

**Defining multi-dimensional associations in fear conditioning: implications for abnormal
fear responses**

Jennifer A. Murray

Psychology Department, University of Michigan

Author Note

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science with Honors in Biopsychology, Cognition and Neuroscience from the University of Michigan 2023

Correspondence concerning this paper should be addressed to Jennifer Murray,
Department of Psychology, University of Michigan, Ann Arbor, MI 48109. E-mail:
jennmurr@umich.edu

Abstract

Post-Traumatic Stress Disorder (PTSD) and anxiety disorders are related to the development of an abnormal fear response that is triggered in situations where a fear response is not appropriate (McSweeney et al., 2014). The current Pavlovian model of fear conditioning contends that animals learn a simple prediction between a cue or context and an aversive stimulus (e.g., footshock). However, in conditions such as PTSD and anxiety the fear response is activated in situations that are not consistent with the prediction, indicating that something other than just a simple prediction may be learned in context fear conditioning. This would seem consistent with research indicating that motivational or affective components are also learned in Pavlovian appetitive conditioning (Berridge, 2018). In this study we aim to lay the groundwork for identifying motivational or affective associative processes in Pavlovian fear conditioning. In addition, our research examined differences between males and females in the acquisition of learned associations, and how these differences may lead to varied generalizations between contexts. We used Pavlovian fear context conditioning to elicit predictions using both high-prediction-training (continuous reinforcement) and low-prediction training (partial reinforcement). A blocking protocol was used to test for memory retrieval and to evaluate the strength of the prediction. Tone testing was conducted to ascertain if there are differences in freezing levels between males and females. In our results, we observed minor differences in blocking levels, and differences in context generalization between males and females, supporting our hypothesis that there are differences in learning between the sexes. Our main finding – that while both high-prediction training and low-prediction training lead to strong predictions only high-prediction training blocks new learning – is evidence that prediction is not sufficient to explain all fear conditioning. This supports our contention that reframing fear conditioning

toward a model based on a multi-dimensional set of associations could have a major impact on the diagnosis and treatment of PTSD and anxiety disorders.

Defining multi-dimensional associations in fear conditioning: implications for abnormal fear responses

At least 18% of people ages 18-54 have an anxiety disorder or Post Traumatic Stress Disorder (PTSD) in any given year (Kessler, 2005). Research into the cause of these disorders suggests that fear conditioning may play a role (McSweeney et al., 2014). Through fear conditioning, animals and humans learn associations between neutral, non-threatening environmental stimuli (the conditional stimuli or CS) and aversive stimuli (the unconditioned stimuli or US) that may be painful, dangerous, or threatening. These learned associations between the CS and the US lead to the development of predictions: specifically, that the CS predicts the US. These predictions in turn produce conditioned responses (CR) (Rescorla & Wagner, 1972). Despite decades of research on fear conditioning and its relationship with PTSD, we have made little progress in understanding the etiology of the disorder based on animal models.

The Pavlovian model of conditioning assumes the development of a prediction as a learned association (McNally et al., 2011) incorporates little to no affective or motivational information. For example, in Pavlov's experiments with dogs the prediction that a light (cue) predicts food (unconditioned stimulus) was assumed to be a learned predictive association between the light and food and little attention was given to other motivational factors or physiological states such as desire or hunger. Fear conditioning based on this Pavlovian model of cue-outcome prediction contends that the development of a fear association is based on the learned association between a cue and an aversive stimulus. This association leads to a strong prediction that the cue leads to the outcome and prevents the acquisition of any further knowledge about the cue (Lis & McNally, 2014).

However, this predictive model does not always explain the behavioral responses observed in fear conditioning. If prediction is all that is learned in a fear association, then recall of that prediction should lead to the strongest fear responses when the prediction is strong. But even in situations where a strong prediction has not been established, we still see strong fear response behaviors, indicating the possibility that something other than just the cue-outcome association is being learned.

While our understanding of the neural mechanisms of the development of normal fear responses using the Pavlovian model of fear conditioning is well established, this has not led to a better understanding of the development of dysregulated or abnormal fear responses such as those that occur in anxiety or PTSD (Richter-Levin, 2019). Though the exact role of fear conditioning in anxiety disorders is not known, it may be that a normal predictive association generalizes to a second neutral cue (Huckleberry et al., 2016), causing a dysregulated fear response in otherwise non-threatening situations. If this is in fact the case, could there be a role for other factors in the acquisition and expression of learned fear associations that contributes to abnormal fear responses such as those seen in PTSD and anxiety disorders?

There is evidence to support the assumption that learning is more than just the development of a simple prediction during context fear conditioning. Research by Leon Kamin (1969) established the concept of blocking, or the inability of an animal to form an association between a cue and a second stimulus if a prior learned association between that cue and a first stimulus has already been formed. A strong blocking response is an indication of a strong prediction and an indication that no further learning can occur. Once a fear response has been established to a particular cue, it is assumed that no further learning to that cue is possible, and the fear association established with that cue cannot be transferred to a second stimulus.

However, we know that even in fear context training protocols that do not lead to a strong fear prediction, such as when an animal is exposed to an inconsistent cue-shock pairing (i.e., partial reinforcement) where the cue is only sometimes followed by a shock, animals exhibit strong fear behaviors – sometimes stronger, and usually more resistant to decreases (Huh et al., 2009). This suggests that there must be other factors influencing the development of the predictive association.

In addition, past research in our laboratory has shown different patterns of generalizations between contexts and different neural circuit activation during retrieval between males and females, suggesting that there may be sex differences in the type of information males and females learn during context fear conditioning (Keiser et al., 2017). The Keiser study demonstrated that females may employ more risk-averse cognitive strategies in fearful situations than males, suggesting that context information may be learned and remembered differently and therefore influence the development of learned fear associations differently (Keiser et al., 2017). If in fact there are individual differences such as sex that influence what and how Pavlovian fear memories are learned, understanding what may be happening beyond the development of a learned associative prediction might provide greater insight into individual risks for development of dysregulated fear responses.

Some clues as to what may be happening can be found in studies of the affective components of appetitive Pavlovian conditioning theories, in which the learned association is between a cue and a positive reinforcer as opposed to a neutral or aversive stimuli. In the Incentive Saliency theory, Kent Berridge (2018) posits that a “wanted” stimuli gains saliency and becomes a motivating factor in learned associations. Over time sensitization to the cue can lead to a blunted response (Berridge, 2018), suggesting that conditioned responses are affected by

additional affective factors. This means that in appetitive Pavlovian associations, prediction is important but not the only information that is acquired or that informs behavioral responses.

Further research by Cogan et al. (2019) demonstrated a disruption of motivational components of memory but not predictive components, suggesting that motivational or emotion-related associative components are independent of predictive associative components.

If appetitive Pavlovian conditioning incorporates both prediction and motivational / affective associative components, we expect that Pavlovian fear conditioning would also be comprised of multiple types of associations. To understand this interaction, we first must be able to separate fear responses as a consequence of predictive associations from fear responses as a consequence of other associations. In this study, we therefore aimed to determine whether we could distinguish between learned fear behaviors as a consequence of high-prediction training, versus fear behaviors as a consequence of low-prediction training.

To initiate learned associations and evaluate the strength of predictions in animals, we used a Pavlovian context fear conditioning protocol. In this protocol, animals were exposed to a context or cue that was paired with an aversive stimulus. For two experimental groups, a context (cue) was paired with an aversive stimulus (footshock) either on a continuous basis or on a partial basis, while a control group was only exposed to the context. The animals were evaluated for freezing behavior, the most common measure of fear.

To evaluate the relationship between learned associations and the development of abnormal fear responses we investigated the impact of high-prediction training and low-prediction training on the formation of these associations between contextual stimuli and aversive outcomes by testing prediction strength through blocking and assessed the generalization of these effects by testing remote memory. We manipulated the incidence of cue-

footshock occurrences to expose animals to high-prediction training and low-prediction training. High-prediction training involved a continuous reinforcement paradigm where animals received a footshock every time they were placed in the testing context leading to the development of a strong prediction between the cue (context) and the aversive stimulus (footshock). The low-prediction training protocol involved partial reinforcement of the aversive stimulus where the animal only received the shock 50% of the time they were placed in the training context.

To test the strength of the learned association we used a blocking protocol. In a strong learned association, there is a strong prediction that the CS (cue) leads to the US (stimulus) and that association should preempt – or block – any further learning about a new CS-US relationship with a second cue when the cues are presented together. If the prediction is not strong, there will be new learning to the second cue. In a blocking protocol, a learned fear association is first induced through context fear conditioning using a cue and an aversive stimulus. Once this association has been learned, the prediction strength is tested by measuring the amount of learning that takes place to a second cue.

In our study, we used context fear conditioning with naïve mice to test the strength of a learned fear association by using either high-prediction training or low-prediction training. In an eight-day study, the animals were exposed to six days of conditioning to context and shock, followed by one day of exposure to a second CS (tone) and shock. On the eighth day they were presented again with the tone in a second, neutral context to determine prediction of the shock and to look for the presence or absence of blocking. Taken together, we expect that the animals exposed to the low-prediction training model will develop a strong prediction and show robust fear behavior but will not show blocking, suggesting that prediction is not the primary memory

retrieved. In addition, we expect to see differences in blocking levels between males and females, indicating differences in learning.

To further investigate whether high-prediction or low-prediction training may be involved in the generalization of learned associations to neutral stimuli, we tested remote memory to the tone. Differences in blocking levels after an extended period of time may be indicative of persistent differences in what has been learned.

We hypothesized that the development of the prediction in context fear conditioning is not purely a learned association between the CS and the US, and sex differences exist in the acquisition of predictions. Contrary to the current model of Pavlovian fear conditioning, we contend that learned predictions are not based solely on the cue-aversive stimulus association, but rather are learned through a complex, multi-dimensional set of associations. This study will provide information that is useful in better understanding what is learned during fear conditioning and whether additional factors may be influencing the development of learned associations. A fear conditioning model that is based on a multi-dimensional set of associations could have a major impact on the diagnosis and treatment of PTSD and anxiety disorders.

Materials and Methods

Animals

In our study we used C57BL/6N naïve mice that were approximately 9 weeks of age. Each group included 24 mice (12 male and 12 female) for a total of 72 mice across three experimental groups. The mice were obtained from Envigo (Indianapolis, IN) and separated into single-animal cages upon arrival, where they were allowed to acclimate for one week to their housing and the colony room before beginning the first session. They remained in individual cages throughout the experiment and were given a standard diet and water ad libitum.

The colony room housing the individual cages was maintained at 20 ± 2 °C with a 12 h 0700: 1900 h light/dark cycle (lights on at 0700 h). The colony room is adjacent to the behavioral testing room and access is restricted to authorized personnel for both areas, reducing noise and sleep-cycle interruptions as much as possible.

All experimental procedures performed in this research were approved by the University of Michigan Committee on the Use and Care of Animals.

Materials

Behavioral testing was conducted in conditioning chambers with clear plastic front panels (9 3/4"× 12 3/4"× 9 3/4"; MedAssociates, VT) that were enclosed in sound-attenuating cubicles, equipped with NIR cameras (VID-CAM-MONO-2A), and maintained at the same temperature as the colony room. The chambers are connected to a shock generator by way of interchangeable grid floors. Chamber odor was controlled using either 70% ethanol or 1% acetic acid, which was reapplied between each animal. The shock level was set at 0.8mA for all trials.

The operant chambers were configured with different floor grids, wall shapes, brightness levels and smells to create two distinct environments (Keiser et al., 2017). Context A was always the fear conditioning chamber and was configured with white walls, an evenly sized grid floor consisting of 36 1/8" diameter stainless steel rods spaced 1/4" apart, 70% ethanol odor applied to the wall and grid surfaces, bright lights, and a fan.

Context B was always the test context. Context B was configured with black angled walls that darkened the chamber, an unevenly sized grid floor with alternating 1/8" and 3/16" stainless steel grid rods, a 1% acetic acid odor applied to the wall surfaces and grid, and no fan.

VFreeze software (MedAssociates, VT) was used to record video of the trials as well as to automatically score % freezing, % locomotor activity and shock reactivity levels.

Behavioral Testing

Behavioral testing was conducted in two experiments using a standard Pavlovian context fear conditioning protocol (McNally et al., 2011). We planned to look at group differences, sex effects and overall differences by sex. In the first experiment, prediction was manipulated and measured by varying the proportion of time an animal was presented with the shock. This procedure was conducted over a period of eight days during which time the animals were placed in Context A for 3 minutes a day for six days, followed by one day of context-tone-shock training (CTS) in Context A (2.5 minutes baseline then 30 second tone) and then one day of tone testing in Context B (2 minutes baseline then (30s tone with 60s ITT) x 3).

In the second experiment, the mice were returned to Context B after a period of four to eight weeks to test for memory retrieval of the fear conditioned CS. We conducted a tone test consisting of 6 minutes in Context B with a tone at 120 sec, 210 sec and 300 sec with no shock.

Each group consisted of 24 mice (12 male and 12 female). The mice were divided randomly into two groups of 12 mice each (males (n = 6) and females (n = 6)). Two groups were selected as Continuous Reinforcement groups (CRft) and received the shock during each day of the training, two groups were selected as Partial Reinforcement groups (PRft) and received the shock 50% of the time (days 1, 4 and 6 of the training phase), and two groups were selected as Control groups (CNTRL). Control groups were only exposed to the neutral context and did not receive any shocks during the fear conditioning training phase.

Experimental Design: Context Fear Conditioning Procedure (Experiment 1)

Group I: Control versus Continuous Reinforcement

Phase I (days 1-6): On days 1-6 the context only CNTRL group (male (n = 6) and female (n = 6)) was placed in the Context A chamber for 6 consecutive days with 3 minutes to explore

and no footshock. Simultaneously, the fear conditioning CRft group (male (n = 6) and female (n = 6)) was placed in the Context A chamber for 6 days with 3 minutes to explore followed by a 2s, 0.8mA footshock. Instances of freezing, locomotor activity and reactivity to the shock were recorded. Following conditioning, mice were returned to their home cage in the colony room.

Phase II (day 7): On day 7 both the male and female CNTRL groups and the male and female CRft groups were exposed to one day of context-tone-shock training. The mice were placed in Context A where they were allowed to explore the environment for 2.5 minutes, followed by a 30s tone and then a 2s, 0.8mA footshock. Instances of freezing, locomotor activity and reactivity to the shock were recorded. Following training, mice were returned to their home cage in the colony room.

Blocking Test: (day 8): On day 8 both the male and female CNTRL groups and the male and female CRft groups were placed in Context B for a total of 6 minutes ((30s tone with 60s ITT) x 3). The mice were allowed to explore for 2.5 minutes, then presented with a 30s tone followed by a -minute inter-trial interval (ITI), 30s tone, 1-minute ITI and a final 30s tone. Instances of freezing were recorded. Following the test, mice were returned to their home cage in the colony room.

Group 2: Continuous Reinforcement versus Partial Reinforcement

Phase I (days 1-6): On days 1-6 the fear conditioning CRft group (male (n = 6) and female (n = 6)) was placed in the Context A chamber for 6 days with 3 minutes to explore followed by a 2s, 0.8mA footshock. Simultaneously, the fear conditioning PRft group (male (n = 6) and female (n = 6)) was placed in Context A for 6 consecutive days where they were either allowed to explore the environment for 3 minutes with no foot shock (days 2, 3 and 5) or explore for 3 minutes and then receive a 2s, 0.8mA footshock (days 1, 4 and 6). Instances of freezing,

locomotor activity and reactivity to the shock were recorded. Following conditioning, mice were returned to their home cage in the colony room.

Phase II (day 7): On day 7 both the male and female CRft groups and the male and female PRft groups were exposed to one day of context-tone-shock training. The mice were placed in Context A where they were allowed to explore the environment for 2.5 minutes, followed by a 30s tone and then a 2s, 0.8mA footshock. Instances of freezing, locomotor activity and reactivity to the shock were recorded. Following training, mice were returned to their home cage in the colony room.

Blocking Test: (Day 8): On day 8 both the male and female CRft groups and the male and female PRft groups were placed in Context B for a total of six minutes ((30s tone with 60s ITT) x 3). The mice were allowed to explore for 2.5 minutes, then presented with a 30 second tone followed by a one-minute inter-trial interval (ITI), 30 second tone, 1 minute ITI and a final 30 second tone. Instances of freezing were recorded. Following the test, mice were returned to their home cage in the colony room.

Group 3: Control vs Partial Reinforcement

Phase I (days 1-6): On days 1-6 the context only CNTRL group (male (n = 6) and female (n = 6)) was placed in the Context A chamber for six days with 3 to explore and no footshock. Simultaneously, the fear conditioning PRft group (male (n = 6) and female (n = 6)) was placed in Context A for six consecutive days where they were either allowed to explore the environment for 3 minutes with no foot shock (days 2, 3 and 5) or explored for three minutes followed by a 2s footshock (days 1, 4 and 6). Instances of freezing, locomotor activity and reactivity to the shock were recorded. Following conditioning, mice were returned to their home cage in the colony room.

Phase II (day 7): On day 7 both the male and female CNTRL groups and the male and female PRft groups were exposed to one day of context-tone-shock training. The mice were placed in Context A where they were allowed to explore the environment for 2.5 minutes, followed by a 30s tone and then a 2s, 0.0mA footshock. Instances of freezing, locomotor activity and reactivity to the shock were recorded. Following training, mice were returned to their home cage in the colony room.

Blocking Test (Day 8): On day 8 both the male and female CNTRL groups and the male and female PRft groups were placed in Context B for a total of six minutes ((30s tone with 60s ITT) x 3). The mice were allowed to explore for 2.5 minutes, then presented with a 30s tone followed by a one-minute inter-trial interval (ITI), 30s tone, 1-minute ITI and a final 30s tone. Instances of freezing were recorded. Following the test, mice were returned to their home cage in the colony room.

Experimental Design: Context Fear Conditioning Retest (Experiment 2)

After a period of 4 – 8 weeks, the animals were returned to Context B to assess levels of freezing to the tone. No shocks were administered during this testing phase.

Blocking Retest (day 1): The mice were returned to Context B for a total of six minutes ((30s tone with 60s ITT) x 3). The mice were allowed to explore for 2.5 minutes, then presented with a 30s tone followed by a one-minute inter-trial interval (ITI), 30s tone, 1-minute ITI and a final 30s tone. Measurements of freezing and locomotion levels were capturing throughout the 6 minutes in 30 second intervals.

Statistical Analysis

All statistical analyses were conducted using SPSS v28. A three-way repeated measures ANOVA (DAY x SEX x GROUP) was used to measure acquisition of context fear conditioning

and sex differences in freezing. Bonferroni correction for post-hoc tests were used to analyze these effects in more detail.

To examine the effect of blocking we used a three-way repeated measures ANOVA (TONE x SEX x GROUP) with Bonferroni correction for post-hoc tests to compare between groups for each sex, for each tone.

To determine differences in baseline freezing levels prior to tone onset at test we used a 2-way repeated measures ANOVA (GROUP x SEX). Bonferroni corrected post-hoc tests were conducted to further examine differences in baseline freezing in either sex or by group.

Results

Both Males and Females Show Acquisition of Context Fear Conditioning

We used a three-way DAY x SEX x GROUP ANOVA with DAY as the repeated measure to examine acquisition of context fear conditioning. Administration of a shock during phase I of the context fear training resulted in increasing levels of freezing across days in both the CRft group and the PRft group in both sexes (Main effect DAY: $F(5,140) = 80.584$, $p < 0.001$, $\eta_p^2 = 0.957$; DAY x GROUP: $F(10,140) = 11.62$, $\eta_p^2 = 0.651$, $p < 0.001$; DAY x GROUP x SEX: $F(5,140) = 1.863$, $p = 0.105$, $\eta_p^2 = 0.286$). As expected, since control groups were exposed only to context without shock, this effect differed across group with CRft and PRft animals showing substantially more freezing than CNTRL animals (GROUP: $F(2,28) = 102.71$, $p < 0.001$). Fear conditioning also differed by SEX (Main effect SEX: $F(1,28) = 12.42$, $p < 0.001$, $\eta_p^2 = 0.476$; GROUP x SEX: $F = 8.92$, $p < 0.01$) with males showing higher freezing levels than females.

We used Bonferroni correction for post-hoc tests to examine these effects in more detail. Both the CRft group and PRft group showed higher levels of freezing when compared to the first

day of training (before the first footshock) (all $p < 0.001$); and both CRft and PRft groups, of both sexes, showed higher levels of freezing on the last day of training than did the control group, that received no footshock (all $p < 0.001$) and females ($p < 0.001$) (Figure 1).

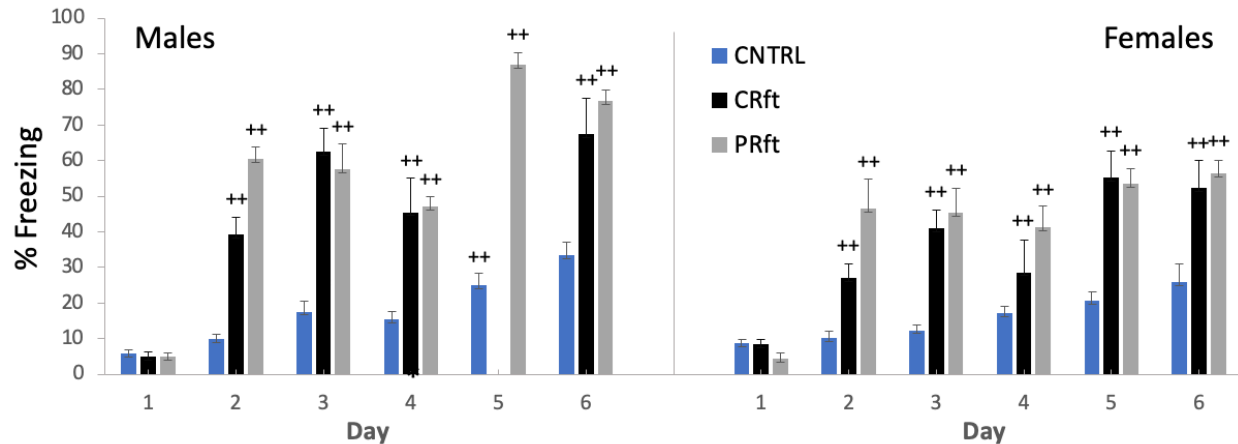


Figure 1: Freezing levels for males (left) and females (right) during training. Both males and females show increased freezing after either CRft (black bars) or PRft (grey bars) training when compared to the first day of training. Both CRft and PRft groups, of both sexes, showed higher levels of freezing on the last day of training than did the control group (CNTRL), that received no footshock. Note: male Day 5 CRft data was lost due to computer malfunction. Error bars represent SEM. ++ $p < 0.001$ vs CNTRL.

Both Males and Females Show Reactivity in Fear Context Training

We used a 3-way SHOCK REACTIVITY x SEX x GROUP ANOVA, with REACTIVITY as the repeated measure, to determine differences across group in reactivity to the shock. We identified a main effect of REACTIVITY ($F(6, 138) = 39.580, p < 0.001, \eta_p^2 = 0.632$); and GROUP ($F(12, 138) = 18.117, p < 0.001, \eta_p^2 = 0.612$), but not REACTIVITY x SEX interaction: $F(6, 138) < 1$; or REACTIVITY x SEX x GROUP: $F(6, 138) < 1$, demonstrating that differences in reactivity levels were seen across groups but not across sex (Figure 2). There were between-subject effects of GROUP: $F(2, 23) = 58.606, p < 0.001$ but not SEX: $F(1, 23) < 1$, or SEX x GROUP: $F(2, 23) < 1$. Pairwise comparisons of groups showed significant differences in reactivity levels where CRft showed higher reactivity than CNTRL animals ($P < 0.001$), and similarly PRft showed higher reactivity than CNTRL animals ($p < 0.001$) – this is expected since

CNTRL animals did not receive footshocks. More importantly, PRft animals showed higher shock responses than CRft animals ($p < 0.001$) suggesting that unpredictable shocks result in sensitized responding. This effect was also observed in pairwise comparisons for both males (CNTRL and PRft: $p < 0.001$) and females (CNTRL and CRft: $p < 0.001$, CNTRL and PRft: $p < 0.001$, CRft and PRft: $p < 0.05$), suggesting that the differences in reactivity to the shock were seen across both sexes.

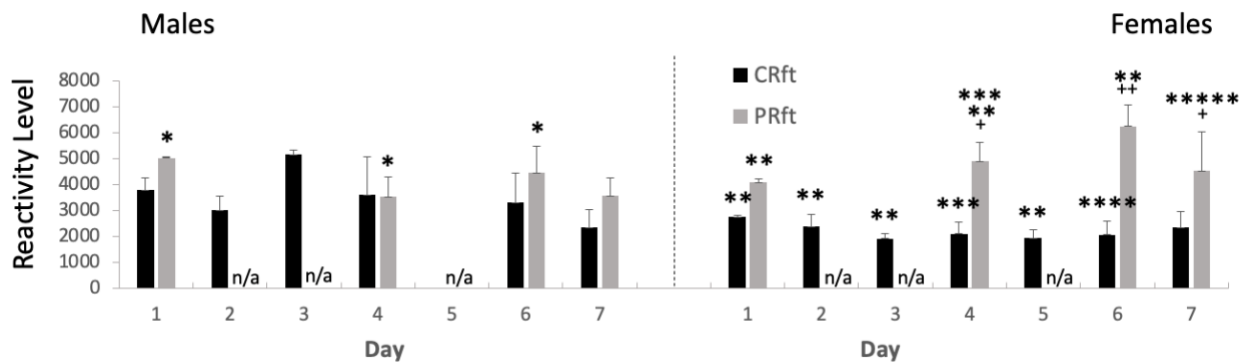


Figure 2: Shock reactivity levels for males (left) and females (right) during training. The difference between the CRft-trained animals (black bars) and PRft-trained (grey bars) was observed in both males and females. This effect was also observed in pairwise comparisons for both males and females. Note 1: the PRft group only receive a footshock on days 1, 4 and 6 and 7. Note 2: male Day 5 CRft data was lost due to computer malfunction. Error bars represent SEM. + $p < 0.05$; ++ $p < 0.001$ between CRft and PRft; * $p < 0.001$ between male CNTRL and PRft; ** $p < 0.001$ between female CNTRL; *** $p < 0.05$ between female CRft; **** $p < 0.01$ between female CNTRL; ***** $p < 0.01$ between female CNTRL.

In addition to Reactivity, we looked at levels of locomotor activity following a shock to evaluate whether males and females showed different locomotor activity levels following a shock. We used a 3-way LOCOMOTORACTIVITY x SEX x GROUP ANOVA, with LOCOMOTOR as a repeated measure, to determine whether both males and females show reactivity to the shock. We identified a main effect of GROUP ($F(2,35) = 34.64$, $p < 0.001$, $\eta_p^2 = 0.664$), but not LOCOMOTOR x SEX interaction: $F(1,35) < 1$; or LOCOMOTOR x GROUP x SEX: $F(2,35) < 1$, and CRft-trained and PRft-trained animals showed lower locomotor activity levels compared to CNTRL animals (Figure 3), an unsurprising effect given increased freezing decreases locomotor activity in shocked groups. There were between-subject effects of SEX:

$F(1,35) = 7.076$, $p < 0.05$ and GROUP: $F(2,35) = 14.23$, $p < 0.001$, but not in SEX x GROUP: $F(2,35) < 1$, with males showing higher locomotor activity levels than females across groups.

There also appeared to be between-subject differences in locomotor activity in the CNTRL animals between males and females ($P < 0.05$) but not in either the CRft males and females ($p < 1$) or the PRft males and females ($p < 1$), suggesting differences in locomotor activity levels between CRft and PRft males and females but not CNTRL males and females.

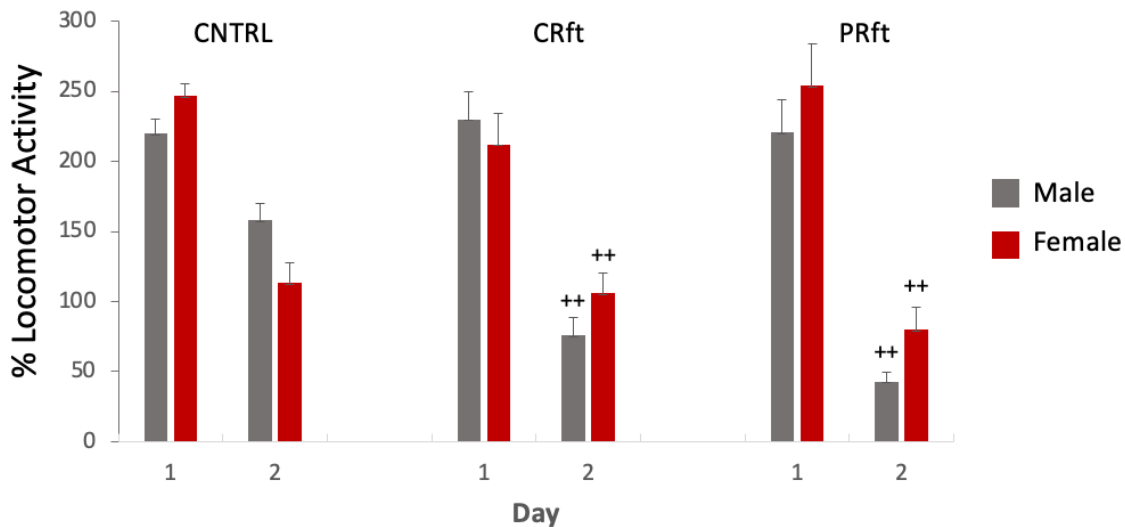


Figure 3: Locomotor activity levels for all groups on day 1 and day 2, following administration of a footshock. CRft-trained and PRft-trained animals showed lower locomotor activity levels compared to CNTRL animals between day 1 and day 2 following administration of a footshock to CRft and PRft animals. Differences in locomotor activity levels were observed between CNTRL and CRft and CNTRL and PRft between day 1 and day 2, but not between CRft and PRft, demonstrating that CRft and PRft animals were reacting to the shock. An overall decrease in locomotor activity over the course of the training was likely due to habituation to the context. Error bars represent SEM. ++ $p < 0.001$ vs CNTRL.

To examine this within group effect further we used Bonferroni-corrected post-hoc tests to compare between groups. Differences were observed between CNTRL and CRft ($p < 0.001$) and CNTRL and PRft ($p < 0.001$) but not between CRft and PRft ($p < 1$), demonstrating that the stronger decrease in locomotor activity in these groups was due to increased freezing.

Continuous Reinforcement but not Partial Reinforcement Leads to Higher Blocking Levels in Males but Not Females

We used a 3-way TONE x SEX x GROUP ANOVA, with TONE as a repeated measure, to examine the effect of CRft and PRft on fear conditioning to the tone, that is, blocking. We identified a main effect of TONE ($F(2,78) = 8.172$, $p < 0.001$, $\eta_p^2 = 0.173$); but not of TONE x GROUP interaction: $F(4,78) < 1$; or TONE x SEX: $F(2,78) < 1$) demonstrating that across groups and sexes, animals showed increasing freezing across the tone-test session (Figure 4). We observed a non-significant but suggestive effect of GROUP ($F(2,39) = 3.01$, $p = 0.061$), suggesting that prior CRft or PRft altered freezing to the tone at test. There were no effects of SEX, and no significant interactions (SEX: $F(1,39) < 1$; GROUP x SEX: $F(2,39) < 1$; TONE x SEX x GROUP: $F(4,78) < 1$). Pairwise comparisons between GROUP showed a significant effect between CNTRL animals and CRft animals ($p < 0.05$), but not between PRft animals and either CNTRL animals ($p = 0.925$) or CRft animals ($p = 0.072$).

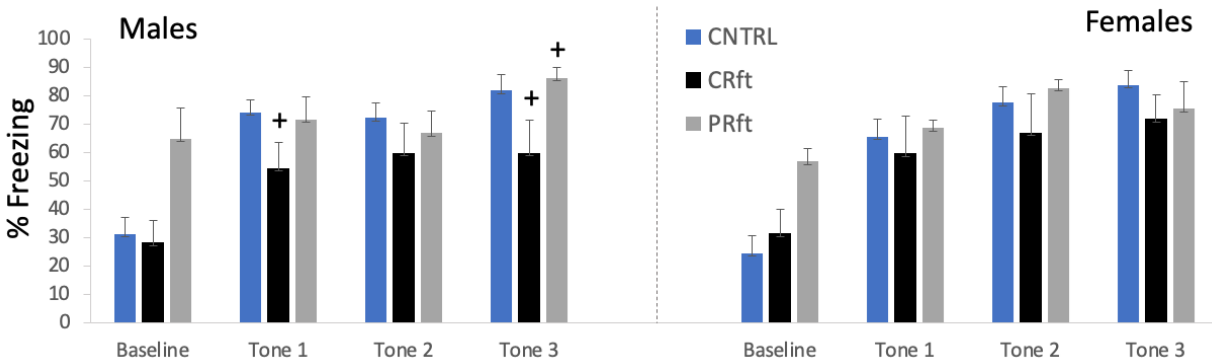


Figure 4: Freezing levels for males (left) and females (right) to the tone in the tone test. Across groups and sexes, animals showed increasing freezing across the tone-test session. A suggestive effect of GROUP indicates that prior CRft (black bars) or PRft (grey bars) may have altered freezing to the tone at test. Pairwise comparisons demonstrate this effect between male CNTRL (blue bars) and CRft animals on tones 1 and 3, and between male CRft and PRft animals on tone 3. Error bars represent SEM. + $p < 0.05$ between CRft and CNTRL or PRft.

To examine this effect further, given our planned comparisons within sex, and the main effect of tone, we used Bonferroni tests to compare between groups for each sex, for each tone.

In males, we observed across all three tones that CRft animals showed significantly less freezing than CNTRL animals ($p < 0.05$), while PRft did not differ from CNTRL animals ($p = 0.89$). In contrast, females exhibited no differences in freezing to the tone between groups (CRft vs CNTRL: $p = 0.26$; PRft vs CNTRL: $p = 0.996$).

This effect was similar across tones in males, where CRft animals showed significantly lower levels of freezing than CNTRL animals to the first and last tones at test (Tone1: $p < 0.05$, Tone2: $p = 0.22$; Tone3: $p < 0.05$); and no differences between PRft and CNTRL animals for any tone (all $p > 0.60$). In females, there were no differences between groups for any tone (all $p > 0.20$).

Baseline Freezing

We used a 2-way GROUP x SEX ANOVA to determine differences in baseline freezing levels prior to tone onset at test. We observed an overall main effect of GROUP ($F(2,40) = 13.162$, $p < 0.001$, $\eta_p^2 = 0.397$) and no main effect of sex ($F(1,40) = 0.804$, $p = 0.375$); and no GROUP x SEX interaction ($F(2,40) = 0.01$, $p = 0.99$), indicating differences in baseline freezing between groups. A significant difference was observed in pairwise comparisons. PRft-trained animals showed significantly higher freezing than CNTRL animals (males: $p < 0.001$; females: $p = 0.001$), while CRft trained animals did not differ from CNTRL animals in baseline freezing in either sex (males: $p = 0.126$; females: $p = 0.094$). Additionally, no significant differences were noted in baseline freezing between CRft animals and PRft animals (males: $p = 0.76$; females: $p = 0.094$).

Bonferroni corrected post-hoc tests showed a significant difference between PRft animals and CNTRL animals ($p < 0.001$) and PRft animals and CRft animals ($p < 0.05$), but not between

CRft and CNTRL ($p = 0.080$). This means that the PRft group showed substantially more generalization between the training context (Context A) and the testing context (Context B).

Continuous Reinforcement but Not Partial Reinforcement Leads to Higher Blocking Levels in Males but Not Females in Remote Tone Test

We used a 3-way TONE x SEX x GROUP ANOVA, with TONE as a repeated measure, to examine the effect of CRft and PRft on fear conditioning to the tone (blocking) after 4-8 weeks. We identified a main between-subject effect of GROUP ($F(2,39) = 8.101$, $p = 0.001$, $\eta_p^2 = 0.025$) but not TONE x GROUP interaction: $F(4,78) < 1$; or TONE x SEX: $F(2,78) < 1$. CRft-trained animals showed lower freezing to the tone than CNTRL or PRft-trained mice (Figure 5), suggesting that prior CRft altered freezing to the tone in the remote tone test. There were no effects of SEX, and no significant interactions (SEX: $F(1,39) < 1$; GROUP x SEX: $F(2,39) < 1$; TONE x SEX x GROUP: $F(4,78) < 1$, $p < 1$).

There were within-subject effects observed in the pairwise comparisons. In males, we observed differences in freezing to the tone between groups, with CRft animals showing significantly less freezing than CNTRL animals ($p < 0.05$) but not PRft animals ($p = 0.134$), while PRft animals did not differ from CNTRL animals ($p = 0.517$). Female CRft mice exhibited differences in freezing to the tone between groups (CRft vs CNTRL: $p < 0.01$; CRft vs PRft: $p < 0.01$), but there was no difference between CNTRL and PRft ($p = 0.721$). Therefore, both males and females in the CRft group appear to show the blocking effect with less freezing to tone.

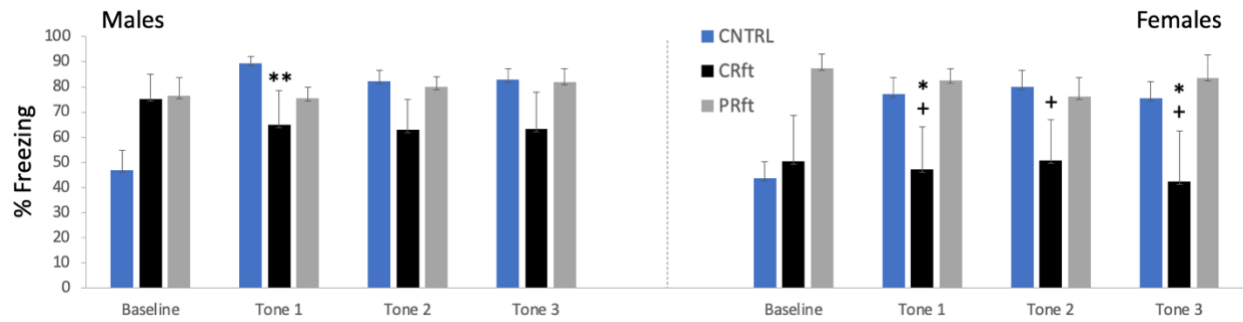


Figure 5: Freezing levels for males (left) and females (right) to the tone in the remote tone test. CRft-trained animals (black bars) showed lower freezing to the tone than CNTRL (blue bars) or PRft-trained animals (grey bars), suggesting that prior CRft altered freezing to the tone in the remote tone test. Males showed differences in freezing to the tone between groups, with CRft animals showing significantly less freezing than CNTRL animals but not PRft animals, while PRft animals did not differ from CNTRL animals. In females, CRft animals exhibited differences in freezing to the tone between groups (CRft vs CNTRL and CRft vs PRft), but there was no difference between CNTRL and PRft. Error bars represent SEM. + $p < 0.05$ between female CRft and CNTRL; * $p < 0.01$ between female CRft and PRft; ** $p < 0.001$ between male CRft and CNTRL.

In pairwise comparisons, there were significant differences in females on tone 1 between CRft and CNTRL ($p < 0.05$) and CRft and PRft ($p < 0.01$); tone 2 between CRft and CNTRL ($p < 0.05$) and suggestive on CRft and PRft ($p = 0.061$); and tone 3 between CRft and CNTRL ($p < 0.05$) and CRft and PRft ($p < 0.01$). In males, there were significant differences on tone 1 between CRft and CNTRL ($p < 0.05$); and a suggestive effect in tone 2 between CRft and CNTRL ($p = 0.058$).

To examine the between-subject effect of GROUP further, we used Bonferroni-corrected post-hoc tests to compare between groups. Differences were observed between CRft and CNTRL ($p < 0.01$) and CRft and PRft ($p < 0.05$) but not between CNTRL and PRft ($p = 1$), demonstrating that CRft animals were showing differences in freezing levels to the tone.

Discussion

In this study, we used Pavlovian context fear conditioning and a blocking paradigm to determine whether predictions are not sufficient to explain all Pavlovian fear conditioning. We observed differences in blocking levels between males and females in continuous reinforcement training, with males showing significantly higher levels of blocking than females, suggesting

there may be differences in learning between the sexes. Our main observation, consistent with our hypothesis, was that both high-prediction training and low-prediction training resulted in strong fear responses, but only high-prediction training led to blocking. This is incongruent with the concept that blocking should occur in all groups with strong context fear conditioning, as would be expected with the traditional Pavlovian association learning model, leading us to conclude that there must be something more being learned in fear associations than just a cue-outcome prediction.

The main finding of our study is that continuous reinforcement, but not partial reinforcement leads to blocking of new learning. Because blocking requires that a previously established CS (in this case a context) predicts a US (footshock), such that when a new CS (in this case a tone) is presented together with the original CS, and both are followed by the original US, there is no prediction error and no new learning (Kamin, 1969), our results support the hypothesis that fear predictions are based on more than just a predictive cue-aversive stimulus association. Our results, which were consistent with prior research findings, demonstrated that in high-prediction training where there is continual reinforcement (CRft) of the cue-aversive stimulus there is a strong prediction (McSweeney, 2014), while in low-prediction PRft training there is not (Huh et al., 2009).

We demonstrated this strong prediction by CRft animals using Pavlovian context fear conditioning. Administration of a shock during Phase I of the context fear training resulted in increasing levels of freezing across days in both the CRft group and the PRft group in both sexes, as would be expected since the learned association is reinforced over the course of six days of training.

We then used a blocking protocol to assess the level of learning acquired by each group. During Phase II training on day 7, we introduced a second neutral cue (tone) in the original Context A and paired it with a footshock. Animals in the CNTRL group, which had no prior exposure to the shock, showed a strong learned association to the second cue when exposed to the tone in a different context (Context B) during the tone test on day 8.

We again saw this blocking effect in the CRft animals in the remote tone test, where CRft-trained animals of both sexes showed significantly less freezing to the tone than either control or PRft-trained animals. CRft-trained animals showed lower freezing to the tone than CNTRL or PRft-trained animals, suggesting that prior CRft training altered freezing to the tone in the remote tone test. Both male and female CRft animals demonstrated significantly less freezing than CNTRL animals and PRft animals, which we would expect to see in a blocking effect driven by high prediction in CRft animals. While the PRft animals did not show lower freezing levels to the tone than CNTRL animals, this may be a result of their higher baseline freezing levels and therefore cannot be interpreted as higher freezing to the tone.

Baseline freezing levels prior to tone onset at test on day 8 showed no difference between CRft trained animals and CNTRL animals in baseline freezing in either sex, but significantly higher freezing levels than CNTRL animals for PRft-trained animals. This significant effect of group where CRft-trained animals of both sexes showed significantly less freezing to the tone than either CNTRL or PRft-trained animals is unsurprising and is suggestive of less specificity of fear to the shock-paired context, less prediction, and more fear or anxiety overall (Grupe & Nitschke, 2013).

In fact, the data also demonstrated that partial reinforcement does not result in strong blocking. Despite being trained in a manner consistent with a failure to form strong predictions

(Huh et al., 2009), in our PRft group, we saw higher levels of freezing than the CRft group across tones in Context B, similar to those of the CNTRL group. But unlike the CNTRL group, the PRft group showed a high baseline level of freezing, indicating a difference in what had been previously learned was resulting in a low prediction.

If the fear response in Pavlovian conditioning is strictly a learned prediction between the cue and the stimulus, we would not expect to see these differences in freezing levels between the CRft group and the PRft group in Context B at test, as both groups were exposed to the shock in the same original context, and both should have demonstrated similar blocking patterns to the tone in the new context.

Additionally, unlike the CRft group and CNTRL group, the PRft animals showed high levels of freezing before the tone at test. If they had acquired a strong prediction to the tone, they would not have exhibited high levels of freezing until after hearing the tone. This suggests that instead of acting on a learned association between the tone and the footshock, they were already behaving as if they were afraid. This fear reaction is therefore not a context-footshock prediction, but rather an indication that additional associative information is driving fearful behavior. This lack of prediction likely facilitates the generalization of fear to a neutral stimulus (Context B), and we hypothesize that some other information was learned by the PRft animals. If Pavlovian fear conditioning is similar to Pavlovian appetitive conditioning, it could be that there are other motivational or emotional components involved in the learning of the fear response that is being expressed in PRft animals. A central goal for future work will be to identify precisely what other – non-predictive – associations are learned in the course of Pavlovian fear conditioning.

In this heightened fear response in PRft animals, the inability to learn an association and develop a high prediction may be manifesting as a generalization of the learned response to a

novel context. But other factors could also be responsible for this heightened fear response, including a fear of any new context or an inability to discern from a given context whether to expect the US, which would cause the animal to remain on high alert.

In all three instances, this suggests what the animal is experiencing is not a prediction but fear. Visual observations of the PRft animals during fear context training noted higher levels of what might be considered anxiety-like behavior, including disordered patterns of behavior before, during and after tones during tone testing. Unlike CNTRL and CRft animals, which tended to freeze to the tone, PRft animals froze between tones and often jumped and darted during the tone (Gruene et al, 2015). This may suggest that a lack of prediction to the shock leads to a disorganized or perhaps an escape response, whereas a strong prediction of the shock preferentially leads to a freezing response (at least in the kinds of boxes we test in).

One confounding factor in understanding what PRft-trained animals learn is that these PRft animals might be freezing at high levels because they learned the tone-shock association in Phase II (which would indicate that they did not acquire blocking during the original context-shock association in Phase 1) or because they are simply freezing to the new context on introduction. Further studies will need to disentangle these effects.

Manipulation of other experimental variables might also yield insight into what is being learned. For example, if the shock level was lowered to be mildly unpleasant but not jarring, would the PRft animals still exhibit exaggerated freezing levels to either the context or the tone trained in blocking protocol? Indeed, if animals are learning motivational/affective associative components in addition to prediction, then low shock with CRft training might lead to more prediction, less affective information learned and a stronger blocking effect than we observed here, whereas PRft animals might still show high freezing levels and high affective/motivational

components since their training precludes prediction, and uncertainty can drive anxiety (Grupe & Nitschke, 2013).

Another option might be to measure alternative behavioral responses such as darting that might be associated with fear (Gruene et al., 2015). For example, during visual observations of PRft animals during training it was noted that they often exhibited different patterns of behavior than CRft or CNTRL animals. While CRft or CNTRL animals would stop to the tone, often the PRft animals would jump or run. Might these different behaviors indicate a heightened fear response? Innovative new methods of discerning specific behaviors are being developed that might shed light on what PRft animals are experiencing (Gruene et al., 2015; Mitchell et al., 2022).

Finally, our data provided additional evidence for our hypothesis that males and females show different freezing levels to fear, indicating a potential difference in the information they learn in context fear training (Keiser, 2017). In the tone test, the CRft males showed a higher level of blocking than the CRft females. The between-subject effects significance level suggests a main effect between males and females in this group. In addition, overall freezing levels were higher for males than females in all groups.

One explanation for this difference between males and females is that across time males are generalizing to the context more than females. This suggests that the males are either losing specific context memory and / or are relying less on context for prediction (Wiltgen & Silva, 2007) and are shifting toward less predictive components of associative memory. While inconclusive, this is suggestive of different behavioral strategies between males and females that may be influenced by factors other than just the learned cue-outcome association and will be investigated further in a follow-up study.

This potential difference in learning between the sexes is important for several reasons: first, if males and females are learning different information during fear prediction acquisition this implies that other factors are influencing the development of these predictions, and second, extinction of those predictions may require different extinction methods. Distinguishing between fear behaviors that are the result of high-prediction training from fear behaviors that are the result of low-prediction training will give us insight into what this additional information may be, how it translates into a fear response behavior and how it comes to be incorporated in the fear response.

Our finding that males and females show different freezing patterns may also point to the salience of additional motivational factors. We can use blocking to aide in differentiating fear behaviors by the level of predictive training. However, we contend that there are sex differences in behaviors exhibited during the learned fear response and believe that there is much more knowledge to be gained about the different ways in which males and females learn and how they express those differences in their behavior, and this defines an area requiring further study.

We surmise that different motivational factors might be eliciting different behaviors observed in mice during context fear training. The current model of Pavlovian fear conditioning is based on the premise that freezing rates are the only method of assessing the strength of the learned association between the cue and stimulus. However, variations in behavior have been observed in animals during fear conditioning experiments. Mitchell et al. (2105) found evidence supporting a second, novel fear response in female rats. The researchers identified darting as a potential expression of a learned fear association and contend that it may represent an alternative fear response behavior to freezing.

We know that mice and other rodents exhibit a much broader array of fear behaviors, including darting, jumping, stretching and nose poking. During visual observation of animals undergoing context fear training, we observed many of these behaviors between the sexes and in different contexts. Future work will need to include more nuanced assessments of fear-related behaviors. This work is already happening on darting (Gruene et al., 2015; Mitchell, et al., 2022) and will be expanded with recent development of AI tools for automated behavioral assessments such as DeepLabCut (Lauer et al., 2022) and LabGym (Hu et al., 2023).

In assessing other behaviors, we observe the animals during the first training trial (i.e., before the first shock) and throughout training to watch for potential confounds. For example, if a mouse is showing high levels of freezing or immobility before the first shock, this would be an exclusion criterion because we cannot then discern whether high levels of freezing at test are due to learning.

Sometimes a set of animals received at the same time will experience some (unknown to us) stressor or event that leaves them with unusual patterns of baseline data – this can dramatically change fear related behaviors and we compare “batches” of animals to ensure consistency in pretraining data and in control groups. In this study, it was determined that a group of animals exhibited behaviors outside of the normal range during the first trial and was therefore excluded from the study. This group of mice was removed from the experiment because of significantly high pre-testing anxiety levels across the group which were evident through different behavioral patterns from the start of the study. When placed into the experimental chamber on day one, many of the mice were overly jumpy and frantic. Before being removed from the box, they demonstrated aggressive and reckless behavior including lunging at the glass and jumping straight up into the ceiling of the experiment box. During the

test they exhibited less of the normal exploring behavior we see when an animal is introduced to new surroundings and taken together this abnormal behavior might be an indication that something other than normal study variables might have influenced their behavior prior to the start of the study.

The wide acceptance of the Pavlovian conditioning paradigm as the definitive statement on fear acquisition and demonstration has found little challenge in the decades since its inception. However, as we learn more about the neural mechanisms behind learning and memory and the impact of affective factors on these processes (Berridge, 2018), combined with growing evidence that learned fear responses are not strictly the result of a cue-stimulus learned association, it becomes more apparent that there is a need for more diverse behavioral considerations in studying and understanding the acquisition of these responses.

Unraveling what is learned in fear conditioning is important in understanding how learned fear associations are connected to innocuous stimuli in abnormal or dysregulated fear responses such as PTSD or anxiety disorders. This may assist in discerning why a similar event leads to a normal fear response in some people (high prediction training / high-prediction) and a dysregulated response in others (high-prediction training / low prediction). This in turn may lead to a better understanding of how and why different motivational and affective factors may differently impact the way in which males and females acquire fear predictions.

While PTSD was once considered chiefly an affliction affecting soldiers returning from conflict zones, it is now recognized as a disorder with a wide array of triggers affecting a broad cross-section of Americans. Research into the physiological causes of PTSD and other anxiety disorders is based on the use of the standard Pavlovian fear conditioning model. We hypothesize that this model is incomplete and does not account for multi-dimensional associations that also

influence learned associations, and these factors may be at the root of PTSD and other anxiety disorders. By designing a new fear conditioning framework, we may be better able to understand the physiological basis for abnormal fear reactions that are at the root of these disorders and develop better treatment options for those living with these conditions.

References

- Berridge, K. C. (2018). Evolving concepts of emotion and motivation. *Frontiers in Psychology*, 9, 1647–1647. <https://doi:10.3389/fpsyg.2018.01647>
- Cogan, E. S., Shapses, M. A., Robinson, T. E., & Tronson, N. C. (2019). Disrupting reconsolidation: memory erasure or blunting of emotional/motivational value? *Neuropsychopharmacology* 44(2), 399–407. <https://doi.org/10.1038/s41386-018-0082-0>
- Gruene, T. M., Flick, K., Stefano, A., Shea, S. D., & Shansky, R. M. (2015). Sexually divergent expression of active and passive conditioned fear responses in rats. *eLife*, 4, 1-7. <https://doi:10.7554/eLife.11352>.
- Grupe D. W., & Nitschke J. B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat Reviews Neuroscience*, 14(7), 488-501. <https://doi:10.1038/nrn3524>
- Huckleberry, K. A., Ferguson, L. B., & Drew, M. R. (2016). Behavioral mechanisms of context fear generalization in mice. *Learning & Memory*, 23(12), 703–09. <https://doi:10.1101/lm.042374.116>
- Hu, Y., Ferrario, C. R., Maitland, A. D., Ionides, R. B., Ghimire, A., Watson, B., Iwasaki, K., White, H., Xi, Y., Zhou, J., & Ye, B. (2023). LabGym: Quantification of user-defined animal behaviors using learning-based holistic assessment. *Cell Reports Methods*, 3(3), 1-14. <https://doi.org/10.1016/j.crmeth.2023.100415>
- Huh, K. H., Guzman, Y. F., Tronson, N. C., Guedea, A. L., Gao, C., & Radulovic, J. (2009). Hippocampal erk mechanisms linking prediction error to fear extinction: roles of shock expectancy and contextual aversive valence. *Learning & Memory*, 16(4), 273–78. <https://doi:10.1101/lm.1240109>

Kamin, L. J. (1967). Predictability, surprise, attention, and conditioning, 1-25.

<https://ntrs.nasa.gov/api/citations/19680014821/downloads/19680014821.pdf>

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 617–627.

<https://doi.org/10.1001/archpsyc.62.6.617>

Keiser, A. A., Turnbull, L. M., Darian, M. A., Feldman, D. E., Song, I., & Tronson, N. C.

(2017). Sex differences in context fear generalization and recruitment of hippocampus and amygdala during retrieval. *Neuropsychopharmacology*, 397–407.

<https://doi:10.1038/npp.2016.174>

Lauer, J., Zhou, M., Ye, S., Menegas, W., Schneider, S., Nath, T., Rahman, M. M., Di Santo, V., Soberanes, D., Feng, G., Murthy, V. N., Lauder, G., Dulac, C., Mathis, M. W., & Mathis, A. (2022). Multi-Animal pose estimation, identification and tracking with DeepLabCut.

Nature Methods, 19(4), 496–504. <https://doi:10.1038/s41592-022-01443-0>

Lis, S. S. Y., & McNally, G. P. (2014). The conditions that promote fear learning:

Prediction error and Pavlovian fear conditioning. *Neurobiology of learning and memory*. 2014(108), 14–21. <https://doi.10.1016/j.nlm.2013.05.002>

McNally G.P., Johansen J.P., & Blair H.T. (2011). Placing prediction into the fear circuit. *Trends in Neurosciences*.34, 283–292. <https://doi:10.1016/j.tins.2011.03.005>

McSweeney, F. K., Murphy, E. S., Faneslow, M. S., & Sterlace, S. R. (2014). Pavlovian fear conditioning function, cause, and treatment. *Wiley Blackwell handbook of operant and classical conditioning*, 118-124. <https://escholarship.org/uc/item/0kp8z40n>

Mitchell, J. R., Trettel, S. G., Li, A. J., Wasielewski, S., Huckleberry, K. A., Fanikos, M.,

- Golden, E., Laine, M. A., & Shansky, R. M. (n.d.) (2022). Darting across space and time: parametric modulators of sex-biased conditioned fear responses. *Learning & Memory*, 29(7), 171–80. <https://doi:10.1101/lm.053587.122>
- Rescorla R.A., & Wagner A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II Current Research and Theory*, 21, 64–99.
- Richter-Levin G., Stork, O., & Schmidt, M. V. (n.d.) (2018). Animal models of PTSD: a challenge to be met. *Molecular Psychiatry*, 24(8), 1135–56. <https://doi:10.1038/s41380-018-0272-5>.
- Tronson, N.C. (2019). Uncertainty versus prediction error in Pavlovian fear conditioning: Commentary on Walker et al. (2019). *European Journal of Neuroscience*, 52(5), 3485–86. <https://doi:10.1111/ejn.14578>
- Wiltgen, B. J., & Silva, A. J. (2007). Memory for context becomes less specific with time. *Learning & Memory*, 14(4), 313–17. <https://doi:10.1101/lm.430907>