Prediction error drives associative olfactory learning and conditioned behavior in a spiking model of *Drosophila larva*

Anna-Maria Jürgensen[a], Panagiotis Sakagiannis[a], Michael Schleyer[b,c], Bertram Gerber[b,d,e], and Martin Paul Nawrot[a]

[a]Computational Systems Neuroscience, Institute of Zoology, University of Cologne, Cologne, Germany
[b]Leibniz Institute for Neurobiology (LIN), Department of Genetics, Magdeburg, Germany
[c]Institute for the Advancement of Higher Education, Faculty of Science, Hokkaido University, Japan
[d]Institute for Biology, Otto-von-Guericke University, Magdeburg, Germany
[e]Center for Brain and Behavioral Sciences (CBBS), Otto-von-Guericke University, Magdeburg, Germany

Abstract

Predicting positive or negative reinforcement from environmental clues is essential to guide decision making and goal-directed behavior. In insect brains the mushroom body is a central structure for learning such valuable associations between sensory signals and reinforcement. We propose a biologically realistic spiking network model of the *Drosophila* larval olfactory pathway for the association of odors and reinforcement to bias behavior towards either approach or avoidance. We demonstrate that prediction error coding through integration of present and expected reinforcement in dopaminergic neurons can serve as a driving force in learning that can, combined with synaptic homeostasis, account for the experimentally observed features of acquisition and extinction of associations that depend on the intensity of odor and reward, as well as temporal features of the odor/reward pairing. To allow for a direct comparison of our simulation results with behavioral data we model learning-induced plasticity over the full time course of behavioral experiments and simulate locomotion of individual larvae towards or away from odor sources in a virtual environment.

Introduction

Goal directed behavior in dynamic situations relies on the ability to predict the future conditions in the environment from the occurrence of sensory clues. In insects, the mushroom body (MB) is the central brain structure for learning valuable associations between different sensory inputs, allowing the animal to predict relevant changes [1–8].

The term associative learning describes the gradual establishment of a relationship between two previously unrelated elements. In classical conditioning the conditioned sensory stimulus (CS) obtains behavioral relevance via its concurrence with the reinforcing unconditioned stimulus (US), an acquisition process depending dynamically on their spatiotemporal proximity. The temporal evolution of this process has been formalized in the Rescorla-Wagner (RW) model (eqn. 1) [9].

$$\Delta V = \alpha * (\lambda_{\text{US}} - V(t)),$$
$$V(t+1) = V(t) + \Delta V. \quad (1)$$

Here, a sensory stimulus (CS) obtains predictive power of concurrent or successive reinforcement (US) [9], that depends on the strength of the association between the CS and US $V(t)$, allowing for anticipatory behavior to the expected US [10, 11]. The acquisition of this association terminates when the US is fully predicted. Until then the change in associative strength ΔV is proportional to the difference between the maximum associative strength (or asymptote) λ_{US} and the current associative strength $V(t)$ (eqn. 1). While the maximum associative strength is a property of the US, determined mainly by the intensity of the reinforcement, the current associative strength $V(t)$ is defined by the shared learning history of CS and US [9]. A derivative of this fundamental theory is the concept of prediction error (PE) [12]. In this, the error signal equals the difference between the current λ_{US} and the predicted value of US $V(t)$. Over the course of the memory acquisition/training phase, the pace of learning, which can be formalized as the slope of the acquisition curve $V(t)$, decreases as the PE is reduced, minimizing the driving force for changes of the association [9, 12]. This difference is multiplied with a learning rate parameter (α), here combined for the CS and the US (eqn. 1).

This optimization of predictions, guided by the PE, allows animals to efficiently adapt their goal-directed behavior in dynamic environments. Among the most relevant associations to be learned are those that enable the prediction of reward or punishment. Dopaminergic neurons (DANs) have long been known to encode information about reward and punishment. These types of neurons respond to the presence of rewards [13–18] as well as punishments [18–22] in the environment and the electrical stimulation or optogenetic activation of DANs induces approach or avoidance both in vertebrates [23–26] and invertebrates [6, 22, 27–31]. Ultimately DANs do not only signal the presence of rewards or punishments but have also been suggested to encode PE in various vertebrate species [19, 32–35] and might have a similar function in insects [3, 18, 22, 36–38].

We utilize our model of the *Drosophila* larva MB that forms associations of odors with reinforcement to test the hypothesis that PE coding takes place in DANs within this circuit, via feedback from the MBONs [22, 39, 40] (see Discussion). We demonstrate that this model can reproduce the experimentally observed findings on the acquisition of associations of odors with reinforcement in a time-resolved manner. To facilitate direct qualitative and quantitative comparisons with animal behavioral data we couple this model with a realistic locomotory model of the larva [41] that captures the effects of learned associations on chemotactic behavior in individual animals.

Results

Connectome-based circuit model of the larval olfactory pathway

The network architecture of our model (fig. 1 A) is based on the anatomy of the olfactory pathway in one *Drosophila* larva brain hemisphere [6, 22, 42, 43] (for details see Materials and Methods). Peripheral processing is carried out by 21 olfactory receptor neurons (ORNs), each expressing a different olfactory receptor type [42, 44, 45]. ORNs form one-to-one excitatory synaptic connections with 21 projection neurons (PNs) and 21 local interneurons (LNs) in the antennal lobe [42]. Each LN connects with all PNs via inhibitory GABAergic synapses [42], establishing the motif for lateral inhibition within the antennal lobe. The 72 mature larval Kenyon cells (KCs) [43] are the excitatory principal cells of the mushroom body. Each KC receives excitatory input from 2-6 randomly selected PNs [43]. The KCs are subjected to feedback inhibition, provided via the GABAergic anterior paired lateral neuron (APL), which receives input from all KCs [6]. Only mature KCs, characterized by a fully developed dendrite are included in this model, yielding a complete convergent synaptic KC >APL connectivity. The output region of the mushroom body is organized in compartments, in which the KC axons converge with the dendrites of mushroom body output neurons (MBONs) [22, 43]. Our

model includes two MBONs that are representative of two different categories of output neurons of the MB that mediate either approach or avoidance [2, 46–53]. Both MBONs receive excitatory input from 100% of the KCs to fully capture the information that is normally represented by the complete set of MBONs. Each compartment is also innervated by a single DAN, signalling either reward or punishment.

Learning through KC>MBON plasticity

We assume that the KC>MBON synapses undergo plasticity, based on strong experimental evidence in larval [22, 46, 47] and adult flies [2, 48, 49, 51–54]. This plasticity requires the convergence of the sensory pathways in form of KC activation and of the reinforcing pathway, mediated by neuromodulatory DAN signaling at the synaptic site. We employ a two factor learning rule (eqn. 2) at each KC>MBON synapse (fig. 1 B,C, see Methods). The first factor is expressed in the pre-synaptic KC activation by an odor, rendering the synapse eligible for modification. This is modeled via an exponentially decaying eligibility trace $e_i(t)$, which is increased step-wise by a fixed amount with each KC spike (fig. 1 B). The decay time constant determines the window of opportunity for synaptic change. The presence of reinforcement (reward or punishment) constitutes the second factor and is signaled by the reward-mediating DAN_+ or punishment-mediating DAN_-, respectively. Spiking of the DAN provides a neuromodulatory reinforcement signal $R(t)$ to the presynaptic site. If a DAN spike coincides with a positive eligibility at the synapse, the respective synaptic weight will be reduced (fig. 1 C). At each synapse i, the reduction of synaptic weight Δw_i depends on the learning rate a (table 1) and is proportional to the amplitude of the eligibility trace $e_i(t)$ (fig. 1 B):

$$\Delta w_i = a * e_i(t) * R(t) \leq 0. \quad (2)$$

Synaptic homeostasis (eqn. 3) modulates the effects of plasticity at KC>MBON synapses. With each MBON spike, the current weight w_i of each respective KC>MBON synapse is increased, proportionally to the extend to which the weight has been reduced from its original value w_{init} (table 1) and multiplied with a homeostatic regulation factor h (table 1). This mechanism slowly counteracts the time-integrated effect of synaptic plasticity and thus ensures continued input to the MBONs, despite the reduction of input weights over the course of the learning process (eqn. 2). The homeostatic factor h hereby serves as an implementation of a time constant of this exponential process. The interaction of synaptic plasticity and homeostatic regulation defines the magnitude of the weight at the next simulation timestep $t + \Delta t$ as

$$w_i(t + \Delta t) = w_i(t) + \Delta w_i + (w_{init} - w_i(t)) * h. \quad (3)$$

It has been shown in behavioral experiments that specific MBONs encode a behavioral tendency to either approach or avoid a currently perceived stimulus, depending on the acquired stimulus valence [2, 46, 48–51]. In the naive state of our model all KC>MBON synapses have the same initial weights w_{init} (table 1) and hence the spiking activity of both MBONs is highly similar. Learning alters the KC>MBON synaptic weights and thus skews the previously balanced MBON output. This acquired imbalance between MBON outputs biases behavior towards approach or avoidance of the conditioned odor. To quantify the effect of learning we compute the behavioral bias BB (eqn. 4) from the firing rates of both MBONs over $t = 1s$ as follows:

$$BB = \frac{MBON_+ - MBON_-}{t}. \quad (4)$$

Implementation of prediction error coding in the KC-MBON-DAN motif

In the model, DANs are activated by external reward/punishment signals and also receive excitatory and inhibitory feedback from both MBONs (fig. 1 A). As the initial balance between the two MBON outputs shifts over the course of the training process, the amount of excitatory and inhibitory feedback that DANs receive continues to diverge, allowing the

DANs to access the model's learning history. Ultimately DAN activation signals the difference between the current external activation and the expected activation based on prior learning, implemented as the difference between excitatory and inhibitory MBON>DAN feedback. Including this feedback leads to learning curves that saturate when the reward is fully predicted and the prediction error approaches zero (fig. 2 A). This effect disappears, when the feedback circuit is disabled (fig. 2 A). Increasing reward intensity, learning rate or odor intensity (see Methods) fosters a faster acquisition of the association and increases the maximum strength of the association at the same time (fig. 2 A).

Increasing the reward intensity after a 2.5 min (black curve) or 5 min (gray curve) of appetitive training, results in a steeper slope of the learning curve and also increases the maximum during training trials of 2.5min duration with increased reward intensity (fig. 2 B). Higher intensity of the reward results in an average DAN spike rate of 39.14Hz($sd = 1.27$) compared to 33.11Hz($sd = 1.34$).

Additionally we tested for memory extinction, an effect that refers to the reduction in behavioral memory expression, over the course of prolonged exposure to the CS without the US, following initial memory acquisition [9, 55]. We test memory extinction in our model experiments by presenting the odor, previously paired with reward, for an extended period of time, in the absence of reinforcement. During the test phase and without the presence of reward to trigger synaptic KC>MBON_ weight reduction the homeostasis is no longer outweighed by learning and drives each individual weight back towards w_{init} (fig. 2 C, upper panel).

Learned preference and behavior generalize to similar odors

We trained our model by pairing reward with a single odor for 4min. After the training procedure we tested the behavioral bias either for the same odor, or a different one, following the experimental approach used in the larva [56]. Mimicking the experimental data we show that the behavioral bias is highest if training odor and testing odor are identical. This is true for training and testing with either amylacetate or 3-octanol (fig. 3 A). With decreasing similarity, quantified as ORN activation pattern overlap (see Methods, fig. 1 B), the generalization effect to a new odor is diminished (fig. 3 A). Figure 3 B shows the network response to 30sec stimulations with amylacetate and 3-octanol in a single exemplary model instance. On the level of the ORNs 3-octanol merely activated a subset of the amylacetate-activated neurons. The uniqueness of the odor identities is enhaced in the KC population [57].

The model reproduces temporal features in trace conditioning experiments

Including an odor-evoked eligibility trace at the KC>MBON synapses allows the model to maintain the sensory odor representation for a time window, during which reinforcement will trigger synaptic change (fig. 1 B). The time window between odor and reward/punishment onset $(0, 10, 20, 30, 40, 50, 60, 120\text{s})$ was varied for experiments with a 30s presentation of odor and reward/punishment that was repeated three times. A small trace Δt_{trace} of 10 or 30s leads to an increase in behavioral bias compared to complete overlap of odor and reinforcement (fig. 3 C), using the extended window of opportunity for synaptic change triggered by each KC spike. Long traces do not lead to learning as the eligibility trace declines back to zero during this time (fig. 3 C). These findings match observations from animal experiments [6, 58, 59] with the caveat that the trace in the real larva brain seems to extend for a slightly longer period of time, compared with our experiments.

The model reproduces paired and unpaired associative conditioning experiments

To validate our model we replicated conditioning experiments performed with larvae [60]. The animals were trained with the odor amylacetate in a single trial of varying duration ($1-8$ min). To this end larvae were placed on a Petri dish coated with an agar-

sugar substrate and the odor in two small containers for diffusion in the air. Either before or following this training protocol larvae underwent a single trial without sugar and odor. Afterwards the animals were transferred to a new dish with two odor containers placed at different sides (one of them contained amylacetate, the other one was empty). Following this paired training protocol (odor and reward are presented concurrently) the animals showed a tendency to approach the previously rewarded odor, as measured by the difference in the number of animals on each side at the end of a 3min test phase, divided by the total number of animals. Following the experimental literature we will refer to this measure as the preference index [60] eqn. 15). The animals' preference is relatively consistent across training trials of different duration. Prolonged training did not lead to further increase in behavioral preference (fig. 4 B1). This paired training was compared with an unpaired protocol with separate (randomized order) presentations of amylacetate and sugar [60, 61]. Here the extent to which animals preferred amylacetate over no odor depended on the duration of the training trial. The longer the duration of the training the more their preference index decreased from an initially high value, but saturated around 2.5min (fig. 4 B1).

We aimed to replicate these behavioral experiments on two levels. Firstly we focused on the direct model output that reflects the strength of the acquired association between amylacetate and reward (behavioral bias, eqn. 4) and also simulated behavior based on these biases. We simulated both the paired and unpaired training protocol (fig. 4 A1). While the unpaired training yielded almost no behavioral bias, the models that underwent the paired training show an increased behavioral bias, that depends on the duration of the training (fig. 4 A1). Based on the observation that even after very little training real larvae show a strong preference towards amylacetate (fig. 4 B1) we readjusted our experiments to include a pre-training period of 10 minutes to start the conditioning experiments with the amylacetate-reward association already established. This adaptation of the protocol lead to the results (fig. 4 A2) that match the results obtained in real behavioral experiments (fig. 4 B1).

Secondly, since the effect of training is quantified behaviorally via spatially defined, group-level metrics (preference index and performance index (eqn. 15,eqn. 16), [60]) we performed behavioral simulations of the testing phase with groups of virtual larvae for both the paired and unpaired condition [60], allowing a straightforward comparison with the animal experiments (fig. 4 B2). To this end we utilized a realistic model for the simulation of larval locomotion and chemotactic behavior [41] that utilizes the behavioral bias at the output of the MB model as a constant gain factor to modulate the locomotory behavior of individual larva towards or away from a spatially placed odor source in a virtual arena (see Methods).

The resulting preference indices, acquired across groups of independently simulated larva, can directly be compared to the experimentally obtained preference indices (fig. 4 B1). We also compare performance indices from our simulated experiments (fig. 4 C2) with those from the lab experiments (fig. 4 C1) and find that the model can replicate these.

Discussion

Seeking rewards and avoiding punishments by predicting change in the environment is a major motivator of animal behavior. Sensory clues can acquire the necessary predictive power to guide behavior through classical conditioning, a process potentially driven by reward/punishment PE [9, 12], as observed in vertebrates [19, 32–35]. To test the biological plausibility of our proposed PE coding motif inn the insect MB we implemented a spiking network model of the *Drosophila* larva olfactory pathway, coupled with a simulation of locomotory behavior [41], to replicate larval conditioning experiments in a time-resolved manner. We demonstrate that our model of PE coding, in combination with synaptic homeostasis, results in saturating group-level and individual learning curves, where the slope and maximum of the learning curve are determined by the intensity of both the reward and the odor signal. Learning is also influenced by the timing of odor and reinforcement and can be extinguished if reinforcement is omitted

during the presentation of the sensory clue (odor).

A mechanistic implementation of the RW model

A number of predictions can be derived from the phenomenological RW model [9] and tested in our mechanistic model thereof. We found that regardless of odor/reward intensity or the model's learning rate the strength of the odor-reward association (quantified as the behavioral bias) saturates over time (fig. 2 A), as the strength of the already acquired association $V(t)$ approaches the maximum value supported by the given reinforcement input (λ_{US}) in eqn. 1. Consequently our model's acquisition curve saturates at a higher value when the intensity of the reinforcement is increased (fig. 2 A), as predicted by the RW model, in which a stronger US will result in a higher value of λ_{US} [9]. According to the RW model, increasing either the intensity of the odor or the reinforcement should also increase the learning rate α [9], leading to a faster acquisition of the association. In our model (eqn. 2) increasing the learning rate α directly or increasing either the intensity of the odor or reinforcement signal all lead to a faster increase in the behavioral bias(fig. 2 B), since the conditions for synaptic weight change are met more often.

Associations between an odor and reinforcement generalize, to a varying extend, to other odors [56, 62]. In our model we demonstrate both generalization to other odors, as well as a loss in strength, compared to the training odor (fig. 3 A). We also show that the extend of generalization depends on the similarity of the training and test odor, as measured by the overlap of the input patterns (fig. 1 D). The larval pathway with its relatively small coding space [42, 44] is prone to such poor discriminative abilities, as can be observed in animal experiments [56, 62]

Comparison of modeling results to experimental findings

A variety of experiments have demonstrated group-level acquisition curves that saturate over multiple training trials or with increasing duration of a single trial in olfactory conditioning [60, 63–66]. To replicate larval behavior in reward learning experiments [60] with varying duration of the learning phase (fig. 3 B1) we trained our model with an odor and reward in a paired vs. unpaired fashion (Methods). Real larvae show a strong odor preference even after very short training and no significant increase in their preference when trained in a paired manner for longer periods of time [60, 67] (fig. 3 B1). Instead the animals trained in an unpaired protocol start out with a similarly high preference, which then decreases over time. This behavior is very counter-intuitive since the coincidence of odor and reward should yield an association of those two stimuli and thus an increased behaviorally expressed preference for the CS [9]. To resolve this contradiction, we propose that animals might not have been naive to the training odor prior to the beginning of the experiments [68]. In that case the animals would enter into the experiment with an already established reward prediction that would be violated during unpaired training. Three scenarios lend themselves as plausible causes of this effect: Firstly accidental conditioning over the course of their lifespan during which they are raised on a food substrate, while being exposed to air that carries many different odorants. Alternatively, or in fact additionally, the animals might exhibit an innate preference for some odors [67–71]. Finally the presence of the reward during reward-only phases might lead to an association of the experimental context with that reward (previously discussed by Saumweber et al. [67]). The resulting reward expectation (solely based on the always present context), unmet during the odor-only phases could lead to a prediction error. All three candidate explanations would yield a similar projection for the unpaired experimental protocol: A reward expectation acquired prior to the actual experiment would cause a violation of that expectation during odor-only trials of the unpaired experiments. In both cases the animal's preferences might also generalize to a broader array of odors, leading to an overall preference of many odors, as observed experimentally. To test this hypothesis we pre-trained our model before simulating conditioning experiments (fig. 3 A2) and observed that this allows us to reproduce the animal experiments (fig. 3 B1). This is in accordance with

the RW model [9] that would predict no change in associative strength in the paired condition, if the curve is saturated prior to the learning phase, while the unpaired condition should lead to an extinction of the association. A possible alternative explanation to extinction could be a sensory habituation process to the odor that might cause odor preferences to decrease over time, resulting in the observed patterns for unpaired learning. In the paired condition this effect might be abolished by the continued presentation of odor and reward together [72].

Thus far we have tested our model in experiments where the CS and US presentation were fully overlapping (paired conditions). We now consider different onset times, with the onset of the CS always preceding the onset of the US by Δt_{trace} (fig. 3,C). For these experiments we used s shorter duration of $30s$ for both CS and US presentation, repeated over three acquisition trial to mimic experimental conditions in larval experiments [6, 59] that used optogenetic activation of DANs as a proxy for sugar reward. Similar to their experiments we show that the behavioral bias clearly depends on the temporal delay between CS and US (fig. 3,C). Complete temporal overlap of CS and US ($\Delta t_{trace} = 0$) does not seem to expend the full potential of learning the association, instead partial overlap yields stronger associations due to the extended window of opportunity for synaptic change triggered by the odor's eligibility trace. In our model the eligibility trace e(t) represents a molecular process that maintains the odor signal locally in the KC>MBON synapses (eqn. 2). Appetitive and aversive trace conditioning experiments have been conducted with larvae [6, 58, 59] and adult flies and other insects [63, 65, 73–75]. All animal and model experiments agree that longer inter-stimulus intervals abolish learning of the CS-US association, when no KC odor representation persists during the reinforcement period. In the cases of shorter intervals the experimental data is not entirely conclusive. Either the odor preference was higher for partial overlap, compared with complete overlap [6] or highest for complete overlap [59, 63, 65]. This observation, contrasting our modeling results, gives reason to believe our implementation of the eligibility trace might not be a perfect match for biological reality.

Comparison with other MB models

Recently a number of modeling approaches have targeted the idea of PE coding in DANs in the adult *Drosophila* [76–79] as well as in the larval MB [22]. In these models some form of MBON>DAN feedback allows for a saturation of the learning curve across trials [77, 78] as well as extinction of an acquired association [22, 77, 78]. All of these approaches seek to compare the preferences acquired by the model, or even simulated motor output [78], with observations from animal experiments. Beyond the scope of these contributions we implemented PE coding mechanistically in a fully spiking network equipped with synaptic eligibility and homeostasis to allow for the continuous assessment of dynamic change in the model's odor preference. In combination with a time-continuous behavioral simulation [41] during memory retention tests this allowed for straightforward comparison with larval experiments.

Outlook

Some experimentally observed effects in insect learning can not be captured by the RW model [9] and are thus not targeted by our model implementation. Among them are CS and US pre-exposure effects [80–83] that might be explained by changes either in attention to the CS or habituation to the CS or the US, caused by prolonged exposure prior to training, rather than changes in associative strength (for a review see [84]). Also interesting, but unpredicted by the RW model [9] is the experimental observation of second order conditioning in adult *Drosophila* [85–88], where a second CS2 is paired with the CS, after this CS has acquired an association with the US. Through the CS2-CS pairing without the US, the CS2 acquires predictive power of the US. Different mechanisms have been proposed to be involved in causing this effect [87, 89]. Among them is an excitatory synaptic KC>DAN connection, strengthened during first order conditioning, that would allow the KC odor representation to activate the DAN as a substitute of reinforcement during the CS2-CS pairing. Exploring this phenomenon using network models could yield valuable insights into the *Drosophila*

circuit, as well as aid in our general understanding of PE coding. Insect experiments have provided mixed evidence for other phenomena that can be predicted from the RW model, such as blocking [90–93] and hints at conditioned inhibition [61, 94, 95] that would be interesting to investigate. Furthermore expanding the model to include different MB output compartments would offer a perspective to explore parallel associations regarding the same stimulus [36]. This could enable extinction learning, while simultaneously retaining parallel memory for recovery. Ultimately more possible directions arise from the major benefit of using a spiking model, which offers the potential to conduct experiments at high temporal resolution, instead of in a trial-based manner [22, 76–78]. In a future closed-loop approach that connects our continual learning MB model with the locomotory model in full temporal resolution we intend to simulate a behaving agent to investigate the temporal dynamics of adaptive behavior in analogy to the tracking experiments of real larva [64, 96–99].

Methods

Network model

All neurons are modeled as leaky integrate-and-fire neurons with conductance-based synapses (parameters provided in table 1). Each neuronal membrane potential v_i is reset to the resting potential whenever a spike occurs, followed by an absolute refractory period of 2 ms, during which the neuron does not integrate any inputs. Any neuron from a given population (v^{O},v^{P},v^{L},v^{A},v^{M},v^{D}) is governed by the respective equation for ORNs, LNs, PNs, KCs, APL, MBONs and DANs (eqs. (5) to (11), fig. 1 A). An additional spike-triggered adaptation conductance was implemented for ORNs, KCs, MBONs and DANs (eqn. 12, [57]). With every spike this conductance g_{Ia} is increased, where τ_{Ia} is the decay time.

$$C_m \frac{d}{dt} v_i^O = g_L(E_L - v_i^O) + g_e^{InputO}(E_E - v_i^O) - g_{Ia}(E_{Ia} - v_i^O) \quad (5)$$

$$C_m \frac{d}{dt} v_i^L = g_L(E_L - v_i^L) + g_e^{OL}(E_E - v_i^L) \quad (6)$$

$$C_m \frac{d}{dt} v_i^P = g_L(E_L - v_i^P) + g_e^{OP}(E_E - v_i^P) - g_i^{LP}(E_I - v_i^P) \quad (7)$$

$$C_m \frac{d}{dt} v_i^K = g_L(E_L - v_i^K) - g_i^{APLK}(E_I - v_i^K) + g_e^{PK}(E_E - v_i^K) - g_{Ia}(E_{Ia} - v_i^K) \quad (8)$$

$$C_m \frac{d}{dt} v_i^A = g_L(E_L - v_i^A) + g_e^{KAPL}(E_E - v_i^A) \quad (9)$$

$$C_m \frac{d}{dt} v_i^M = g_L(E_L - v_i^M) + g_e^{KM}(E_E - v_i^M) \quad (10)$$

$$C_m \frac{d}{dt} v_i^D = g_L(E_L - v_i^D) + g_e^{InputD}(E_E - v_i^D) \quad (11)$$

$$\frac{d}{dt} g_{Ia} = -\frac{g_{Ia}}{\tau_{Ia}}. \quad (12)$$

All code for the model implementation is accessible via
https://github.com/nawrotlab/PEcodingDosophilaMB

We based our circuit model on the larval connectome [22, 42, 43] and introduced simplifications to support the mechanic investigation of the MBON>DAN feedback circuit and its role in PE coding and excluded a number of connections that have been demonstrated in the larva. Due to a limited availability of anatomical, functional and behavioral data most of our circuit implementation is based on the fist instar larva [22, 42, 43], while the information on the APL connectivity within the circuit originates from studies on the third instar larva [6]. Behavioral experiments used for comparison with our simulation results were also performed with third instar larvae [6, 59, 60]. We demonstrate that our model based

on the less developed circuit in the first instar larva is sufficient to reproduce animal behavior as observed in the older animals. From the anatomy of first instar larva we excluded DAN>KC [43] and DAN>MBON synapses [43] that may play an additional role in learning-induced plasticity at KC>MBON synapses [43], the details of which are not fully known. Instead we induce plasticity purely via the simulation of a neuromodulatory effect of the DANs onto the KC>MBON synapses ([43]). We also neglect recurrent interactions among KC themselves [43]. Many of these interactions affect KC that encode different sensory modalities, which are not included in our purely olfactory model. Furthermore we simplified the connectivity between PNs and KCs to $2-6$ PN inputs per KC, which excludes the set of KCs in the larva that receives exclusive input from only one PN [43]. This modification supported model robustness with respect to odor encoding within the small set of 72 KCs. Finally, from the population of ≈ 25 larval MBONs we only modeled two and correspondingly adapted KC>MBON synapses to provide both MBONs with input from all KCs.

Sparse odor representation

We implemented four mechanisms supporting population- and temporal sparseness in the MB odor representation [57]. Population sparseness is defined as activation of only a small subset of neurons by any given input [100]. In this circuit population sparseness is enhanced through lateral inhibition (via LNs), inhibitory APL feedback and the divergent connectivity from PNs to a larger number of KCs [57]. Temporal sparseness indicates that an individual neuron responds with only a few spikes to a specific stimulus configuration [101–103], which supports encoding dynamic changes in the sensory environment [104, 105]. In our model temporal sparseness is facilitated by SFA in ORNs and KCs and by inhibitory feedback via the APL[57].

Sensory input

In the olfactory pathway of larval *Drosophila* any odor activates up to $\approx 1/3$ of ORNs, depending on its concentration [106, 107]. We implemented receptor input with stochastic point processes to ORNs via synapses to mimic the noise in a transduction process at the receptors. Each of the 21 receptor inputs is modelled according to a gamma process (shape parameter k= 3). The spontaneous firing rate of larval ORNs has been measured in the range of $0.2-7.9\,\mathrm{Hz}$, depending strongly on odor and receptor type [107, 108]. ORNs in our model exhibit an average spontaneous firing rate of 8.92Hz (sd=0.2). We constructed realistic olfactory input across the ORN population for amylacetate and 3-octanol by estimating ORN spike frequency from the calcium signals measured in the receptor neurons [106] (dilution of 10^{-4} [106]), ensuring the spike rates would not exceed the rates reported by [107]. They showed that using an even stronger odor concentration (dilution 10^{-2}) ORN never exceeded a frequency of 200Hz. Due to lower concentration used for amylacetate and 3-octanol (fig. 1 D) [106] in our experiments and because Kreher et al, 2005 measured only the first 0.5s after odor onset when the effects of SFA in ORNs are the weakest (leading to higher spike rates) we decided to use a maximum of 150Hz in odor activated ORNs. After generating the gamma process realizations we clipped multiple spikes occurring in each time step of the simulation discarding all but the first spike in each time step. Similar as for the odor input, the presence of either reward or punishment in the experimental context was implemented as input to the $\mathrm{DAN}_{+}/\mathrm{DAN}_{-}$. Regular gamma spike trains ($k=10$) were generated and clipped as for the odor input.

To assess the effects of odor similarity on generalization we in addition created four artificial odors (A,B,C,D) (fig. 1 D) and quantified the pair-wise distances in ORN coding space using the cosine distance (eqn. 13), where vectors a and b each represent the input spike rate of two odors.

$$D_{cos} = 1 - \frac{\sum_{i=1}^{n} a_i \cdot b_i}{\sqrt{\sum_{i=1}^{n} a_i^2} \cdot \sqrt{\sum_{i=1}^{n} b_i^2}}. \qquad (13)$$

The cosine distance between odors A and B equals 0.21, 0.77 between odors A and C and 0.99 between odors A and D. The comparison of amylacetate and 3-octanol yields a distance of 0.16.

Experimental protocols

The experiments reported here belong to one of three categories. The first were performed to provide insight into the model and the effects of specific circuit functions on synaptic plasticity, prediction error coding and extinction learning (fig. 1;fig. 2). To this end we used amylacetate as the primary odor input. We varied the intensity of the reward via the frequency of gamma input into the DANp (either 500Hz or 550Hz, resulting in an average output spike rate of 33.11/39.14Hz),and the learning rate α(0.6nS or 0.8nS). Additionally MBON>DAN feedback was either enabled or disabled (fig. 2).

Experiments belonging to the second category were designed to replicate larva lab experiments to allow for a direct comparison with our model results. With these comparisons we aim to validate the model and show to what extent our assumptions about the circuit functions allow us to recreate experimental data (fig. 3;fig. 4). Replicating lab experiments also provides more insights into the circuit mechanisms and and offers alternative interpretations of the phenomena observed in data from animal experiments (fig. 3). Our implementations of the lab experiments were set up following the general procedure described in the Maggot Learning Manual [109]. Regardless of the specific protocol used in different experiments, larvae are placed into Petri dishes in groups of 30 animals. They are allowed to move around freely on the substrate that contains reinforcing substances, such as sugar or bitter tastants. During the entire time they are subjected to specific odorants, emitted from two small containers in the dish to create permanent and uniformly distributed odor exposure within the dish. In analogy of the experimental setting, in our simulated experiments each model instance is trained individually through concurrent presentation of olfactory stimulation and reward/punishment. One minute intervals with only baseline ORN stimulation were included between training trials to simulate the time needed in the lab experiments for transferring larvae between Petri dishes. Unless otherwise specified and test phases refer to 3 min, during which only odors are presented. All simulations were implemented in the network simulator Brian2 [110].

Realistic modeling of larva locomotion

Behavior during the testing phase of the olfactory learning experiment is simulated via the freely available python-based simulation platform Larvaworld (https://github.com/nawrotlab/larvaworld, [41]). A group of 30 virtual larvae is placed with random initial orientation around the center of a 100 mm diameter Petri dish and left to freely behave for 3 minutes. The previously conditioned odor is placed at one side of the dish, 10 mm from the arena's boundary. Each larva features a bisegmental body, supervised by a layered control architecture [41]. The basic layer of the control architecture is a locomotory model, capable of realistic autonomous exploration behavior. It consists of two coupled oscillators, one of which represents the crawling apparatus that generates forward velocity oscillations, resembling consecutive peristaltic strides [41]. The other oscillator generates alternating left and right lateral bending, manifested as oscillations of angular velocity [111]. The crawling and the bending oscillators are coupled via phase-locked suppression of lateral bending to capture the bend dependency on the stride-cycle phase during crawling (weathervaning). Finally, intermittent crawling is achieved by a superimposed intermittency module that generates alternating epochs of crawling and stationary pauses, with more headcasts for orientation during the latter [41].

Modulation of behavior due to sensory stimulation is introduced at the second, reactive layer of the control architecture. An odor signal can transiently alter both, the amplitude and frequency of the lateral bending oscillator, which biases free exploration towards approach or avoidance along an olfactory chemical gradient. This modulation of behavior is directly influenced via top-down signaling from the third, adaptive layer of the control architecture. In our approach, the spiking MB model populates the adaptive layer and its learning-dependent output, defined as the behavioral bias BB (i.e. the difference in MBON firing rates, eqn. 4), provides the top down signal. We formalize the gain of behavioral modulation as

$$G = g * BB. \tag{14}$$

which is directly proportional to the behavioral bias and the additional proportionality factor $g = 0.5$.

A set of $10 * 30$ trained MB model instances is used to generate 10 groups of 30 simulated larvae. The Preference index and the Performance index [60] for these simulations is illustrated in figure 4.

Preference indices (Pref) are computed individually for the paired and the unpaired experiments [60], based on the number of animals on each side (odor vs. empty) of the Petri dish at the end of the test phase.

$$Pref = \frac{count_{odor} - count_{empty}}{count_{odor} + count_{empty}}. \tag{15}$$

The Performance indices (PI) are computed from the preference indices of the paired and unpaired experiments [60].

$$PI = \frac{Pref_{paired} - Pref_{unpaired}}{2}. \tag{16}$$

References

[1] Martin Heisenberg. "Mushroom body memoir: from maps to models". In: *Nature Reviews Neuroscience* 4.4 (2003), pp. 266–275. DOI: 10.1038/nrn1074.

[2] Yoshinori Aso et al. "Mushroom body output neurons encode valence and guide memory-based action selection in Drosophila". In: *Elife* 3 (2014), e04580. DOI: /10.7554/eLife.04580.

[3] Yoshinori Aso and Gerald M Rubin. "Dopaminergic neurons write and update memories with cell-type-specific rules". In: *Elife* 5 (2016), e16135. DOI: /10.7554/eLife.16135.

[4] Scott Waddell. "Dopamine reveals neural circuit mechanisms of fly memory". In: *Trends in neurosciences* 33.10 (2010), pp. 457–464. DOI: /10.1016/j.tins.2010.07.001.

[5] Aarti Sehdev et al. "Olfactory object recognition based on fine-scale stimulus timing in Drosophila". In: *Iscience* 13 (2019), pp. 113–124. DOI: 10.1016/j.isci.2019.02.014.

[6] Timo Saumweber et al. "Functional architecture of reward learning in mushroom body extrinsic neurons of larval Drosophila". In: *Nature communications* 9.1 (2018), p. 1104. DOI: 10.1038/s41467-018-03130-1.

[7] Randolf Menzel. "The honeybee as a model for understanding the basis of cognition". In: *Nature Reviews Neuroscience* 13.11 (2012), pp. 758–768. DOI: 10.1038/nrn3357.

[8] Martin Fritz Strube-Bloss, Martin Paul Nawrot, and Randolf Menzel. "Mushroom body output neurons encode odor–reward associations". In: *Journal of Neuroscience* 31.8 (2011), pp. 3129–3140. DOI: 10.1523/JNEUROSCI.2583-10.2011.

[9] Robert A; Rescorla and Allan R Wagner. "A theory of Pavlovian conditioning : Variations in the effectiveness of reinforcement and non- reinforcement". In: *Classical Conditioning 2: Current Theory and Research*. Ed. by

William F. Black, Abraham H.; Prokasy. January 1972. New York, NY: Appelton-century-Crofts, 1972. Chap. 3, pp. 64–99.

[10] Peter D Balsam and C Randy Gallistel. "Temporal maps and informativeness in associative learning". In: *Trends in neurosciences* 32.2 (2009), pp. 73–78. DOI: 10.1016/j.tins.2008.10.004.

[11] Peter S Kaplan. "Importance of relative temporal parameters in trace autoshaping: From excitation to inhibition." In: *Journal of Experimental psychology: Animal behavior processes* 10.2 (1984), p. 113.

[12] Leon Kamin. "Predictability, surprise, attention and conditioning". In: *Punsihment and aversive behavior*. Ed. by Byron A. Campbell. New York: Appleton-Century-Crofts, 1969, 279–298.

[13] Scott Waddell. "Reinforcement signalling in Drosophila; dopamine does it all after all". In: *Current opinion in neurobiology* 23.3 (2013), pp. 324–329. DOI: 10.1016/j.conb.2013.01.005.

[14] Wolfram Schultz. "Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey". In: *Journal of neurophysiology* 56.5 (1986), pp. 1439–1461.

[15] Takemasa Satoh et al. "Correlated Coding of Motivation and Outcome of Decision by Dopamine Neurons". In: *Journal of Neuroscience* 23.30 (2003), pp. 9913–9923. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.23-30-09913.2003.

[16] Yoriko Takikawa, Reiko Kawagoe, and Okihide Hikosaka. "A possible role of midbrain dopamine neurons in short- and long-term adaptation of saccades to position-reward mapping". In: *Journal of Neurophysiology* 92.4 (2004), pp. 2520–2529. DOI: 10.1152/jn.00238.2004.

[17] Genela Morris et al. "Coincident but Distinct Messages of Midbrain Dopamine and Striatal Tonically Active Neurons". In: *Neuron* 43.1 (2004), pp. 133–143. ISSN: 0896-6273. DOI: 10.1016/J.NEURON.2004.06.012.

[18] Thomas Riemensperger et al. "Punishment prediction by dopaminergic neurons in Drosophila". In: *Current Biology* 15.21 (2005), pp. 1953–1960. ISSN: 09609822. DOI: 10.1016/j.cub.2005.09.042.

[19] Jeremiah Y Cohen et al. "Neuron-type-specific signals for reward and punishment in the ventral tegmental area". In: *Nature* 482.7383 (2012), pp. 85–88. DOI: 10.1038/nature10754.

[20] Martin Schwaerzel et al. "Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in Drosophila". In: *Journal of Neuroscience* 23.33 (2003), pp. 10495–10502. DOI: /10.1523/JNEUROSCI.23-33-10495.2003.

[21] Ariel Y Deutch, See-Ying Tam, and Robert H Roth. "Footshock and conditioned stress increase 3, 4-dihydroxyphenylacetic acid (DOPAC) in the ventral tegmental area but not substantia nigra". In: *Brain research* 333.1 (1985), pp. 143–146.

[22] Claire Eschbach et al. "Recurrent architecture for adaptive regulation of learning in the insect brain". In: *Nature Neuroscience* 23.4 (2020), pp. 544–555. ISSN: 15461726. DOI: 10.1038/s41593-020-0607-9.

[23] Dale Corbett and Roy A. Wise. "Intracranial self-stimulation in relation to the ascending dopaminergic systems of the midbrain: A moveable electrode mapping study". In: *Brain Research* 185.1 (1980), pp. 1–15. DOI: 10.1016/0006-8993(80)90666-6.

[24] William R. Stauffer et al. "Dopamine Neuron-Specific Optogenetic Stimulation in Rhesus Macaques". In: *Cell* 166.6 (2016), 1564–1571.e6. DOI: 10.1016/J.CELL.2016.08.024.

[25] R A Wise and P.-P Rompre. "Brain Dopamine and Reward". In: *Ann. Rev. Psychol* 40 (1989), pp. 191–225.

[26] Ilana B. Witten et al. "Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement". In: *Neuron* 72.5 (2011), pp. 721–733. DOI: 10.1016/j.neuron.2011.10.028.

[27] Chang Liu et al. "A subset of dopamine neurons signals reward for odour memory in Drosophila". In: *Nature 2012 488:7412* 488.7412 (2012), pp. 512–516. ISSN: 1476-4687. DOI: 10.1038/nature11304.

[28] Adam Claridge-Chang et al. "Writing Memories with Light-Addressable Reinforcement Circuitry". In: *Cell* 139.2 (2009), pp. 405–415. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2009.08.034.

[29] Christian König et al. "Reinforcement signaling of punishment versus relief in fruit flies". In: *Learning & Memory* 25.6 (2018), pp. 247–257. DOI: 10.1101/lm.047308.118.

[30] Michael Schleyer et al. "Identification of dopaminergic neurons that can both establish associative memory and acutely terminate its behavioral expression". In: *Journal of Neuroscience* 40.31 (2020), pp. 5990–6006. DOI: 10.1523/JNEUROSCI.0290-20.2020.

[31] Christian Schroll et al. "Light-induced activation of distinct modulatory neurons triggers appetitive or aversive learning in Drosophila larvae". In: *Current biology : CB* 16 (17 2006), pp. 1741–1747. DOI: 10.1016/j.cub.2006.07.023.

[32] Wolfram Schultz, Paul Apicella, and Tomas Ljungberg. "Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task". In: *Journal of neuroscience* 13.3 (1993), pp. 900–913.

[33] Wolfram Schultz, Peter Dayan, and P Read Montague. "A neural substrate of prediction and reward". In: *Science* 275.5306 (1997), pp. 1593–1599.

[34] Kareem A Zaghloul et al. "Human substantia nigra neurons encode unexpected financial rewards". In: *Science* 323.5920 (2009), pp. 1496–1499. DOI: 10.1126/science.1167342.

[35] Wei-Xing Pan et al. "Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network". In: *Journal of Neuroscience* 25.26 (2005), pp. 6235–6242. DOI: /10.1523/JNEUROSCI.1478-05.2005.

[36] Johannes Felsenberg et al. "Integration of parallel opposing memories underlies memory extinction". In: *Cell* 175.3 (2018), pp. 709–722. DOI: /10.1016/j.cell.2018.08.021.

[37] Makoto Mizunami, Kanta Terao, and Beatriz Alvarez. "Application of a prediction error theory to Pavlovian conditioning in an insect". In: *Frontiers in psychology* 9 (2018), p. 1272. DOI: /10.3389/fpsyg.2018.01272.

[38] Maria E. Villar et al. "Differential coding of absolute and relative aversive value in the Drosophila brain". In: *Current Biology* 32.21 (2022), 4576–4592.e5. ISSN: 0960-9822. DOI: 10.1016/j.cub.2022.08.058.

[39] Louis K Scheffer et al. "A connectome and analysis of the adult Drosophila central brain". In: *Elife* 9 (2020), e57443. DOI: /10.7554/eLife.57443.

[40] Nils Otto et al. "Input connectivity reveals additional heterogeneity of dopaminergic reinforcement in Drosophila". In: *Current Biology* 30.16 (2020), pp. 3200–3211. DOI: 10.1016/j.cub.2020.05.077.

[41] Panagiotis Sakagiannis, Anna-Maria Jürgensen, and Martin Paul Nawrot. "A realistic locomotory model of Drosophila larva for behavioral simulations". In: *bioRxiv* (2021). DOI: 10.1101/2021.07.07.451470.

[42] Matthew E Berck et al. "The wiring diagram of a glomerular olfactory system". In: *Elife* 5 (2016). DOI: /10.7554/eLife.14859.

[43] Katharina Eichler et al. "The complete connectome of a learning and memory centre in an insect brain". In: *Nature* 548.7666 (2017), pp. 175–182. ISSN: 1476-4687. DOI: 10.1038/nature23455.

[44] Africa Couto, Mattias Alenius, and Barry J Dickson. "Molecular, anatomical, and functional organization of the Drosophila olfactory system". In: *Current Biology* 15.17 (2005), pp. 1535–1547. DOI: /10.1016/j.cub.2005.07.034.

[45] Leslie B Vosshall and Reinhard F Stocker. "Molecular architecture of smell and taste in Drosophila". In: *Annual review of neuroscience* 30 (2007), pp. 505–533. DOI: 10.1146/annurev.neuro.30.051606.094306.

[46] Claire Eschbach et al. "Circuits for integrating learned and innate valences in the insect brain". In: *Elife* 10 (2021), e62567. DOI: /10.7554/eLife.62567.

[47] Birgit Michels et al. "Cellular site and molecular mode of synapsin action in associative learning". In: *Learning & Memory* 18.5 (2011), pp. 332–344. DOI: 10.1101/lm.2101411.

[48] Yoshinori Aso et al. "The neuronal architecture of the mushroom body provides a logic for associative learning". In: *elife* 3 (2014), e04577. DOI: 10.7554/eLife.04577.

[49] David Owald and Scott Waddell. "Olfactory learning skews mushroom body output pathways to steer behavioral choice in Drosophila". In: *Current opinion in neurobiology* 35 (2015), pp. 178–184. DOI: /10.1016/j.conb.2015.10.002.

[50] Julien Séjourné et al. "Mushroom body efferent neurons responsible for aversive olfactory memory retrieval in Drosophila". In: *Nature neuroscience* 14.7 (2011), pp. 903–910. DOI: /10.1038/nn.2846.

[51] Toshihide Hige et al. "Heterosynaptic plasticity underlies aversive olfactory learning in Drosophila". In: *Neuron* 88.5 (2015), pp. 985–998. DOI: /10.1016/j.neuron.2015.11.003.

[52] Clare E Hancock, Florian Bilz, and André Fiala. "In vivo optical calcium imaging of learning-induced synaptic plasticity in Drosophila melanogaster". In: *JoVE (Journal of Visualized Experiments)* 152 (2019), e60288. DOI: 10.3791/60288.

[53] Clare E Hancock et al. "Visualization of Learning-induced Synaptic Plasticity in Output Neurons of the Drosophila Mushroom Body γ-Lobe". In: *bioRxiv* (2022). DOI: 10.21203/rs.3.rs-1296152/v1.

[54] Feng Li et al. "The connectome of the adult Drosophila mushroom body provides insights into function". In: *Elife* 9 (2020), e62576. DOI: 10.7554/eLife.62576.

[55] Karyn M Myers and Michael Davis. "Behavioral and neural analysis of extinction". In: *Neuron* 36.4 (2002), pp. 567–584. DOI: 10.1016/S0896-6273(02)01064-4.

[56] Yi-chun Chen et al. "A behavioral odor similarity "space" in larval Drosophila". In: *Chemical senses* 36.3 (2011), pp. 237–249. DOI: /10.1093/chemse/bjq123.

[57] Anna-Maria Jürgensen et al. "A neuromorphic model of olfactory processing and sparse coding in the Drosophila larva brain". In: *Neuromorphic Computing and Engineering* 1.2 (2021), p. 024008. DOI: 10.1088/2634-4386/ac3ba6.

[58] Juliane Thoener et al. "Optogenetically induced reward and 'frustration'memory in larval Drosophila melanogaster". In: *Journal of Experimental Biology* 225.16 (2022), jeb244565. DOI: 10.1242/jeb.244565.

[59] Aliće Weiglein et al. "Aversive teaching signals from individual dopamine neurons in larval Drosophila show qualitative differences in their temporal "fingerprint"". In: *Journal of Comparative Neurology* 529.7 (2021), pp. 1553–1570. DOI: /10.1002/cne.25037.

[60] Aliće Weiglein et al. "One-trial learning in larval Drosophila". In: *Learning & Memory* 26.4 (2019), pp. 109–120. DOI: /10.1101/lm.049106.118.

[61] Michael Schleyer et al. "Associative learning of stimuli paired and unpaired with reinforcement: evaluating evidence from maggots, flies, bees, and rats". In: *Frontiers in psychology* 9 (2018), p. 1494. DOI: 10.3389/fpsyg.2018.01494.

[62] Dushyant Mishra, Matthieu Louis, and Bertram Gerber. "Adaptive adjustment of the generalization-discrimination balance in larval Drosophila". In: *Journal of neurogenetics* 24.3 (2010), pp. 168–175. DOI: /10.3109/01677063.2010.498066.

[63] Chang Zhao et al. "Predictive olfactory learning in Drosophila". In: *Scientific reports* 11.1 (2021), pp. 1–17. DOI: 10.1038/s41598-021-85841-y.

[64] Amanda Lesar et al. "Switch-like and persistent memory formation in individual Drosophila larvae". In: *Elife* 10 (2021), e70317. DOI: /10.7554/eLife.70317.

[65] Tim Tully and William G Quinn. "Classical conditioning and retention in normal and mutantDrosophila melanogaster". In: *Journal of Comparative Physiology A* 157.2 (1985), pp. 263–277.

[66] Kirsa Neuser et al. "Appetitive olfactory learning in Drosophila larvae: Effects of repetition, reward strength, age, gender, assay type and memory span". In: *Animal Behaviour* 69 (4 Apr. 2005), pp. 891–898. ISSN: 00033472. DOI: 10.1016/j.anbehav.2004.06.013.

[67] Timo Saumweber, Jana Husse, and Bertram Gerber. "Innate attractiveness and associative learnability of odors can be dissociated in larval Drosophila". In: *Chemical senses* 36.3 (2011), pp. 223–235. DOI: /10.1093/chemse/bjq128.

[68] Scott A Kreher et al. "Translation of sensory input into behavioral output via an olfactory system". In: *Neuron* 59.1 (2008), pp. 110–124. DOI: 10.1016/j.neuron.2008.06.010.

[69] Jennefer Boyle and Matthew Cobb. "Olfactory coding in Drosophila larvae investigated by cross-adaptation". In: *Journal of Experimental Biology* 208.18 (2005), pp. 3483–3491. DOI: 10.1242/jeb.01810.

[70] Michael Schleyer et al. "The impact of odor–reward memory on chemotaxis in larval Drosophila". In: *Learning & Memory* 22.5 (2015), pp. 267–277. DOI: /10.1101/lm.037978.114.

[71] KP Siju et al. "Valence and state-dependent population coding in dopaminergic neurons in the fly mushroom body". In: *Current Biology* 30.11 (2020), pp. 2104–2115. DOI: 10.1016/j.cub.2020.04.037.

[72] Isabell Twick, John Anthony Lee, and Mani Ramaswami. "Olfactory habituation in Drosophila—odor encoding and its plasticity in the antennal lobe". In: *Progress in Brain Research* 208 (2014), pp. 3–38. DOI: 10.1016/B978-0-444-63350-7.00001-2.

[73] Kristina V Dylla et al. "Trace conditioning in insects—keep the trace!" In: *Frontiers in physiology* 4 (2013), p. 67. DOI: /10.3389/fphys.2013.00067.

[74] Paul Szyszka et al. "Mind the gap: olfactory trace conditioning in honeybees". In: *Journal of Neuroscience* 31.20 (2011), pp. 7229–7239. DOI: 10.1523/JNEUROSCI.6668-10.2011.

[75] Dana Shani Galili et al. "Olfactory trace conditioning in Drosophila". In: *Journal of Neuroscience* 31.20 (2011), pp. 7240–7248. DOI: 10.1523/JNEUROSCI.6667-10.2011.

[76] James E.M. Bennett, Andrew Philippides, and Thomas Nowotny. "Learning with reinforcement prediction errors in a model of the Drosophila mushroom body". In: *Nature Communications* 12.1 (2021). ISSN: 20411723. DOI: 10.1038/s41467-021-22592-4.

[77] Magdalena Springer and Martin Paul Nawrot. "A mechanistic model for reward prediction and extinction learning in the fruit fly". In: *Eneuro* 8.3 (2021). DOI: /10.1523/ENEURO.0549-20.2021.

[78] Linnie Jiang and Ashok Litwin-Kumar. "Models of heterogeneous dopamine signaling in an insect learning and memory center". In: *PLoS Computational Biology* 17.8 (2021), e1009205. DOI: /10.1371/journal.pcbi.1009205.

[79] Evripidis Gkanias et al. "An incentive circuit for memory dynamics in the mushroom body of Drosophila melanogaster". In: *Elife* 11 (2022), e75611. DOI: 10.7554/eLife.75611.

[80] Andrew B Barron and Sarah A Corbet. "Pre-exposure affects the olfactory response of Drosophila melanogaster to menthol". In: *Entomologia experimentalis et applicata* 90.2 (1999), pp. 175–181.

[81] Vanesa M Fernández et al. "Latent inhibition in an insect: the role of aminergic signaling". In: *Learning & Memory* 19.12 (2012), pp. 593–597. DOI: /10.1101/lm.028167.112.

[82] Pedro F Jacob et al. "An opposing self-reinforced odor pre-exposure memory produces latent inhibition in Drosophila". In: *BioRxiv* (2021). DOI: 10.1101/2021.02.10.430636.

[83] Sathees BC Chandra, Jay S Hosler, and Brian H Smith. "Heritable variation for latent inhibition and its correlation with reversal learning in honeybees (Apis mellifera)." In: *Journal of Comparative Psychology* 114.1 (2000), p. 86. DOI: 10.1037/0735-7036.114.1.86.

[84] RE Lubow, I Weiner, and Paul Schnur. "Conditioned attention theory". In: *Psychology of learning and motivation*. Vol. 15. Elsevier, 1981, pp. 1–49. DOI: 10.1146/annurev.neuro.30.051606.094306.

[85] Christopher J Tabone and J Steven de Belle. "Second-order conditioning in Drosophila". In: *Learning & Memory* 18.4 (2011), pp. 250–253. DOI: /10.1101/lm.2035411.

[86] Christian König et al. "An optogenetic analogue of second-order reinforcement in Drosophila". In: *Biology Letters* 15.7 (2019), p. 20190084. DOI: 10.1101/lm.047308.118.

[87] Daichi Yamada et al. "Hierarchical architecture of dopaminergic circuits enables second-order conditioning in Drosophila". In: *bioRxiv* (2022).

[88] Syed Abid Hussaini et al. "Forward and backward second-order Pavlovian conditioning in honeybees". In: *Learning & Memory* 14.10 (2007), pp. 678–683. DOI: 10.1101/lm.471307.

[89] Isaac Cervantes-Sandoval et al. "Reciprocal synapses between mushroom body and dopamine neurons form a positive feedback loop required for learning". In: *Elife* 6 (2017), e23789. DOI: /10.7554/eLife.23789.

[90] Kanta Terao and Makoto Mizunami. "Roles of dopamine neurons in mediating the prediction error in aversive learning in insects". In: *Scientific reports* 7.1 (2017), pp. 1–9. DOI: /10.1038/s41598-017-14473-y.

[91] Brian H Smith and Susan Cobey. "The olfactory memory of the honeybee Apis mellifera. II. Blocking between odorants in binary mixtures." In: *The Journal of experimental biology* 195.1 (1994), pp. 91–108.

[92] Robert S Thorn and Brian H Smith. "The olfactory memory of the honeybee Apis mellifera. III. Bilateral sensory input is necessary for induction and expression of olfactory blocking." In: *The Journal of experimental biology* 200.14 (1997), pp. 2045–2055.

[93] JS Hosler and Brian H Smith. "Blocking and the detection of odor components in blends". In: *Journal of Experimental Biology* 203.18 (2000), pp. 2797–2806. DOI: 10.1242/jeb.203.18.2797.

[94] K Takeda. "Classical conditioned response in the honey bee". In: *Journal of Insect Physiology* 6.3 (1961), pp. 168–179.

[95] Kevin C Daly and Brian H Smith. "Associative olfactory learning in the moth Manduca sexta". In: *Journal of Experimental Biology* 203.13 (2000), pp. 2025–2038. DOI: 10.1242/jeb.203.13.2025.

[96] David Tadres and Matthieu Louis. "PiVR: An affordable and versatile closed-loop platform to study unrestrained sensorimotor behavior". In: *PLoS biology* 18.7 (2020), e3000712. DOI: /10.1371/journal.pbio.3000712.

[97] Benjamin Risse et al. "FIMTrack: An open source tracking and locomotion analysis software for small animals". In: *PLoS computational biology* 13.5 (2017), e1005530. DOI: /10.1371/journal.pcbi.1005530.

[98] Isabell Schumann and Tilman Triphan. "The pedtracker: An automatic staging approach for drosophila melanogaster larvae". In: *Frontiers in Behavioral Neuroscience* (2020), p. 241. DOI: /10.3389/fnbeh.2020.612313.

[99] Tomoko Ohyama et al. "High-throughput analysis of stimulus-evoked behaviors in Drosophila larva reveals multiple modality-specific escape strategies". In: *PloS one* 8.8 (2013), e71706. DOI: /10.1021/acscatal.9b04293.

[100] Bruno A. Olshausen and David J. Field. "Sparse coding of sensory inputs". In: *Current Opinion in Neurobiology* 14.4 (2004), pp. 481–487. ISSN: 09594388. DOI: 10.1016/j.conb.2004.07.007.

[101] Horace B Barlow. "Sensory mechanisms, the reduction of redundancy, and intelligence". In: *Mechanisation of thought processes*. London: Her Majesty's Stationery Office, 1959, pp. 535–539.

[102] Iori Ito et al. "Sparse odor representation and olfactory learning". In: *Nature neuroscience* 11.10 (2008), pp. 1177–1184. ISSN: 1546-1726. DOI: 10.1038/nn.2192.

[103] Roger Herikstad et al. "Natural movies evoke spike trains with low spike time variability in cat primary visual cortex". In: *Journal of Neuroscience* 31.44 (2011), pp. 15844–15860. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.5153-10.2011.

[104] Bilal Haider et al. "Synaptic and Network Mechanisms of Sparse and Reliable Visual Cortical Activity during Nonclassical Receptive Field Stimulation". In: *Neuron* 65.1 (2010), pp. 107–121. ISSN: 08966273. DOI: 10.1016/j.neuron.2009.12.005.

[105] Chris Häusler, Alex Susemihl, and Martin Paul Nawrot. "Natural image sequences constrain dynamic receptive fields and imply a sparse code". In: *Brain research* 1536 (2013), pp. 53–67. DOI: /10.1016/j.brainres.2013.07.056.

[106] Guangwei Si et al. "Structured Odorant Response Patterns across a Complete Olfactory Receptor Neuron Population". In: *Neuron* 101.5 (2019), 950–962.e7. ISSN: 0896-6273. DOI: 10.1016/J.NEURON.2018.12.030.

[107] Scott A. Kreher, Jae Young Kwon, and John R. Carlson. "The molecular basis of odor coding in the Drosophila larva". In: *Neuron* 46.3 (2005), pp. 445–456. ISSN: 08966273. DOI: 10.1016/j.neuron.2005.04.007.

[108] Derek J Hoare et al. "Modeling peripheral olfactory coding in Drosophila larvae". In: *PLoS One* 6.8 (2011), e22996. DOI: /10.1371/journal.pone.0022996.

[109] Birgit Michels et al. "Pavlovian conditioning of larval Drosophila: An illustrated, multilingual, hands-on manual for odor-taste associative learning in maggots". In: *Front. Behav. Neurosci.* 11.April (2017), pp. 1–6. ISSN: 16625153. DOI: 10.3389/fnbeh.2017.00045.

[110] Marcel Stimberg, Romain Brette, and Dan FM Goodman. "Brian 2, an intuitive and efficient neural simulator". In: *Elife* 8 (2019), e47314. DOI: /10.7554/eLife.47314.

[111] Antoine Wystrach, Konstantinos Lagogiannis, and Barbara Webb. "Continuous lateral oscillations as a core mechanism for taxis in Drosophila larvae". In: *Elife* 5 (2016). DOI: 10.7554/elife.15504.

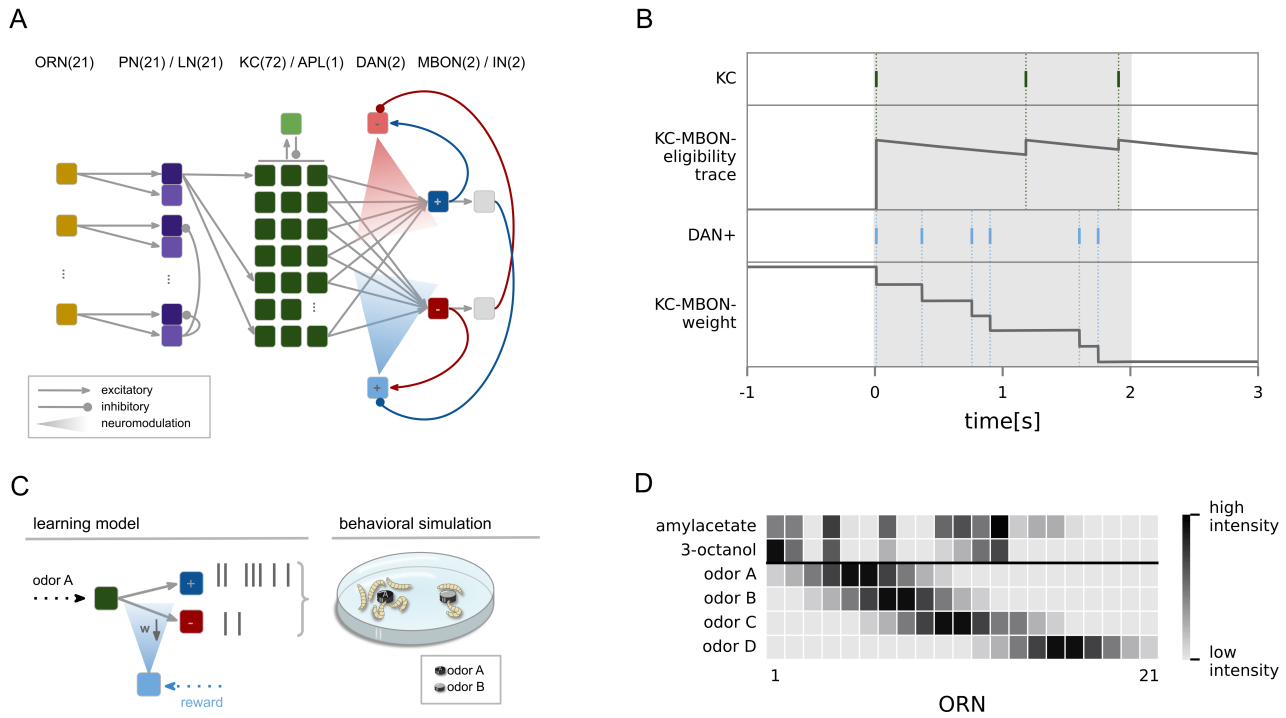

Figure 1: **Network mechanisms**. (A) Network model of the *Drosophila* larva olfactory pathway including all neurons and connections implemented. One-to-one feed-forward connections between 21 olfactory receptor neurons (ORN) and 21 projection neurons (PN)/local interneurons (LN) and from 2-6 PN to each of the 72 Kenyon cells (KC). Lateral inhibition from each LN innervates all PNs and recurrent feedback inhibition from the anterior paired lateral neuron (APL) is provided onto all KCs. The MB output region is organized in two distinct compartments. The upper compartment holds the approach encoding MBON+ and is innervated by the punishment mediating DAN-, the lower compartment holds the avoidance mediating MBON- and is innervated by the reward mediating DAN_+. Each DAN can exert a neuromodulatory effect on the plastic KC>MBON synapses within its compartment. MBONs provide excitatory and inhibitory (via gray interneurons) feedback to the DANs. (B) Sketch of synaptic weight change at a single KC>MBON synapse with respect to the synaptic eligibility trace elicited by KC spikes and the occurrence of reward-triggered spikes in DAN_+. Amylacetate is paired with reward for 2s (gray shaded area). (C) Simulation of locomotory behavior during a testing situation following training, based on the ratio of MBON+/MBON- spikes. (D) All odors (see Methods) used in the experiments. Naturalistic odor patterns for amylacetate and 3-octanol as well as four artificial patterns (odorA,odorB,odorC,odorD) with varying distance (see Methods) from odorA. Each odor activates a different set of input neurons with a different spike rate, as indicated by the color bar.

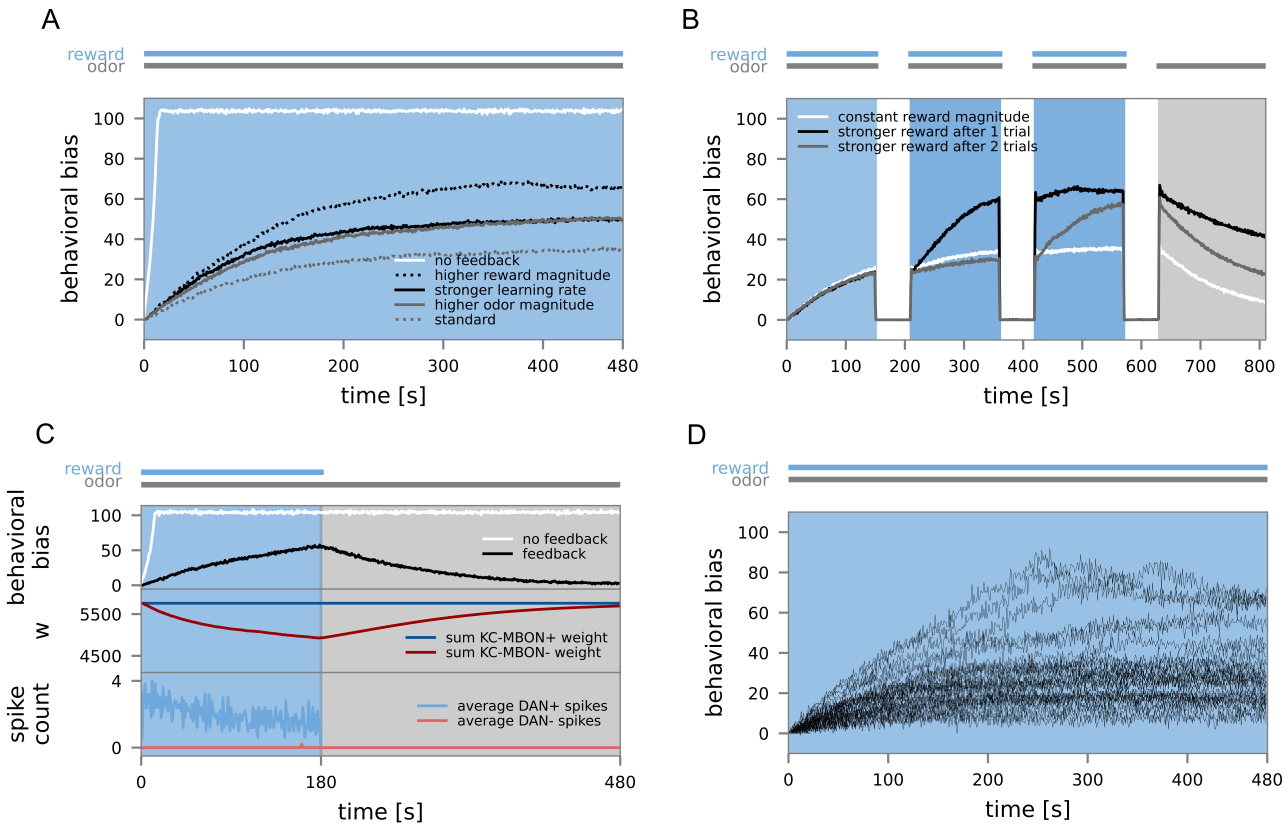

Figure 2: **Learning with prediction errors**. (A) $N = 30$ model instances were trained with the odor amylacetate (CS) and reward (US, blue background). MBON>DAN feedback, the reward/odor intensity and the learning rate were manipulated in separate experiments. The odor preference (behavioral bias) was measured continuously in windows of 1 sec and averaged over all model instances. (B) $N = 30$ model instances were trained during three trials with amylacetate and reward (blue background). Reward intensity was either constant across the three training trials (white curve), or enhanced during the third (gray) or the second and the third trial (black). The training was followed by a 3 min test phase with odor only (greay background). (C)$N = 30$ model instances were trained with amylacetate and reward (blue background) and then underwent an extended test phase (gray background).(D) Individual acquisition curves for $N = 30$ model instances (standard experiment fig. 2 A).

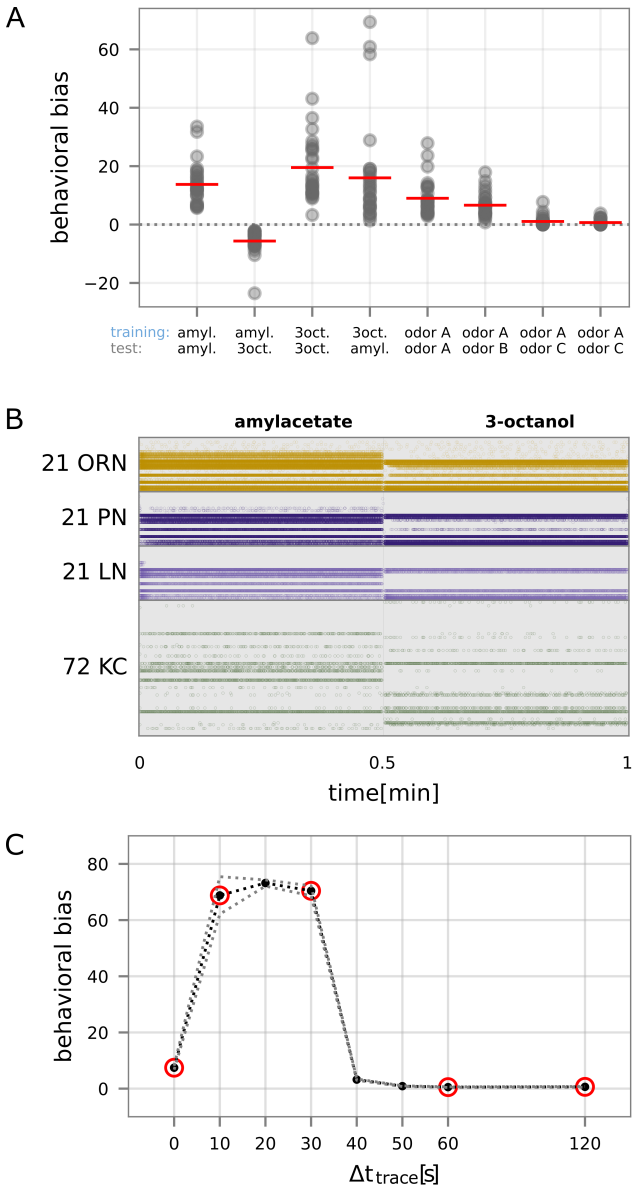

Figure 3: **Reward generalization and trace conditioning**. (A) The behavioral bias generalizes to odors that differ from the training odor after a 4min training (3min test phase). We conducted experiments with different combinations of training and testing odor, each for 10 groups (gray circles represent the mean of a group) of $N = 30$ larvae, red lines indicate the mean between groups. During testing it is highest when the training and the testing odor are the same. (B) Spiking activity in the network during the presentation of amylacetate (left) and 3-octanol (right) in a single naive model instance. (C) Trace conditioning with odor (amylacetate) and reward. Δt_{trace} indicates the time between odor and reward onset. The black line displays the mean, gray lines the std over $N = 10$ groups of 30 model instance each.Conditions circled in red correspond to the animal experiments [6, 59]

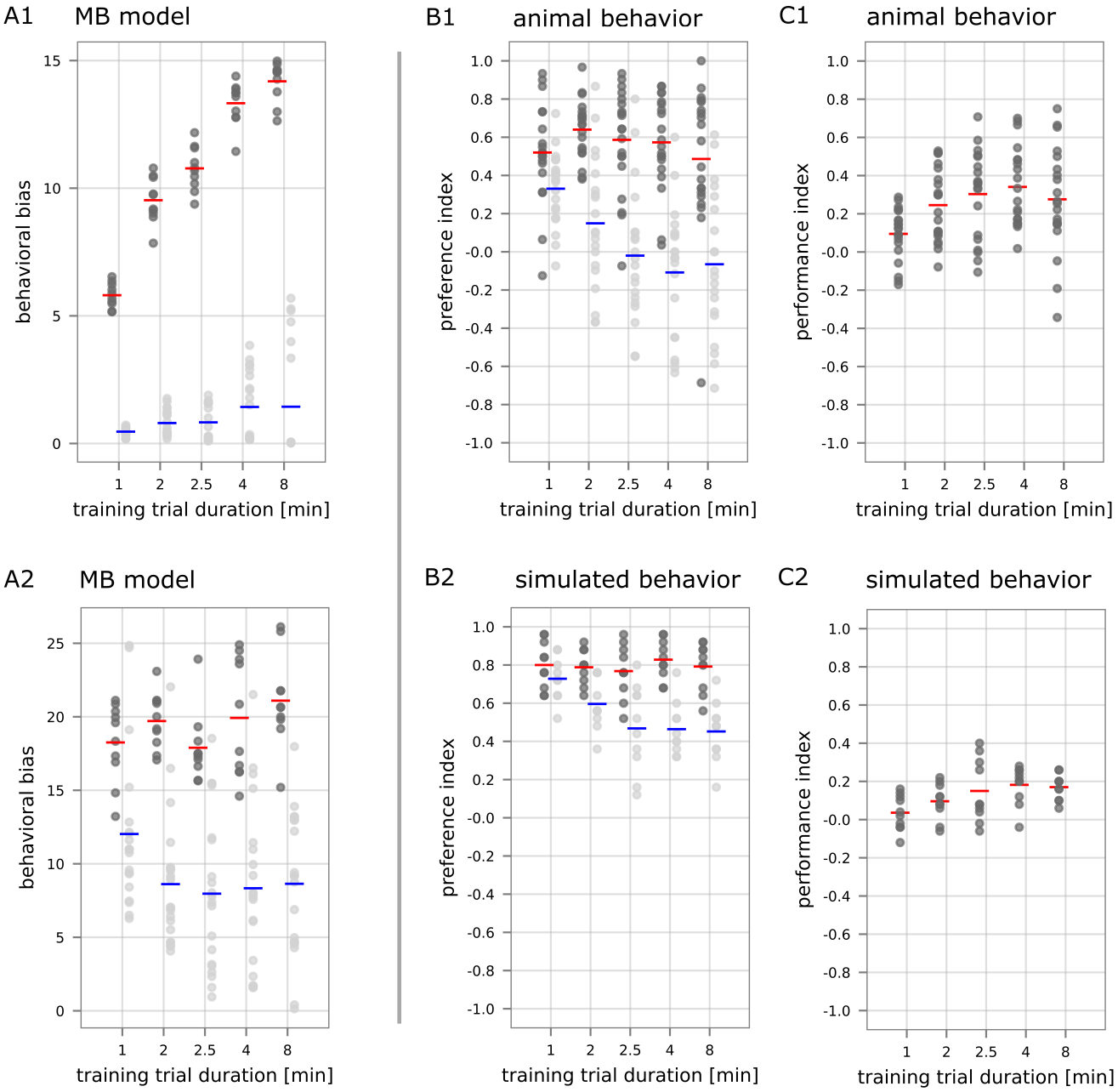

Figure 4: **Replicating behavioral experiments of paired and unpaired training**. (A1) The model's behavioral bias for training with amylacetate and reward for $N = 10$ paired and $N = 10$ unpaired experiments with groups of 30 modeled larvae each. In the unpaired condition half of the groups were trained with the odor preceding the reward. (B1) Experimental preference indices for amylacetate for 10 groups of 30 animals each for paired and unpaired experiments with randomized order of odor and reward [60]. (C1) Experimental performance indices for amylacetate for 20 groups of 30 animals in paired and unpaired experiments with randomized order of odor and reward [60]. (A2) Model behavioral bias for amylacetate for $N = 30$ paired and $N = 30$ unpaired experiments with randomized order of odor and reward. Prior to the conditioning experiment the model instances underwent a 10min pre-training. (B2) The simulated behavior is based on the protocol in A2. Simulated preference indices for amylacetate for $N = 10$ paired and $N = 10$ unpaired group experiments with varied order of odor and reward. (C2) Simulated performance indices for amylacetate based on B2.

Neuron Parameters		
Capacitance ORN	C_m	100pF
Capacitance PN	C_m	30pF
Capacitance LN	C_m	50pF
Capacitance KC	C_m	30pF
Capacitance APL	C_m	200pF
Capacitance MBON	C_m	100pF
Capacitance DAN	C_m	100pF
Leak Conductance ORN	g_L	5nS
Leak Conductance PN	g_L	2.5nS
Leak Conductance LN	g_L	2.5nS
Leak Conductance KC	g_L	5nS
Leak Conductance APL	g_L	5nS
Leak Conductance MBON	g_L	5nS
Leak Conductance DAN	g_L	5nS
Leak Potential ORN	E_L	-60mV
Leak Potential PN	E_L	-59mV
Leak Potential LN	E_L	-59mV
Leak Potential KC	E_L	-55mV
Leak Potential APL	E_L	-60mV
Leak Potential MBON	E_L	-60mV
Leak Potential DAN	E_L	-60mV
Threshold Potential ORN	V_T	-35mV
Threshold Potential PN	V_T	-30mV
Threshold Potential LN	V_T	-30mV
Threshold Potential KC	V_T	-35mV
Threshold Potential APL	V_T	-30mV
Threshold Potential MBON	V_T	-30mV
Threshold Potential MBIN	V_T	-30mV
Resting Potential ORN	V_r	-60mV
Resting Potential PN	V_r	-59mV
Resting Potential LN	V_r	-59mV
Resting Potential KC	V_r	-55mV
Resting Potential APL	V_r	-60mV
Resting Potential MBON	V_r	-60mV
Resting Potential DAN	V_r	-60mV
Refractory Time	τ_{ref}	2ms

Synaptic Parameters		
Excitatory Potential	E_E	0mV
Inhibitory Potential	E_I	-75mV
Excitatory Time Constant	τ_e	5ms
Inhibitory Time Constant	τ_i	10ms

Plasticity Parameters		
Eligibility Trace Time Constant	$\tau_{eligibility}$	5s
Learning Rate	α	0.3nS
Synaptic Weights		
Weight Input-ORN	wORNinputORN	3nS
Weight ORN-PN	wORNPN	10nS
Weight ORN-LN	wORNLN	4nS
Weight LN-PN	wLNPN	1nS
Weight PN-KC	wPNKC	1nS
Weight KC-APL	wKCAPL	20nS
Weight APL-KC	wAPLKC	50nS
Weight KC-MBON	wKCMBON	80nS
Weight Input-DAN	wDANinputDAN	2.5nS
Excitatory Weight MBON-DAN	wMBONDANex	4nS
Excitatory Weight MBON-local interneuron	$wMBONMBON_LN$	35nS
Inhibitory Weight local interneuron-DAN	wMBONDANin	70nS
Normalization Factor KC-MBON	$normalization_factor$	0.0001
Adaptation Parameters		
Adaptation Time Constant	τ_{Ia}	1000ms
Adaptation Reversal Potential	E_{Ia}	-90mV
Increase of Adaptation Conductance ORN	ORN_{SFA}	0.1nS
Increase of Adaptation Conductance KC	KC_{SFA}	0.02nS
Increase of Adaptation Conductance MBON	$MBON_{SFA}$	0.1nS
Increase of Adaptation Conductance DAN	DAN_{SFA}	0.1nS
Simulation Parameters		
Time Step	dt	0.1ms

Table 1: Network parameters.

Acknowledgements

This project is funded by the German Research Foundation (DFG) within the Research Unit 'Structure, Plasticity and Behavioral Function of the *Drosophila* mushroom body' (DFG-FOR 2705, grant no. 403329959, https://www.uni-goettingen.de/en/601524.html to BG and MN). BG and MN received additional funding from the Federal Ministry of Education and Research (BMBF, grant no. 01GQ2103A, 'DrosoExpect'). AMJ received additional support from the Research Training Group 'Neural Circuit Analysis' (DFG-RTG 1960, grant no. 233886668).