instrumental Contingency Degradation Is Mediated by the Entorhinal Cortex and Its Efferents via the Dorsal Hippocampus. *Journal of Neuroscience*, *22*(24), 10976–10984. https://doi.org/10.1523/JNEUROSCI.22-24-10976.2002

Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioural Brain Research*, *146*(1–2), 167–174. https://doi.org/10.1016/j.bbr.2003.09.025

Cremer, A., Kalbe, F., Gläscher, J., & Schwabe, L. (2021). Stress reduces both model-based and model-free neural computations during flexible learning. *NeuroImage*, *229*, 117747. https://doi.org/10.1016/j.neuroimage.2021.117747

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

de Wit, S., Kindt, M., Knot, S. L., Verhoeven, A. A. C., Robbins, T. W., Gasull-Camos, J., Evans, M., Mirza, H., & Gillan, C. M. (2018). Shifting the Balance Between Goals and Habits: Five Failures in Experimental Habit Induction. *Journal of Experimental Psychology. General*, *147*(7), 1043–1065. https://doi.org/10.1037/xge0000402

Dezfouli, A., & Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *The European Journal of Neuroscience*, *35*(7), 1036–1051. https://doi.org/10.1111/j.1460-9568.2012.08050.x

Dezfouli, A., & Balleine, B. W. (2013). Actions, Action Sequences and Habits: Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized. *PLOS Computational Biology*, *9*(12), e1003364. https://doi.org/10.1371/journal.pcbi.1003364

Dezfouli, A., & Balleine, B. W. (2019). Learning the structure of the world: The adaptive nature of state-space and action representations in multi-stage decision-making. *PLOS Computational Biology*, *15*(9), e1007334. https://doi.org/10.1371/journal.pcbi.1007334

Dezfouli, A., Lingawi, N. W., & Balleine, B. W. (2014). Habits as action sequences: Hierarchical action control and changes in outcome value. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130482. https://doi.org/10.1098/rstb.2013.0482

Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., Costa, R. M., & Sousa, N. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science (New York, N.Y.)*, *325*(5940), 621–625. https://doi.org/10.1126/science.1171203

Dickinson, A. (1985). Actions and habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 308(1135), 67–78. https://doi.org/10.1098/rstb.1985.0010

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1–18. https://doi.org/10.3758/BF03199951

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995). Motivational control after extended instrumental training. *Animal Learning & Behavior*, *23*(2), 197–206. https://doi.org/10.3758/BF03199935

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Dougherty, R., Thrailkill, E. A., Mohammed, Z., VonDoepp, S., Hilton-Vanosdall, E., Charette, S., Van Horn, S., Quirk, A., Kraus, A., & Toufexis, D. J. (2024). Acute

stress facilitates habitual behavior in female rats. *Physiology & Behavior*, *275*, 114456. https://doi.org/10.1016/j.physbeh.2024.114456
Favila, N., Gurney, K., & Overton, P. G. (2024). Role of the basal ganglia in innate and learned behavioural sequences. *Reviews in the Neurosciences*, *35*(1), 35–55. https://doi.org/10.1515/revneuro-2023-0038
Ferguson, L. A., Matamales, M., Nolan, C., Balleine, B. W., & Bertran-Gonzalez, J. (2024). Adaptation of sequential action benefits from timing variability related to lateral basal ganglia circuitry. *iScience*, *27*(3), 109274. https://doi.org/10.1016/j.isci.2024.109274
Field, A.P., 2007. Discovering Statistics Using SPSS: (And Sex, Drugs and Rock'n'roll), 2. ed. SAGE Publ. reprinted.
Frölich, S., Esmeyer, M., Endrass, T., Smolka, M. N., & Kiebel, S. J. (2023). Interaction between habits as action sequences and goal-directed behavior under time pressure. *Frontiers in Neuroscience*, *16*. https://www.frontiersin.org/articles/10.3389/fnins.2022.996957
Gobet, F., & Simon, H. A. (1998). Expert Chess Memory: Revisiting the Chunking Hypothesis. *Memory*, *6*(3), 225–255. https://doi.org/10.1080/741942359
Goodman, J., Leong, K.-C., & Packard, M. G. (2012). Emotional modulation of multiple memory systems: Implications for the neurobiology of post-traumatic stress disorder. *Reviews in the Neurosciences*, *23*(5–6). https://doi.org/10.1515/revneuro-2012-0049
Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, *70*(1–2), 119–136. https://doi.org/10.1006/nlme.1998.3843
Graybiel, A. M. (2008). Habits, Rituals, and the Evaluative Brain. *Annual Review of Neuroscience*, *31*(1), 359–387. https://doi.org/10.1146/annurev.neuro.29.051605.112851
Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, *4*(1), Article 1. https://doi.org/10.1038/ncomms3264
Griffiths, K. R., Morris, R. W., & Balleine, B. W. (2014). Translational studies of goal-directed action as a framework for classifying deficits across psychiatric disorders. *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00101
Halbout, B., Marshall, A. T., Azimi, A., Liljeholm, M., Mahler, S. V., Wassum, K. M., & Ostlund, S. B. (2019). Mesolimbic dopamine projections mediate cue-motivated reward seeking but not reward retrieval in rats. *eLife*, *8*, e43551. https://doi.org/10.7554/eLife.43551
Harms, M. B. (2017). Stress and Exploitative Decision-Making. *Journal of Neuroscience*, *37*(42), 10035–10037. https://doi.org/10.1523/JNEUROSCI.2169-17.2017
Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement,. *Journal of the Experimental Analysis of Behavior*, *4*(3), 267–272. https://doi.org/10.1901/jeab.1961.4-267
Hogarth, L. (2020). Addiction is driven by excessive goal-directed drug choice under negative affect: Translational critique of habit and compulsion theory.

Neuropsychopharmacology, *45*(5), 720–735. https://doi.org/10.1038/s41386-020-0600-8
Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, *1*(4), 304–309. https://doi.org/10.1038/1124
Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., Dayan, P., & Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3098–3103. https://doi.org/10.1073/pnas.1414219112
JASP Team (2023). JASP (Version 0.17.3)[Computer software].
Jin, X., & Costa, R. M. (2010). Start/Stop Signals Emerge in Nigrostriatal Circuits during Sequence Learning. *Nature*, *466*(7305), 457–462. https://doi.org/10.1038/nature09263
Kalueff, A. V., Stewart, A. M., Song, C., Berridge, K. C., Graybiel, A. M., & Fentress, J. C. (2016). Neurobiology of rodent self-grooming and its value for translational neuroscience. *Nature Reviews. Neuroscience*, *17*(1), 45–59. https://doi.org/10.1038/nrn.2015.8
Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, *70*(6, Pt.1), 387–403. https://doi.org/10.1037/h0026739
Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex (New York, N.Y.: 1991)*, *13*(4), 400–408. https://doi.org/10.1093/cercor/13.4.400
Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A Neostriatal Habit Learning System in Humans. *Science*, *273*(5280), 1399–1402. https://doi.org/10.1126/science.273.5280.1399
Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, *1*(2), 106–120. https://doi.org/10.1101/lm.1.2.106
Lenow, J. K., Constantino, S. M., Daw, N. D., & Phelps, E. A. (2017). Chronic and Acute Stress Promote Overexploitation in Serial Decision Making. *The Journal of Neuroscience*, *37*(23), 5681–5689. https://doi.org/10.1523/JNEUROSCI.3618-16.2017
MacLennan, A. J., & Maier, S. F. (1983). Coping and the Stress-Induced Potentiation of Stimulant Stereotypy in the Rat. *Science*, *219*(4588), 1091–1093. https://doi.org/10.1126/science.6681679
Matisz, C. E., Badenhorst, C. A., & Gruber, A. J. (2021). Chronic unpredictable stress shifts rat behavior from exploration to exploitation. *Stress*, *24*(5), 635–644. https://doi.org/10.1080/10253890.2021.1947235
McDowell, A. L., Fransen, K. M. H., Elliott, K. S., Elghouche, A., Kostylev, P. V., O'Dea, P. K., & Garraghty, P. E. (2015). Sex Differences and the Impact of Chronic Stress and Recovery on Instrumental Learning. *Neuroscience Journal*, 2015, 1–12. https://doi.org/10.1155/2015/697659
Meier, J. K., Staresina, B. P., & Schwabe, L. (n.d.). Stress diminishes outcome but enhances response representations during instrumental learning. *eLife*, *11*, e67517. https://doi.org/10.7554/eLife.67517

Miles, O. W., Thrailkill, E. A., Linden, A. K., May, V., Bouton, M. E., & Hammack, S. E. (2018). Pituitary Adenylate Cyclase-Activating Peptide in the Bed Nucleus of the Stria Terminalis Mediates Stress-Induced Reinstatement of Cocaine Seeking in Rats. *Neuropsychopharmacology*, *43*(5), Article 5. https://doi.org/10.1038/npp.2017.135

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*, *20*(9), Article 9. https://doi.org/10.1038/nn.4613

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2022). Value representations in the rodent orbitofrontal cortex drive learning, not choice. *eLife*, *11*, e64575. https://doi.org/10.7554/eLife.64575

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without Values. *Psychological Review*, *126*(2), 292–311. https://doi.org/10.1037/rev0000120

Mosher, L. J., Godar, S. C., Nelson, M., Fowler, S. C., Pinna, G., & Bortolato, M. (2017). Allopregnanolone mediates the exacerbation of Tourette-like responses by acute stress in mouse models. *Scientific Reports*, *7*(1), 3348. https://doi.org/10.1038/s41598-017-03649-1

Nelson, A. J. D., & Killcross, S. (2013). Accelerated habit formation following amphetamine exposure is reversed by D1, but enhanced by D2, receptor antagonists. *Frontiers in Neuroscience*, *7*, 76. https://doi.org/10.3389/fnins.2013.00076

Nelson, A., & Killcross, S. (2006). Amphetamine Exposure Enhances Habit Formation. *Journal of Neuroscience*, *26*(14), 3805–3812. https://doi.org/10.1523/JNEUROSCI.4305-05.2006

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Ostlund, S. B., Winterbauer, N. E., & Balleine, B. W. (2009). Evidence of Action Sequence Chunking in Goal-Directed Instrumental Conditioning and Its Dependence on the Dorsomedial Prefrontal Cortex. *Journal of Neuroscience*, *29*(25), 8280–8287. https://doi.org/10.1523/JNEUROSCI.1176-09.2009

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The Curse of Planning: Dissecting multiple reinforcement learning systems by taxing the central executive. *Psychological Science*, *24*(5), 10.1177/0956797612463080. https://doi.org/10.1177/0956797612463080

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. https://doi.org/10.1073/pnas.1312011110

Packard, M. G., Gadberry, T., & Goodman, J. (2021). Neural systems and the emotion-memory link. *Neurobiology of Learning and Memory*, *185*, 107503. https://doi.org/10.1016/j.nlm.2021.107503

Park, H., Lee, D., & Chey, J. (2017). Stress enhances model-free reinforcement learning only after negative outcome. *PLOS ONE*, *12*(7), e0180588. https://doi.org/10.1371/journal.pone.0180588

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Pew, R. W. (1966). Acquisition of hierarchical control over the temporal organization of a skill. *Journal of Experimental Psychology*, *71*(5), 764–771. https://doi.org/10.1037/h0023100

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), Article 6863. https://doi.org/10.1038/35107080

Radenbach, C., Reiter, A. M. F., Engert, V., Sjoerds, Z., Villringer, A., Heinze, H.-J., Deserno, L., & Schlagenhauf, F. (2015). The interaction of acute and chronic stress impairs model-based behavioral control. *Psychoneuroendocrinology*, *53*, 268–280. https://doi.org/10.1016/j.psyneuen.2014.12.017

Raio, C. M., Konova, A. B., & Otto, A. R. (2020). Trait impulsivity and acute stress interact to influence choice and decision speed during multi-stage decision-making. *Scientific Reports*, *10*(1), 7754. https://doi.org/10.1038/s41598-020-64540-0

Reis, D. G., Scopinho, A. A., Guimarães, F. S., Corrêa, F. M. A., & Resstel, L. B. M. (2011). Behavioral and Autonomic Responses to Acute Restraint Stress Are Segregated within the Lateral Septal Area of Rats. *PLoS ONE*, *6*(8), e23171. https://doi.org/10.1371/journal.pone.0023171

Ridley, R. M. (1994). The psychology of perseverative and stereotyped behaviour. *Progress in Neurobiology*, *44*(2), 221–231. https://doi.org/10.1016/0301-0082(94)90039-6

Schepers, S. T., & Bouton, M. E. (2019). Stress as a context: Stress causes relapse of inhibited food seeking if it has been associated with prior food seeking. *Appetite*, *132*, 131–138. https://doi.org/10.1016/j.appet.2018.10.016

Schoenberg, H. L., Bremer, G. P., Carasi-Schwartz, F., VonDoepp, S., Arntsen, C., Anacker, A. M. J., & Toufexis, D. J. (2022). Cyclic estrogen and progesterone during instrumental acquisition contributes to habit formation in female rats. *Hormones and Behavior*, *142*, 105172. https://doi.org/10.1016/j.yhbeh.2022.105172

Schoenberg, H. L., Sola, E. X., Seyller, E., Kelberman, M., & Toufexis, D. J. (2019). Female rats express habitual behavior earlier in operant training than males. *Behavioral Neuroscience*, *133*(1), 110–120. https://doi.org/10.1037/bne0000282

Schreiner, D. C., Renteria, R., & Gremel, C. M. (2020). Fractionating the all-or-nothing definition of goal-directed and habitual decision-making. *Journal of Neuroscience Research*, *98*(6), 998–1006. https://doi.org/10.1002/jnr.24545

Schultz, W., & Dickinson, A. (2000). Neuronal Coding of Prediction Errors. *Annual Review of Neuroscience*, *23*(1), 473–500. https://doi.org/10.1146/annurev.neuro.23.1.473

Schwabe, L., Dickinson, A., & Wolf, O. T. (2011). Stress, habits, and drug addiction: A psychoneuroendocrinological perspective. *Experimental and Clinical Psychopharmacology*, 19(1), 53–63. https://doi.org/10.1037/a0022212

Schwabe, L., Höffken, O., Tegenthoff, M., & Wolf, O. T. (2011). Preventing the stress-induced shift from goal-directed to habit action with a β-adrenergic antagonist. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(47), 17317–17325. https://doi.org/10.1523/JNEUROSCI.3304-11.2011
Schwabe, L., Tegenthoff, M., Höffken, O., & Wolf, O. T. (2010). Concurrent glucocorticoid and noradrenergic activity shifts instrumental behavior from goal-directed to habitual control. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(24), 8190–8196. https://doi.org/10.1523/JNEUROSCI.0734-10.2010
Schwabe, L., Tegenthoff, M., Höffken, O., & Wolf, O. T. (2012). Simultaneous Glucocorticoid and Noradrenergic Activity Disrupts the Neural Basis of Goal-Directed Action in the Human Brain. *Journal of Neuroscience*, *32*(30), 10146–10155. https://doi.org/10.1523/JNEUROSCI.1304-12.2012
Schwabe, L., Tegenthoff, M., Höffken, O., & Wolf, O. T. (2013). Mineralocorticoid Receptor Blockade Prevents Stress-Induced Modulation of Multiple Memory Systems in the Human Brain. *Biological Psychiatry*, *74*(11), 801–808. https://doi.org/10.1016/j.biopsych.2013.06.001
Schwabe, L., & Wolf, O. T. (2009). Stress Prompts Habit Behavior in Humans. *The Journal of Neuroscience*, *29*(22), 7191–7198. https://doi.org/10.1523/JNEUROSCI.0979-09.2009
Schwabe, L., & Wolf, O. T. (2010). Socially evaluated cold pressor stress after instrumental learning favors habits over goal-directed action. *Psychoneuroendocrinology*, *35*(7), 977–986. https://doi.org/10.1016/j.psyneuen.2009.12.010
Shipman, M. L., Trask, S., Bouton, M. E., & Green, J. T. (2018). Inactivation of prelimbic and infralimbic cortex respectively affects minimally-trained and extensively-trained goal-directed actions. *Neurobiology of Learning and Memory*, *155*, 164–172. https://doi.org/10.1016/j.nlm.2018.07.010
Shors, T. J., Beylin, A. V., Wood, G. E., & Gould, E. (2000). The modulation of Pavlovian memory. *Behavioural Brain Research*, *110*(1), 39–52. https://doi.org/10.1016/S0166-4328(99)00183-7
Smeets, T., Ashton, S. M., Roelands, S. J. A. A., & Quaedflieg, C. W. E. M. (2023). Does stress consistently favor habits over goal-directed behaviors? Data from two preregistered exact replication studies. *Neurobiology of Stress*, *23*, 100528. https://doi.org/10.1016/j.ynstr.2023.100528
Smith, K. S., & Graybiel, A. M. (2013). A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics. *Neuron*, *79*(2), 361–374. https://doi.org/10.1016/j.neuron.2013.05.038
Smith, K. S., & Graybiel, A. M. (2016). Habit formation. *Dialogues in Clinical Neuroscience*, *18*(1), 33–43.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. 352.
Taylor, S. B., Anglin, J. M., Paode, P. R., Riggert, A. G., Olive, M. F., & Conrad, C. D. (2014). Chronic stress may facilitate the recruitment of habit- and addiction-

related neurocircuitries through neuronal restructuring of the striatum. *Neuroscience*, *280*, 231–242. https://doi.org/10.1016/j.neuroscience.2014.09.029

Thorn, C. A., Atallah, H., Howe, M., & Graybiel, A. M. (2010). Differential Dynamics of Activity Changes in Dorsolateral and Dorsomedial Striatal Loops During Learning. *Neuron*, *66*(5), 781–795. https://doi.org/10.1016/j.neuron.2010.04.036

Thrailkill, E. A., & Bouton, M. E. (2015). Contextual control of instrumental actions and habits. *Journal of Experimental Psychology. Animal Learning and Cognition*, *41*(1), 69–80. https://doi.org/10.1037/xan0000045

Thrailkill, E. A., Trask, S., Vidal, P., Alcalá, J. A., & Bouton, M. E. (2018). Stimulus Control of Actions and Habits: A Role for Reinforcer Predictability and Attention in the Development of Habitual Behavior. *Journal of Experimental Psychology. Animal Learning and Cognition*, *44*(4), 370–384. https://doi.org/10.1037/xan0000188

Trask, S., Shipman, M. L., Green, J. T., & Bouton, M. E. (2020). Some factors that restore goal-direction to a habitual behavior. *Neurobiology of Learning and Memory*, *169*, 107161. https://doi.org/10.1016/j.nlm.2020.107161

Vaghi, M. M., Vértes, P. E., Kitzbichler, M. G., Apergis-Schoute, A. M., van der Flier, F. E., Fineberg, N. A., Sule, A., Zaman, R., Voon, V., Kundu, P., Bullmore, E. T., & Robbins, T. W. (2017). Specific Frontostriatal Circuits for Impaired Cognitive Flexibility and Goal-Directed Planning in Obsessive-Compulsive Disorder: Evidence From Resting-State Functional Connectivity. *Biological Psychiatry*, *81*(8), 708–717. https://doi.org/10.1016/j.biopsych.2016.08.009

van Elzelingen, W., Warnaar, P., Matos, J., Bastet, W., Jonkman, R., Smulders, D., Goedhoop, J., Denys, D., Arbab, T., & Willuhn, I. (2022). Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. *Current Biology*, *32*(5), 1163-1174.e6. https://doi.org/10.1016/j.cub.2021.12.027

van Marle, H. J. F., Hermans, E. J., Qin, S., & Fernández, G. (2009). From Specificity to Sensitivity: How Acute Stress Affects Amygdala Processing of Biologically Salient Stimuli. *Biological Psychiatry*, *66*(7), 649–655. https://doi.org/10.1016/j.biopsych.2009.05.014

van Timmeren, T., Piray, P., Goudriaan, A. E., & van Holst, R. J. (2023). Goal-directed and habitual decision making under stress in gambling disorder: An fMRI study. *Addictive Behaviors*, *140*, 107628. https://doi.org/10.1016/j.addbeh.2023.107628

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., Schreiber, L. R. N., Gillan, C., Fineberg, N. A., Sahakian, B. J., Robbins, T. W., Harrison, N. A., Wood, J., Daw, N. D., Dayan, P., Grant, J. E., & Bullmore, E. T. (2015). Disorders of compulsivity: A common bias towards learning habits. *Molecular Psychiatry*, *20*(3), Article 3. https://doi.org/10.1038/mp.2014.44

Watson, P., & de Wit, S. (2018). Current limits of experimental research into habits and future directions. *Current Opinion in Behavioral Sciences*, 20, 33–39. https://doi.org/10.1016/j.cobeha.2017.09.012

Watson, P., O'Callaghan, C., Perkes, I., Bradfield, L., & Turner, K. (2022). Making habits measurable beyond what they are not: A focus on associative dual-process models. *Neuroscience & Biobehavioral Reviews*, *142*, 104869. https://doi.org/10.1016/j.neubiorev.2022.104869

Wingard, J. C., & Packard, M. G. (2008). The amygdala and emotional modulation of competition between cognitive and habit memory. *Behavioural Brain Research*, *193*(1), 126–131. https://doi.org/10.1016/j.bbr.2008.05.002

Wirz, L., Bogdanov, M., & Schwabe, L. (2018). Habits under stress: Mechanistic insights across different types of learning. *Current Opinion in Behavioral Sciences*, *20*, 9–16. https://doi.org/10.1016/j.cobeha.2017.08.009

Wood, G. E., Beylin, A. V., & Shors, T. J. (2001). The contribution of adrenal and reproductive hormones to the opposing effects of stress on trace conditioning males versus females. *Behavioral Neuroscience*, *115*(1), 175–187. https://doi.org/10.1037/0735-7044.115.1.175

Wood, G. E., & Shors, T. J. (1998). Stress facilitates classical conditioning in males, but impairs classical conditioning in females through activational effects of ovarian hormones. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(7), 4066–4071.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, *19*(1), 181–189. https://doi.org/10.1111/j.1460-9568.2004.03095.x

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2005). Blockade of NMDA receptors in the dorsomedial striatum prevents action–outcome learning in instrumental conditioning. *European Journal of Neuroscience*, *22*(2), 505–512. https://doi.org/10.1111/j.1460-9568.2005.04219.x

Yoshida, K., Drew, M. R., Kono, A., Mimura, M., Takata, N., & Tanaka, K. F. (2021). Chronic social defeat stress impairs goal-directed behavior through dysregulation of ventral hippocampal activity in male mice. *Neuropsychopharmacology*, *46*(9), Article 9. https://doi.org/10.1038/s41386-021-00990-y