Bar-Ilan University

# The Role of the Cerebellum in Reinforcement Learning by Prediction Error.

David Slobodiansky

**Submitted in partial fulfilment of the requirements for the Master's Degree in the Interdisciplinary Studies unit, the Gonda Multidisciplinary Brain Research Center, Bar Ilan University**

**Ramat Gan, Israel** **2025**

*This work was carried out under the supervision of Prof. Dana Cohen, Gonda Multidisciplinary Brain Research Center, Bar Ilan University.*

***Table of Content***

**List of abbreviations**

The following table provides a comprehensive list of the most significant abbreviations and acronyms frequently encountered throughout this thesis. These terms are essential for understanding the content and are consistently used to maintain clarity and brevity in the text.

| ***Abbreviation*** | ***Meaning*** |
|---:|---|
| *ANOVA* | Analysis of variance |
| *BG* | Basal ganglia |
| *CFs* | Climbing fibers |
| *CN* | Cerebellar nuclei |
| *DA* | Dopaminergic |
| *IO* | Inferior olive |
| *ITI* | Inter-trial interval |
| *KW* | Kruskal-Wallis H-test |
| *PC's* | Purkinje cells |
| *RPEs* | Reward prediction errors |
| *SEM* | Standard error of the mean |
| *SNc* | Substantia nigra pars compacta |
| *SPEs* | Sensory prediction errors |
| *STD* | Standard deviation |
| *TD* | Temporal difference |
| *VTA* | Ventral tegmental area |
| *VR* | Virtual reality |

**Abstract: The Role of the Cerebellum in Reinforcement Learning by Prediction Error**

Two decades ago, Doya hypothesized that the cerebellum utilizes supervised learning that relies on sensory prediction errors (SPEs) for generating sensorimotor adaptation[1] and can facilitate motor learning while basal ganglia (BG) utilize reinforcement learning that relies on reward prediction errors (RPEs) for maximizing future reward. This hypothesis has guided research into how distinct brain structures utilize specific learning algorithms, emphasizing the importance of understanding neuronal correlates of behavior to comprehend brain function. However, recent findings suggest that the cerebellum may also play a pivotal role in reinforcement learning by encoding reward-related information, challenging the long-standing division between these two structures.

SPEs are conveyed by climbing fibers originating in the inferior olive, and innervating neurons in the cerebellar cortex and nuclei, thus modulating cerebellar activity. RPEs, on the other hand, are conveyed by dopaminergic neurons from the ventral tegmental area and substantia nigra pars compacta (SNc), targeting striatal neurons. Recent evidence of reward signals in the cerebellum blurs the functional distinction between the cerebellum and basal ganglia, especially given their reciprocal connections via disynaptic pathways indicating that they may not be as distinct as initially thought.

This study aims to develop tools and methods for investigating the role of the cerebellum in reinforcement learning, specifically its potential function in generating prediction error signals for unexpected rewards and forming new reward associations to adapt future expectations. To this end, an experimental system has been developed, which integrates head-fixed behavioral tracking, virtual reality, and auditory cues to assess predictive and associative learning in mice. In this setup, mice were placed on a running wheel that was synchronized with visually displayed virtual corridor, while their motor activity and licking behavior were recorded. Auditory cues predicting specific reward sizes were delivered through a speaker, and reward timing and size were precisely delivered and registered.

Mice were trained on one of two tasks: (1) a long-term association task, in which mice learned to run a predetermined distance in the virtual corridor. Here, an auditory cue at the start of the track predicted the reward size awaiting them at the end, and (2) a short-term association task, in which rewards followed an auditory cue that indicated reward size provided at a fixed delay.

In the long-term task, we successfully demonstrated the system's capability to facilitate learning in a challenging setup. With a carefully selected inter-trial interval (ITI), mice showed a clear anticipatory behavior by adjusting their running speed, movement percentage, stop durations and trial completion time based on reward size cues. In contrast, mice trained with shorter ITIs primarily relied on motivation derived from the previous trial's outcome rather than forming tone-specific associations. However, when the ITI was extended, these mice transitioned from reliance on prior trial outcome to mastering the task through genuine predictive learning, ultimately learning the correct association between the tone and the reward size. This result underscores the importance of precise ITI calibration for fostering robust associative learning in reinforcement learning tasks. This task examined the mice's ability to form predictions over a long-term association, managing sudden changes in setup to facilitate predictive learning- a process hypothesized in this study to be driven by the cerebellum.

In the short-term task, diverse learning strategies emerged across mice. All subjects initially acquired reward prediction, associating auditory cues with impending rewards. They then progressed to tone prediction, anticipating auditory cue timing based on time assessment of the ITIs. While some mice focused primarily on tone prediction, others advanced to tone discrimination, differentiating between cues for small and large rewards. A subset of mice required an increase in the ITI to transition from time-based tone prediction to tone discrimination, ultimately achieving reward discrimination with distinct licking patterns reflecting anticipation of the large rewards. This paradigm highlights the mice's ability to predict and respond to specific cues that signal an impending reward size, effectively demonstrating fast associative learning.

Together, these findings illustrate how mice adaptively respond to auditory and timing cues, transitioning from basic time-based predictions to more complex reward-based discriminations. This study provides insight into multiple learning pathways that underlie predictive and reward-based behaviors in reinforcement learning. The customized system developed here enables the exploration of both short- and long-term associations under various conditions, offering a comprehensive understanding of associative and predictive learning in mice. This groundwork paves the way for the next phase of research, where optogenetic manipulations at key cue times and cerebellar recordings will be employed. These steps aim to uncover the cerebellum's critical role in reinforcement learning by facilitating prediction mechanisms, ultimately shedding light on its enigmatic contributions to the fascinating learning processes.

## Introduction

Learning and Prediction: In the world of uncertainty in order to survive, the animal must make a prediction that will dictate how to act. The brain must establish internally generated predictions that can be compared against feedback from the external world in order to guide anticipatory actions and perceptions[2]. Prediction is a fundamental component that has driven the evolution of intelligence in animals and is central to most types of learning and brain function. Reinforcement learning is one type of learning, it is an adaptive process in which an animal utilizes its previous experience to improve the outcomes of future choices by receiving rewards[3,4]. Another type is supervised learning which its feedback about a system's performance is used to adjust internal parameters and thereby improve future performance[1,5].

The cerebellum and its anatomy: Doya has suggested that the cerebellum utilizes supervised learning that relies on sensory prediction errors (SPEs) for generating sensorimotor adaptation[1] and can facilitate motor learning [1,3,6,7,8,9]. The SPE occurs when an initial motor command is generated but the predicted sensory consequences do not match the observed values[10,11]. The climbing fibers (CFs) originating in the inferior olive (IO), are thought to carry those error signals that ultimately alter the activity of Purkinje cells (PC's) in the cerebellar cortex that target neurons in the cerebellar nuclei (CN) [6,8,12]. The CN neurons respond to the mismatch between expected and actual movements indicating prediction error is encoded also by these neurons [13]. The PCs, the sole output from the cerebellar cortex, inhibit the CN, which also receive excitatory inputs from collaterals of both CFs and mossy fibers (Fig. 1) [14].

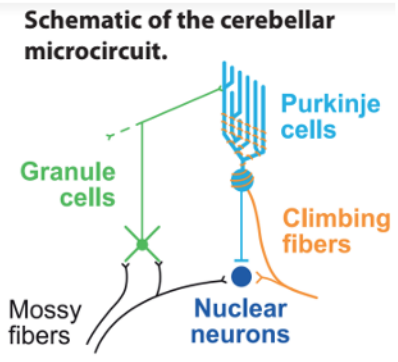



***Fig.1. Schematic of the cerebellar microcircuit.*** *Excitatory input arrives via the mossy fiber and the CFs pathways. Mossy fibers activate granule cells in the cerebellar cortex, which via the parallel fiber system excite Purkinje cells, which also receive excitatory input from a single CF. Purkinje cell (PC), the sole output of the cerebellar cortex, inhibit the CN, which also receive excitatory inputs from collaterals of both mossy fibers CFs. Adapted from [14].*

The SPE is thought to be conveyed by the CFs, that innervate neurons in the cerebellar cortex and nuclei [1], allowing modulation of cerebellar activity which contribute to cerebellar processing that

influences motor and cognitive function throughout the brain [14]. Moreover, it has been found that rewarding successful actions alone sufficed to enable the learning of a new sensorimotor mapping[10, 12]. Furthermore, reward delivery has been shown to facilitate the rate of motor adaptation[15] and the retention of the adapted state[16].

Basal ganglia (BG) and the ventral tegmental area (VTA): Doya suggested that the BG utilize reinforcement learning that relies on reward prediction errors (RPEs) for maximizing future reward[1]. Reward signals are carried by dopaminergic (DA) neurons in the VTA; these neurons increase their activity in response to unpredicted rewards, and decrease their activity in response to the absence of expected reward. This type of signals are thought to be a part of the reward prediction error signals that enable the BG to maximize future rewards by selecting the best actions [1,17,18]. It has been shown that if there is a cue that predicts an expected reward, its quantity or its probability, DA neurons in the VTA would shift their activity in time and respond to that cue while decreasing their response to the reward itself. The activity in DA neurons would be proportional to the reward size or its probability[1,19,18].
Carta et al. found direct projections from the deep cerebellar nuclei (CN) to the VTA[20,21]. Using anatomical tracing, they showed that axonal projections from the CN form synapses with dopaminergic neurons in the VTA. These data suggest that the cerebellum has the ability to transfer information to the VTA and specifically to the dopaminergic neurons which are thought to be an essential part of the brain reward system[1,20,21].

Rewards in the cerebellum: Recently, reward signals have been measured in the cerebellum in many of its components (e.g. CFs, Purkinje cells, granule cells, CN neurons, etc.) [2,14,12,22,23]. It has been found that reward context is encoded in CFs input to Purkinje cells. CFs are responsible for reward expectation, delivery and evaluation [23-24]. The granule cells also participate in expectation of reward, they develop reward predictive activity in both operant and Pavlovian conditioning as shown using calcium imaging in mice[25]. Unexpected rewards evoked CF activity which may be due to violation of expectations[24], and they fire in anticipation of reward[14]. They are possibly activated in order to eliminate previously created associations since the expectations were not fulfilled[24].
It has been shown that CF activity patterns are well-suited to drive learning by providing predictive instructional input that is consistent with an unsigned reinforcement learning signal but does not rely exclusively on motor errors as previously thought (e.g. SPE)[12,26]. Kostadinov and Hausser[14] suggested that the cerebellum may combine supervised learning with reinforcement learning to

optimize goal directed actions. In contrast, Kim et-al[27] claimed that the cerebellum does not support an architecture in which both processes operate in parallel but rather suggested that reward signals may serve as a gain on implicit adaptation or provide a distinct error signal for a second, independent implicit learning process. Reward signals were found also in the primate cerebellum. In monkeys, when visual cue predicted reward size, climbing fibers in the cerebellar flocculus had signaled upcoming reward size[28] .Likewise, when reward-related task timing is predictable, climbing fiber inputs signal these events predictively[29].In humans the cerebellum involved a reward anticipation and outcome processing as part of a predictive coding routine. Additionally, complex spikes from the CF reflect events that successfully predicted the upcoming rewards in a scalar manner that was proportional to the reward expectation and reported violation in expectation[14,12,23]. Based on these evidences, it is possible that the cerebellum contributes to the learning process and the performance of reinforcement learning [14,12,22,23,25,28,30] which raises the possibility that the roles of the cerebellum and the basal ganglia are not as distinct as initially thought. Specifically, it remains unclear whether the reward-related signals in the cerebellum carry motivational value and/or contextual information and what information is conveyed by the error signals of the cerebellum.

A different point of view of the cerebellum- The prediction: We know that the cerebellum operates predictively to support motor control, motor adaption and motor learning [1,3,6-9,31,32]and anticipates the expected outcome of motor commands in order to refine future movements[6- 9,31]. Additionally, most of the papers that recorded reward coding in the cerebellum in fact had some connection to reward prediction (expectation, violation and predictive coding) [13-12,22-25,28,30,33]. Moreover, time dimension reduction that represent the cerebellum as a clock that precisely encodes and predicts upcoming stimulation[11,12,24,25,30,33,34] or its violation [11,31,34,35] has been shown. It is possible that the link between the following distinct cerebellar functions: supervised learning to support motor learning, predictive reward signals that contributes to reinforcement learning and temporal difference (TD) coding, may be prediction. It is possible that the role of the cerebellum is to make predictions in order to facilitate different kinds of learning throughout the brain. The prediction can be made by error signals similar to SPE described by Doya[1] but in a more general manner that relies on prediction errors such as sensory, reward and time errors. These error signals activated during violation of expected outcome (from various components) and by that “waving a red flag” indicating the violation of estimation in order to eliminate previously created associations since

the expectations were not fulfilled. By generating internal prediction models and actively comparing anticipated and actual outcomes in order to reach a desired end state. Thus, the cerebellum may have a pivotal role in reinforcement learning by providing error prediction signals about rewards that do not match the anticipated outcome, and establishing new reward associations in order to make new expectations. In that case, the cerebellum may receive information about given rewards from several parallel pathways of inputs[14,21] make a prediction[14,12,23] and then send this information to the VTA[14,20,21] directly and indirectly. Thus, this pathway may enable the prediction of rewards and influence the generation of RPE in VTA and facilitate reinforcement learning.

In order to investigate the cerebellum's role in reinforcement learning and its potential contribution to generating prediction error signals, I developed a novel behavioral setup. This setup was designed to test the hypothesis that the cerebellum influences learning by providing predictive error signals. Two tasks were created for the study. In the first, a long-term association task, mice were trained to run a predetermined distance on a virtual reality track, guided by an auditory cue predictive of reward size given at the end of the run. This task assessed how changes in reward size affected motivation and performance. The second task was a short-term association task, where rewards were delivered after a fixed delay from an auditory cue, again predictive of reward size. These tasks allowed us to explore the behavioral characteristics of reinforcement learning and how the learning depends on task parameters such as ITI. This research lays the foundation for understanding the cerebellum's role in learning, prediction, and reinforcement processes.

## Research objectives

The central goal of my work has been to develop behavioral tasks that will enable me to identify and characterize the cerebellar role in reinforcement learning. These tasks will enable me to address the question of whether the cerebellum generates a reward prediction error signal having a motivational value similar to that described in the BG, or alternatively, the cerebellum generates a sensory prediction error that lacks motivational value to facilitate different types of learning. To achieve my goal, I have set 2 aims:

**Aim 1: Develop and design a novel system for studying reinforcement learning**

**Objective**: To create an innovative experimental system that enables the study of the cerebellum's role in reinforcement learning, with a particular focus on its hypothesized role in prediction. This system was designed to test both short-term and long-term associations by examining how variations in reward size influence the behavior of mice, providing insights into predictive and associative learning processes within the cerebellum.

**Rationale**: Traditional approaches often emphasize the cerebellum's role in motor control and temporal processing; however, evidence suggests it may also contribute to reinforcement learning, which, according to my hypothesis, occurs through the management of predictive signals. To explore this, a novel system was essential-one capable of systematically assessing how mice modify their behavior based on expected reward size. By developing a system that combines predictive cues with associative learning across different timescales, I aimed to directly assess the cerebellum's involvement in reinforcement learning tasks requiring flexible, prediction-based behaviors.

**Hypothesis**: The newly designed system would effectively capture behavioral modifications driven by predictive cues and varying reward sizes, enabling precise measurements of reinforcement learning dynamics in both short- and long-term contexts. By providing flexible, finely-tuned behavioral assessments, this system was expected to reveal distinct patterns of task engagement, particularly the way mice adjust their actions based on anticipated rewards. It was hypothesized that the system's design, incorporating predictive cues and tracking real-time responses, would successfully capture nuanced behavioral adaptations in response to both immediate (short-term) and delayed (long-term) rewards. This design, capable of supporting future optogenetic and neuronal recording studies, would offer a unique tool for isolating and measuring the cerebellum's

role in prediction-driven learning and behavior modification across varying timescales. the primary hypothesis was that this novel system would provide an essential platform for observing, analyzing, and manipulating prediction-based learning processes, facilitating a deeper exploration of cerebellar involvement in reinforcement learning.

**Approach**: The system incorporated a virtual reality setup in which head-fixed mice were trained to run along a virtual corridor, receiving auditory cues that predicted reward size (small vs. large). The setup included a running wheel with precise speed and distance tracking, three screens displaying a synchronized virtual reality (VR) corridor, and a reward delivery mechanism calibrated for accurate timing and amount. This comprehensive system enabled the measurement of key behavioral responses, including running speed, trial duration, and movement patterns in response to reward size. Additionally, it provided real-time tracking of licking patterns and timing, video recording for detailed movement analysis, ultrasonic sound capture, and dynamic adjustment capabilities- allowing for a robust examination of reinforcement learning and predictive behavior. Crucially, the system was designed with future studies in mind, accommodating neuronal recording (using Neuropixels probes) and optogenetic manipulation to explore neural mechanisms underlying reinforcement and predictive learning.

**Aim 2: Investigate learning dynamics in short- and long-term association tasks**

**Objective**: To characterize the behavioral adaptations of mice in both short- and long-term associative learning tasks in reinforcement learning paradigm, focusing on how they form and utilize predictive associations between auditory cues and reward timing or size. This aim seeks to comprehensively profile mice behavior across different learning phases and task complexities, highlighting both fast and gradual learning adaptations.

**Rationale**: By analyzing behavioral patterns across short- and long-term association tasks, this study aims to deepen understanding of animal learning processes, particularly in the context of predictive and associative learning mechanisms in reinforcement learning. This dual-task approach, which includes both immediate and delayed rewards, not only reveals the phases of learning and adaptation in mice but also paves the way for future studies using optogenetic manipulation and neuronal recordings to investigate the neural underpinnings of these behaviors.

**Hypothesis**: In the long-term association task, mice were expected to adjust multiple aspects of their performance, such as the speed of task completion, running velocity, movement percentage, and the number and duration of stops, based on anticipated reward size. It was hypothesized that when expecting a larger reward, mice would complete trials more quickly and with higher running velocity due to increased motivation, in line with incentive-based motivation theory[42]. This adaptive behavior would indicate the use of predictive cues to optimize task performance. Given the longer delay between cue and reward, as well as the added complexity of motor activity (running), it was expected that the optimal settings for this task would require careful adjustment, making it a more challenging association to form.

In contrast, the short-term association task was hypothesized to be learned more rapidly and with a higher success rate, as the short interval between the auditory cue and reward required minimal effort between events. This immediate reward structure was expected to enable quicker, stable predictive responses, allowing mice to form the association with relative ease. In this task, it was further hypothesized that mice would also learn to discriminate between tones, establishing all predictive elements of the task, including reward prediction and tone prediction. It was anticipated that different learning strategies might emerge, with variations in learning pace and accuracy influenced by the timing between trials, possibly affecting the mice's adaptive responses.

**Approach**: Two tasks were developed: In the **short-term association task**, mice learn to predict immediate rewards following an auditory cue, with a focus on rapid adaptation and cue-response learning. By recording and analyzing licking behaviors in specific time windows around the cue and reward, this task assesses the mice's ability to make quick associations and predictions based on short-term cues. In the **long-term association task**, mice learn to run a predetermined distance in a virtual environment to reach a reward, with an auditory cue signaling reward size at the trial's start. The extended time interval allows mice to adapt to a more complex, delayed reward structure, facilitating the exploration of cerebellar involvement in long-term predictive learning.

Both tasks were designed for later application of optogenetic manipulation to identify the specific contributions of cerebellar circuits. By examining the changes in behavioral metrics and identifying the distinct learning phases in each task, this study aims to delineate the cerebellum's role in reinforcement learning and prediction across different time scales and complexities.

## Methods

*Animals:* All procedures were approved by the Bar-Ilan University Animal Care and Use Committee and performed in accordance with the National Institutes of Health (NIH) guidelines. Adult wild-type (male and female) mice (C57BL/6) aged >8 weeks were used; 13 mice for the short term association task (10 females, 3 males), and 8 mice for the long term association task (all males). All animals were maintained on a 12/12 h light/dark cycle and had ad libitum access to food until the beginning of experiments that were performed during the light phase.

*The surgical procedure:* All the mice underwent head-plate surgery (shown in Fig.2.). Mice were anaesthetized with 1.5-2% isoflurane and injected with ketamine and xylazine (12.5ml : kg−1, subcutaneous) for analgesia, their head was shaved and they were placed in a stereotaxic apparatus on a heat pad. Their scalp was cleaned with Povidone-iodine and alcohol, followed by exposure and cleaning of the skull. The incision was sealed with tissue glue, and a custom head plate was securely attached to the frontal bone using strong adhesive [24] . Grounding was established by either drilling a screw into the skull or placing a metal wire on it.

***Fig.2. Head-plate attachment****. The image shows the head-plate securely affixed to the exposed skull of the mouse. A screw is visible, providing grounding, and the surrounding tissue is sealed with adhesive.*

*Statistical analysis:* All analyses were conducted following an assessment of normality. For small sample sizes, the Shapiro-Wilk test was used, while the Kolmogorov-Smirnov test was applied for larger samples. Normally distributed data were analyzed with t test and ANOVA. Non-normally distributed data were analyzed using permutation tests for means and bootstrapping methods to calculate the standard error of the mean (SEM), standard deviation (STD), and medians. Variability across animals was assessed using Levene's test for homogeneity of variances, ensuring a robust evaluation of the equality of variances in licking behavior across animals[36]. The means of sessions were combined only after confirming consistent behavioral patterns across all animals and ensuring no significant variability in behavior between animals. This verification was crucial to justify the aggregation of session means from different animals and from the animal itself.

*Data analysis:* Data from the rotary encoder and licking activity were collected and processed through a custom-designed system. For the long-term association task, ViRMEn software (Princeton

University) was adapted and fine-tuned to meet task-specific requirements. For the short-term association task, a custom "Behavioral system" was developed in C++ to manage analog and digital signals, enabling specific trials and sessions. An ESP32-WROOM-32 chip integrated with the Arduino IDE software provided microsecond-accurate calculations for speed and distance. Data analyses were conducted using MATLAB (R2018b, MathWorks Inc.), Python (Van Rossum & Drake, 2009), and C++.

## Results

The associations and predictions that drive the VTA to act in a specific manner may depend on predictive instructional signals originating from the cerebellum. To explore this, a task that integrates reinforcement learning with association and prediction processes is essential. Therefore, I had to developed a novel system to address this question. The system incorporates measurable elements of reinforcement learning, which are assessed through mouse motivation and motor activity.

The tasks: The role of the cerebellum in contributing to reinforcement learning through prediction can be tested within two distinct time frames: a short-term window and a long-term window, each probing different aspects of predictive behavior. Mice were trained on two tasks: (1) a long-term association task, where they learned to run a predetermined distance in a virtual reality environment. At the start of the track, an auditory cue signals the size of the reward that will be delivered at the end (Fig.3. left); and (2) a short-term association task, where the reward is delivered after a set delay following an auditory cue, with the cues similarly predicting reward size (Fig.3. right).


***Fig.3. Schematic representation of behavioral tasks for short- and long-term association experiments.*** *The schematic illustrates two distinct tasks. In the short-term association task (left), an auditory cue (blue or yellow musical note) predicts the size of a reward (blue or yellow water droplet), which is delivered after a fixed delay. In the long-term association task (right), an auditory cue similarly predicts the reward size, but the reward is delivered only after the mouse completes a predetermined run in a virtual reality environment.*

**The system's design:** The mice were head fixed (see methods) and placed on the running wheel, allowing them to run while the running speed and distance were calculated from a rotary encoder placed on the axis of the wheel. The licking activity was obtained using a lickometer attached to the licking port. In front of the wheel 3 screens were placed that can display the visual flow of a patterned virtual corridor synchronously with the running speed of the mice. The mice had received an auditory cue and after completing the task, mice were rewarded. Throughout the sessions, the mice behavior was recorded via a digital camera and taken for further analysis using DeepLabCut[37], and the associated sounds were also recorded.

The running wheel: The running wheel was constructed following the KineMouse protocol as outlined in[38]. Running speed and distance were measured, and then converted into relevant units. A separate system utilizing Arduino calculated the distance traveled and the speed of the mice. The data was then transmitted to the appropriate software, including ViRMEn software (Princeton University) for virtual reality flow, as well as a custom-developed software for behavioral analysis. All the measured variables were transmitted to a DAQ system (N114, National Instruments) and written to the same file to ensure synchrony. The wheel was cleaned and maintained daily.

Screens and Settings: Three Samsung T45F 24-inch screens were placed 26 cm in front of the mice, displaying a visual flow of a patterned virtual pathway synchronized with the running speed and distance traveled by the mice. This setup was powered and programmed using ViRMEn software (Princeton University).

Virtual reality design: The virtual reality pathway was designed with specific colors to align with the visual capabilities of mice. Research shows that mice use color information to guide behavior[39–41], relying on dichromatic vision through short (S) and middle (M) wavelength opsins, with peak sensitivities at 370 nm (UV) and 510 nm (green)[42] .Notably, mice can discriminate colors primarily in their upper visual field, where retinal neural circuits are adapted for color detection, supporting survival-oriented tasks such as predator detection [43].Furthermore, studies have shown that mice can distinguish colors across a broader range of the visual field than previously expected based on cone-opsin distribution, with specific wavelength sensitivities affecting luminance perception[44]. By using distinct colors along the corridor walls, this design enhances visual cues within the mouse's perceptual range, facilitating spatial and behavioral learning in the virtual environment (Fig.4.). Moreover, to ensure the synchronization of the mice's speed with the virtual reality flow, a 10 ms buffer was implemented. The steps taken by the mice were analyzed within this buffer, which

subsequently sent signals to the ViRMEn system for further analysis. This setup aids mice in distinguishing distance traveled and orienting themselves within the corridor.
The virtual reality environment included three distinct areas: (1) the ITI Room - a dark room between trials lasting 5, 12, 15, 20 seconds (depending on the animal and paradigm); (2) the Running Room - where the predictive reward auditory cue was given at the start, and visual indicators on the walls showed distance traveled; and (3) the Reward Room - where the mice received their reward.

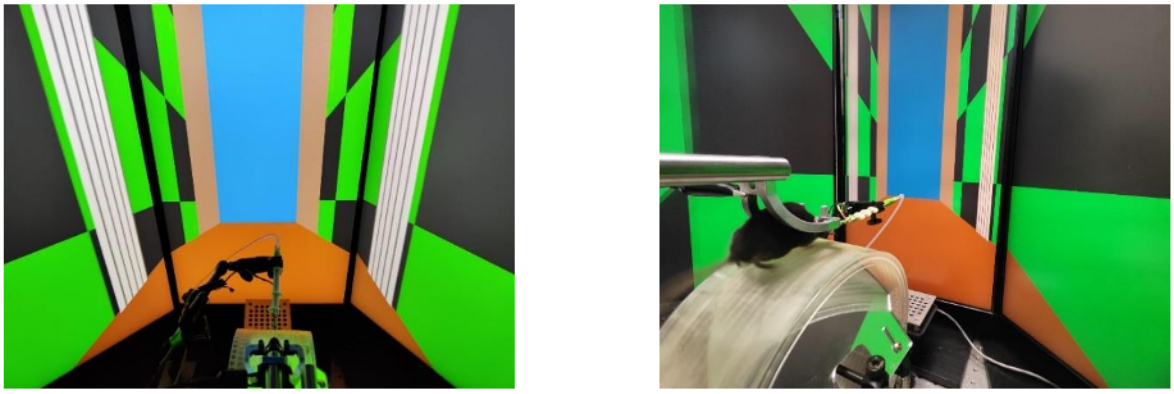


***Fig.4. Virtual reality setup**. The setup features a three-screen display presenting a visually guided pathway with black-and-green checkered walls that move in real-time based on the mouse's speed, creating an immersive environment. The running wheel allows the mouse to navigate the virtual corridor, restricted to forward or backward movement, toward a blue reward room at the end of the track. A reward delivery system is positioned in front of the mouse and delivers the liquid reward at the designated reward room. The speaker that provides auditory cues during the experiment is placed in front of the mouse.*

Reward Delivery: Rewards were administered via a 2-Way Normally Closed Isolation Valves (24V, NResearch), which were programmed to open and close at predefined durations. Small rewards consisted of 1 microliter of liquid per trial, with valve opening times between 6-8 ms, while larger rewards of 6 microliters per trial required valve opening times of 15-25 ms. Valve opening times was calibrated weekly for precision and maintained daily by thorough cleaning with 70% ethanol, double-sterilized water, and air pressure. The reward solution consisted of water with 5% sucrose, and mice were water-restricted to enhance the reward system's effectiveness.

Licking data acquisition and analysis: Licking activity was recorded using a licking port equipped with an electrical circuit. When the animal licked the port, it closed the circuit, which ran from the animal's tongue through the reward delivery needle and back to the head-fixed system, creating a

closed loop that included the headplate. An LED indicator was installed to monitor the functionality of the circuit, which was checked before each session to ensure proper operation. The design was constructed by Deuteron Technologies LTD.

Auditory cues: Sounds were delivered 16 cm away from the mice using an ENV-224BM speaker (Med Associates), capable of producing precise frequencies. The frequencies used were 8 kHz and 14 kHz, which fall within the hearing range of both humans and mice, and are particularly sensitive for the mice [45,46]. The tones were played for either 200 ms or 1 second, during the short- and long-term association tasks, respectively. Audio recordings were captured at a sampling rate of 144,000 Hz using an Ultramic UM200K ultrasonic microphone (DODOTRONIC).

Video Data Acquisition: Animal monitoring and positional tracking were conducted using a GS3-U3 FLIR camera, along with FlyCap digital camera software, Spinnaker SDK and Bonsai software[47]. Infrared (IR) illumination was provided by an SSC-IR-850W Advanced Security light. Video was recorded at 120 frames per second (fps) to capture each mouse's movements and ensure a smooth flow of action in order to later analyze it. The recordings were made with a resolution of 576 x 488 pixels. The recording allowing both top and bottom views of the animal's body and paws (Fig.5.). The visual recordings involved capturing 3D information using a mirroring of the animal body and tracking the animal's movements at high speed. The captured video data was analyzed using DeepLabCut[37].

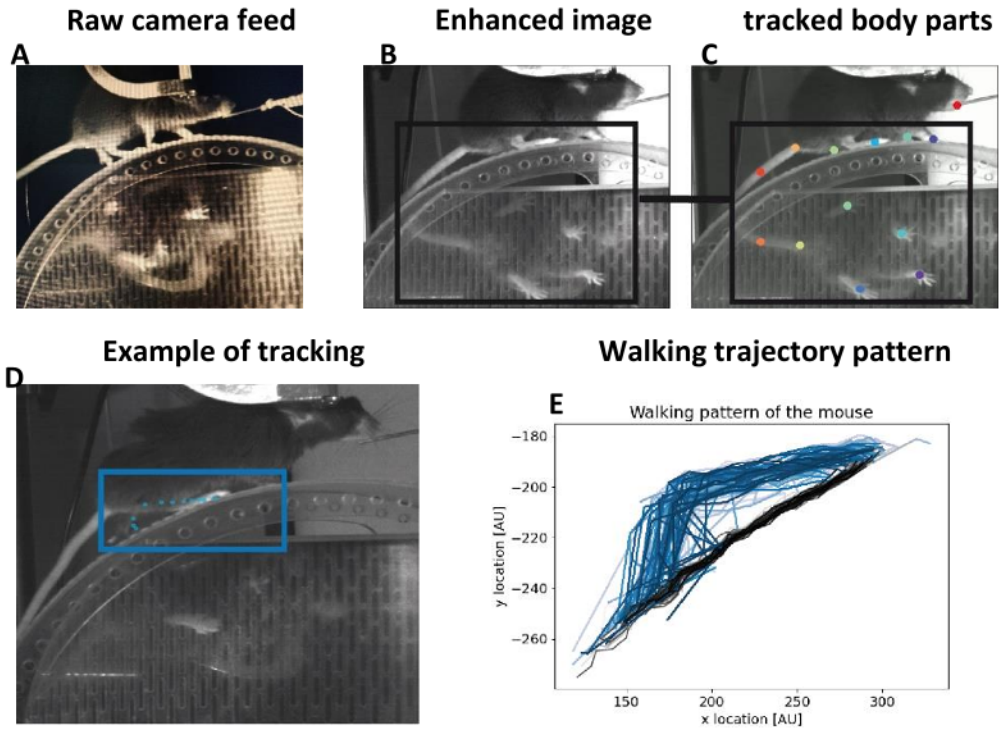


***Fig.5. Visual tracking and walking pattern analysis of mice.***
*Top left: (A) Raw camera feed of the mouse showing the experimental setup. (B) Enhanced image with infrared (IR) adjustments to reduce noise and increase clarity of movement. (C) Processed image showing tracked body parts identified by DeepLabCut after training, with each body part marked in a distinct color. (D) Example of tracking the right hind leg using DeepLabCut, with individual frames highlighted to show movement across time. (E) Walking trajectory pattern, where blue lines represent forward movement and black lines represent backward movement, illustrating the mouse's overall movement path during trials.*

The software was adjusted to identify the specific body parts of the mice within the experimental environment, utilizing Napari for advanced labeling (Fig.5.C). This labeling process facilitated the training and validation of a new artificial neural network in DeepLabCut, tailored to meet the requirements of the tasks. Training the algorithm and fine tuning it enabled detailed behavioral analysis of the mice (Fig.5.D, E). This dual-view setup allows for precise, high-speed tracking of both body position and paw movements, enabling detailed analysis of locomotor behavior.

The system settings: Figure 6 presents a comprehensive view of the experimental system used for mouse behavioral experiments, including components for virtual reality navigation, neuronal recording, and optogenetic manipulation. The setup comprises multiple interconnected components, detailed across different views (Fig.6.). The head holder secures the mouse in position while the reward delivery system is aligned with the licking port, providing liquid rewards based on task performance. The speaker delivers auditory cues as part of the experimental design. The virtual reality environment, displayed on three screens, allows the mouse to navigate through the VR, while the camera monitors the behavior during task execution. The running wheel enables the mouse to mov through the virtual pathway in a controlled manner. The microscope is positioned to allow close observation for the optogenetical manipulation and neurological recording. The IR light enables low-light imaging, while the microphone records any sounds, providing additional sensory information for analysis. The DAQ system manages data flow, while the ESP-32 chip operates the various devices within the system. The entire setup is arranged on a stable platform to minimize vibrations, ensuring precision in both behavioral and neural measurements. This integrated system enables real-time tracking of mouse behavior in a virtual reality environment, precise reward delivery, optogenetic manipulation, and high-resolution neuronal recordings.

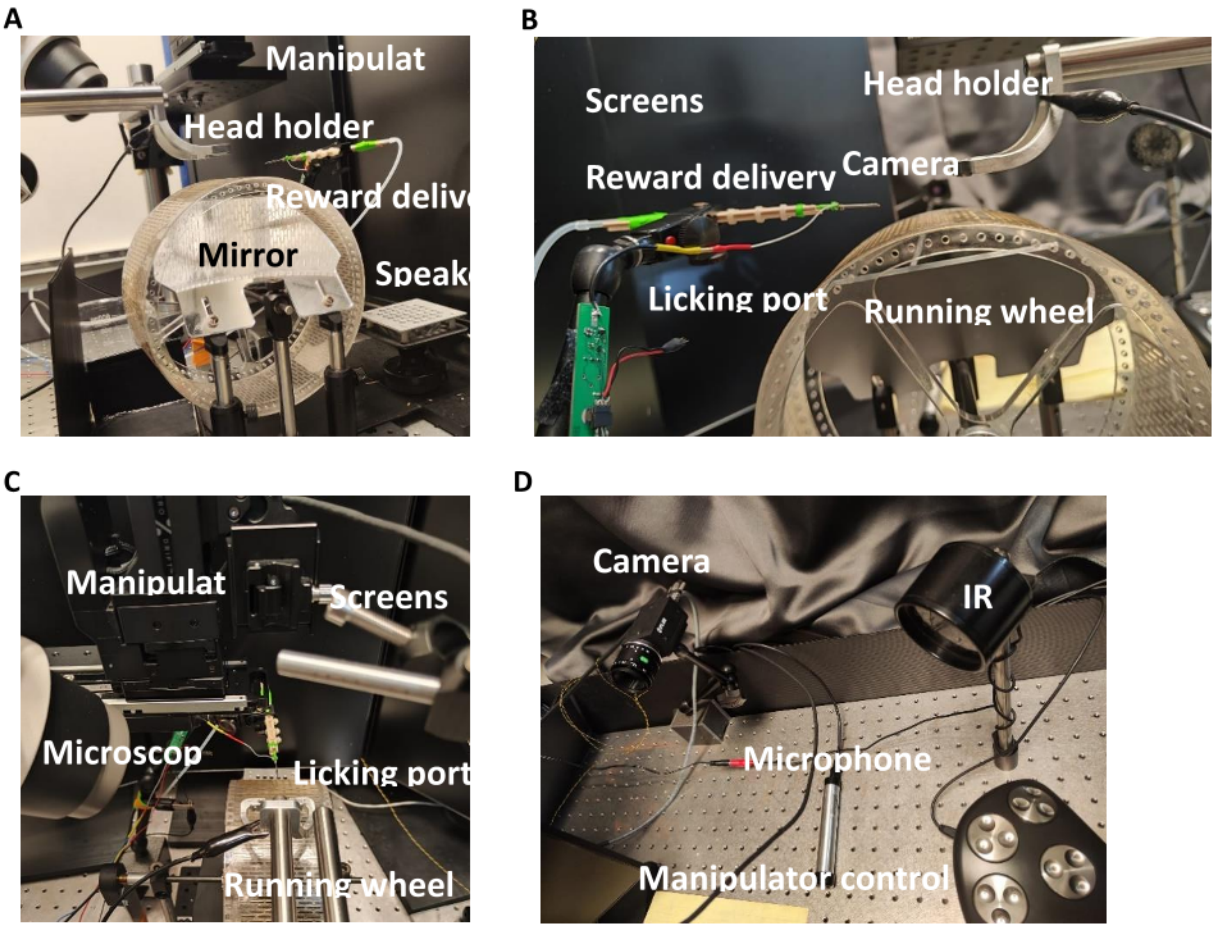
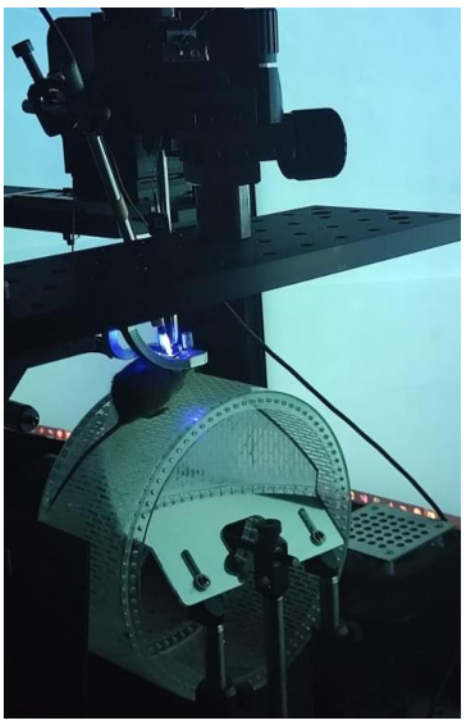


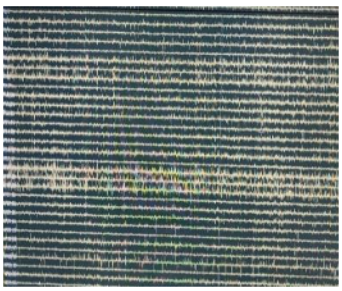

***Fig.6. Components of the novel experimental system.*** *(A) A picture of the manipulator, head holder, reward delivery system, mirror, and speaker. (B) A picture of the screens of the virtual reality display, reward delivery, licking port, running wheel, and camera. (C) Top view of the manipulator, microscope, licking port, and running wheel, offering a perspective on how each component is spatially arranged. (D) The position of the infrared (IR) light source, the microphone, and the manipulators controller. (E) A picture of the complete setup capturing the data acquisition (DAQ) system, ESP-32 system-on-chip (SoC), power sources, and screens, cameras, and IR lighting.*

Optogenetic Manipulation and System Adjustments: The system is equipped to support precise optogenetic manipulation, essential for future studies exploring the cerebellum's role in reinforcement learning (Fig.7.). Using a Plexon LED driver (LD-1) with a PlexBright blue (465 nm) LED, optogenetic stimulation can be applied with a light output of 3.5 mW to 6 mW at the fiber tip,

as measured with a Thorlabs PM100D power meter. An optogenetic fiber (OPT 200/230HP, Plexon) is carefully positioned above the dura near simplex lobule in the cerebellar cortex with a Sensapex MicroMp-4 manipulator and a Zeiss Stemi-305 microscope for precise targeting. Optogenetic stimulation is delivered at 10 Hz, chosen to effectively activate Purkinje cells (PCs)[48] in adult transgenic mice expressing channelrhodopsin-2 (ChR2) in the cerebellar cortex through the PCP2 line, where Cre-recombinase is expressed[49–53]. This functionality enables the system to isolate the cerebellum's predictive role in reinforcement learning through targeted optogenetic inhibition of its output pathways during specific learning phases.

Electrophysiological recordings can be carried out using a NI PXIe 1071 system (National Instruments) with Neuropixels 1.0 probes (imec). Each probe is mounted on a four-dimensional micromanipulator (Sensapex MicroMp-4 guided by a Zeiss Stemi-305 microscope), allowing precise adjustments to the angle and position, followed by controlled lowering of the probes into the brain (Fig.7.). Neuronal activity can be recorded from the distal 384 channels at a 30 kHz sampling rate using SpikeGLX software (http://billkarsh.github.io/SpikeGLX/) with a 300 Hz high-pass filter. Spike sorting is automatically performed with Kilosort 2.5 (https://github.com/MouseLand/Kilosort2.5) and manually curated using the 'Phy' graphical user interface (https://github.com/kwikteam/phy). Only well-isolated single neurons will be included for further analysis in Matlab (R2018b, MathWorks Inc).





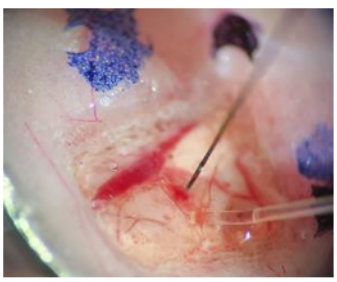


***Fig.7. Example system setup for optogenetic manipulation and Neuronal recording in future studies.*** *(left), Blue light is directed onto the cerebellar lobule simplex of the mouse brain. (Top right) neuronal recording captured using a Neuropixel probe positioned in the cerebellum. (Bottom right) An image of the Neuropixel probe is shown, with the LED targeting lobule simplex, demonstrating the integration of optogenetic stimulation and high-resolution neural recording.*

Training protocol for running: The system allows fine-tuning of each mouse's position on the running wheel, adjusting height, distance from the wheel center, and body angle. Mice initially receive random rewards to get accustomed to the reward needle, learning to lick upon hearing a click that indicated reward delivery. Early in training, mice are placed toward the back of the wheel and are gradually moved closer to the center as they adapt. For pre-training mice can receive rewards after they propagate a predetermined distance (2-10 cm) on the wheel.

## long-term association task

Behavioral Task: To investigate the cerebellum's role in reinforcement learning and prediction during the learning process, I designed a task with three key components: (1) movement assignment, (2) outcome prediction, and (3) reinforcement learning. In the first stage, head-fixed mice (as detailed in the surgical procedure and shown in Fig.2.) were trained to complete a 50 cm track on a running wheel, starting with rewarded distances of 2 cm that increased by 2-10 cm every 10-20 successful trials (Fig.8.). Upon successful completion of the track, mice received a liquid reward, with the goal of associating running with the reward until achieving continuous, uninterrupted 50 cm runs within approximately seven sessions. The next phase involves training the mice to run on a virtual 50 cm pathway corridor, which they practice for one week.

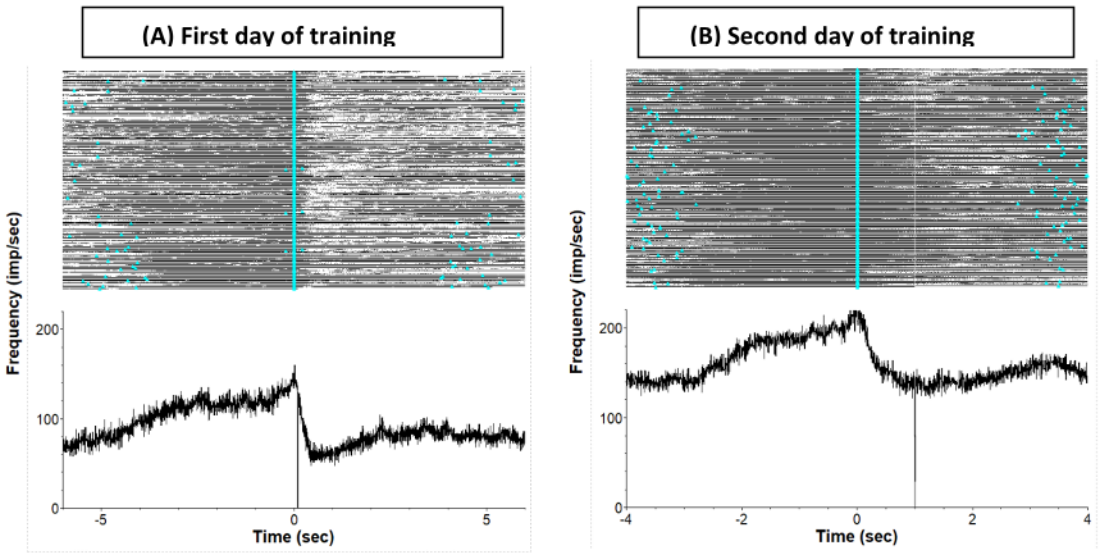


***Fig.8. Mice learned to run a predetermined distance of 50 cm over two training days.*** *A & b, Shown is the average speed of a mouse on the first training day (A) and the second training day (B). The black dots represent the rotary encoder bits (each bit is equivalent to a distance of 5.85 mm). Trials are presented from top to bottom and aligned to reward delivery (light blue). On the first training day it took the mouse 58 minutes to complete 200 trials whereas on the second training it completed 200 trials in 19 minutes. On both training days the mouse reduced its speed after it received the reward.*

In the advanced phase of the task, an auditory cue was introduced before each trial to signal the mice about the reward size (1 µl for a small reward or 6 µl for a large reward of water with 5% sucrose). The continuous tones, at 8 kHz and 14 kHz respectively, lasted 1 second and indicated the size of the reward awaiting the mice at the end of the track. At this stage, the task evolved from simple association to prediction, as the mice learned to anticipate specific rewards based on the auditory cues. The mice were required to run a predetermined distance of 50 cm on the treadmill to reach the reward room, where the reward was delivered through a precisely timed valve system. Visual feedback from the virtual environment's walls indicated their position along the track, enhancing their spatial orientation. Following reward delivery, a darkened ITI lasting 5, 12, 15, or 20 seconds began, after which the next trial reset to the starting position with the selected auditory cue played again. This setup allowed for detailed analysis of how reward size influenced the mice's motivation, as reflected in their running behavior. Improvements were observed as training progressed, with the mice increasing their running speed and completing sessions more quickly. Notably, the mice exhibited anticipatory behavior, accelerating in response to the auditory cues that predicted larger rewards.

The results varied based on the ITI duration. When the ITI was short (5 seconds, N=4), mice did not associate the auditory cue with the size of the upcoming reward, as trial durations showed no significant differences between the two tones (Fig.9, two bars on the right). Instead, their performance and motivation to complete the task faster were influenced by the outcome of previous trials. Statistical analyses confirmed this observation: a Kruskal-Wallis H-test (KW) revealed significant differences across reward conditions ($p=0.0001$). Post hoc analysis showed that mice ran faster after receiving a large reward in the prior trial (pink) and slower after receiving a small reward (light blue, $*p=0.0237$). This effect was amplified when trials of the same reward size occurred consecutively. For instance, trial durations were shortest when the last few rewards were

consistently large and longest when they were consistently small (Fig.9, six bars on the left). Specifically, trials following one or more small rewards resulted in significantly longer durations (p=0.0018 for one prior small reward, p=0.0027 for two prior small rewards), while trials following one or more large rewards were consistently shorter, even compared to a single small reward. Overall, these findings suggest that the mice's behavior was guided by the cumulative evaluation of prior rewards rather than by the auditory cues predicting reward size.

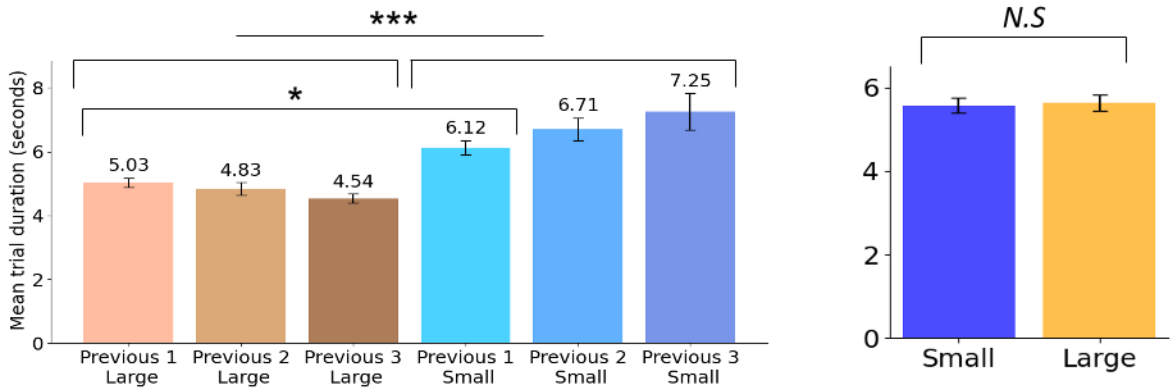



***Fig.9. Previous trials reward size influence the current trial performance during short ITIs of 5 s.*** *(Left) Trial durations were averaged for 1,2 and 3 previous trials with a similar outcome (large or small reward). The bars shows mean trial durations based on the previous reward size across different sequences, with error bars representing MSE, and the values written above. Trials following large rewards (left three bars) are shorter in duration compared to those following small rewards (fourth to sixth bars). A Kruskal-Wallis H-test indicated significant differences in trial durations across reward conditions (***p = 0.0001). Post hoc comparisons revealed that trials following small rewards were significantly longer than those following large rewards (*p = 0.0237), with durations increasing progressively after consecutive small rewards. Trials following large rewards did not differ significantly based on subsequent reward sizes (N.S. p = 0.2677).* Shown on the right are the average durations of the trials based on the tone played. *Small reward trials are shown in blue and large reward trials are shown in orange*.

I then increased the ITI to 15 seconds to assess whether this change in the paradigm was sufficient for eliminating the incorrect association between the outcome of the previous trial and the animal's running speed. The extended ITI successfully eliminated the erroneous association (Fig.10. left); however, the mice were unable to distinguish between the tones (N=4, Fig.10. right). The previously observed differences in trial durations following consecutive large or small rewards, completely disappeared with the longer ITI. This outcome suggests that a sufficiently long ITI prevents the mice from relying on prior trials to guide their motivation. Instead, it may enable the

mice to focus on each individual trial, allowing them to better associate the auditory cue with the reward magnitude they will receive at the end of the trial.

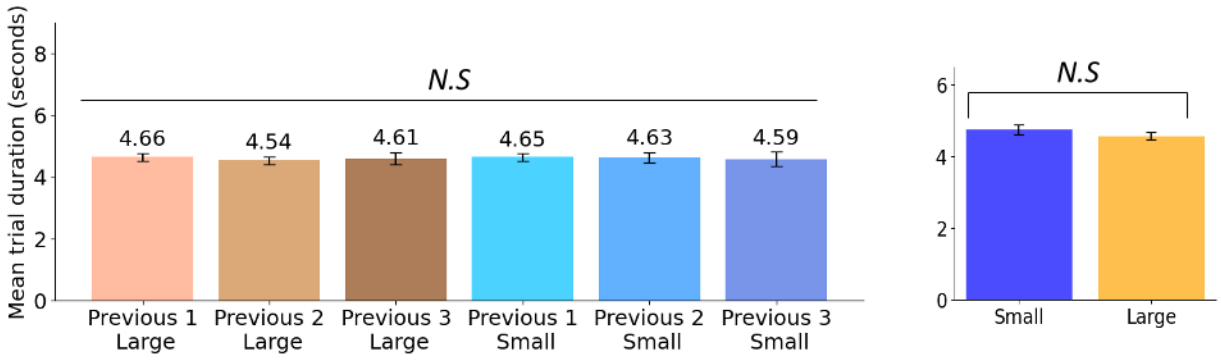



***Fig.10. The association between the previous trial outcome and running speed was eliminated by using an ITI of 15 s.*** *The average trial duration is written above each category on the left (similar to figure 9). Nonetheless, the average duration of the trials separated according to the played tones remained similar (2 bars on the right; KW p = 0.4626).*

Finding the balance in ITI duration is crucial for effective task learning. If the ITI is too short, mice tend to perceive the current trial as a continuation of the previous one, with their behavior influenced by motivational factors linked to prior rewards. Conversely, an ITI that is too long can result in reduced motivation, as the mice may lose interest in completing the task. An optimally balanced ITI helps the mice treat each trial independently, allowing them to focus on associating the auditory cue at the beginning of the virtual corridor with the reward at the end, rather than relying on outcomes from previous trials.

When training naïve mice, determining the optimal ITI is essential and may vary between individuals. For example, among three animals trained with a 12-second ITI, two initially associated their performance with the outcomes of previous trials, although one eventually overcame this association, after shifting to longer ITI. The third animal, however, immediately learned to associate the tones with reward size without relying on prior trials. This mouse exemplifies successful task acquisition during a 12 s ITI, as it formed a direct link between the tone and the reward at the end of the trial (Fig.11.). The analysis was therefore focused on this animal as a model for efficient learning.

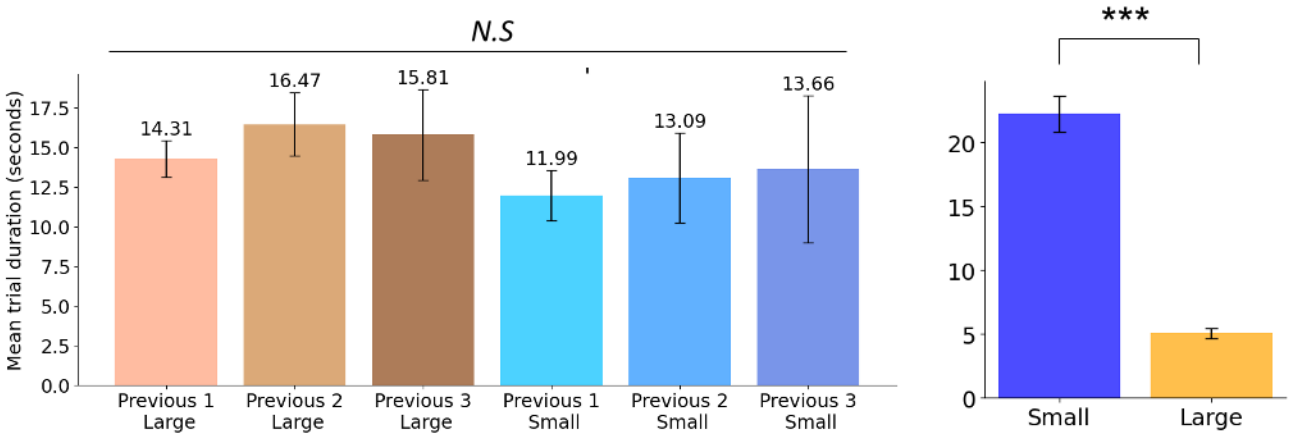



***Fig.11. A naïve mouse learned to associate the tones with reward size when trained on ITI = 12 s.*** *An example of one animal that managed to learn the task with optimal ITI (12 seconds) and form the correct association between tone and trial outcome. The average duration of the trials depended on the tone played was significantly different (2 bars on the right; The KW test revealed highly significant differences in trial durations ,*** $p < 0.0001$*). The large rewarded trials were faster compared to small rewarded trials. Pairwise comparisons showed no significant differences between different previous large reward trials nor between different previous small reward trials.*

The learning exhibited by the mouse in this task resembles an "insight learning" process, that is, the mouse appears to integrate all task parameters and within an instant start distinguishing between small and large reward trials. Key parameters, including trial duration (Fig.12.), average velocity per trial (Fig.13.), movement percentage (Fig.15.), reveal this rapid learning. When examining trial duration, Figure 12 shows that the mouse rapidly distinguished between large rewarded trials and smaller ones. For example, focusing on the sessions where learning first became evident (session 11), all measured parameters showed significant differences. A distinct pattern in trial duration based on reward size emerged, with small-reward trials averaging 35.02 seconds duration, while large-reward trials were significantly shorter at an average of 3.89 seconds (Fig.12.A). Interestingly, the size of the previous trial's reward (irrespective of size) does not impact the current trial's duration. For current large-reward trials, the duration averages 3.82 seconds after a small previous trial and 3.98 seconds after a large one. Similarly, for current small-reward trials, the average duration following a small reward is 34.35 seconds, compared to 35.54 seconds following a large reward. Thus, the trial duration depends on the expected reward size of the current trial as signaled by the different tones. Notably, the same effect was observed in the previous unlearned session (session 10); trials with a small reward have an average duration of 3.64 seconds, while trials with a large reward average 2.60 seconds; this difference is not statistically significant. As well the

previous trial sizes have no impact, with trials following a small reward averaging 3.12 seconds and those following a large reward averaging 3.13 seconds. The primary factor facilitating learning is the mice's ability to treat the current trial independently from previous ones, resulting after achieving the optimal individual's ITI.

It is possible to consolidate data from all sessions, as they exhibit similar patterns and significant differences across various metrics (Fig.12.B), including trial duration (and later also velocity without stops, movement percentage, and stop duration). The results of the statistical tests reveal marked differences between learned and unlearned sessions in both trial duration (after ensuring normality). For learned sessions, a highly significant difference in trial duration is evident as presented in Fig.12.D (t-statistic of -11.90, p-value < 0.0001). This indicates that trial durations are meaningfully different, highlighting a robust effect in these sessions. In contrast, unlearned sessions show no significant difference in trial duration seen in Fig.12.C ( t-statistic of 4.78 , p-value approaching 1 (0.99)). This suggests that trial durations remain similar regardless of conditions, indicating an absence of learned differentiation in trial timing.

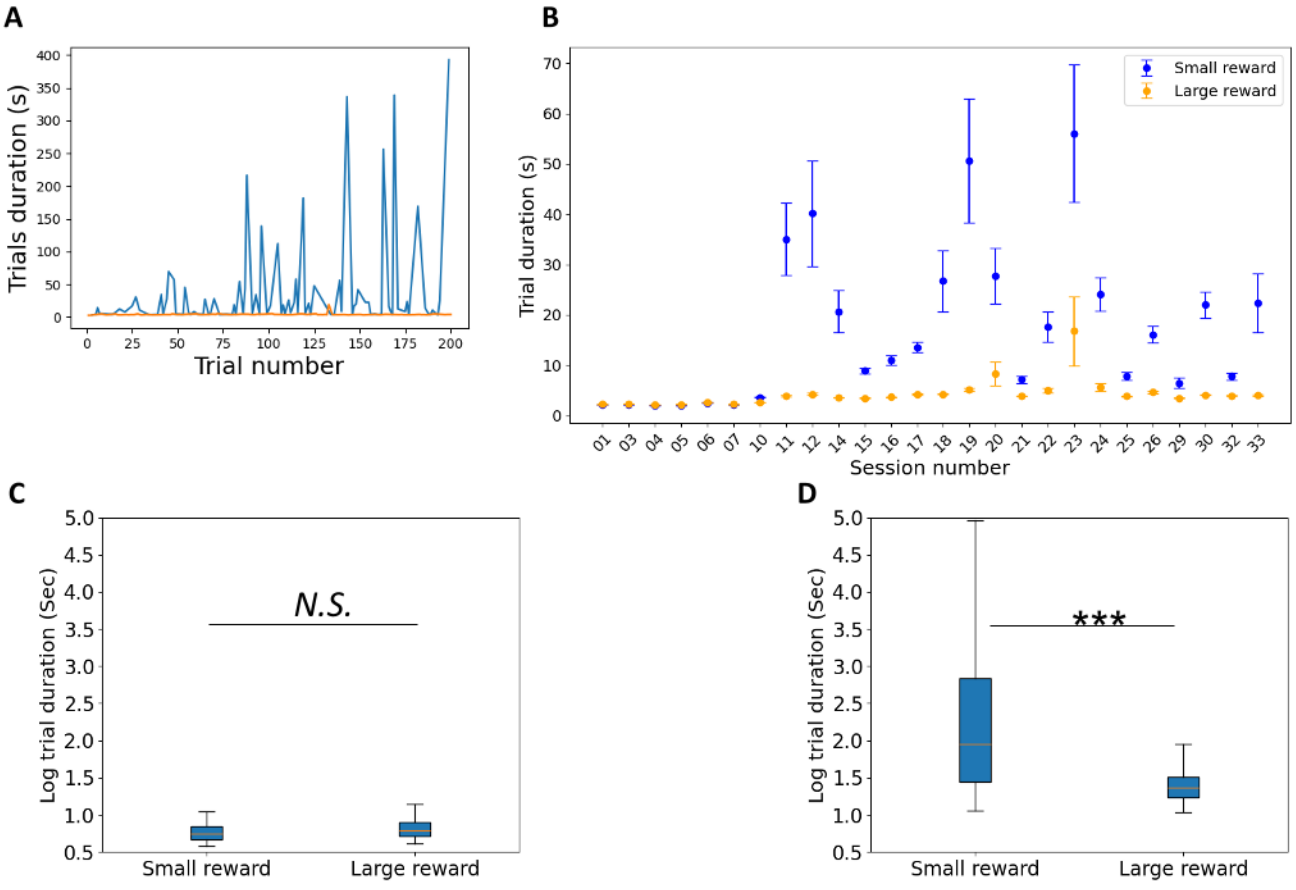


***Fig.12. Large-reward trials are completed faster than small-reward trials.*** *(A) Sessions example of the first session learned - 11. Here shown are small (blue) and large (orange) rewarded trial durations across 200 trials. (B) The trial duration (seconds) is plotted against session number for small rewarded trials (blue) and large rewarded trials (orange). Each data point represents the mean trial duration for a session, with error bars indicating SEM. (C,D) Those figures compares trial durations for small (left) and large (right) rewards in unlearned (C) and learned (D) sessions in a log transformation form. In unlearned sessions (C) trial durations are similar for both reward sizes. Statistical analysis confirms no significant difference (N.S. t-statistic = 4.78, p = 0.99), with mean durations of 2.18 s (small) and 2.29 s (large). In learned sessions (D), a clear difference emerges: small-reward trials are significantly longer than large-reward trials (t-statistic = -11.90, ***p < 0.0001), with mean durations of 22.21 s (small) and 5.05 s (large). This reflects a behavioral shift, with mice allocating more time to small-reward trials post-learning.*

The differences in trial duration result from variations in velocity, the number of stops, and stop durations, which contributing to the movement percentage. Additionally, the correlation between trial duration and mice velocity during the task makes examining velocity differences particularly intriguing. The mouse altered its velocities as it learned to distinguish between the rewarded trials. Initially, the velocities were the same, but after learning, the mouse differentiated between the trials. Interestingly, both velocities decreased following the learning process (Fig.13.B). It is intriguing to examine the differences in velocity between trials with large and small rewards across the virtual corridor track. In Figure 13.A, the graph illustrates the mean velocity (cm/s) of the mice as they progress along the corridor (0 to 50 cm) in session 11. At the beginning of the trials, the velocity remains relatively consistent across both reward conditions, indicating that the mice are initially running at a steady pace regardless of the reward size (we see increment in velocity indicating that running starts from the trial start). However, as the trials progress, notable differences in velocity emerge. The graph shows that, after the initial phase, the mice exhibit a significantly higher velocity in large reward trials compared to small reward trials. This disparity may be attributed to the auditory cue presented at the start of each session (lasting 1 second) as the mice begin to run. As they learn to associate the tone with the larger reward, their behavior and velocity adjust accordingly, leading to a more rapid pace for the large reward trials. The variability in the shaded regions also indicates that the performance of the mice can vary, further highlighting the adaptability in their responses based on the anticipated reward. Notably, the mean and median velocities for trials with large rewards are significantly higher compared to those with small rewards. Additionally, as illustrated in Figure 13 C and D, a permutation test performed on velocity revealed highly significant differences between the conditions while the velocity difference was -

5.430, p < 0.0001 for learned sessions (Fig.13.D) and was not significant for the unlearned (Fig.13. C. velocity diffrences: 0.8900 , p = 0.276). This suggests that learning significantly influences the mice's performance, leading to enhanced behavior in response to anticipating reward size with prediction factor.

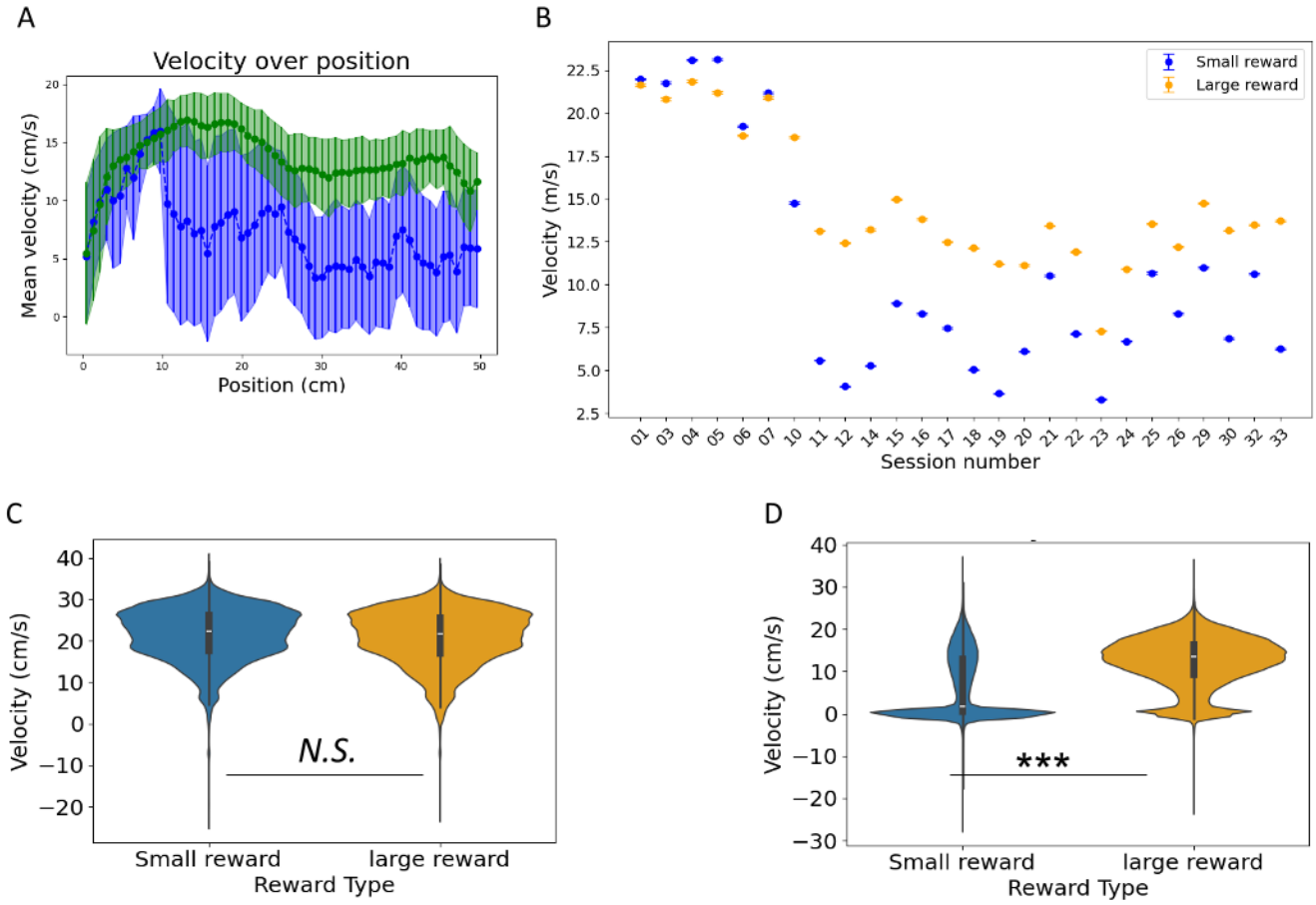



***Fig.13. Increased speed after learning in large rewarded trials compared to small.*** *(A) Sessions example of session 11. This plot shows the mean velocity (in cm/s) across different spatial positions (in cm) for trials associated with small rewards (blue) and large rewards (green) without including the stops (zero cm\s). Mean Velocity and error range: The green line represents the mean velocity for large-reward trials, with an error range shaded in light green indicating the standard deviation. Similarly, the blue line indicates the mean velocity for small-reward trials, with an error range shaded in light blue representing the standard deviation. (B) The trial velocity (meter/seconds) is plotted against session number for small rewarded trials (blue) and large rewarded trials (orange). Each data point represents the mean trial velocity for a session. Violin plots (C, D) compares velocity distributions for small and large rewards in unlearned (C) and learned (D) sessions. In unlearned sessions (C), velocities are similar for small (21.64 cm/s) and large rewards (20.78 cm/s), with no significant difference. While in learned sessions (D),mice reduced velocity for small rewards ( 6.33 cm/s) while maintaining higher speeds for large rewards (12.27 cm/s), resulting in a significant difference (***velocity difference = -5.43,permutation p value < 0.0001).*

A closer examination of the running behavior reveals that the movement percentage of the mice varies significantly between trial types (Fig.14.A, B). This movement percentage reflects the frequency of stops made during the trials, providing valuable insights into the mice's behavior. Further analysis of the stop patterns, illustrated in figures 14 C and D, shows that the mice make more stops and spend a longer duration in the small reward trials compared to the large reward trials. In the large reward trials, large percentage of stops occur at the beginning, indicating the mice are hearing and processing the tone. These findings highlight the impact of reward size on movement dynamics, with larger rewards encouraging continuous movement and fewer stops, while smaller rewards are associated with frequent stopping and longer stop durations, those influence the mouse's percentage of movement during the trial.

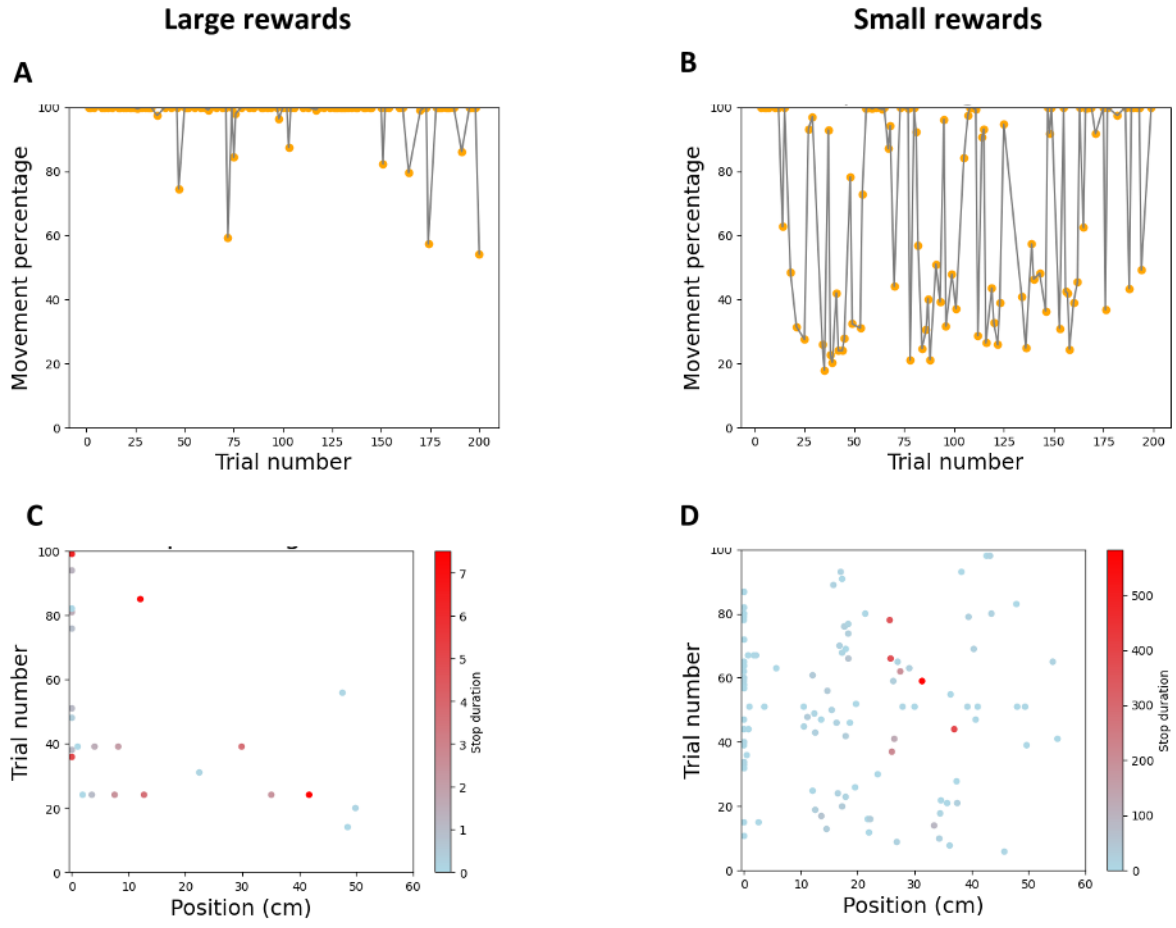



***Fig.14. Increased movement and reduced stopping in large rewarded trials compared to small rewarded trials.*** *The top panels (A,B) display the movement percentage per trial in a specific session, for large (A) and small (B) rewards. The bottom panels (C,D) illustrate the spatial distribution of stops along the track for large rewards (C) and small rewards (D). The color scale in the stop pattern plots indicates stop duration, with darker red representing longer stops (heatmaps).*

The learning process for movement percentage mirrors the patterns observed in previous learning related to trial completion and velocity. Once learning became apparent, it persisted throughout the entire session (Fig.15.A). When comparing learned and unlearned sessions; there is an extremely significant difference in movement percentage between large and small rewards (Fig.15.C: t-statistic of -36.10, p-value <0.0001). As well as movement the stops duration also differ significantly (t statistic=42.73, p value<0.0001). This finding further supports a substantial distinction in behavior during learned trials, where movement percentage varies significantly with reward size. In contrast, unlearned sessions show no significant difference in movement percentage between large and small rewards (Fig.15.B: N.S. t-statistic of 1.960, p-value of 0.974) and neither in stop durations (N.S. t statistic=1.81, p value=0.077). Learning has led to the establishment of distinct behavioral patterns in response to reward size. Conversely, unlearned sessions do not reveal significant differences, suggesting that these behaviors have yet to differentiate concerning reward size in the absence of learning.

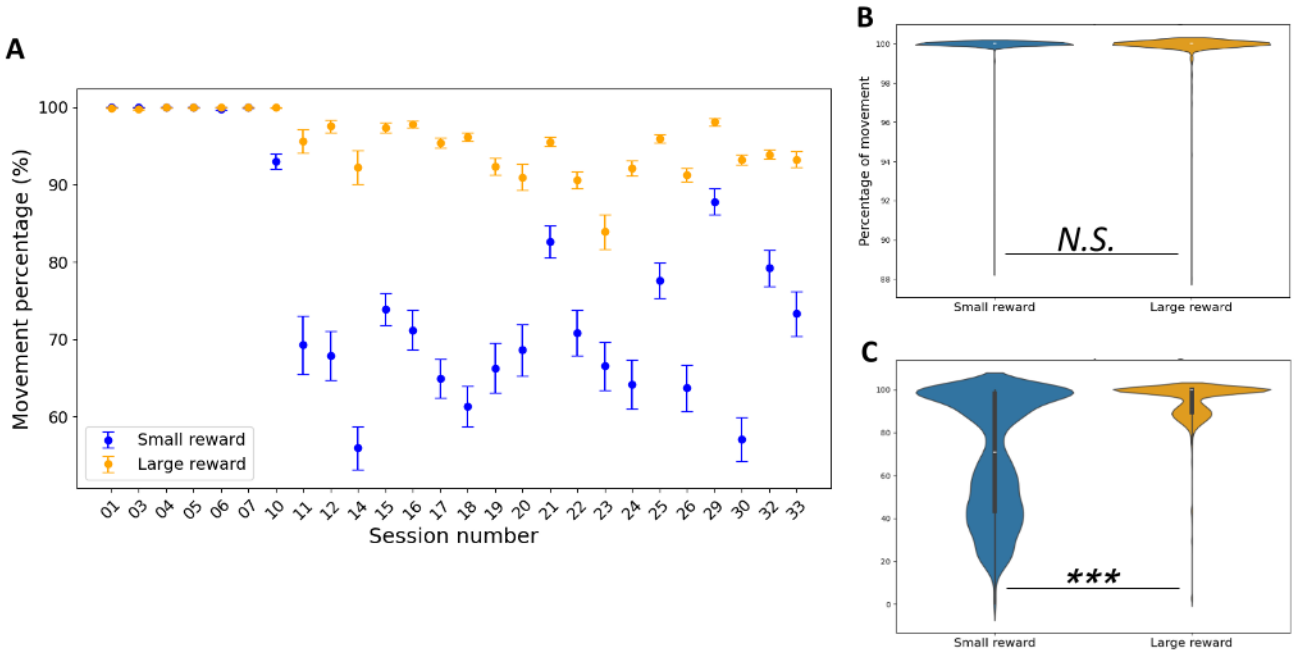



***Fig.15. Movement percentage decrease for small rewarded trials after learning.*** *(A) The trial's movement percentage is plotted against session number for small rewarded trials (blue) and large rewarded trials (orange). Each data point represents the mean movement percentage for a session. The right panels compare movement percentages for small and large rewards in unlearned sessions (B) and learned sessions (C). (B) Before learning, movement percentages are nearly identical for both reward types, showing no significant difference (t-statistic = 1.960, p = 0.974). (C) After learning, movement patterns shift significantly, with large-reward trials showing a much higher movement percentage compared to small-reward trials (t-statistic = -36.10, ***p < 0.0001).*

The significant results indicate that the mouse has successfully learned to associate the auditory cue with the reward delivered at the end of the virtual corridor. Notably, there are significant differences across various conditions, including trial duration (Fig.12.), velocity (Fig.13.), stop duration (Fig.14.), and movement percentage (Fig.15.), between trials that result in small versus large rewards, underscoring the unequivocal learning of this association. In contrast, unlearned sessions show no significant differences, further emphasizing the importance of this association to correct task performance.

When examining the possibility of breaking the connection to previous trials' outcome, all mice in the experiment successfully eliminated this association when shifted to longer ITIs, specifically from 5 seconds to 15 s and from 12 s to 20 s. Among the mice tested with an ITI of 12 s, one mouse learned the task, while another mouse learned it successfully only after the ITI was changed from 12 to 20 s.

In conclusion, the results from this study highlight the effectiveness of the learning paradigm established through careful manipulation of the ITI. By optimizing the ITI, the mice successfully learn to distinguish between the rewards associated with different tones, forming strong associations with the auditory cues that signal reward magnitudes. The statistical analyses reveal significant differences in trial duration, movement percentage, stops duration and velocity, underscoring the impact of learning on behavioral patterns. Specifically, learned sessions demonstrate robust differentiation in both trial duration and movement behavior in response to reward size, while unlearned sessions fail to show these distinctions. This indicates that the mice not only adapt their behaviors based on the rewards received but also develop the ability to process information independently from previous trials. The findings from this paradigm provide valuable insights into the mechanisms of learning and decision-making in mice, with implications for understanding how reward structures influence behavior in broader contexts.

**Short - term association task: Predictive licking**

The goal of this short-term association task was to investigate the mice's ability to form associations and predictions, characterizing both their learning process and their behavior in response to the task. I analyzed the licking patterns of the mice, specifically, their responses to different reward cues such as the clicking sound signaling reward delivery, and how this behavior evolved with time (Fig.16.). The aim was not only to observe their actions but to track learning patterns over multiple days. Mice were tested to determine whether they can learn to predict when to lick and whether this behavior was purely sensory or involved predictive elements.

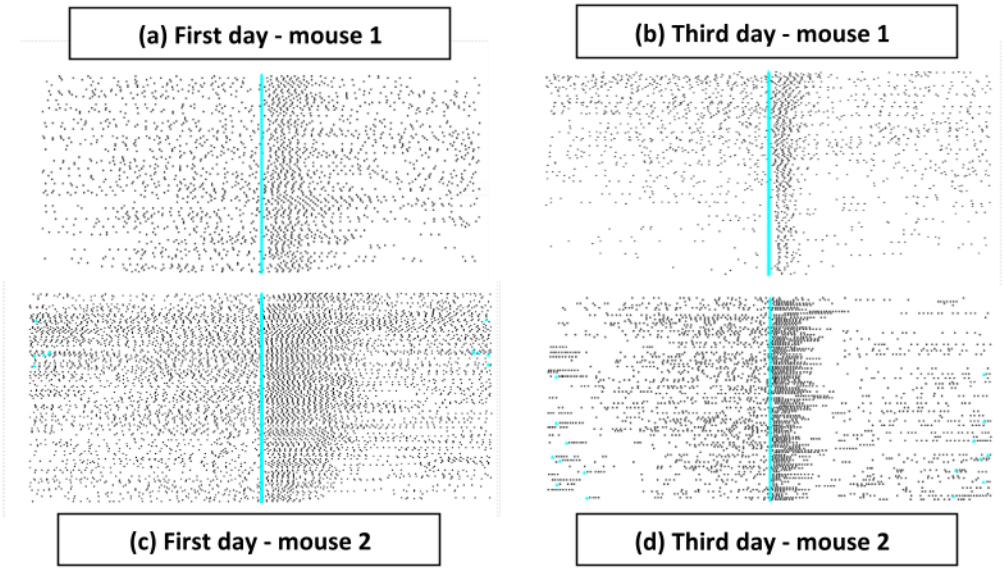



***Fig.16. Mice's licking pattern change with training. The licking patterns of two mice (top and bottom) is shown for the first (left) and third (right) training sessions.*** *In the raster plots, black dots represent licking times, and light blue indicates reward delivery. Trials are presented in order from top to bottom*

After confirming that the mice can learn to associate the clicking sound with the reward, a more complex task was introduced. This task involved an auditory cue predicting the reward, the clicking sound indicating reward delivery, and variable ITIs. All head-fixed mice performed the task using the previously described system. Rewards were delivered after a predetermined delay following the auditory cue. Two 200 ms tones signaled both the reward and its magnitude, which was delivered 500 ms after the tone. These cues indicated either a small (1 µL) or a large (6 µL) reward, with ITIs of 5-7 s or 7-9 s, depending on the group. Specific time windows were used in all trials to assess the mice's behavior and learning patterns. Each time window was 250 ms in duration, during which licking activity was measured: (1) Tone prediction - this window captured the 250 ms preceding the

tone's appearance and was analyzed to investigate whether the mice can predict the appearance of the auditory cue signaling the reward using time assessment (Fig.17. gray rectangle). (2) Reward prediction following the auditory cue - this window spanned 250 ms before reward delivery and 250 ms after the tone was presented to the animal, allowing analysis of the mice's ability to anticipate the reward based on the tone (Fig.17. red rectangle). (3) Licking behavior after reward delivery - this window measured licking activity 250 ms after the reward was delivered, with the clicking sound serving as the indicator that the valve had been opened (Fig.17. black rectangle). (4) Control licking - licking was measured during 250 ms and measured at a random time during the ITI, occurring between 2-3 seconds after the reward delivery or 2-6 seconds before the next tone.

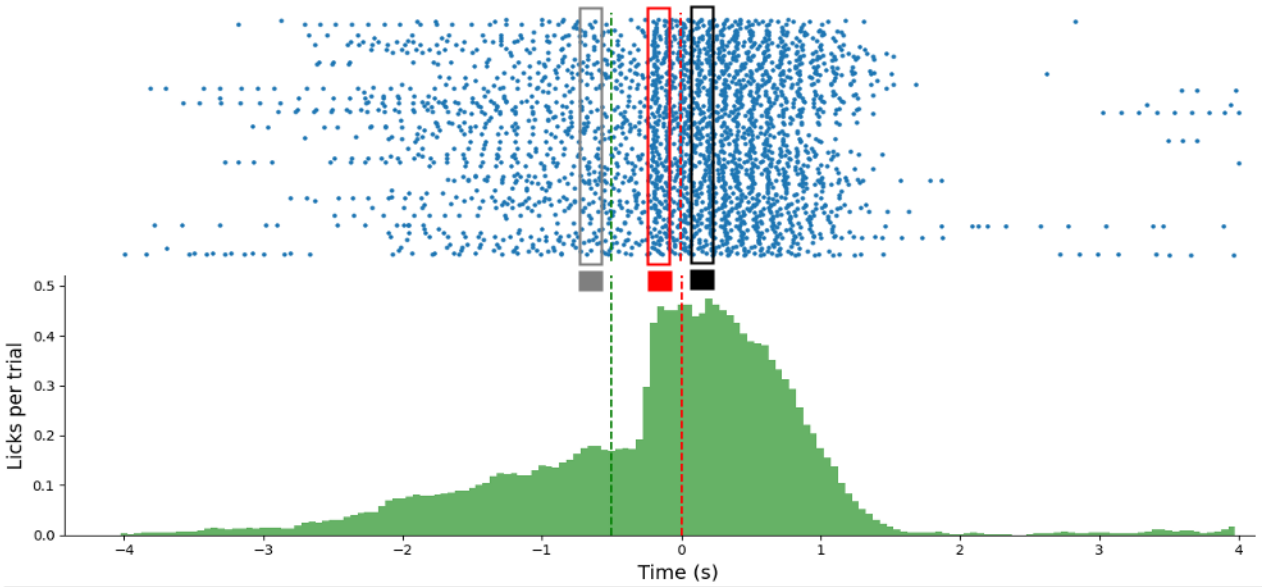



***Fig.17. Mice's licking behavior during the trial.*** *An example of the licking pattern of a mouse during one session. Shown are a raster plot of the licks (top) and a peristimulus time histogram (PSTH, bottom). In the raster plot, each blue dot represents a lick, with trials displayed from top to bottom. Key events are indicated by vertical dashed lines: the green line marks tone delivery, the red line marks reward delivery. The PSTH below illustrates the average number of licks per trial (in green) across all trials at specific time points during the trial. Three distinct time windows are highlighted by rectangles, each lasting 250 ms: the gray rectangle represents the tone prediction window before the tone, the red rectangle shows the reward predictive licking window following the tone, and the black rectangle represents the post-reward licking window. Control licking is randomly measured before and after the trial end.*

Mice learned to associate the auditory tones with rewards, as evidenced by specific predictive licking behaviors. Some of them could distinguish between different tones and associate each tone with a specific reward size. The mice also learned to predict the timing of the cues themselves, forming a complex association between the tone, the reward, and the timing of both.

A total of thirteen mice were trained on the task. They exhibited different learning strategies and progressed through distinct phases, though not necessarily in the same order. Based on their licking pattern, all the mice learned to expect a reward following the tone and therefore started licking extensively before the actual reward was delivered (Fig.17. red rectangle). At this point, none of the mice was able to distinguish between tones or reward sizes. This **reward prediction phase** occurred mostly on the first or second training days. After several days, all the animals started to lick prior to the tones (Fig.17. grey rectangle) suggesting that they learned to estimate the ITIs. This **tone prediction phase occurred** mostly during the first 4 training days and always after reward prediction phase occurred. The majority of the mice (n=10) learned to distinguish between the two tones associated with reward size as shown in their licking pattern which was different after the tones and prior to reward. This **tone discrimination phase**, occurred later in the training and was modulated by the ITI. Finally, in the **reward discrimination phase**, the mice (n=10) learned to distinguish between different reward sizes, and their licking behavior after receiving the reward reflected this distinction. Importantly, these learning strategies did not follow a strict sequence, and individual mice could exhibit different learning strategies.

In this paradigm, a total of 13 animals were tested. Out of them, 10 started training on ITIs of 5-7 s, and then 8 of them were switched to it is 7-9 s to study how their behavior changed with a longer ITI. 3 mice were trained on ITIs of 7-9 s from the first training session (shown in table.1.).

| ITI | N animals | Learned reward prediction | Learned tone prediction | Learned tone discrimination | Learned reward discrimination |
|---|---|---|---|---|---|
| 5-7 sec ITI | **2 animals** | All (2) | All (2) | 1 - 50% | 1 - 50% |
| 7-9 sec ITI | **3 animals** | All (3) | All (3) | All (3) | All (3) |
| 5-7 → 7-9 sec ITI | **8 animals** | All (8) | All (8) | 2 → 5 (25% → 62.5%) | 2 → 5 (25% → 62.5%) |

***Table.1. Summary of learning outcomes across different ITIs.***
*This table presents the progression of learning in reward prediction, tone prediction, tone discrimination, and reward discrimination across groups of animals subjected to different ITI conditions. Each row corresponds to a group with specific ITI settings, while the columns represent different categories of learned behavior. The last group (5-7 → 7-9 sec ITI) is the group that their ITI had been shifted, the results of the later percentages represented by arrow (→).*

In order to confirm that learning had occurred, significance levels were assessed for relevant variables during each phase across all animals. Mice were classified as having learned only if they demonstrated significant performance over two consecutive sessions ($p < 0.05$, using a t-test, all normally distributed). After the learning phase, a comprehensive analysis was performed by aggregating the data from mice that successfully learned the task, separating it from data reflecting unlearned behavior.

**Reward prediction learning:**

The predictive licking in response to the tone was measured within a 250 ms time window following the auditory cue in each trial (250 ms after the onset of the tone and 50 ms after it's offset). Licking behavior was analyzed by combining data for both small and large rewards. The resulting licking pattern can be observed in Fig.18.

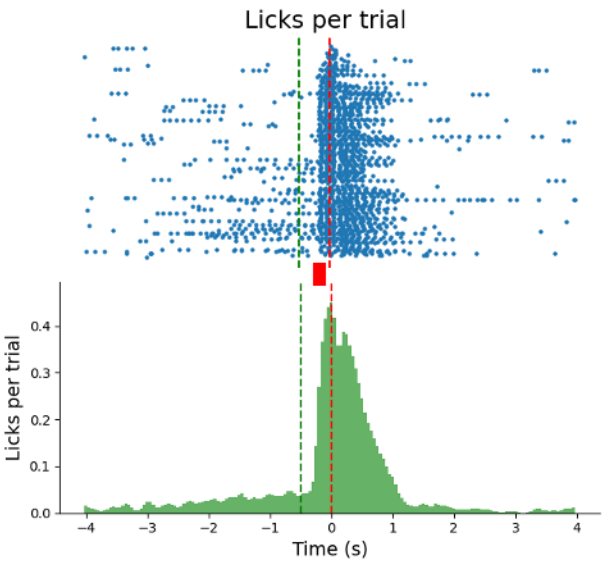



***Fig.18. Reward prediction licking behavior.*** *Shown is an example of one animal's licking behavior, with a raster plot (top) and PSTH (bottom) aligned to reward delivery. Conventions as in Fig.17. The number of licks was counted within a 250 ms window before the reward as indicated by the red rectangle.*

All 13 mice exhibited a 100% success rate in predicting the reward after hearing the tone, marking the initial stage of learning. The mice learned this behavior within 1-2 days, and notably, they displayed a consistent learning strategy across the group.

A Levene's test for homogeneity of variance showed that the variances in licking behavior across animals was similar (Fig.19. $p = 0.2547$), indicating no significant difference in variability between animals. This result suggests that the variance in both predictive and control licking behavior was

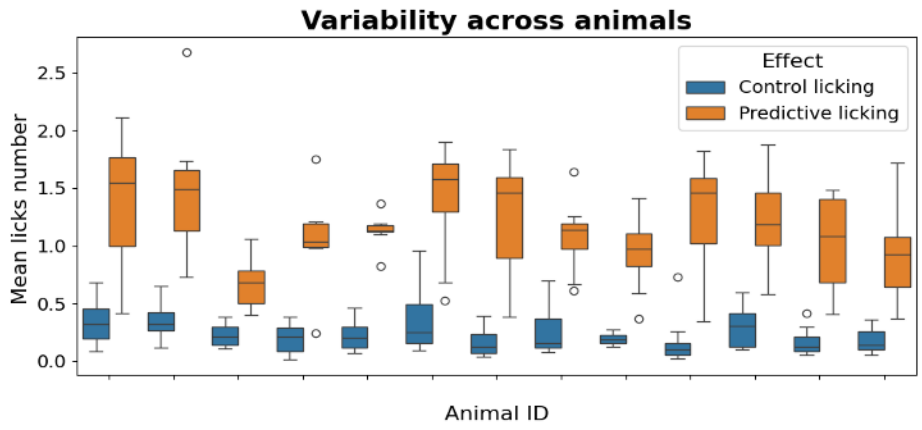



***Fig.19. Variability in licking behavior across animals.*** *The box plots show the mean number of licks per session for each animal (Animal ID on the x-axis) categorized into two conditions: control licking (blue) and predictive licking (orange). The control licking reflects baseline licking during random times in the inter-trial interval, while predictive licking represents licking behavior in anticipation of reward following the tone. The variability across animals was not significant, with all animals exhibiting higher levels of predictive licking compared to control licking.*

consistent across all subjects, with no individual animal displaying disproportionately higher or lower variability in their licking behavior compared to others.

With no significant difference in variances between control and predictive licking across all animals, session means were combined across animals. The results were highly significant, showing that reward predictive licking was significantly higher than control licking (Fig.20. paired t-test combining all trials; t-statistic = -42.22, $p < 0.0001$). A comparison of the means from the first 8 sessions also yielded highly significant results (t-test: t-statistic = -29.33, $p < 0.0001$).

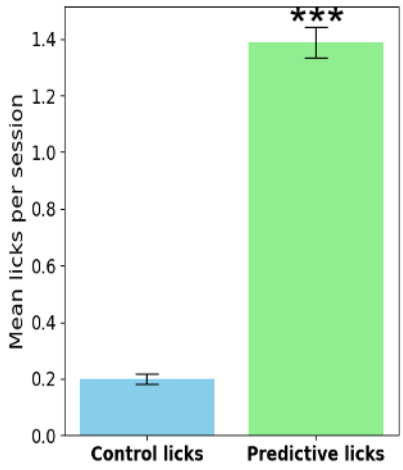



***Fig.20.* Reward prediction learning.** *The bar graph shows the mean number of licks per session during the control licking period (blue) and the reward predictive licking period following the tone (green). The control licking reflects spontaneous licking during a random ITI, while the predictive licking reflects anticipatory behavior after the auditory cue but before the reward. The error bars indicate the SEM. Mice exhibit significantly higher licking rates during the predictive phase compared to the control phase, demonstrating learned anticipation of the reward following the tone.*

A mixed-effect model was employed to analyze the effects of session number, licking condition, and inter-animal variability. Session number significantly affected licking behavior (ANOVA: $F = 9.57$, $p < 0.001$), indicating that the animals' behavior changed notably across sessions. A significant effect was also observed for licking condition (ANOVA: $F = 293.74$, $p < 0.001$), with reward predictive

licking triggered by the tone resulting in significantly higher licking rates than control licking. This effect was consistent across sessions and animals, highlighting the association of the auditory cues with expected reward. Additionally, a significant interaction between session number and licking condition (Fig.21. left, ANOVA: $F = 32.67$, $p < 0.001$) showed that the difference between control and predictive licking evolved over time. Predictive licking became more pronounced with each session, while control licking remained stable (Fig.21.). These findings support the conclusion that the differences in licking behavior between control and reward predictive conditions are not confounded by large inter-animal variability. This consistency further strengthens the observation that reward-predictive licking is reliably learned across the group.

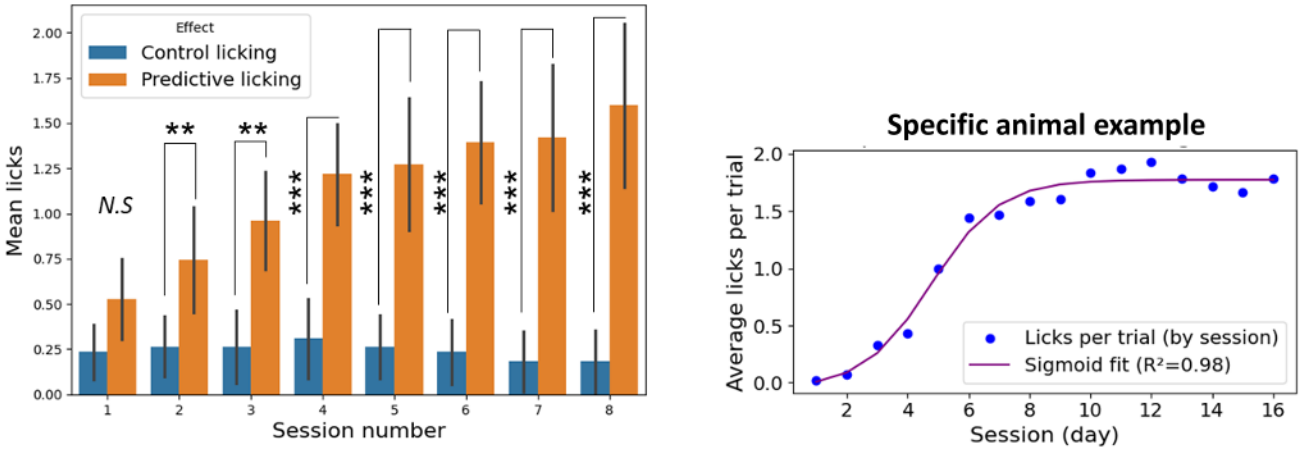



***Fig.21. Increase in reward-predictive licking over the course of sessions.*** *(left) The bar graph shows the average number of licks per session, comparing control licking (blue) and reward predictive licking (orange) across eight sessions. The error bars represent the standard error of the mean (SEM). Predictive licking, triggered by the auditory cue, consistently increases over sessions, indicating a learning effect, while control licking remains relatively stable. The mixed-effects model analysis confirmed significant effects of session number ($F = 9.57$, $p < 0.001$), licking condition ($F = 293.74$, $p < 0.001$), and a significant interaction between session number and licking condition ($F = 32.67$, $p < 0.001$). From the second day onward, all learning is statistically significant. (right) The scatter plot illustrates the average number of licks per trial (y-axis) across sessions (x-axis) for an individual animal. Each blue dot represents the average licks per trial for a given session. The purple regression line represents the sigmoid fit ($R^2=0.98$) modeling the relationship between session number and average licks per trial. Additionally, the Spearman correction ($r = 0.8559$, $p= 0.0001$) indicates significant monotonic association between session number and licking behavior.*

**Tone prediction learning**

Tone predictive licking in response to time assessment was measured in a 250 ms window preceding the auditory cue in each trial. Licking behavior was analyzed for both small and large rewards during this window, as shown in Figure 22. The primary aim of this analysis was to determine whether mice learn to estimate the appearance of a conditioned cue using different ITIs. The data show that licking activity prior to the tones increases significantly over the course of training sessions (Fig.22.).

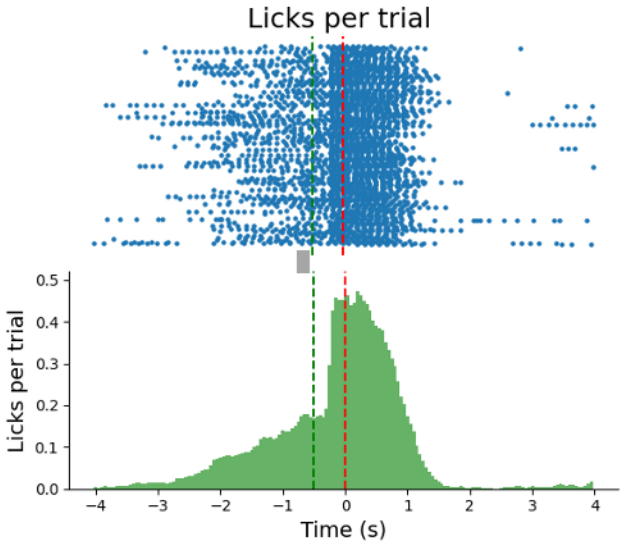



***Fig.22. Tone predictive licking by time assessment.*** *This figure shows an example of one animal's licking behavior, with a raster plot (top) and PSTH (bottom), presented as in Fig.17. Tone predictive licking was measured in a 250 ms window preceding the auditory cue (the dashed green line), indicating the mice's anticipation of the upcoming reward (represented by grey rectangle). Pearson correlation analysis shows a strong relationship between session progression and predictive licking (r = 0.83, p <0.0001), further supporting this learning effect.*

All 13 mice successfully learned to predict the appearance of the tones. The mice required 1 to 4 days to learn the strategy using time assessment to predict the played tone. This learning consistently occurred after the mice had learned reward prediction, irrespective of the interval duration (5-7 s or 7-9 s). Statistical analysis confirmed the significance of these results (Fig.23. paired t-test= -19.31 , p< 0.0001). The licking behavior increased in anticipation of the tone, illustrating the development of time-based predictive behavior as mice learned to anticipate reward delivery.

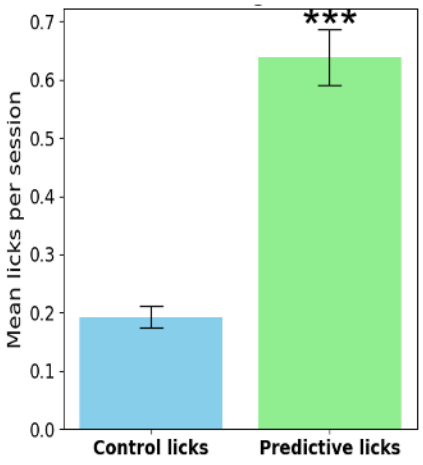



***Fig.23. Tone prediction learning.*** *The bar graph shows the mean number of licks per session for control licking (blue) and tone predictive licking (green) as the mice learned to predict the trial start time based on the interval preceding the tone and reward. Error bars are in SEM. Statistical analysis revealed a significant difference in licking behavior between control and predictive conditions (Paired t-test = -19.3142. ***P < 0.0001).*

The interaction between session number and condition (ANOVA) was statistically significant (Fig.24. F = 5.4002, a ***p = 0.0004). This result suggests that the effect of condition (control vs. tone predictive licking) varied across sessions, indicating that the difference in licking behavior between the two conditions evolved over time. Although Tukey's Honestly Significant Difference (HSD) test, which compares individual sessions, did not reveal significant differences between specific session pairs, this result suggests that while the overall population trend is strong, the effect within individual animals may not be pronounced enough to show significant differences at the session level. This likely reflects a more gradual or distributed learning process that becomes evident when averaged across all animals.

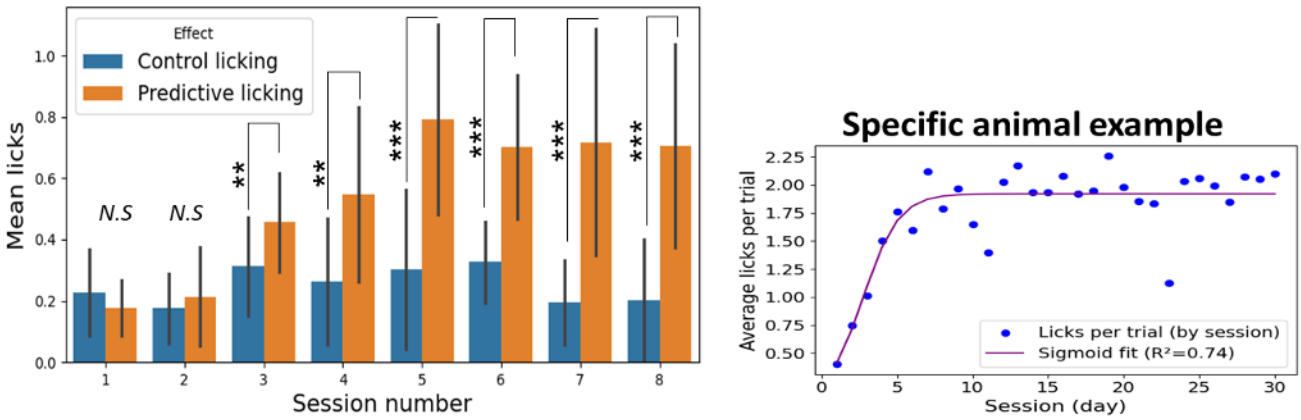



***Fig.24. Tone prediction learning over the course of sessions.*** *(left) The bar graph compares the mean number of licks per session for control licking (blue) and tone predictive licking (orange) across the first eight sessions. Error bars represent SEM. A significant interaction between session number and licking condition was observed (ANOVA: F = 5.4002, p = 0.0004), indicating that the difference between control and tone predictive licking changed over time. From the third session, significant differences are observed across animals (t = -3.03, **P = 0.01), with the fifth session being the first to show a significant difference compared to day 1 (p = 0.0005). This indicates that the animals' behavior evolved across sessions, reflecting their learning and adaptation to the task. (right) an example of one animal showing the average number of licks per trial (y-axis) across sessions (x-axis). The purple curve is the regression line representing a sigmoidal relationship between session number and licks per trial ($R^2$ = 0.7358 for the sigmoid fit, Spearman correlation: r = 0.5496, p = 0.001).*

Levene's test was conducted to assess the consistency of variance in licking behavior across animals. During the initial eight days of learning, the variability in licking behavior was not statistically significant (Fig.25. p = 0.053), indicating similar levels of consistency among animals. However, six days after all animals had completed the learning process, the variability became significant (p = 0.0031). This suggests that although variability was consistent during the learning

process, it diverged post-learning, with animals showing differing levels of response consistency. Predictive licking showed an increase in variability as the experiment progressed, with the standard deviation growing from 0.1921 (Session 1 -after learning ) to 0.3768 (Session 6 -after learning). In contrast, control licking exhibited relatively stable variability, with minor fluctuations across sessions (0.1715). Thus, perhaps there are several learning strategies of tone prediction that are formed after learning. This effect will be further analyzed below.

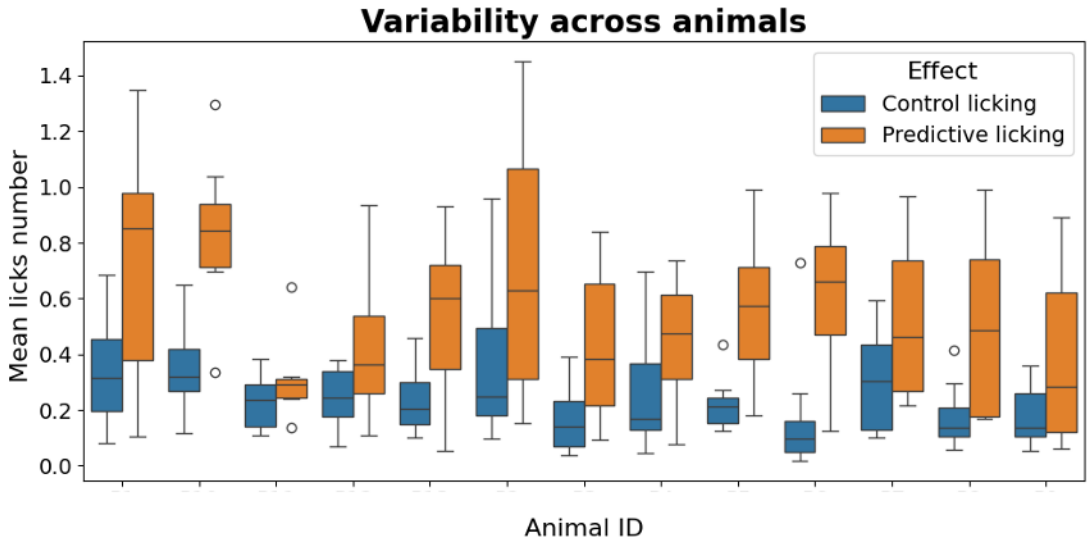



***Fig.25. Variability in tone predictive licking behavior across animals.*** *The box plots show the mean number of licks per session for individual animals (Animal ID on the x-axis) under control licking (blue) and tone predictive licking (orange). The variability in licking behavior across animals is evident, with some animals showing pronounced differences between control and predictive licking, while others displayed more balanced behavior. Levene's test for the first eight days returned a p-value of 0.0523, while for the six days after learning had accrued a p value of 0.0031.*

**Tone discrimination learning**

In order to determine whether the animals were able to distinguish between the two predictive tones I compared between their licking responses measured following each of the tones. The time window was 250 ms as defined before for reward predictive licking. This way, the licking behavior was analyzed separately for predictions of small and large rewards (Fig.26.).

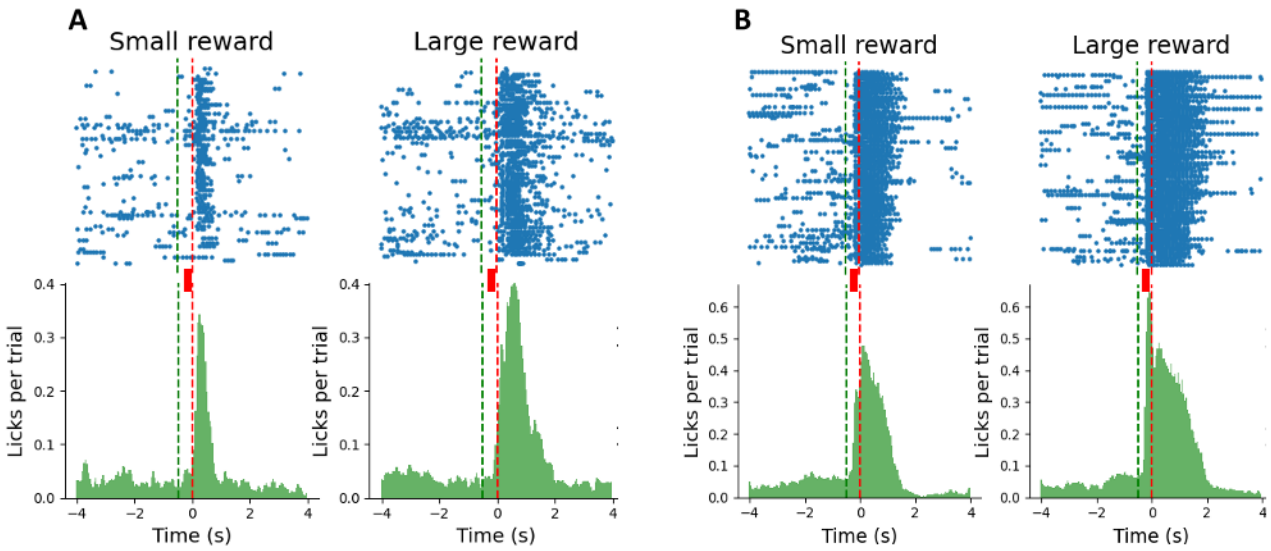



***Fig.26. Mice learn to distinguish between tones that predict different reward amounts.***
*Shown is an example of one animal's licking responses during the first (A) and 15$^{th}$ (B) training sessions. Each panel shows the licking response in small reward (left) and large reward (right) trials. Raster plots and PSTHs as in Fig.17. Red rectangles mark the 250ms window before reward delivery as in as in Fig.17. At session 15, there is evident tone discrimination, with more pronounced reward predictive licking after the tone and before reward delivery for large rewards compared to small rewards.*

In tone discrimination learning, not all animals successfully learned the association. Out of the 13 animals, 10 (76%) discriminated between the tones and their associated meaning. For those that learned this phase, the number of licks was significantly higher following the tone predicting large rewards compared to the tone predicting small rewards (Fig.27.A ). 3 out of 10 animals, did not initially learn tone discrimination and were able to learn it after the ITI was shifted from 5-7 s to 7-9 s. The effect of tone discrimination (large vs. small rewards) was highly significant, indicating a substantial difference in licking behavior in response to the two tones (Fig.27.A, paired t-test, t-statistic = 8.47, $p < 0.0001$). In addition to the increase in the number of licks, the licking pattern prior to reward delivery was different for the large reward compared to the licking prior to the small reward. When the animal expected a large reward, the licking pattern occurred more consistently compared to the licking during small reward expectation. Magnifying the scale displaying the licking activity for large rewards shows a typical pattern where licking before reward delivery closely mirrors licking after reward delivery (Fig.27.C). In contrast, for the small reward, licking behavior was less organized, with fewer licks per trial and a more dispersed pattern (Fig.27.B). The robust and focused licking bouts for the large reward after the tone and before

reward delivery, highlights the mice's learned association between the tone and reward size (see Fig.27. B, C).

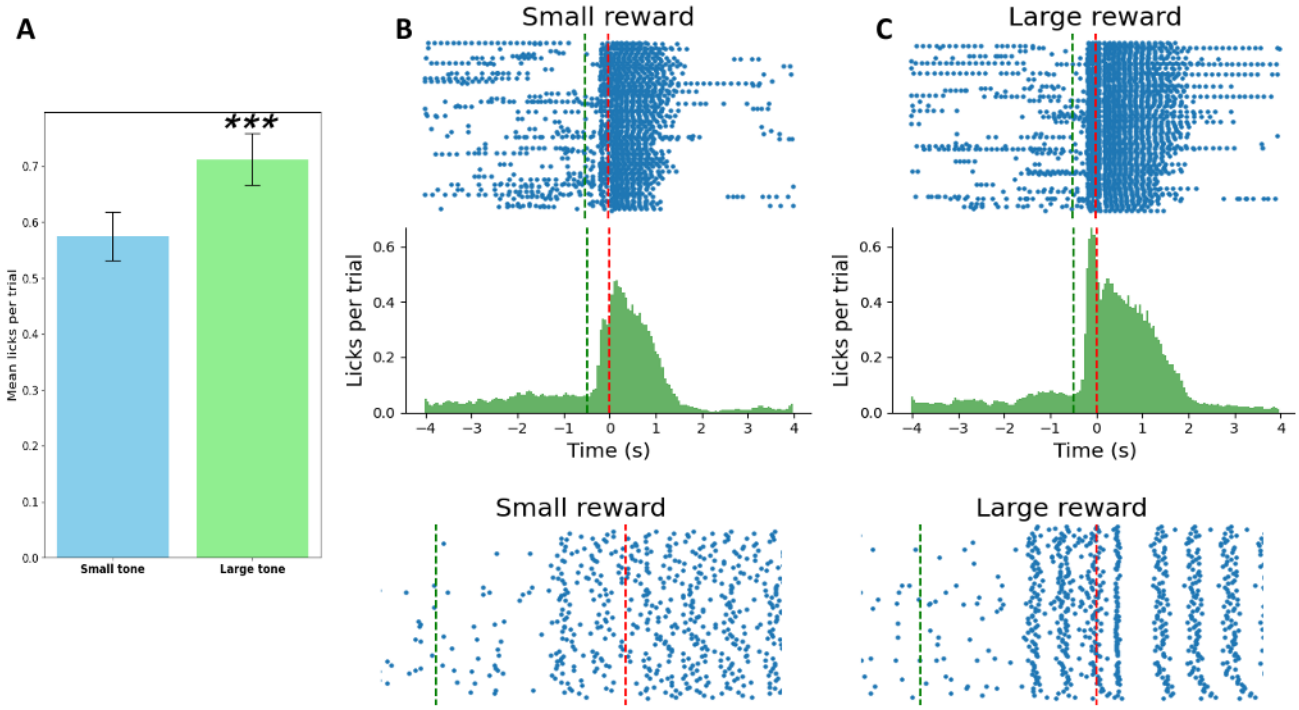



***Fig.27. Typical reward predictive licking activity in response to the different tones.***
*(A) Shown are the mean number of licks per trial during reward predictive licking for greater licking number for large rewards (green) compared to small rewards (blue) - t-statistic = 8.47, ***p < 0.0001. The error bars are SEMs. (B,C) The figure shows an example of one animal's licking behavior using raster plots (top) and PSTH (below), presented similarly to Fig.17. The figure compares predictive licking behavior for rewards following a specific tone, distinguishing between small (B) and large (C) rewards. The green dashed line marks tone onset, and the red dashed line indicates reward delivery, occurring 500 ms after the tone. Expansion of the x axis (bottom) demonstrates the difference in predictive licking behavior between small (B) and large (C) rewards.*

The effect of session number on licking behavior, however, was not statistically significant (ANOVA: F = 2.4641, p = 0.0839). Neither, the interaction between session number and tone effect (ANOVA: F = 2.6190, p = 0.0713). There was a high degree of similarity between large and small tone trials in the mean number of licks (std- large: 0.2309, small: 0.2252). The similarity in the standard deviations suggests that, although there was a significant difference in mean licking behavior (as shown by ANOVA), the variability in responses to both tones was comparable. Levene's test showed that the variance in licking behavior was similar across animals (p = 0.2432). This suggests consistent licking behavior across animals when distinguishing between tones, with no individual showing extreme variability. The similarity in licking activity for predictive licking in response to

specific tones indicates that once animals learned to distinguish between tones, their licking behavior became similar. As a result, it is statistically valid to collapse all the sessions and represent the data as a single group in further analysis.

**Reward discrimination learning**

Licks following reward delivery were measured within a 250 ms window immediately after the reward, and licking behavior was analyzed for both small and large rewards. The resulting licking patterns are shown in Fig. 28. The primary goal of this analysis was to determine whether the mice licked as a sensory response to the reward or as a result of their ability to associate the tone with a specific reward, anticipating the appropriate licking behavior.

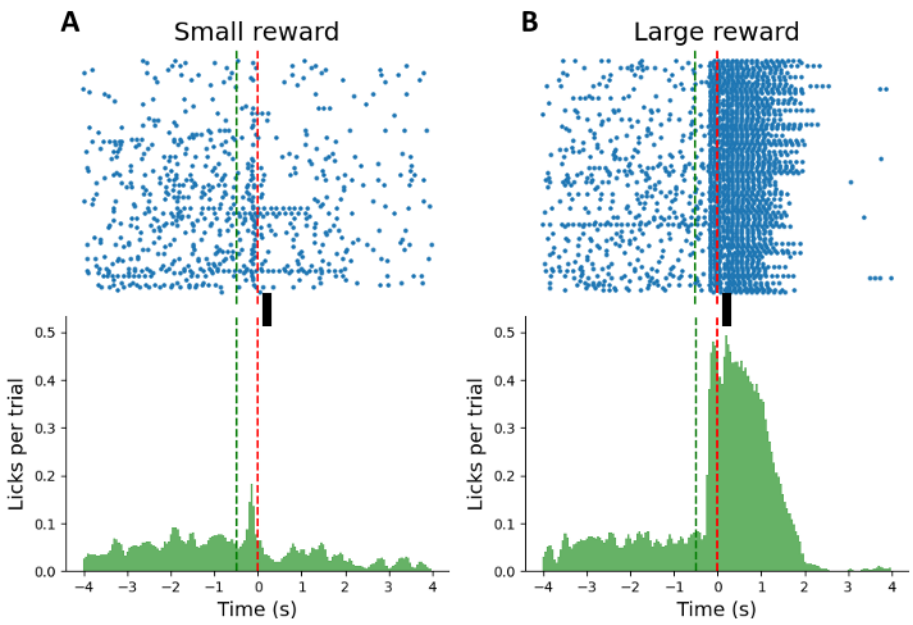



***Fig.28. Licks per trial in response to reward delivery for different reward sizes.*** *The figure shows an example of one animal's licking behavior with a raster plot (top) and PSTH (bottom), presented similarly to Fig.17. comparing small rewards (left) and large rewards (right). The black rectangle denotes the time window used to measure the licks in response to reward delivery (dashed red lines).*

In the reward discrimination phase, 10 of the mice (76%) successfully learned to differentiate between large and small rewards, as evidenced by the number of licks immediately following the reward (delivered after a click signal). This stage of reward discrimination was typically acquired after the tone discrimination learning and often emerged later in the learning process. The statistical analysis produced a high significance (Fig.29. t-statistic of -10.7223, p-value< 0.0001).

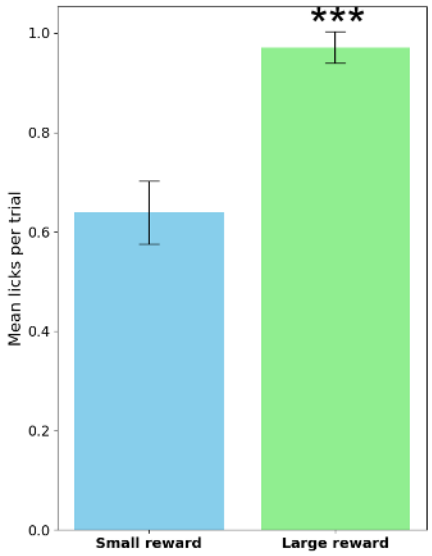



***Fig.29. Licks after reward for small and large rewards.*** *The bar graph compares the mean number of licks per trial after the delivery of small rewards (blue) and large rewards (green). Error bars are SEMs. Licking behavior following large rewards was significantly higher than for small rewards (*t-statistic = -10.72, ***p<0.0001*).*

Reward discrimination learning primarily occurred after tone discrimination was fully established. This pattern was consistent across all mice, with tone discrimination being completely learned before reward discrimination could be effectively achieved (presented in Fig.30.).

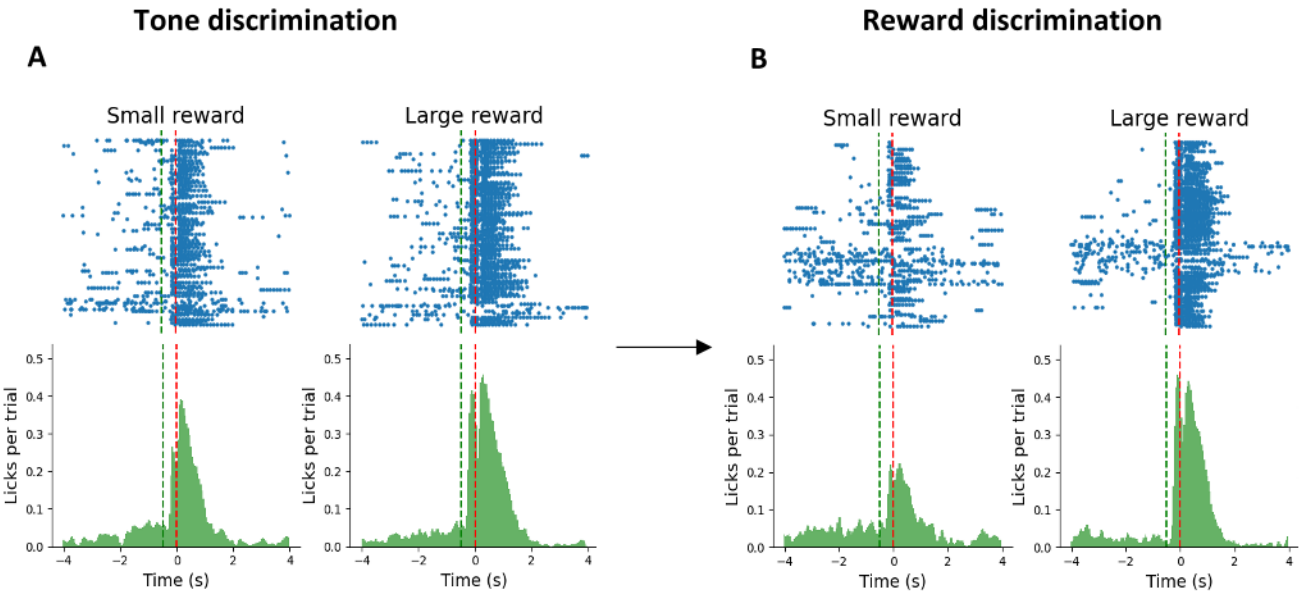



***Fig.30. Progression from tone discrimination to reward discrimination.*** *The figure illustrates predictive licking behavior for small rewards (left) and large rewards (right), during Day 11 (A) and Day 13 (B). (A) In session 11, tone discrimination learning was observed, characterized by increased predictive licking for the large reward, without any subsequent change in reward licking. (B) In session 13, reward discrimination learning occurred, as apparent from the increased number of licks following reward delivery. Raster plots (top) and PSTHs (bottom) conventions, as in Fig.17.*

The effect of session number on licking behavior was not statistically significant ($F = 1.1846$, $p = 0.3341$). However, the effect of reward size (large vs. small) on licking behavior was highly significant ($p < 0.001$), indicating a strong and meaningful difference. Animals exhibited significantly

more licking following larger rewards compared to smaller ones. The influence of reward size on licking behavior remained consistent across sessions after the mice's learning (Fig.31.).

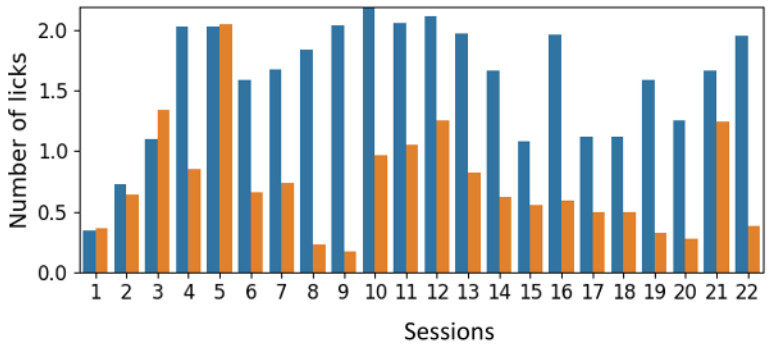



***Fig.31. Licking behavior following reward delivery across sessions, with increased licks for large rewards.*** *An animal example - the bar graph shows the number of licks after reward delivery across the sessions. Licking behavior following large rewards is represented by blue bars, while licking behavior following small rewards is shown in orange bars. The number of licks for large rewards is consistently higher than for small rewards after sessions 6 (P<0.01). This indicates that mice exhibit stronger licking behavior after receiving larger rewards compared to smaller rewards.*

**Classification into groups:**

Thirteen mice were included in this study, divided based on ITI duration into two groups: 10 mice trained on a 5-7 s ITI, of which 8 transitioned to a 7-9 s ITI after 8-20 days of stabilized behavior, and 3 trained exclusively on a 7-9 s ITI. This separation aimed to investigate distinct behavioral patterns associated with each ITI duration. In the 5-7 s group (n=2), one mouse exhibited strong tone prediction without tone discrimination, while the other showed the reverse, with tone discrimination and reward discrimination but minimal tone prediction. In the 7-9 s group (n=3), all mice demonstrated tone and reward discrimination with reduced tone prediction, indicating that longer ITIs may hinder tone prediction and shift focus toward tone discrimination.

To explore how disrupting time assessment affects tone prediction, 8 mice were shifted from a 5-7 s to a 7-9 s ITI after establishing reward and tone prediction learning. This shift disrupted stable behavior and altered licking patterns. Post-shift, tone predictive licking behavior no longer varied significantly across sessions when considering all conditions together ($F(16, 112) = 0.8978$, $p = 0.5733$), and the interaction between session number and condition (control vs. tone predictive licking) was not significant ($F(16, 112) = 0.7983$, $p = 0.6848$). However, the difference between tone

prediction licking and control licking remained highly significant (F(1, 7) = 14.8660, p = 0.0062), indicating a persistent distinction between these behaviors despite the disruption.

Following the ITI shift, variability in tone-predictive licking increased significantly (SD = 0.456), while control licking remained consistent (SD = 0.167) (Fig.32.A). Levene's test confirmed this significant variability post-shift (Fig.32.B, p < 0.0001) compared to pre-shift levels (p = 0.14). Moreover, a one-tailed bootstrap analysis revealed a significant decrease in tone predictive licking after the time shift (Fig.32.C, mean difference = -0.1679 ± 0.0881, p = 0.0280; Fig.32.D, median difference = -0.1873 ± 0.1188, p = 0.0493), further confirming the disruption of time-based associations.

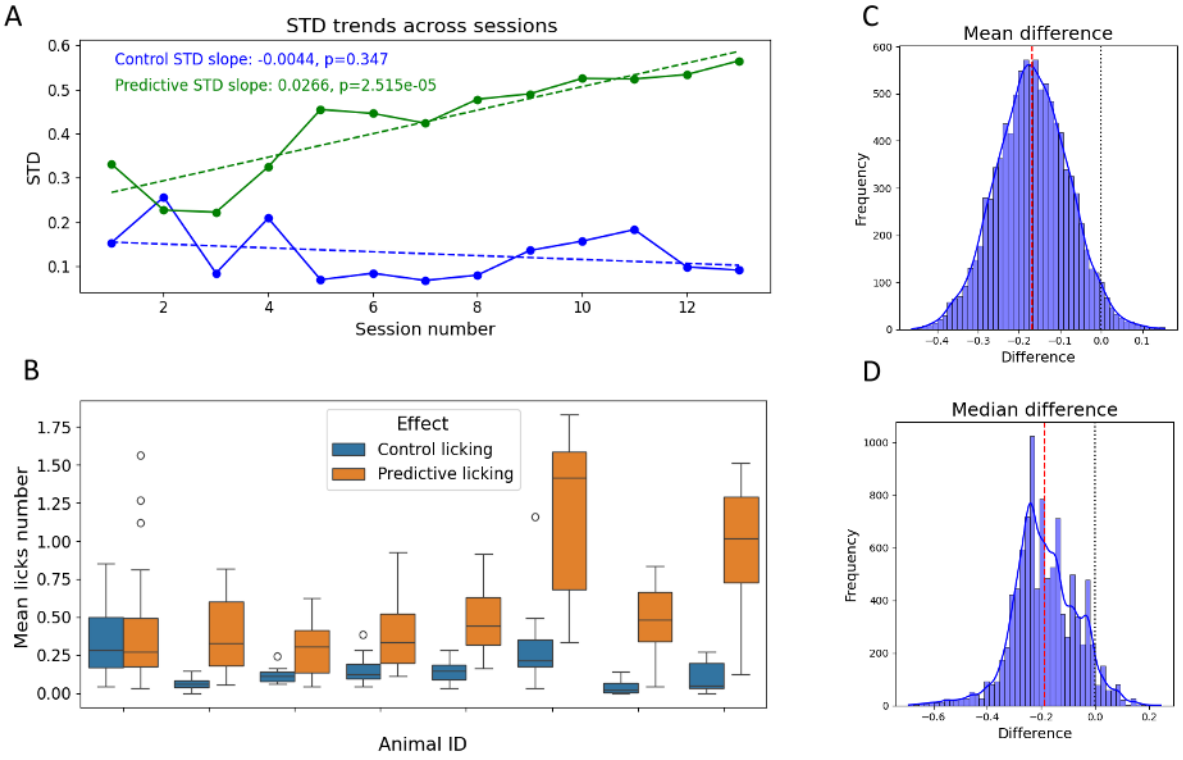



***Fig.32. ITI shift causes the mice to display different tone prediction licking behavior.*** *(A) Trends in the standard deviation (STD) of licking behavior across sessions for control (blue) and predictive (green) conditions. The control condition shows a non-significant STD slope (-0.0044, p = 0.347), while the predictive condition shows a significant positive slope (0.0266, p = 2.515e-05). Dashed lines represent linear regression, indicating increased variability in predictive licking behavior over sessions. (B) Box plots depict the mean number of licks per session for individual animals (Animal ID on the x-axis) in the control (blue) and predictive (orange) conditions after the shift from 5-7 s to 7-9 s ITI. Levene's test revealed significant variability (p < 0.0001) among ITI-shifted animals (N=8), emphasizing individual differences in responses to the shift. (C,D) The histograms display the distribution of differences in mean (C) and median (D) predictive licking behavior before and after the ITI shift. The red dashed lines represent the observed differences, while the black dashed lines show the zero-difference reference.*

The increased variability following the ITI shift demonstrated non-uniform responses among the animals, disrupting tone prediction. Categorizing the animals based on their behavior revealed distinct adaptation strategies: some maintained time associations, while others shifted focus to tone discrimination. Of the 8 mice that underwent the ITI shift, 3 remained in the tone prediction group, 2 stayed in the tone discrimination group, and 3 transitioned from tone prediction to tone discrimination (Table.2.). This shift underscores the transition from time-based to tone-based learning.

| Group | Pre- shift | Post- shift | |
|---|---|---|---|
| **Tone prediction focus (Group 1)** | 6 animals | 3 animals (-50%) | 3 Maintained tone prediction |
| **Tone discrimination focus (Group 2)** | 2 animals | 5 animals (+150%) | 3 Shifted to tone discrimination |

***Table.2. Prolonging the ITI from 5-7 s to 7-9 s altered the learned associations.*** *This table represents the distribution of animals across learning focus groups following an ITI shift, illustrating how changes in ITI influence group composition and learning outcomes. The table classifies animals based on their performance on tone prediction (Group 1) or tone discrimination (Group 2) prior to (pre-shift) and following (post-shift) change in ITI. 3 animals transitioned from tone prediction association to tone discrimination association.*

**Group 1: Emphasis on tone prediction**

This group of mice demonstrated reward prediction learning and later developed tone prediction abilities, typically within 1-2 days of training, as shown by increased licking behavior as the tone that predicts reward approached. However, they did not develop tone discrimination or reward discrimination (Fig.33.). Among the 13 animals, 7 displayed this learning strategy, and 6 of these animals underwent an interval time shift. After the shift out of 6 animals 3 remained in the group (Table.2.).

Notably, this group maintained their time evaluation for tone prediction even when the ITI was shifted from 5-7 s to 7-9 s, though adaptation to the new ITI took several days (Fig.34.). For this group, the primary learning was time-based rather than tone-based discrimination between

different tones or rewards. That are the mice did not discriminate between the tones or the resultant reward.

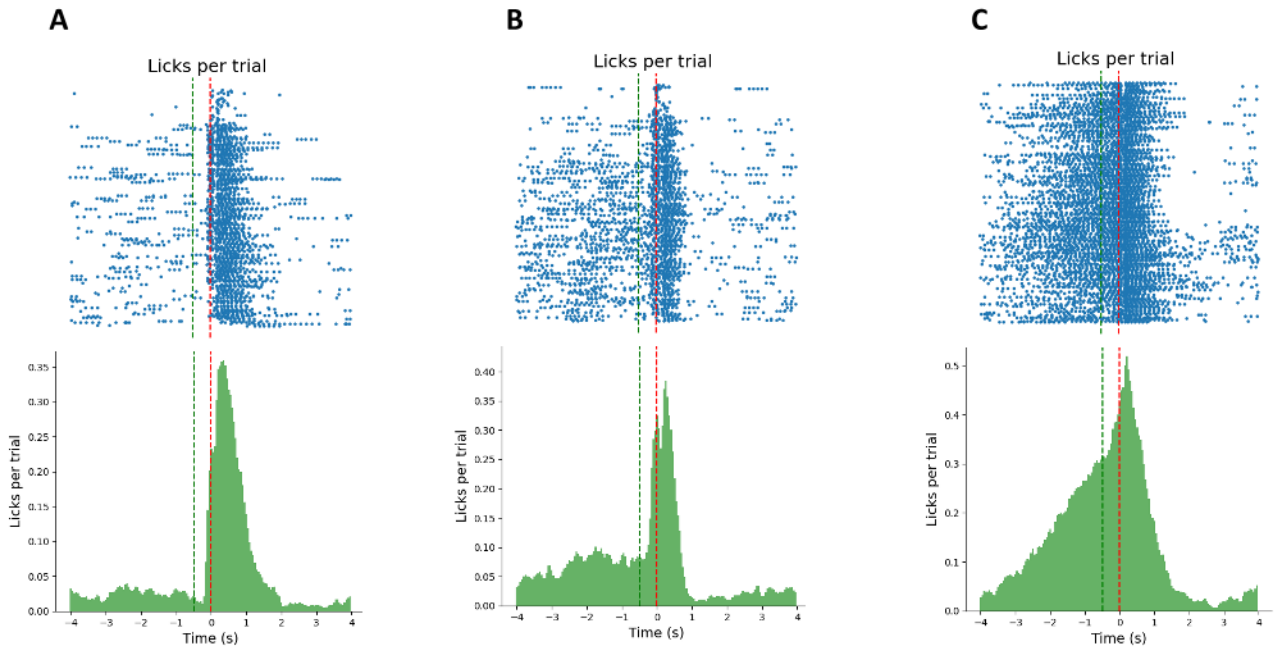



***Fig.33. Behavioral characteristics of group 1 over sessions without ITI shift.*** *The raster plots (top) and PSTH (bottom) display the development of predictive licking behavior in response to the tone (green line) and reward (red line), similar to Fig.17, across three time points: (A) Session 1, (B) Session 3, and (C) Session 14.*

Following ITI change from 5-7 seconds to 7-9 seconds, tone predictive licking initially decreased. However, the mice adapted by gradually learning a new time association over subsequent sessions, highlighting the mice's strong reliance on time assessment for prediction (Fig.34.). Animals in this group consistently improved their ability to anticipate tone timing over days, reinforcing the strength of their time association learning. This strong focus on time-based tone prediction may have limited their ability to differentiate between tones (further discussed in the discussion section). Before the time shift, strong tone-predictive licking was evident (Fig.34.A), with a clear peak just before reward delivery. Although the predictive licking behavior was initially disrupted after the time shift (Fig.34.B, first session), the mice began to recover this behavior over time, albeit less robustly than before the shift (Fig.34.B, 8$^{th}$ session). This indicates that while the original time association was broken, the animals were capable of adapting to the new ITI through relearning.

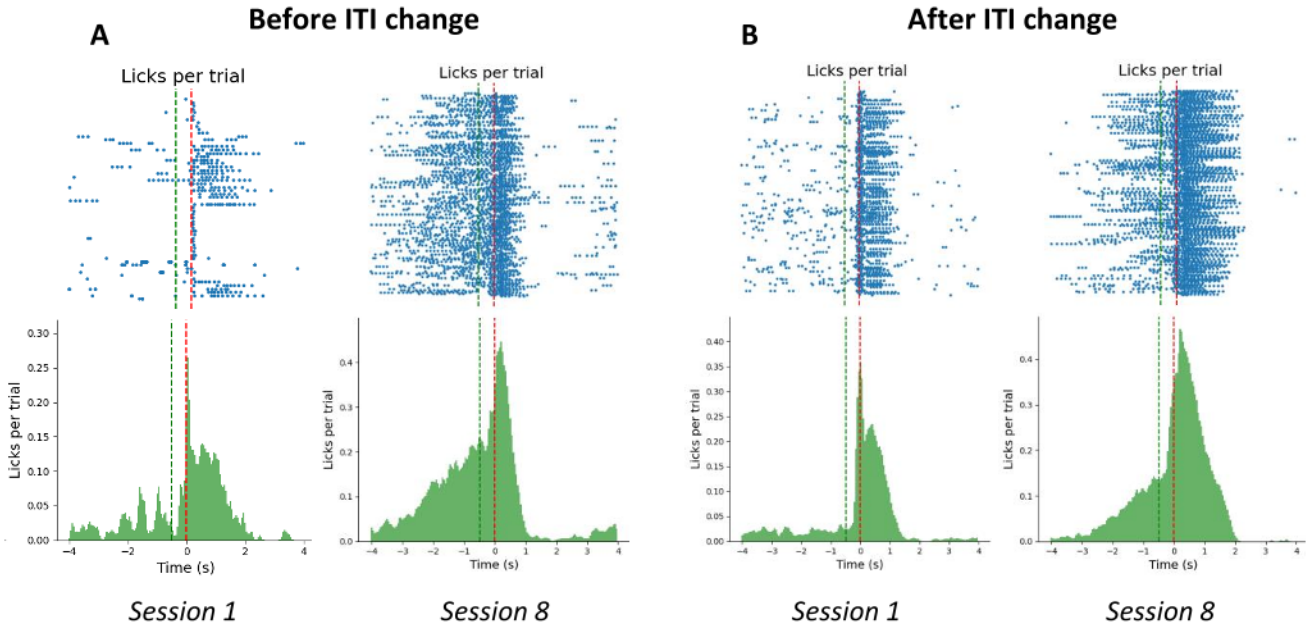



***Fig.34. Example of restoration of tone-predictive licking behavior across multiple sessions following ITI shift.*** *This figure presents an example of one animal's licking behavior using raster plots (top) and Peri-Stimulus Time Histograms (PSTH, bottom), displayed as in Fig.17. In panel (A), the left two graphs show licking behavior before the time shift, with the first graph representing Session 1 and the second showing Session 8. (B) Licking behavior following the time shift to a 7-9 s ITI, showing session 1 post-shift (left) and session 8 post-shift (right). In the first session after the time shift, tone predictive licking is noticeably disrupted, with fewer anticipatory licks occurring before reward delivery. By Session 8 post-shift, some recovery in tone predictive licking is evident; however, the peak remains lower compared to the pre-shift behavior.*

The tone discrimination ratio reflects the mice's ability to differentiate between tones associated with different reward sizes ($\frac{\{large - small\}}{\{large + small\}}$), while tone prediction shows the accuracy of their timing in predicting tone onset licking (after normalization to control licking). Tone prediction shows variability but trends upward, demonstrating a gradual improvement in the mice's ability to anticipate tone timing despite the ITI change. This graph (Fig.35.) captures the distinct yet interacting processes of tone discrimination and timing adaptation in response to the shifted ITI. And can be well seen that this mouse can not distinguish between tones.

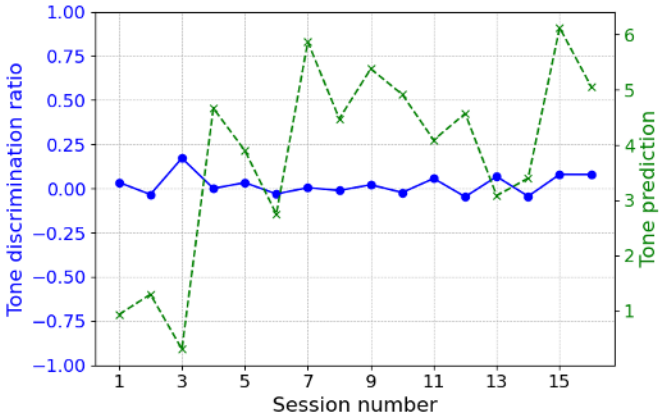



***Fig.35. Increase in tone-predictive licking behavior across sessions following ITI change.*** The graph illustrates an example of one animal of the progression of tone discrimination ratio (left y-axis, blue line) and tone prediction (right y-axis, green dashed line) across sessions following the ITI change.

**Group 2: Early tone discrimination**

This group progressed beyond basic tone-reward associations and quickly learned to distinguish between tones, showing differences in licking behavior to the two trial types. 6 of the 13 mice followed this learning path (3 of the 7-9 second ITI mice and 3 of the 5-7 second ITI mice). Following the ITI shift, 3 additional animals from Group 1 were reclassified into Group 2 ($N_{total}$=10). All mice in this group learned both tone- reward prediction and tone discrimination. While tone prediction was present, it was less pronounced than in Group 1, with a greater emphasis on tone discrimination and reward discrimination (Fig.36 A, B). From Fig.36. it can be seen that on the first day of training, tone and reward predictive licking does not exist (A). In the last session before ITI change, tone predictive licking is being well shown, also the reward prediction is higher for the large reward comparing the small (B).

The learning sequence in these animals was clear: reward prediction, followed by tone prediction based on time association, and then tone and reward discrimination. Even after the time shift, these mice retained their ability to distinguish tones. However, the time association was disrupted, and tone predictive licking for the trial start decreased. In Fig.36. C,D the figure shows raster plots and PSTHs comparing licking behavior for small and large rewards after the time shift. Prior to the shift, tone predictive licking becomes evident with learning, especially for large rewards after the tone (Fig.36.B). After the shift from 5-7 to 7-9 s, tone predictive licking decreases, indicating a disruption in the mice's ability to anticipate tones timing. On the first day after the time shift, the mice appear to anticipate tones that do not arrive, continuing to lick until the tone is presented (Fig.36.C). As the days progress, this time association breaks down further, indicating that the mice's predictive licking becomes less accurate over time, while tone discrimination remained intact (Fig.36.D). A scatter plot (Fig.36.E) illustrates the average number of licks per trial for one

animal, showing a decrease in time-associated licking behavior post-shift (Spearman r = -0.8023, p=0.0001), reflecting the breakdown in time association.

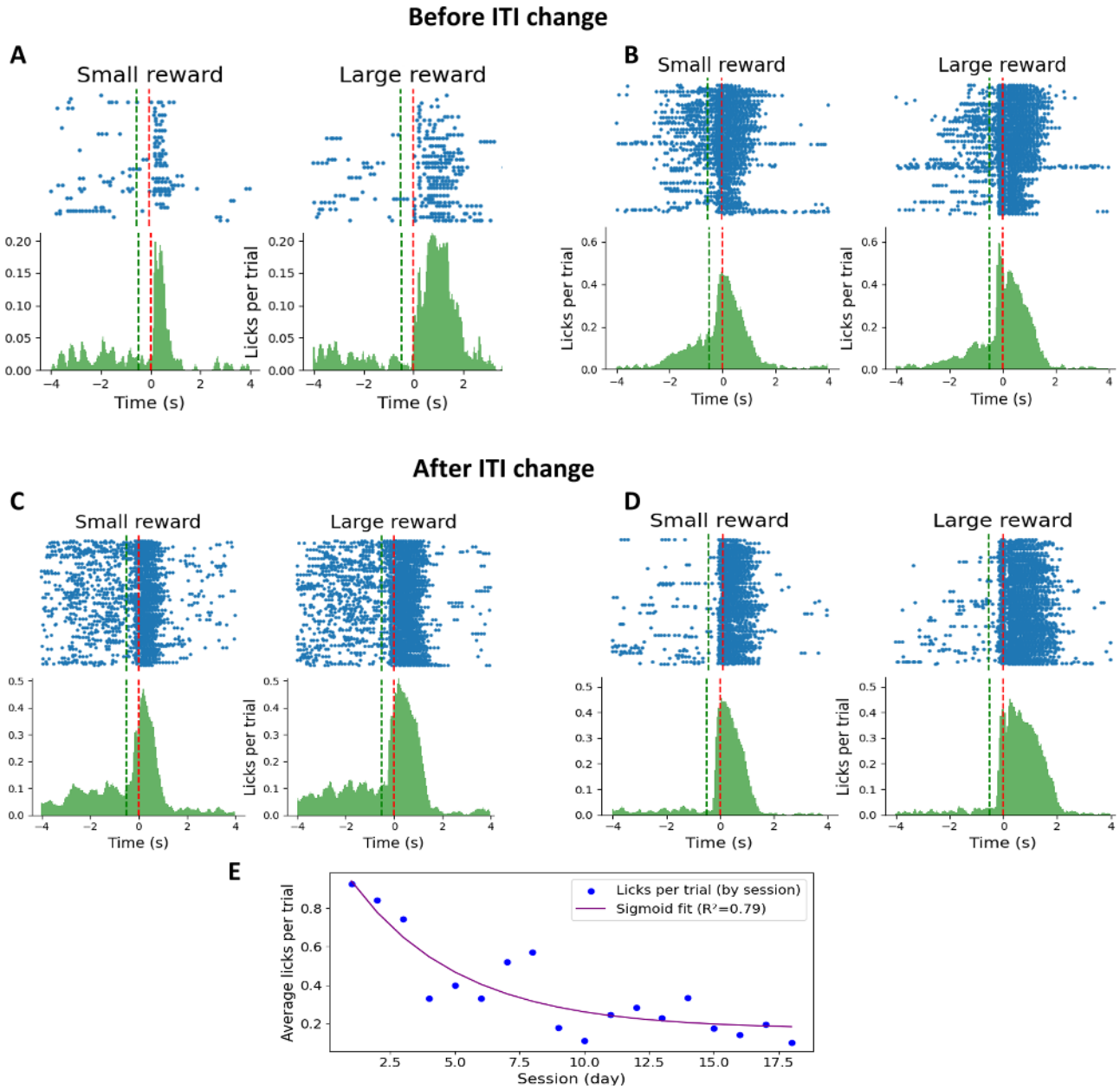



***Fig.36. Effect of ITI change on licking behavior for small and large rewards, showing extinciton of tone prediction.*** *This figure shows an example of one animal's licking behavior with a raster plot (top) and PSTH (bottom), presented as in Fig.17. Licking behavior is compared across two stages: before the ITI shift (top) and after the ITI shift (bottom). (A & C) Shown are licking responses on the first training session before ITI shift (A) and after the ITI shift (C). (B & D) Shown are licking responses after learning stabilized with short it is (B) and long it is (D). Each panel shows responses to small (left) and large (right) rewards. (E) This scatter plot illustrates the average number of licks per trial for one animal, highlighting the decrease in tone predictive licking after the interval shift with with a sigmoid fit (purple line, $R^2 = 0.79$) capturing the downward trend (Spearman r = -0.8023, p = 0.0001).*

**The transition group: Breakdown of tone prediction leading to tone discrimination**

This sub group consists of mice originally from Group 1 (3 out of 6) that were reclassified after the ITI shift. Initially, these mice learned both tone prediction and reward prediction with a 5-7 s ITI. However, when the ITI was extended to 7-9 s, this change caused them to shift from time-based tone prediction to tone discrimination, similar to the learning pattern seen in Group 2. Over time, they also developed reward discrimination, showing increased licking for larger rewards and distinct licking patterns in response to different tones. Among the 8 mice that experienced the ITI shift (6 from Group 1 and 2 from Group 2), only these 3 mice originally from Group 1 displayed this transition from tone prediction to tone discrimination (see Table.2.). As the ITI shifted, their previously established time association weakened, leading to a reduction in licks associated with tone prediction across sessions. This disruption enabled tone discrimination to emerge-a change that only became apparent after the ITI adjustment. Figure 37 presents an example of one animal from this group, which initially showed strong tone prediction rather than tone discrimination.

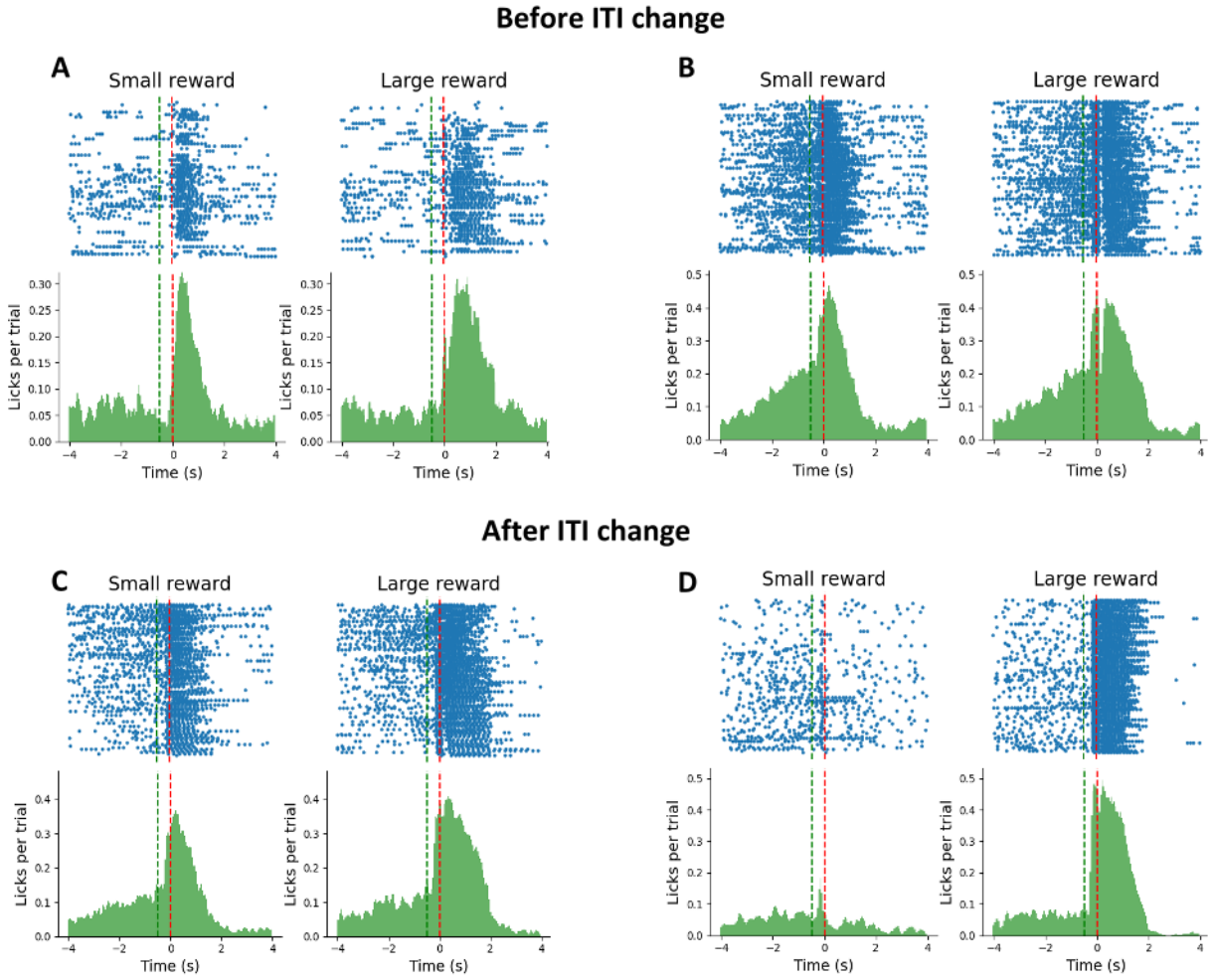


***Fig.37. Transition from tone prediction to Tone discrimination following ITI change.***
*This figure shows an example of one animal's licking behavior with a raster plot (top) and PSTH (bottom), presented as in Fig.17. Licking behavior is compared across two stages: before the ITI shift (top) and after the ITI shift (bottom). (A & C) Shown are licking responses on the first training session before ITI shift (A) and after the ITI shift (C). (B & D) Shown are licking responses after learning stabilized before ITI shift (B) and after ITI shift (D).*

Before the ITI shift, these mice displayed behavior characteristic of Group 1, relying primarily on time-based tone prediction. However, extending the ITI disrupted this timing-based association, leading to a decline in tone prediction. As a result, the animal adapted by shifting its behavior toward tone discrimination, which eventually progressed into reward discrimination- a hallmark of Group 2. Prior to the ITI adjustment, licking behavior did not differ significantly between trials with small and large rewards, nor was there predictive licking linked to specific rewards. Following the ITI shift, distinct differences in licking patterns emerged, reflecting the animal's adaptation to the new interval. This behavioral shift illustrates a transition from dependence on timing cues to reliance on reward-related cues for decision-making, as demonstrated in two mouse examples shown in Figure 38.

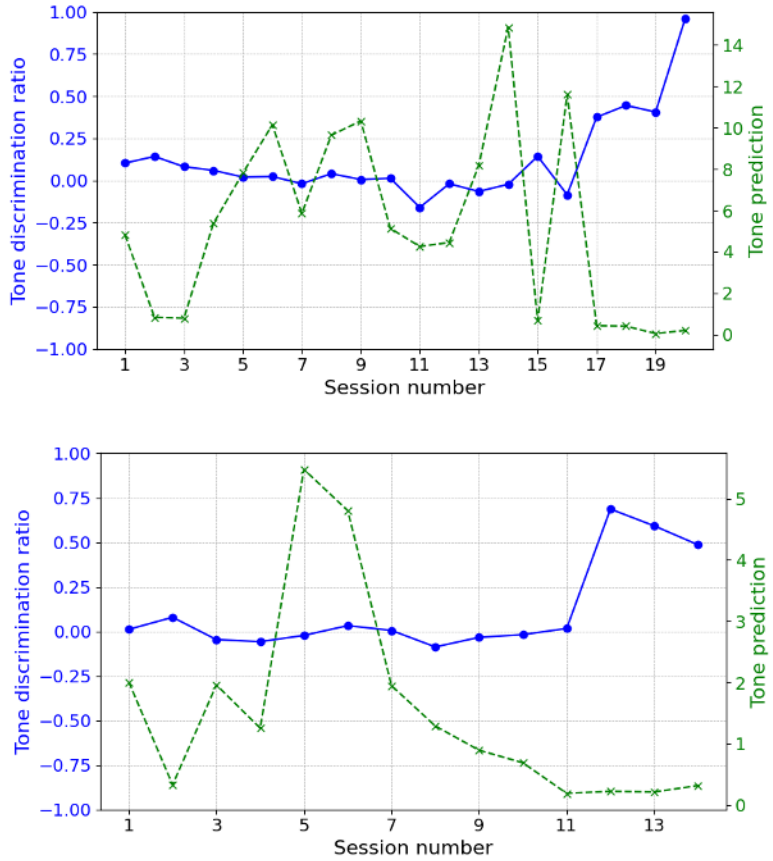



***Fig.38. Tone prediction decrease leads to tone discrimination over course of sessions.*** *The graph illustrates an example of two animals of the progression of tone discrimination ratio (left y-axis, blue line) and tone prediction (right y-axis, green dashed line) across sessions following the ITI change.*

***DISSCUSSION:***

This study aimed to develop methods for investigating the cerebellum's role in reinforcement learning, focusing on its influence on prediction error signals, timing-based associations, and adaptive responses to reward cues. Using a custom-designed experimental system, I examined mice behavior in short- and long-term associative paradigms to isolate key aspects of predictive learning, including reward expectation, timing, and cue identification. Distinct learning strategies, such as time-based tone prediction and reward or tone discrimination, were measured to understand how timing, reward magnitude, and ITI shifts modulate reliance on time- or cue-based predictions. The experimental system integrates motor tracking, auditory cues, precise reward delivery, and a VR environment tailored to the visual capabilities of mice. This design provides a structured and engaging platform for studying reinforcement learning, enabling mice to perform motor responses based on distance while forming reward predictions. The reward delivery system and licking port offer insights into motivation-driven behaviors, while the VR environment ensures alignment of sensory cues with associative learning models. Incorporating DeepLabCut for movement tracking further enhances the system's ability to analyze locomotor adaptations, offering precise control over sensory cues, motor behavior, and reward timing. These features enable rigorous investigation of reinforcement learning processes and highlight the cerebellum's contributions to predictive learning and adaptive behavior.

The long-term association task: The long-term association task designed in this study demonstrates the intricate learning dynamics driven by reinforcement structures. Through this task, I was able to observe how mice adapted their behaviors in response to reward-based cues and timing intervals, thereby shedding light on the underlying mechanisms of associative learning and how task parameters influence the learning process. The results reveal that with optimal ITI settings, mice were able to overcome previous learning biases, indicating that ITI manipulation plays a pivotal role in modulating task engagement. The manipulation of ITI durations significantly influenced how mice formed associations between auditory cues and rewards. With shorter ITIs (e.g., 5 s), mice exhibited behaviors affected by rewards from previous trials, indicating a "carry-over" motivation where recent reward history influenced subsequent task performance. This phenomenon aligns with findings that animals with brief ITIs may perceive each trial as a continuation of the previous one, thereby blurring the distinctions between individual trials. Such carry-over effects can interfere with associative learning and the independent assessment of each trial[54,55]. Upon increasing the ITI

duration, the influence of previous trials on current performance diminished, enabling mice to treat each trial independently. This shift in behavior illustrates a critical point: a sufficiently long ITI allows mice to “reset” between trials, enabling them to focus on the auditory cues in the current trial rather than relying on previous trials’ rewards. These findings highlight the importance of balancing the ITI duration to foster clear associations with the auditory cue, ensuring that reward prediction and motivation are guided by task-specific signals rather than the influence of previous trials. However, an excessively long ITI may lead to disengagement, where mice lose interest in the task. This result highlights how adjusting ITI can serve as a powerful tool in modulating learning, creating conditions that either foster or inhibit associative learning based on the duration and timing of reinforcement signals. Indeed, one of the most compelling observations in this task was the distinct behavioral shift that emerged once an optimal ITI was established. With an ITI set at 12 s for some mice and extended to 15 or 20 seconds for others, effective learning was achieved, allowing the animals to reliably associate auditory cues with reward size. Notably, upon reaching the optimal ITI, mice displayed behaviors suggesting “insightful learning”, where they seemed to suddenly understand the task’s requirements, demonstrating an intuitive grasp of the association between cue and reward. This phase of insightful learning is essential to investigate the cerebellum’s role in prediction, particularly in the hypothesized mechanism where the cerebellum acts as a “red flag”, signaling a discrepancy when expectations are unmet, thereby helping to reset prior associations. To explore this hypothesis, the task was designed in stages. Initially, mice were trained to run 50 cm in the virtual reality corridor without differentiating between reward sizes or tones at the start of the track. Only after seven days of training was the complete task introduced, allowing the mice to associate a specific tone with a corresponding reward. This progression lays the groundwork for later testing with optogenetic interventions during the tone reception phase to pinpoint the cerebellum's role in prediction and neuropixel recording.

Behavioral adaptations in response to large vs. small rewards: The marked behavioral distinctions between large and small reward trials offer significant insights into the motivational adaptations linked to reward size. Mice ran faster, stopped less frequently, and completed trials more quickly when anticipating larger rewards. These adjustments in velocity and stop duration suggest a robust motivational response to reward magnitude. By running faster and minimizing stops (both number of stops and their durations), mice displayed an efficient, goal-directed behavior, indicating that they not only recognized the cues associated with larger rewards but also adapted their motor

responses to maximize reward acquisition speed. In contrast, small reward trials were associated with lower speeds, reduced movement percentage, more stops, and prolonged stop durations, reflecting a less motivated and enhanced exploratory approach. Interestingly, this behavior that emerged suddenly during the “enlightenment process”, had velocities decrease after learning, both for trials with large and small rewards. While a velocity decrease for small-reward trials was expected, the surprising decline in velocity for large-reward trials suggests a shift in the mice's behavior. This reduction in running speed, alongside a decrease in movement percentage, may indicate that the mice associated the tone with the endpoint rewards. Instead of running at a specific speed purely to obtain rewards (as they did before), the mice appeared to slow down, potentially to better discern differences and refine their responses based on the learned association.

Contributions to associative learning: The decision to use a long-term association task was driven by its capacity to reveal the layered process of forming durable, task-specific associations. Unlike short-term associations, which often emphasize immediate reward prediction, the long-term paradigm allowed mice to navigate a virtual environment and integrate motor actions over a longer period, creating an association that required sustained engagement. The task’s design facilitated a gradual buildup of anticipatory behavior, enabling us to isolate the moments when mice transitioned from responding to past rewards to independently associating cues with current rewards. In addition, the long-term association task provides a valuable framework for future studies aiming to dissect the cerebellum’s role in more complex learning scenarios, such as those that involve delayed or compound reinforcement signals. This task enables studying predictive learning across extended timeframes, mirroring real-world learning situations that require sustained attention and the integration of multiple cues. The behavioral data collected from velocity profiles, stop duration, and movement patterns offer a rich basis for further analyses, potentially revealing neural mechanisms underlying reward-based motivation and timing behavior.

The short-term association task: The short-term association task sheds light on the nuanced learning strategies mice employ to link cues with rewards, offering insights into predictive learning processes. This task revealed how mice adaptively responded to auditory cues, transitioning from basic reward prediction to tone prediction, tone discrimination, and finally, for some, to reward discrimination. By grouping the mice, we were able to investigate both individual and group

learning strategies, noting distinct differences in how animals processed cue-reward associations and the impact of ITI adjustments.

Learning stages in predictive and reward-based behavior: Mice exhibited rapid acquisition of reward prediction, associating an auditory cue with reward delivery within just one to two days, as evidenced by synchronized predictive licking behavior. This quick adaptation supports theories that short-term associative learning may depend on immediate reward signals, with the cerebellum- or possibly basal ganglia- facilitating rapid adjustments by processing prediction errors. Following reward prediction, mice advanced to anticipate the tone's timing based on ITI, with all animals mastering this time-based prediction within a few days. For some animals, learning progressed beyond time-based tone prediction to tone discrimination, where they differentiated between cues associated with different reward sizes. This stage revealed individual variability, as 76% of animals (10 out of 13) learned to distinguish between tones, displaying selective licking in response to cues signaling larger rewards. Following tone discrimination, reward discrimination emerged, with animals adapting distinct licking patterns based on reward size, reflecting a deeper association between cues and rewards. Notably, only animals that initially acquired tone discrimination demonstrated this reward-specific behavior, suggesting that their responses were driven by learned associations rather than thirst alone. An exceptional case involved an animal that transitioned from tone discrimination to focusing solely on reward size, disregarding the cue altogether. This shift suggests individual differences in learning strategies, with some animals prioritizing immediate reward signals over predictive cues. Mice that successfully completed tone discrimination demonstrated the ability to differentiate between rewards during reward discrimination learning. After this brief window, licking transitioned to being more sensory-driven, gradually tapering off as the reward was consumed. This behavior suggests that animals engaging in predictive licking prioritized learned associations over instinctual thirst. For mice in Group 1, particularly before and after the ITI shift, the behavior appeared to focus on time-based predictions (tone prediction) rather than reward discrimination. These mice prioritized obtaining the reward itself, with no significant differentiation in their licking behavior based on reward size. In contrast, mice focused on tone discrimination (group 2) developed a more nuanced understanding, distinguishing between tones associated with different reward sizes and adjusting their licking behavior accordingly.

Two alternative explanations for the observed behavior are worth considering. First, if licking was driven solely by the perception of varying liquid amounts, all mice would be expected to lick more for larger rewards. However, this was not the case, as only mice that successfully learned tone discrimination consistently demonstrated reward discrimination, particularly within the initial 250 ms window. Second, if the mice did not prioritize reward size, no differences in licking behavior would be expected. However, the data show that mice undergoing tone discrimination also acquired reward discrimination later in the learning process. The close correlation between these two forms of learning indicates that the ability to distinguish tones was strongly linked to the subsequent capacity for reward discrimination.

Group transition: When mice transitioned from group 1 (tone prediction focus) to group 2 (tone discrimination focus), the ITI shift played a pivotal role. The disruption in time association caused by the ITI change led to a decline in tone-predictive licking over subsequent sessions. This breakdown allowed the mice to shift their focus toward distinguishing between tones and their associated meanings. Over time, licking behavior for tone prediction at the trial's start decreased, while the distinction in licking behavior between tones for large and small rewards became more pronounced. These mice not only learned to differentiate between tones but also progressed to reward learning, displaying stronger licking behavior for larger rewards. This complete learning process highlights their ability to associate tones with reward magnitude effectively. All animals demonstrated a learning sequence starting with reward prediction, followed by tone prediction as a time-based association. While some mice (group 1) remained focused on tone prediction throughout the experiment, others (group 2) advanced to tone discrimination and ultimately reward discrimination. I assume that the mice in group 2, which focused on learning the association between the tone and specific rewards, were motivated by the relationship between the tone and the nature of the reward. In contrast, the mice in group 1 were primarily concerned with the timing of the reward, irrespective of its magnitude. These mice did not differentiate between small or large liquid rewards, suggesting their focus was on when the reward would appear, rather than its association with the tone. This behavior may be due to their indifference to reward magnitude, with the primary concern being the arrival of the reward itself. A subset required the ITI shift to transition between these phases. The diversity in learning strategies across groups reflects varied approaches to task demands. Some animals stayed in the early phases of reward and tone prediction, while others developed more complex behaviors, such as differentiating between tones

and reward sizes. These findings underscore the distinct learning trajectories among the groups, ranging from stable tone-to-reward predictions to nuanced behaviors involving reward size anticipation and post-reward adaptations. This diversity provides valuable insights into the neural processes underlying predictive learning and reward-driven behaviors. Manipulating the ITI revealed critical insights into time-based associations and the emergence of distinct learning strategies. Initially, with a 5-7 s ITI, most mice (6 out of 8) relied on time-based tone prediction, using timing cues to anticipate reward delivery. However, extending the ITI to 7-9 s resulted in a divergence of strategies: some mice (Group 1, N=3) continued to rely on timing cues, while others (Group 2, N=5) shifted to tone discrimination, including three mice who had previously been time-focused. This shift demonstrates how altering the temporal framework of a task can redirect associative learning, with timing consistency breakdowns promoting a focus on cue discrimination. Group 1 re-established tone prediction, while Groups 2 stayed or transitioned to tone discrimination. These distinctions highlight predictive learning as a flexible process that adapts to changes in task structure, potentially engaging different neural circuits. The division into these groups emphasizes the complexity of learning strategies, perhaps suggesting distinct roles for cerebellar and basal ganglia circuits in time- and cue-based associations. By carefully adjusting ITI, this study highlights how the brain may flexibly support either temporal predictions or cue-based reward discrimination. The emergence of distinct learning groups offers valuable insights into the diversity of associative learning pathways, laying the groundwork for future research on cerebellar functions in reinforcement learning.

Future Studies: Building on these findings, this system offers a strong platform for future studies using optogenetic manipulations to dissect cerebellar contributions to learning and reward processing. The data collected on motor and behavioral responses will serve as a valuable baseline, enabling direct comparisons between control and optogenetically manipulated mice to identify cerebellar-specific roles. Moreover, this research highlights the cerebellum's integrative role in adaptive behavior, suggesting that reinforcement paradigms developed here could be translated to study human learning and decision-making processes. Such investigations may provide insights into the cerebellum's contributions to cognitive and affective functions, paving the way for broader applications in neuroscience and translational research.

Short-term association task manipulation: to probe the cerebellum’s involvement in reinforcement learning traditionally attributed to basal ganglia circuits, future studies will employ optogenetic

manipulations to inhibit CN by activation of PCs. Task performance between control and manipulated mice will be compared. Disruption in performance due to inhibition would indicate an active cerebellar role; if unaffected, it may suggest basal ganglia dominance. Neuropixels recordings will be used to analyze cerebellar and basal ganglia contributions to prediction and learning. In short ITI tasks, cerebellar inhibition during the reward phase may have minimal impact, as basal ganglia circuits likely dominate this stage. Conversely, in longer ITI tasks, inhibition during cue presentation could disrupt learning by preventing association formation, resulting in indistinguishable behaviors across trials with varying reward sizes. This approach will clarify the cerebellum's role in linking cues to reward predictions and its interaction with basal ganglia circuits.

Long- term association task manipulation: In the long-term task, optogenetic inhibition of CN at specific phases will examine whether the cerebellum generates predictive signals critical for reinforcement learning. Key phases under investigation include: (1) the cue-receiving phase: Auditory cues signal anticipated reward sizes. Inhibition during this phase, particularly with optimal ITIs, may disrupt association formation, impairing the mice's ability to link cues to rewards and resulting in uniform behaviors (e.g., similar running speeds) across trials. (2) Reward-receiving phase, Inhibition during reward delivery will test whether the cerebellum contributes to motivation and prediction error processing. This study hypothesizes that cerebellar signals detect discrepancies between expected and actual outcomes, "flagging" estimation errors to dissolve outdated associations and form new ones. By generating internal predictive models, the cerebellum may actively compare anticipated and actual outcomes, providing error signals essential for adaptive learning.

Prediction error and reward processing: To explore how cerebellar output influences motivation and prediction error, Neuropixel recordings from the cerebellum and the VTA will track neural responses during reinforcement learning. Experiments will compare performance under conditions of predicted and unpredicted reward omissions, both with and without CN inhibition. Observing behavioral responses to omission cues (e.g., reduced running speed or licking activity) will determine whether the cerebellum encodes sensory, reward, or generalized prediction errors. Randomized rewards or tones will further assess cerebellar and basal ganglia responses, offering a comprehensive view of cerebellar contributions to adaptive learning and cue-based versus time-based predictions.

Summary: This study investigates the cerebellum's role in reinforcement learning, focusing on prediction error processing and adaptive responses to reward cues. The findings emphasize the cerebellum's role in modulating behavior based on timing and reward expectations. Short ITIs appear to sustain motivation across trials, while longer ITIs facilitate clearer cue-reward associations, enabling adaptive learning driven by cerebellar error signals. The integration of motor tracking, sensory cues, optogenetics, and Neuropixels recordings provides a powerful framework for future research. Planned experiments will clarify cerebellar contributions to prediction errors, reward processing, and its interaction with basal ganglia circuits. Ultimately, this research establishes a foundation for understanding the cerebellum's pivotal role in adaptive reinforcement learning and its broader implications for motivation and cognitive processing.

## ***References***


1. Doya, K. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks* **12**, (1999).

2. Keller, G. B. & Mrsic-Flogel, T. D. Predictive Processing: A Canonical Cortical Computation. *Neuron* **100**, 424–435 (2018).

3. Bond, K. M. & Taylor, J. A. Flexible explicit but rigid implicit learning in a visuomotor adaptation task. *J Neurophysiol* **113**, (2015).

4. Lee, D., Seo, H. & Jung, M. W. Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience* vol. 35 Preprint at https://doi.org/10.1146/annurev-neuro-062111-150512 (2012).

5. Raymond, J. L. & Medina, J. F. Computational Principles of Supervised Learning in the Cerebellum. *Annu Rev Neurosci* **41**, 233–253 (2018).

6. Popa, L. S., Streng, M. L., Hewitt, A. L. & Ebner, T. J. The Errors of Our Ways: Understanding Error Representations in Cerebellar-Dependent Motor Learning. *The Cerebellum* **15**, 93–103 (2016).

7. Kim, O. A., Ohmae, S. & Medina, J. F. A cerebello-olivary signal for negative prediction error is sufficient to cause extinction of associative motor learning. *Nat Neurosci* **23**, 1550–1554 (2020).

8. Catz, N., Dicke, P. W. & Thier, P. Cerebellar complex spike firing is suitable to induce as well as to stabilize motor learning. *Current Biology* **15**, (2005).

9. Ito, M. Neural design of the cerebellar motor control system. *Brain Res* **40**, 81–84 (1972).

10. Izawa, J., Criscimagna-Hemminger, S. E. & Shadmehr, R. Cerebellar Contributions to Reach Adaptation and Learning Sensory Consequences of Action. *The Journal of Neuroscience* **32**, 4230–4239 (2012).

11. O'Reilly, J. X., Mesulam, M. M. & Nobre, A. C. The cerebellum predicts the timing of perceptual events. *Journal of Neuroscience* **28**, (2008).

12. Heffley, W. *et al.* Coordinated cerebellar climbing fiber activity signals learned sensorimotor predictions. *Nat Neurosci* **21**, 1431–1441 (2018).

13. Brooks, J. X., Carriot, J. & Cullen, K. E. Learning to expect the unexpected: Rapid updating in primate cerebellum during voluntary self-motion. *Nat Neurosci* **18**, (2015).

14. Kostadinov, D. & Häusser, M. Reward signals in the cerebellum: Origins, targets, and functional implications. *Neuron* **110**, 1290–1303 (2022).

15. Nikooyan, A. A. & Ahmed, A. A. Reward feedback accelerates motor learning. *J Neurophysiol* **113**, 633–646 (2015).

16. Galea, J. M., Mallia, E., Rothwell, J. & Diedrichsen, J. The dissociable effects of punishment and reward on motor learning. *Nat Neurosci* **18**, 597–602 (2015).

17. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science (1979)* **275**, (1997).

18. Keiflin, R., Pribut, H. J., Shah, N. B. & Janak, P. H. Ventral Tegmental Dopamine Neurons Participate in Reward Identity Predictions. *Current Biology* **29**, (2019).

19. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* vol. 482 Preprint at https://doi.org/10.1038/nature10754 (2012).

20. Carta, I., Chen, C. H., Schott, A. L., Dorizan, S. & Khodakhah, K. Cerebellar modulation of the reward circuitry and social behavior. *Science (1979)* **363**, (2019).

21. Pisano, T. J. *et al.* Homologous organization of cerebellar pathways to sensory, motor, and associative forebrain. *Cell Rep* **36**, (2021).

22. Sendhilnathan, N., Semework, M., Goldberg, M. E. & Ipata, A. E. Neural Correlates of Reinforcement Learning in Mid-lateral Cerebellum. *Neuron* **106**, 188-198.e5 (2020).

23. Hull, C. Prediction signals in the cerebellum: Beyond supervised motor learning. *Elife* **9**, (2020).

24. Kostadinov, D., Beau, M., Pozo, M. B. & Häusser, M. Predictive and reactive reward signals conveyed by climbing fiber inputs to cerebellar Purkinje cells. *Nat Neurosci* **22**, (2019).

25. Wagner, M. J., Kim, T. H., Savall, J., Schnitzer, M. J. & Luo, L. Cerebellar granule cells encode the expectation of reward. *Nature* **544**, 96–100 (2017).

26. Popa, L. S., Streng, M. L. & Ebner, T. J. Long-term predictive and feedback encoding of motor signals in the simple spike discharge of purkinje cells. *eNeuro* **4**, (2017).

27. Kim, H. E., Parvin, D. E. & Ivry, R. B. The influence of task outcome on implicit motor learning. *Elife* **8**, (2019).

28. Larry, N., Yarkoni, M., Lixenberg, A. & Joshua, M. Cerebellar climbing fibers encode expected reward size. *Elife* **8**, (2019).

29. Sendhilnathan, N., Ipata, A. & Goldberg, M. E. Mid-lateral cerebellar complex spikes encode multiple independent reward-related signals during reinforcement learning. *Nat Commun* **12**, (2021).

30. Heffley, W. & Hull, C. Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. *Elife* **8**, (2019).

31. Ohmae, S. & Medina, J. F. Climbing fibers encode a temporal-difference prediction error during cerebellar learning in mice. *Nat Neurosci* **18**, 1798–1803 (2015).

32. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn Sci* **2**, 338–347 (1998).

33. McCormick, D. A. & Thompson, R. F. Cerebellum: Essential involvement in the classically conditioned eyelid response. *Science (1979)* **223**, (1984).

34. Tseng, Y. W., Diedrichsen, J., Krakauer, J. W., Shadmehr, R. & Bastian, A. J. Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *J Neurophysiol* **98**, (2007).

35. Kameda, M., Ohmae, S. & Tanaka, M. Entrained neuronal activity to periodic visual stimuli in the primate striatum compared with the cerebellum. *Elife* **8**, (2019).

36. Gastwirth, J. L., Gel, Y. R. & Miao, W. The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice. *Statistical Science* **24**, 343–360 (2009).

37. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* **21**, (2018).

38. Warren, R. A. *et al.* A rapid whisker-based decision underlying skilled locomotion in mice. *Elife* **10**, (2021).

39. Williams, G. A. & Jacobs, G. H. Cone-based vision in the aging mouse. *Vision Res* **47**, 2037–2046 (2007).

40. Jacobs, G. H., Williams, G. A. & Fenwick, J. A. Influence of cone pigment coexpression on spectral sensitivity and color vision in the mouse. *Vision Res* **44**, 1615–1622 (2004).

41. Jacobs, G. H., Williams, G. A. & Fenwick, J. A. Influence of cone pigment coexpression on spectral sensitivity and color vision in the mouse. *Vision Res* **44**, 1615–1622 (2004).

42. Calderone, J. B. & Jacobs, G. H. *Regional Variations in the Relative Sensitivity to UV Light in the Mouse Retina*. *Visual Neuroscience* vol. 12 (1995).

43. Szatko, K. P. *et al.* Neural circuits in the mouse retina support color vision in the upper visual field. *Nat Commun* **11**, (2020).

44. Daniel J Denman Is a corresponding authorJennifer A LuvianoDouglas R OllerenshawSissy CrossDerric WilliamsMichael A BuiceShawn R OlsenR Clay Reid. Mouse color and wavelength-specific luminance contrast sensitivity are non-uniform across visual space.

45. Ehret, G. Age-dependent hearing loss in normal hearing mice. *Naturwissenschaften* **61**, 506–507 (1974).

46. Heffner, H. E. & Heffner, R. S. Hearing ranges of laboratory animals. *Journal of the American Association for Laboratory Animal Science* vol. 46 Preprint at (2007).

47. Lopes, G. *et al.* Bonsai: An event-based framework for processing and controlling data streams. *Front Neuroinform* **9**, (2015).

48. Brown, A. M. *et al.* Purkinje cell misfiring generates high-amplitude action tremors that are corrected by cerebellar deep brain stimulation. *Elife* **9**, (2020).

49. Zhang, F., Wang, L. P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nat Methods* **3**, (2006).

50. Zhao, S. *et al.* Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. *Nat Methods* **8**, (2011).

51. Mahn, M. *et al.* High-efficiency optogenetic silencing with soma-targeted anion-conducting channelrhodopsins. *Nat Commun* **9**, (2018).

52. Prestori, F., Montagna, I., D'angelo, E. & Mapelli, L. The optogenetic revolution in cerebellar investigations. *International Journal of Molecular Sciences* vol. 21 Preprint at https://doi.org/10.3390/ijms21072494 (2020).

53. Gruver, K. M. & Watt, A. J. Optimizing Optogenetic Activation of Purkinje Cell Axons to Investigate the Purkinje Cell – DCN Synapse. *Front Synaptic Neurosci* **11**, (2019).

54. Goltstein, P. M., Reinert, S., Glas, A., Bonhoeffer, T. & Hübener, M. Food and water restriction lead to differential learning behaviors in a head-fixed two-choice visual discrimination task for mice. *PLoS One* **13**, (2018).

55. Sanchez-Roige, S., Peña-Oliver, Y. & Stephens, D. N. Measuring impulsivity in mice: The five-choice serial reaction time task. *Psychopharmacology* vol. 219 253–270 Preprint at https://doi.org/10.1007/s00213-011-2560-5 (2012).

**תקציר התיזה: תפקיד המוחון בלמידת חיזוק באמצעות שגיאת חיזוי.**

לפני למעלה משני עשורים הציע דויה כי המוחון (הצרבלום) מנצל למידה מונחית הנשענת על שגיאות חיזוי תחושתיות (SPEs) ליצירת הסתגלות סנסורימוטורית, בעוד הגרעינים הבזאליים (BG) משתמשים בלמידת חיזוק הנשענת על שגיאות חיזוי תגמול (RPEs) למיקסום תגמולים עתידיים. הצעה זו הנחתה מחקר אודות האופן שבו מבנים מוחיים שונים מנצלים אלגוריתמים ספציפיים ללמידה. עם זאת, ממצאים עדכניים מצביעים על כך שהמוחון עשוי למלא תפקיד גם בלמידת חיזוק על ידי קידוד מידע הקשור לתגמול, מה שמערער את ההבחנה המסורתית בין שני מבנים אלה. שגיאות חיזוי תחושתיות (SPEs) מועברות דרך סיבים (CF) בקליפת המוחון, בעוד ששגיאות חיזוי תגמול (RPEs) מועברות דרך נוירונים דופמינרגיים מאזור הVTA והחלק הקומפקטי של הסובסטנסיה ניגרה (SNc). ממצאים חדשים על אותות תגמול במוחון מטשטשים עוד יותר את ההבחנה התפקודית בין המוחון לגרעינים הבזאליים, במיוחד לאור הקשרים הדו-סינפטיים ההדדיים ביניהם.

המחקר בוחן את תפקיד המוחון בלמידת חיזוק, במיוחד ביכולתו להפיק אותות שגיאת חיזוי עבור תגמולים בלתי צפויים וליצור קשרי תגמול חדשים לצורך התאמת ציפיות עתידיות. לשם כך פותחה מערכת ניסויית מותאמת, שמשלבת מעקב התנהגותי, מציאות מדומה ואותות קוליים כדי להעריך למידה אסוציאטיבית ותחזיתית בעכברים. העכברים אומנו בשתי מטלות: (1) מטלת אסוציאציה ארוכת-טווח שבה עכברים למדו לרוץ במרחק קבוע במסדרון וירטואלי שבו צליל בראש המסלול מנבא את גודל התגמול שיקבלו בסופו, ו-(2) מטלת אסוציאציה קצרת- טווח, שבה תגמולים הוצגו לאחר עיכוב קבוע מהצליל המנבא את גודל התגמול.

במטלת האסוציאציה ארוכת-הטווח, הצלחנו להדגים את יכולת המערכת לעודד למידה בסביבה מאתגרת. באמצעות בחירת מרווח בין ניסוי (ITI) מותאם בקפידה, העכברים הפגינו התנהגות ציפייה ברורה שהתבטאה בהתאמת מהירות הריצה, אחוזי התנועה, משכי העצירה וזמן השלמת הניסוי בהתאם לרמזים על גודל התגמול. לעומת זאת, עכברים שאומנו עם ITIs קצרים הסתמכו בעיקר על מוטיבציות שנגזרו מתוצאת הניסוי הקודם, במקום לפתח אסוציאציות ספציפיות לצלילים. עם זאת, כאשר ה-ITI הוארך, עכברים אלו עברו מהסתמכות על תוצאות הניסוי הקודם ללמידה תחזית מותנת ניסוי עכשווי, ולבסוף למדו את הקשר הנכון בין הצליל לגודל התגמול. ממצא זה מדגיש את החשיבות של כיוונון מדויק של ה-ITI לצורך עידוד למידה אסוציאטיבית חזקה במשימות לבחינת למידת תגמול. מטלה זו בחנה את יכולת העכברים ליצור תחזיות במסגרת אסוציאציה ארוכת-טווח, תוך התמודדות עם שינויים פתאומיים בסביבה שנועדו להקל על למידה תחזיתית - תהליך אשר במחקר זה מיוחס למוחון.

במטלת האסוציאציה קצרת-הטווח, נצפו מגוון אסטרטגיות למידה בקרב העכברים. בתחילה, כל העכברים רכשו יכולת חיזוי תגמול על ידי יצירת אסוציאציה בין צלילים לאותות תגמול קרובים. לאחר מכן, הם התקדמו לחיזוי צליל, כלומר ציפייה לעיתוי הצליל בהתבסס על ה-ITI. בעוד שחלק מהעכברים התרכזו בעיקר בחיזוי מועד הצליל, אחרים התקדמו לאבחנה בין הצלילים, כשהם מבחינים בין רמזים לתגמול קטן ולתגמול גדול. חלק מהעכברים נדרשו להארכת ה-ITI כדי לעבור מחיזוי מבוסס זמן לאבחנה בין הצלילים, ולבסוף הצליחו להבחין בין התגמולים, כשהם מפגינים דפוסי ליקוק

שונים המשקפים ציפייה לתגמול הגדול. משימה זו מדגישה את יכולתם של העכברים לחזות ולהגיב לרמזים ספציפיים המאותתים על גודל תגמול קרב, ובכך ממחישה למידה אסוציאטיבית קצרת-טווח באופן אפקטיבי.

ממצאים אלו יחדיו ממחישים כיצד עכברים מגיבים באופן אדפטיבי לרמזים שמיעתיים ותזמוניים, כשהם עוברים מתחזיות בסיסיות המבוססות על זמן לאבחנות מורכבות יותר המבוססות על תגמולים. המחקר מספק תובנות על מגוון מסלולי למידה המניעים התנהגויות תחזיתיות ומבוססות תגמול בלמידת חיזוק. המערכת המותאמת שפותחה כאן איפשרה חקירה של אסוציאציות קצרות וארוכות-טווח בתנאים מגוונים, וסיפקה הבנה רחבה של למידה אסוציאטיבית ותחזיתית בעכברים. תשתית זו סוללת את הדרך לשלב הבא במחקר, שבו יבוצעו מניפולציות אופטוגנטיות בזמני הרמז המרכזיים והקלטות מהמוחון. צעדים אלו נועדו לחשוף את התפקיד הקריטי של המוחון בלמידת חיזוק על ידי ביצוע תחזיות, ובסופו של דבר לשפוך אור על תרומתו המסתורית לתהליכי הלמידה המרתקים.

עבודה זו נעשתה בהדרכתה של פרופ׳ דנה כהן,

מן היחידה ללימודים בין תחומיים,
המרכז הרב תחומי לחקר המוח ע״ש לסלי וסוזן גונדה.

**אוניברסיטת בר -אילן**

# תפקיד המוחון ותרומתו ללמידת תגמול על ידי חיזוי טעות

**דוד סלובודיאנסקי**

עבודה זו מוגשת כחלק מהדרישות לשם קבלת תואר מוסמך
ביחידה ללימודים בין תחומיים, המרכז הרב תחומי לחקר המוח
ע"ש לסלי וסוזן גונדה של אוניברסיטת בר- אילן

רמת גן תשפ"ה