The incentive circuit: memory dynamics in the mushroom body of *Drosophila melanogaster*

Evripidis Gkanias[1*], **Li Yan McCurdy**[2], **Michael N Nitabach**[2,3,4], **Barbara Webb**[1*]

***For correspondence:**
ev.gkanias@ed.ac.uk (EG);
b.webb@inf.ed.ac.uk (BW)

[1]Institute of Perception Action and Behaviour, School of Informatics, University of Edinburgh, UK; [2]Department of Cellular and Molecular Physiology, Yale University, New Haven, CT, USA; [3]Department of Genetics, Yale University, New Haven, CT, USA; [4]Department of Neuroscience, Yale University, New Haven, CT, USA

Abstract Insects adapt their response to stimuli, such as odours, according to their pairing with positive or negative reinforcements, such as sugar or shock. Recent electrophysiological and imaging findings in *Drosophila melanogaster* allow detailed examination of the neural mechanisms supporting acquisition, forgetting and assimilation of memories. Drawing on this data, we identify a series of microcircuits within the mushroom bodies that reveal, for each motivational state, three different roles of dopaminergic and mushroom body output neurons in the memory dynamics. These microcircuits share components and form a unified system for rapid memory acquisition, transfer from short-term to long-term, and exploration/exploitation trade-off. We show that combined with a novel biologically plausible learning rule, a computational model of the full circuit reproduces the observed changes in the activity of each of these neurons in conditioning paradigms and can be used for flexible behavioural control.

Introduction

Animals deal with a complicated and changing world and they need to adapt their behaviour according to their recent experience. Rapid changes in behaviour to stimuli accompanied by intense reinforcement requires memories in the brain that are readily susceptible to alteration. Yet associations experienced consistently should form long-term memories that are hard to change. Memories that are no longer valid should be forgotten. Every neuron cannot have all of these properties, but they must be connected in a circuit, playing different roles such as supporting short- or long-term memory, and enabling processes to form, retain and erase memories. This complex interaction of memory processes is familiar in principle but its implementation at the single neuron level is still largely a mystery.

The fruit fly *Drosophila melanogaster* is able to form, retain and forget olfactory associations with reinforcements, e.g. electric shock. The key neural substrate is known to lie in neuropils of their brain called the *mushroom bodies* (MB) (***Davis, 1993***; ***Heisenberg, 2003***; ***Busto et al., 2010***). There are two MBs in the insect brain, one in each hemisphere, composed of intrinsic and extrinsic neurons. Extrinsic *projection neurons* (PN) deliver sensory input to the only intrinsic neurons of the mushroom bodies, the *kenyon cells* (KC), whose long parallel axons travel through the pendunculus and then split forming the vertical (α/α') and medial (β/β' and γ) mushroom body lobes (see ***Figure 1***). The extrinsic *mushroom body output neurons* (MBON) extend their dendrites in different regions of the lobes, receiving input from the KCs and forming 15 distinct compartments (***Turner et al., 2008***;

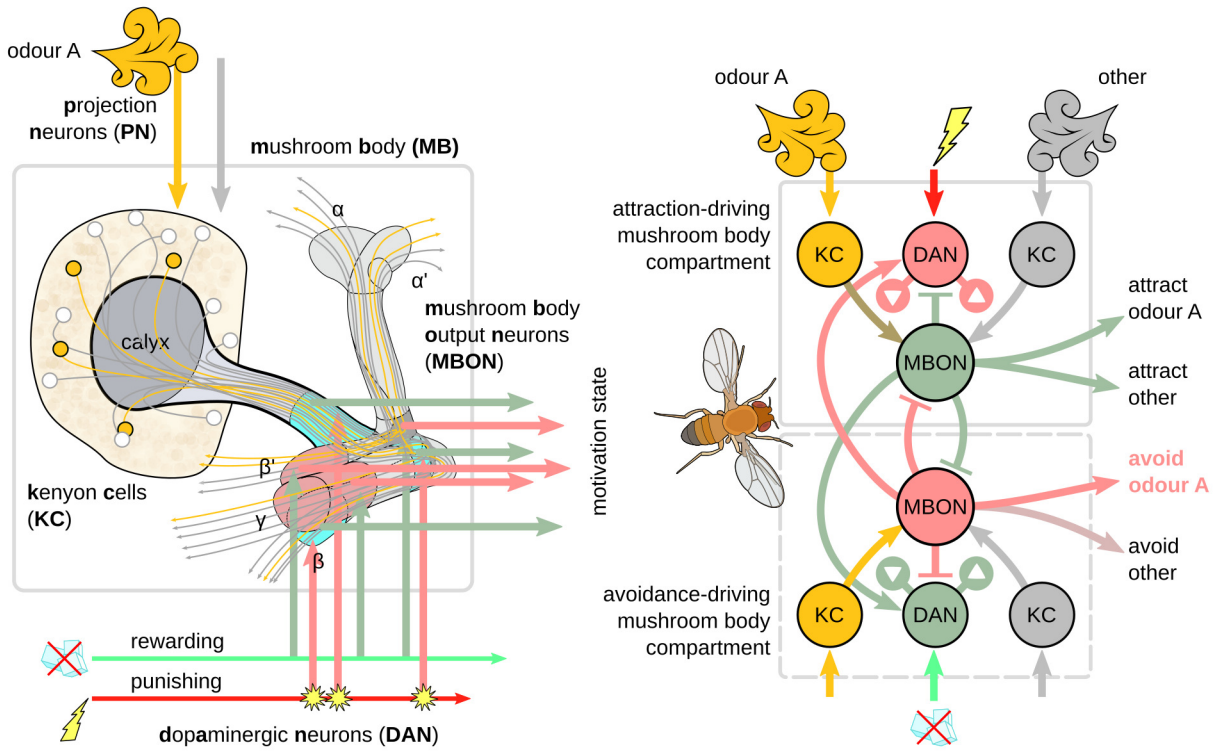


Figure 1. Overview of the MB circuit. Left, the main anatomical pathways. In the illustration odour A activates the kenyon cells (KCs) through the projection neurons (PNs). The parallel axons of KCs propagate this signal to the lobes of the mushroom body. The mushroom body output neurons extend their dendrites in the mushroom body lobes, receiving input from the KCs. Electric shock creates a punishing signal that excites some dopaminergic neurons (DANs), whose axons terminate in the lobes and modulate the synaptic weights between KCs and MBONs. Right, schematic connections between punishment/reward DANs and approach/avoidance MBONs. Note that although DANs transferring punishing signals modulate the KC activation of MBONs that encode positive motivations [decreasing attraction to the presented odour and increasing attraction to odours not present (The dopaminergic learning rule)], MBONs that encode negative motivations will also gain higher responses due to release of inhibition between MBONs, and the feedback connections from MBONs to other DANs. In our model, we further decompose these functions using three DANs and three MBONs for each motivation (positive or negative), and map these units to specific identified neurons and microcircuits in the brain of Drosophila.

Tanaka et al., 2008; ***Campbell et al., 2013***; ***Aso et al., 2014a***). Their activity is thought to provide motivational output that modulates the default behaviour of the animal (***Aso et al., 2014b***). Different groups of extrinsic *dopaminergic neurons* (DAN) terminate their axons in specific compartments of the MB, and modulate the connections between KCs and MBONs (***Aso et al., 2014a***). Many of the DANs respond to a variety of reinforcement signals (***Mao and Davis, 2009***; ***Schwaerzel et al., 2003***; ***Claridge-Chang et al., 2009***; ***Liu et al., 2012***; ***Lin et al., 2014***) and therefore they are considered the main source of reinforcement signals in the mushroom body. Finally, many of the MBON axons and DAN dendrites meet in the convergence zones (CZ), where they create interconnections, such that the motivational output can also influence the activity of the reinforcement neurons (***Li et al., 2020***).

Several computational models have tried to capture the structure and function of the mushroom bodies, usually abstracting the common features of this network across various insect species. Modellers have treated the mushroom bodies as performing odour discrimination (***Huerta et al., 2004***), olfactory conditioning (***Balkenius et al., 2006***; ***Smith et al., 2008***; ***Finelli et al., 2008***; ***Young et al., 2011***; ***Wessnitzer et al., 2012***; ***Peng and Chittka, 2017***; ***Faghihi et al., 2017***; ***Zhao et al., 2021***; ***Springer and Nawrot, 2021***; ***Eschbach et al., 2020***; ***Bennett et al., 2021***) or calculating the scene familiarity (***Wu and Guo, 2011***; ***Baddeley et al., 2012***; ***Arena et al., 2013***; ***Bazhenov et al., 2013***; ***Ardin et al., 2016***). However, it seems like they can subserve all these functions, depending on context (or experience), i.e. activity of the KCs (***Cohn et al., 2015***). This suggests that the output neurons of

the mushroom body do not just inform the animal whether an odour is known or attractive, or if a scene is familiar, but they actually motivate the animal to take an action like approach, avoid, escape or forage. There is emerging evidence supporting this idea of the MBONs driving non-binary but antagonistic motivations (***Schwaerzel et al., 2003***; ***Krashes et al., 2009***; ***Gerber et al., 2009***; ***Waddell, 2010***; ***Lin et al., 2014***; ***Perisse et al., 2016***; ***Senapati et al., 2019***), which has started to be explored in recent models.

In addition to the structural and functional depiction of the mushroom bodies, a variety of learning rules have been used in order to explain the effect of dopamine emitted by the DANs on the KC→MBON synapses. Although the best supported biological mechanism is that coincidence of DAN and KC activity depresses the output of KCs to MBONs, most of the models mentioned before use variations of the Hebbian rule (***Hebb, 2005***), where the coincidence of the input (KCs) and output (MBONs) activation strengthens the synaptic weight (or weakens it for the anti-Hebbian case) and this is gated by the reinforcement (DANs). More recent approaches that try to model the activity of DANs and MBONs in the brain have used learning rules (***Zhao et al., 2021***) or circuit structures (***Springer and Nawrot, 2021***; ***Bennett et al., 2021***; ***Eschbach et al., 2020***) that implement a reward prediction error (RPE) (***Rescorla and Wagner, 1972***), which is the most widely accepted psychological account of associative learning with strong validation evidence in the vertebrate brain (***Niv, 2009***). For the mushroom body, this learning rule is interpreted as the output (MBON) being the prediction of the reinforcement (DAN), so their difference (gated by the activity of the input, KC) drives the synaptic plasticity. However, details of neuronal dynamics in fruit flies (***Dylla et al., 2017***; ***Berry et al., 2018***) suggest that neither of these learning rules captures the plasticity dynamics in the mushroom bodies, which highlights the importance of investigating new learning rules that are a closer approximation to the actual dopaminergic function.

In this work, we propose such a novel learning rule, named the *dopaminergic learning rule* (DLR), which reflects recent understanding of the role of dopamine in depression and potentiation of synapses. Based on evidence of existing MBON→DAN connections, we build a 12-neuron computational model, that we call the *incentive circuit* (IC), which uses the proposed learning rule. In this model, we identify 3 types of DANs ('discharging', 'charging' and 'forgetting') and 3 types of MBONs ('susceptible', 'restrained' and 'long-term memory') for each of two opposing motivational states. We demonstrate that the neural responses generated by this model during an aversive olfactory learning paradigm closely replicate those observed in the animal; and that simulated flies equipped with the incentive circuit generate learned odour preferences comparable to real flies. Finally, we show that such a model could work as motif that extends the set of motivations from binary (e.g. avoidance vs attraction) to a spectrum of motivations whose capabilities are equivalent to 'decision making' in mammals.

Results

The dopaminergic learning rule

We emulate the function of dopamine (and co-transmitters) released in the KC→MBON synaptic weights as an abstracted learning rule. ***Handler et al.*** (***2019***) suggest that ER-Ca^{2+} and cAMP play a decisive role in the dynamics of forward and backward conditioning. More specifically, they suggest that the combination of their responses are proportional to the KC→MBON synaptic change, ΔW^{ij}_{k2m}, which can be written formally as

$$\Delta W^{ij}_{k2m}(t) \propto -\text{ER-Ca}^{2+} - \text{cAMP}. \tag{1}$$

We assume that ER-Ca^{2+} and cAMP can be calculated by using any available information in the local area of the target KC axon (pre-synaptic terminal). These are the dopamine (DA) level emitted by the respective (j^{th}) DAN, $D^{ji}(t) \geq 0$, the activity of the target (i^{th}) KC, $k^i(t) \geq 0$, the respective KC→MBON synaptic weight, $W^{ij}_{k2m}(t) \geq 0$ (assumed always positive, exciting the MBON) and the resting synaptic weights, $w^{ij}_{\text{rest}} \in \mathbb{R}$, which we assume are a constant parameter of the synapse.

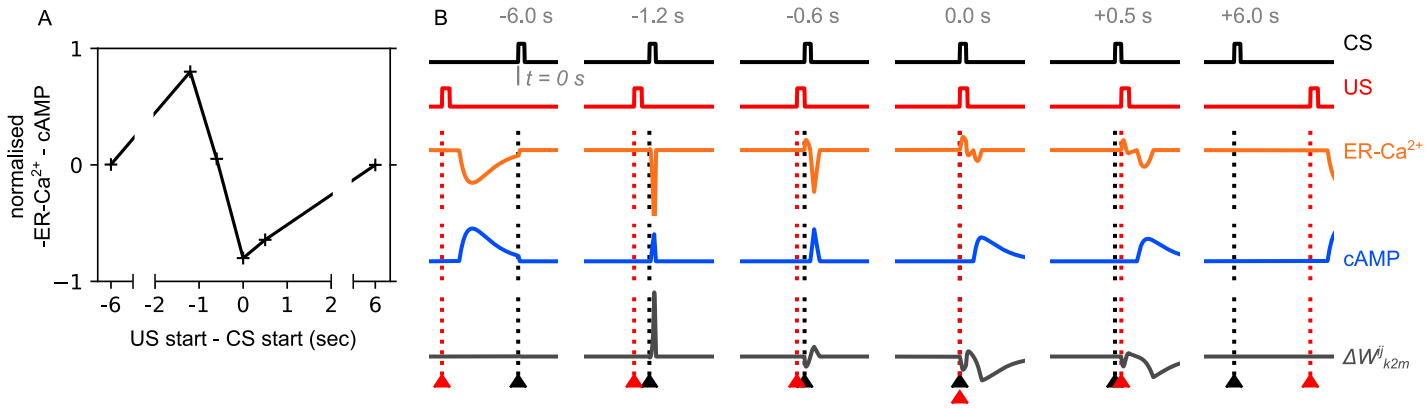


Figure 2. The effect of the ER-Ca^{2+} and cAMP based on the order of the conditional (CS) and unconditional stimuli (US). (A) Normalised mean change of the synaptic weight plotted as a function of the Δs (US start - CS start), similar to ***Handler et al.*** (***2019***) - Figure 2E. (B) Detailed ER-Ca^{2+} and cAMP responses reproduced for the different Δs, and their result synaptic-weight change. Black arrowhead marks time of the CS (duration $0.5\,\text{sec}$); red arrowhead marks time of the US (duration $0.6\,\text{sec}$), similar to ***Handler et al.*** (***2019***) - Figure 2D.

Figure 2–Figure supplement 1. All the chemical levels and neural activities calculated based on the order of the conditional (CS) and unconditional stimuli (US).

Tuning the above quantities in order to reproduce the ER-Ca^{2+} and cAMP levels, we are able to mathematically formulate them as a function of the available parameters

$$\text{ER-Ca}^{2+} \propto -D_2^j(t)\cdot[k^i(t)-w_{\text{rest}}]-[D_2^j(t)-D_1^j(t)]\cdot W_{k2m}^{ij}(t) \tag{2}$$

$$\text{cAMP} \propto D_1^j(t)\cdot[k^i(t)-w_{\text{rest}}]. \tag{3}$$

where D_1 and D_2 are two components of the DA [assumed to correspond to DopR1 and DopR2 receptors (***Handler et al., 2019***), or potentially to involve co-transmitters released by the DAN such as Nitric Oxide (***Aso et al., 2019***)]. D_1 makes a higher peak in its activity followed by a faster diffusion than D_2, which seems to be the key for the backward conditioning. ***Figure 2*** shows the ER-Ca^{2+} and cAMP levels during forward and backward conditioning (see ***Figure 2–Figure Supplement 1*** for the responses of all the terms used), which are comparable to the data shown in ***Handler et al.*** (***2019***).

By replacing ***Equation 2*** and ***Equation 3*** in ***Equation 1***, we can write the update rule as a function of known quantities, forming our *dopaminergic learning rule* (DLR)

$$\Delta W_{k2m}^{ij}(t)=\delta^j(t)[k^i(t)+W_{k2m}^{ij}(t)-w_{\text{rest}}] \qquad \text{where} \qquad \delta^j(t)=D_2^j(t)-D_1^j(t) \tag{4}$$

The above learning rule alters the connection weight between each KC and MBON on each time step depending on the *dopaminergic factor*, $\delta^j(t)$, which is the difference between the D_1 and D_2 levels. The dopaminergic factor can be positive ($D_2 > D_1$) or negative ($D_2 < D_1$), and combined with two possible states of KC activity (active or inactive) results in four different weight modulation effects: *depression*, *potentiation*, *recovery* and *saturation*.

When D_1 and D_2 are equal or both zero (DAN is inactive), $\delta^j(t)=0$ and no learning occurs. If $D_1 > D_2$, $\delta^j < 0$ and the KC is active, the KC→MBON synaptic weight is decreased (*depression* effect of the learning rule, see ***Figure 3***A). However, if the synaptic weight is already low [$k^i(t)=w_{\text{rest}}^{ij}-W_{k2m}^{ij}(t)$], the synaptic weight does not change. The *recovery* effect takes place when the KC is inactive, in which case the synaptic weights tend to reset to the resting weight [$k^i(t)=w_{\text{rest}}^{ij}-W_{k2m}^{ij}(t)=0$; see ***Figure 3***C]. When $D_1 < D_2$, $\delta^j > 0$ and the KC is active, we have the *potentiation* effect, which causes increase in the synaptic weights (see ***Figure 3***B). As the synaptic weight becomes stronger, it further enhances this effect. If the KC is inactive, then we have the *saturation* effect, where if the current synaptic weight is higher than its resting weight, the synaptic weight continue increasing, while if it is lower then it continues decreasing (see ***Figure 3***D). This effect enhances diversity in the responses of the MBON to the different CS, which is essential for memory consolidation and for the formation of long-term memories.

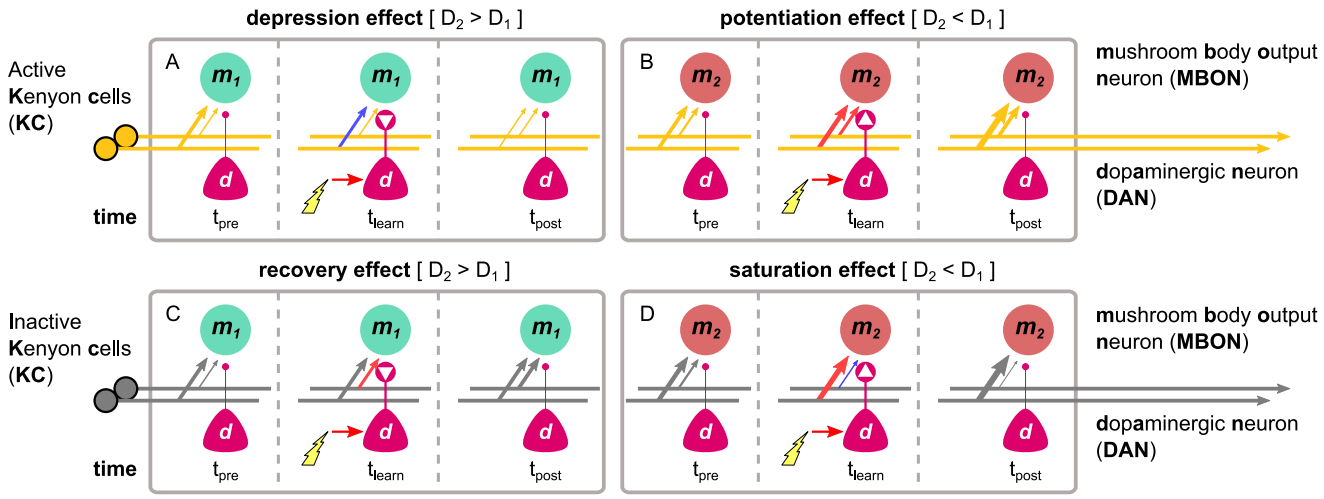


Figure 3. The different effects of the dopaminergic learning rule, depending on the activity of the KC (orange indicates active) and the sign of the dopaminergic factor (white arrowheads in dots). In each box, time-step $t = t_{\text{pre}}$ shows the initial KC→MBON synaptic weights (arrow thickness); electric shock activates the DAN in time-step $t = t_{\text{learn}}$ causing modulation of the synaptic weights (red: increase, blue: decrease), while time-step $t = t_{\text{post}}$ shows the synaptic weights after the shock delivery. (A) Example of the *depression effect* - the synaptic weight decreases when $D_1 > D_2$ and the KC is active. (B) Example of the *potentiation effect* - the synaptic weight increases when $D_1 < D_2$ and the KC is active. (C) Example of the *recovery effect* - the synaptic weight increases when $D_1 > D_2$ and the KC is inactive. (D) Example of the *saturation effect* - the synaptic weight increases further (when $W^{ij}_{k2m}(t) > w_{\text{rest}}$) or decreases further (when $W^{ij}_{k2m}(t) < w_{\text{rest}}$) when $D_1 < D_2$ and the KC is inactive.

Note that in the later simulations in this paper, we simplify the dopaminergic factor to have a net positive or negative value for the time-step in which it influences the synaptic weight change (see Implementation of the incentive circuit).

Mushroom body microcircuits

Although the learning rule above already exhibits some interesting dynamics, more sophisticated memory effects emerge when it is combined with MBON-DAN feedback connections in microcircuits such as those identified in ***Felsenberg et al.*** (***2018***); ***Ichinose et al.*** (***2015***); ***McCurdy et al.*** (***2021***). Starting from these, we identify connections between them and hypothesise, based on neuroanatomy, additional functional connections which we demonstrate through modelling can contribute to memory acquisition, forgetting and assimilation.

Susceptible memory (SM)

Pavlowsky et al. (***2018***) identified a microcircuit in the mushroom body, where PPL1-γ1pedc depresses the KC→MBON-γ1pedc synapses discharging the activity of that MBON, but also MBON-γ1pedc inhibits PPL1-γ1pedc, which they argue is a long-term memory consolidation mechanism (see ***Figure 4***A). We computationally reconstructed this microcircuit showing that, indeed, during the CS and US coincidence the memory is rapidly formed by discharging the KC→MBON-γ1pedc synaptic weight. This lowers the MBON's activity, reducing the inhibition to the discharging DAN, whose activity helps consolidate this memory. However, due to our learning rule, if the US subsequently occurs without the CS, the MBON-γ1pedc synaptic weights reset due to the recovery effect, which results in a memory that is easily created and erased: a 'susceptible memory'. Therefore, we call MBON-γ1pedc an attraction-driving susceptible MBON, s_{at}, whose motivation is discharged by the PPL1-γ1pedc aversive discharging DAN, d_{av}, which creates and erases susceptible memories due to the depression/recovery effects of the learning rule. ***Figure 4***E visualises this microcircuit, and ***Figure 4***F shows an example of its memory dynamics during memory creation.

Similar microcircuits in the mushroom body can be extracted from the connectome described in ***Aso et al.*** (***2014a***) and ***Li et al.*** (***2020***) [also identified in larvae (***Eichler et al., 2017***)]; which leads us to the assumption that there is an exactly corresponding susceptible memory with opposite valence,

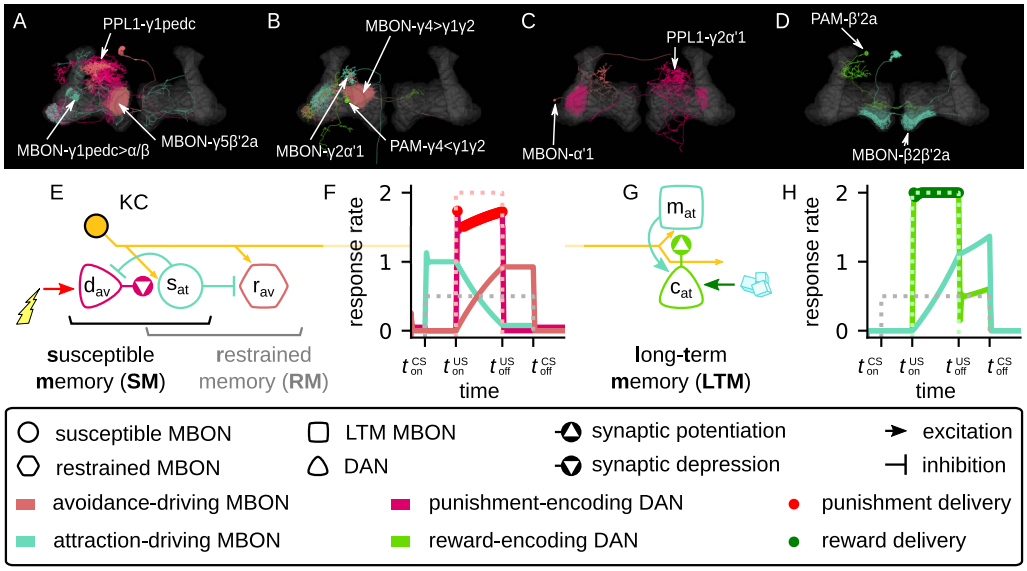


Figure 4. microcircuits of the mushroom body. (A) The attraction-driving susceptible and avoidance-driving restrained memory microcircuits made of the PPL1-$\gamma 1$pedc, MBON-$\gamma 1$pedc and MBON-$\gamma 5\beta'2$a neurons. (B) The avoidance-driving susceptible and attraction-driving restrained memory microcircuits made of the PAM-$\gamma 4 < \gamma 1\gamma 2$, MBON-$\gamma 4 > \gamma 1\gamma 2$ and MBON-$\gamma 2\alpha' 1$ neurons. (C) The avoidance-driving long-term memory microcircuit made of the PPL1-$\gamma 2\alpha' 1$ and MBON-$\alpha' 1$ neurons. (D) The attraction-driving long-term memory microcircuit made of the PAM-$\beta' 2$a and MBON-$\beta 2\beta' 2$a neurons. (E) Schematic representation of the Susceptible memory (SM) and the Restrained memory (RM) microcircuits connected via the susceptible MBON. (F) The responses of the punishment-encoding discharging DAN (magenta), d_{av}, the attraction-driving susceptible MBON (green), s_{at} and the avoidance-encoding restrained MBON (orange), r_{av}, during an experiment. (G) Schematic representation of the Long-term memory (LTM) microcircuit. (H) The responses of the reward-encoding charging DAN (light green), c_{at}, and the attraction-driving long-term memory MBON (green), m_{at}, during an experiment. Images of the neurons in A, B, C and D were created using the Virtual Fly Brain software (***Milyaev et al., 2012***).

i.e. s_{av} and d_{at}, which drives avoidance to odour but has decreased output to odours experienced in the presence of a reward, e.g. sugar, and will recover its response if reward is subsequently experienced in the absence of the odour. Indeed, it is possible for the reward-encoding discharging DAN, d_{at} = PAM-$\gamma 4 < \gamma 1\gamma 2$ and the avoidance-driving susceptible MBON, s_{av} = MBON-$\gamma 4 > \gamma 1\gamma 2$, to build the avoidance-depressing SM microcircuit as shown in ***Figure 4***B (***Li et al., 2020***).

Restrained memory (RM)

Felsenberg et al. (***2018***) took the susceptible microcircuit one step further, by showing that the activity of MBON-$\gamma 5\beta' 2$a increases (driving avoidance) when the activity of MBON-$\gamma 1$pedc decreases and that this is due to direct inhibition from the latter to the former MBON (see ***Figure 4***A, E and F). We call the memory stored in the KC→MBON-$\gamma 5\beta' 2$a synaptic weights a 'restrained memory', because it is only expressed when the restriction put by the MBON-$\gamma 1$pedc is lifted (due to its reduced activity), and we call MBON-$\gamma 5\beta' 2$a an avoidance-driving restrained MBON, r_{av}.

Aso et al. (***2014a***) and ***Cohn et al.*** (***2015***) suggest that it is possible to have inhibiting connections from MBONs in the $\gamma 4$ compartment to the ones in the $\gamma 2$ compartment. Therefore, we assume that there is an equivalent attraction-driving restrained MBON, r_{at} = MBON-$\gamma 2\alpha' 1$, inhibited by the corresponding avoidance-driving susceptible MBON, s_{av} = MBON-$\gamma 4 > \gamma 1\gamma 2$, as shown in ***Figure 4***B. Although MBON-$\gamma 4 > \gamma 1\gamma 2$ is characterised by the Glutamate neurotransmitter, it is possible that it can inhibit MBON-$\gamma 2\alpha' 1$ though glutamate-gated chloride channels (***Cleland, 1996***; ***Liu and Wilson, 2013***; ***McCarthy et al., 2011***).

Long-term memory (LTM)

Ichinose et al. (***2015***) describes a microcircuit, where PAM-$\alpha 1$ potentiates the KC→MBON-$\alpha 1$ synapses, and MBON-$\alpha 1$ in turn excites PAM-$\alpha 1$ (see ***Figure 4***G and H). This potentiation is translated in our learning rule as the potentiation/saturation effects for when the KC is active or not respectively, which suggests that the D_2 concentration should be higher than the D_1 in this case. This means that as long as we have reinforcement and the KC is active the synaptic weight increases, but even when the KC becomes inactive it will continue increasing during the saturation effect, setting a charging momentum for the synaptic weights which is difficult to reverse. For this reason we call this a 'long-term memory' microcircuit with PAM-$\alpha 1$ as its reward-encoding charging DAN, c_{at}, and MBON-$\alpha 1$ as its attraction-driving long-term memory (LTM) MBON, m_{at}.

Li et al. (***2020***) have identified numerous similar microcircuits and so we assume two long-term memory microcircuits corresponding to the attraction and the avoidance motivations respectively. As our goal is to create a unified model that includes all the microcircuit types we describe here, we tried to find two opposing-valence long-term memory microcircuits that can be part of it. As shown in ***Figure 4***C, we suggest the avoidance-driving long-term memory microcircuit is created by the avoidance-driving LTM MBON, m_{av} = MBON-$\alpha' 1$, which extends its dendrites to $\gamma 2$ or $\alpha' 1$ compartments (***Aso et al., 2014a***), and its axon towards the dendrites of the punishment-encoding charging DAN, c_{av} = PPL1-$\gamma 2 \alpha' 1$ (***Aso et al., 2014a***; ***Li et al., 2020***). Similarly, for the opposite motivation (attraction) the attraction-driving LTM MBON, m_{at} = MBON-$\beta 2 \beta'$2a, excites the reward-encoding charging DAN (***Li et al., 2020***), c_{at} = PAM-$\beta' 2a$, and is visualised in ***Figure 4***D.

Reciprocal restrained memories (RRM)

McCurdy et al. (***2021***) suggest that the attraction-driving restrained MBON-$\gamma 2 \alpha' 1$ (i.e. r_{at}) indirectly decreases the synaptic weights from KCs to the avoidance-driving restrained MBON-$\gamma 5 \beta'$2a (i.e. r_{av}) via the reward-encoding charging PAM-β'2a (i.e. c_{at}). This microcircuit is also supported by ***Felsenberg et al.*** (***2018***) and ***Berry et al.*** (***2018***). ***Cohn et al.*** (***2015***) and ***Li et al.*** (***2020***) suggest that MBON-$\gamma 5 \beta'$2a excites PPL1-$\gamma 2 \alpha' 1$, which closes the loop of the 'reciprocal restrained memories' microcircuit as shown in ***Figure 5***A and G (note we have not illustrated the KCs in ***Figure 5***, but the DAN connections are always assumed to be acting on the weights between the KCs and the relevant MBONs). In addition to charging the long-term memories, the charging DANs now also depress KC→MBON synaptic weights of the restrained memories. When electric shock is delivered along with an odour (active KC), c_{av} start depressing the KC→ r_{at} synaptic weights, also reducing its activity. This will also decrease the activity of c_{at}, stopping the depression of KC→ r_{av} synaptic weights (see ***Figure 5***B and C). The increase in the activity of r_{av} in ***Figure 5***C, is due to the reduced inhibition of the susceptible MBON, s_{at}. However, notice that if the electric shock is delivered when the KC is inactive, the KC→r_{at} synaptic weights will recover due to the recovery effect of the learning rule.

Reciprocal forgetting memories (RFM)

As described so far, once the Long-term memory (LTM) microcircuit begins to charge, it will have a self-sustaining increase in the weights during odour delivery, preventing any subsequent adaptation to altered reward contingencies. To allow these weights to decrease, specifically, to decrease in response to charging of the LTM of opposite valence, we connect the two LTM MBONs via respective forgetting DANs, f_{at} and f_{av}, in a similar reciprocal microcircuit to the RRM (see ***Figure 5***D). Note these forgetting DANs do not receive any direct reinforcement signals. Instead, as long as an LTM MBON is active, its respective forgetting DAN is also active and enables the depression effect on the opposite LTM MBON. This counteracts any potentiation effect due to its respective charging DAN (see ***Figure 5***E and F). As a consequence, sustained reinforcement of one valence can gradually overcome the positive feedback of the LTM circuit of opposite valence, causing the charging momentum to drop and eventually negate. The long-term memories are thus in long-term competition.

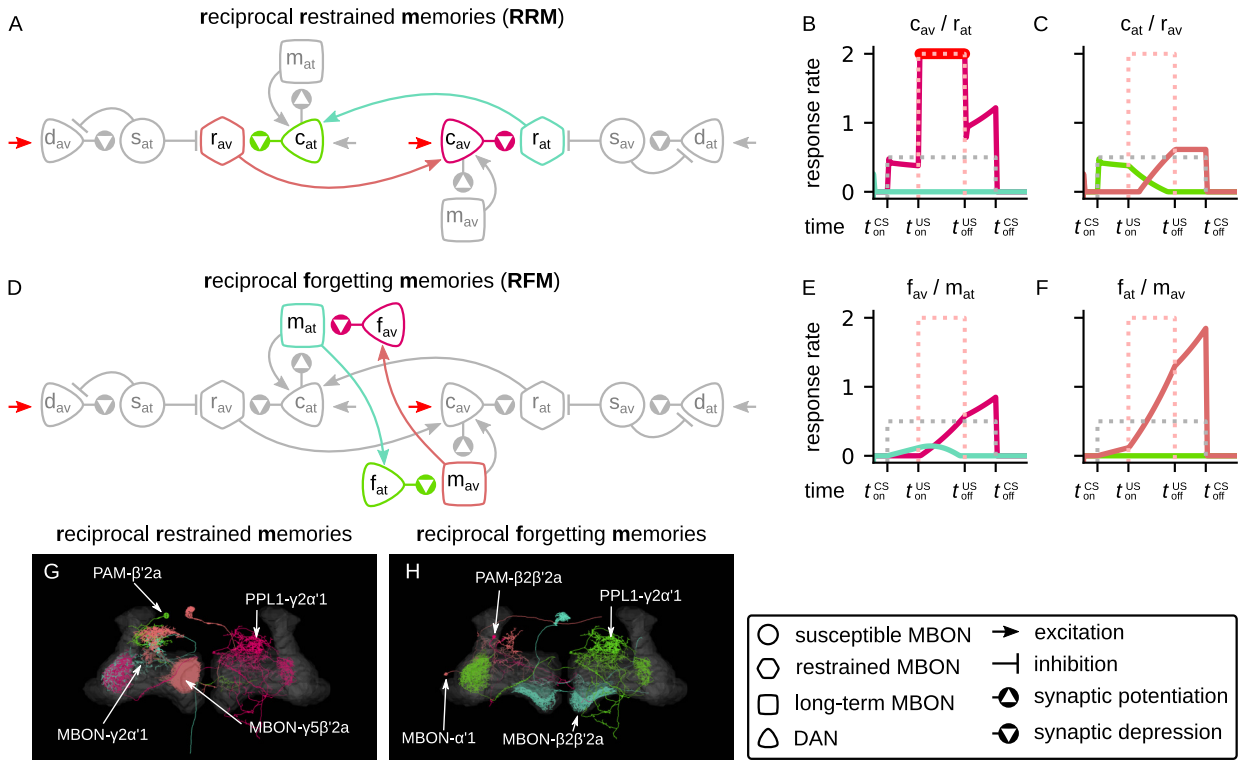


Figure 5. The reciprocal microcircuits of the mushroom body. (A) The circuits in ***Figure 4***E and ***Figure 4***G are combined (left side) by allowing the charging DAN in the LTM circuit to also affect (with a depressing/recovering effect) the restrained MBON. The whole motif is reflected (right side) with opposite valence and the two halves connected by excitatory connections from restrained MBONs to charging DANs, forming the Reciprocal restrained memories (RRM) microcircuit. (B) The responses of the punishment-encoding charging DAN (magenta), c_{av}, and the attraction-driving restrained MBON (green), r_{at}, from the RRM microcircuit to odour and shock. (C) The responses of reward-encoding charging DAN (light green), c_{at}, and the avoidance-driving restrained MBON (orange), r_{av}. (D) Two additional forgetting DANs connect the long-term memory MBONs, forming the Reciprocal forgetting memories (RFM) microcircuit to balance relative LTM strength. (E) The responses of avoidance-driving forgetting DAN (magenta), f_{av}, and the attraction-driving LTM MBON (green), m_{at}, from the RFM microcircuit. (F) The responses of attraction-driving forgetting DAN (light green), f_{at}, and the avoidance-driving LTM MBON (orange), m_{av}. (G) The reciprocal restrained memory microcircuit in the mushroom body is made of the MBON-$\gamma 2\alpha' 1$, PAM-$\beta' 2$a, MBON-$\gamma 5\beta' 2$a and PPL1-$\gamma 2\alpha' 1$ neurons. (H) The reciprocal forgetting memory microcircuit in the mushroom body is made of the MBON-$\alpha' 1$, PAM-$\beta 2\beta' 2$a, MBON-$\beta 2\beta' 2$a and PPL1-$\gamma 2\alpha' 1$ neurons. Images of the neurons in G and H were created using the Virtual Fly Brain software (***Milyaev et al., 2012***).

We have identified the reciprocal forgetting memories microcircuit of ***Figure 5***D in the descriptions of ***Aso et al.*** (***2014a***) and ***Li et al.*** (***2020***), where MBON-$\alpha' 1$ is the avoidance-driving LTM MBON, m_{av}, MBON-$\beta 2\beta' 2$a is the attraction-driving LTM MBON, m_{at}, PAM-$\beta 2\beta' 2$a is the avoidance-driving forgetting DAN, f_{av}, and PPL1-$\gamma 2\alpha' 1$ is the attraction-driving forgetting DAN, f_{at}, as shown in ***Figure 5***H. Unfortunately, there is only one PPL1-$\gamma 2\alpha' 1$ per hemisphere and it has been used as the c_{av}, which is both a different type of neuron in our model and represents a different motivation. However, there are multiple axon terminals of this neuron in the mushroom body (e.g. MB296B$_1$ and MB296B$_2$) and each one of them seems to communicate a different response (see ***Figure 8–Figure Supplement 1***). Interestingly, in a following section, we will see that the responses communicated by the MB296B$_1$ terminal are close to the ones produced by c_{av} and the ones of the MB296B$_2$ are close to the ones produced by f_{at}. This implies that different axons of the same DA neuron might create responses that depend on where the axon terminates and actually work as separate processing units in our model. We find the fact that the same neuron can actually represent two neurons in our model interesting and convenient, as the charging, c_{at}, and forgetting DANs, f_{at}, should terminate their axons at the compartment where both the long-term memory, m_{av}, and the restrained MBONs, r_{at}, extend their dendrites, which aligns nicely with this result.

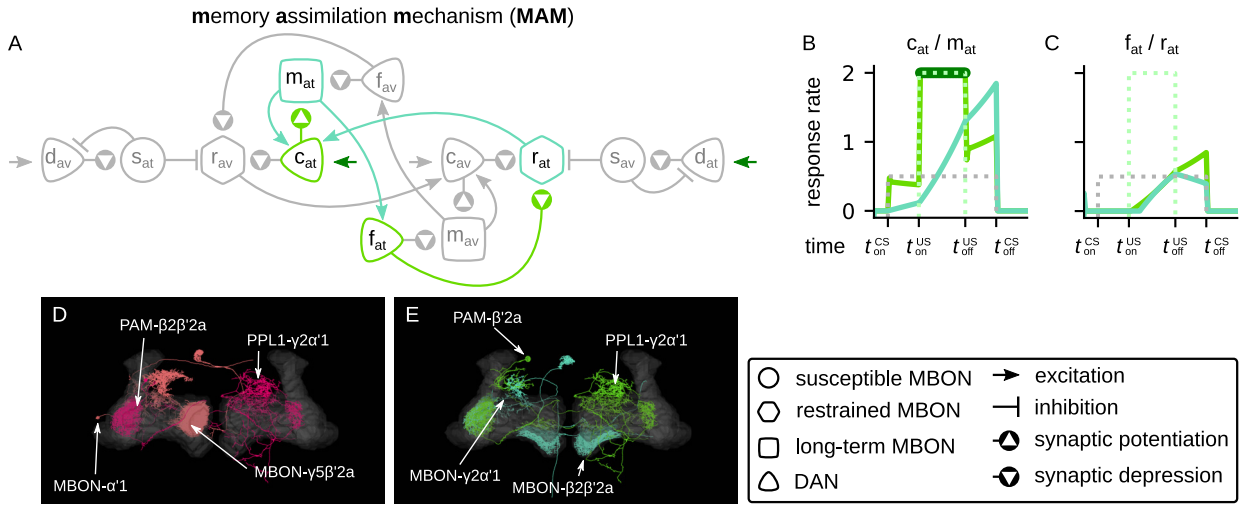


Figure 6. The memory assimilation mechanism (MAM) microcircuit. (A) The forgetting DANs connect to the restrained MBONs of the same valence, hence increasing LTM strength reduces (assimilates) the restrained memory, constituting the Memory assimilation mechanism (MAM) microcircuit. (B) The responses of the reward-encoding charging DAN (light green), c_{at}, and the attraction-driving LTM MBON (green), m_{at}, from the MAM microcircuit. (C) The responses of the attraction-driving forgetting DAN (light green), f_{at}, and the attraction-driving restrained MBON (green), r_{at}. (D) The avoidance-specific MAM is made of the MBON-$\gamma 5\beta'$2a, PPL1-$\gamma 2\alpha' 1$, MBON-$\alpha' 1$ and PAM-$\beta 2\beta'$2a. (E) the attraction-specific MAM is made of the MBON-$\gamma 2\alpha' 1$, PAM-β'2a, MBON-$\beta 2\beta' 2a$ and PPL1-$\gamma 2\alpha' 1$. Images of the neurons in D and E were created using the Virtual Fly Brain software (***Milyaev et al., 2012***).

Table 1. Connections among neurons in the Drosophila mushroom body mapped to the connections of the incentive circuit. Connection types: '⊸', modulates the synaptic weights of the KC→MBON connections terminating in that MBON; '→', excitatory connection; '⊣', inhibitory connection.

Connection in the MB	Connection in the IC	Microcircuit	References
PPL1-$\gamma 1$ped ⊸ MBON-$\gamma 1$ped	d_{av} ⊸ s_{at}	SM	(***Pavlowsky et al., 2018***)
MBON-$\gamma 1$ped ⊣ PPL1-$\gamma 1$ped	s_{at} ⊣ d_{av}	SM	(***Pavlowsky et al., 2018***)
PAM-$\gamma 4 < \gamma 1\gamma 2$ ⊸ MBON-$\gamma 4 > \gamma 1\gamma 2$	d_{at} ⊸ s_{av}	SM	(***Aso et al., 2014a***)
MBON-$\gamma 4 > \gamma 1\gamma 2$ ⊣ PAM-$\gamma 4 < \gamma 1\gamma 2$	s_{av} ⊣ d_{at}	SM	(***Aso et al., 2014a***; ***Cohn et al., 2015***)
MBON-$\gamma 1$ped ⊣ MBON-$\gamma 5\beta'$2a	s_{at} ⊣ r_{av}	RM	(***Felsenberg et al., 2018***)
MBON-$\gamma 4 > \gamma 1\gamma 2$ ⊣ MBON-$\gamma 2\alpha' 1$	s_{av} ⊣ r_{at}	RM	(***Aso et al., 2014a***)
PPL1-$\gamma 2\alpha' 1_2$ ⊸ MBON-$\gamma 2\alpha' 1$	c_{av} ⊸ r_{at}	RRM	(***Aso et al., 2014a***; ***McCurdy et al., 2021***)
MBON-$\gamma 2\alpha' 1$ → PAM-β'2a	r_{at} → c_{at}	RRM	(***McCurdy et al., 2021***)
PAM-β'2a ⊸ MBON-$\gamma 5\beta'$2a	c_{at} ⊸ r_{av}	RRM	(***Aso et al., 2014a***; ***McCurdy et al., 2021***)
MBON-$\gamma 5\beta'$2a → PPL1-$\gamma 2\alpha' 1_2$	r_{av} → c_{av}	RRM	(***Li et al., 2020***)
PPL1-$\gamma 2\alpha' 1_2$ ⊸ MBON-$\alpha' 1$	c_{av} ⊸ m_{av}	LTM	(***Aso et al., 2014a***)
MBON-$\alpha' 1$ → PPL1-$\gamma 2\alpha' 1_2$	m_{av} → c_{av}	LTM	(***Aso et al., 2014a***; ***Li et al., 2020***)
PAM-β'2a ⊸ MBON-$\beta 2\beta'$2a	c_{at} ⊸ m_{at}	LTM	(***Aso et al., 2014a***)
MBON-$\beta 2\beta'$2a → PAM-β'2a	m_{at} → c_{at}	LTM	(***Aso et al., 2014a***; ***Li et al., 2020***)
PPL1-$\gamma 2\alpha' 1_1$ ⊸ MBON-$\alpha' 1$	f_{at} ⊸ m_{av}	RFM	(***Aso et al., 2014a***)
MBON-$\alpha' 1$ → PAM-$\beta 2\beta' 2a$	m_{av} → f_{av}	RFM	(***Li et al., 2020***)
PAM-$\beta 2\beta'$2a ⊸ MBON-$\beta 2\beta'$2a	f_{av} ⊸ m_{at}	RFM	(***Aso et al., 2014a***)
MBON-$\beta 2\beta'$2a → PPL1-$\gamma 2\alpha' 1_1$	m_{at} → f_{at}	RFM	(***Li et al., 2020***)
PPL1-$\gamma 2\alpha' 1_1$ ⊸ MBON-$\gamma 2\alpha' 1$	f_{at} ⊸ r_{at}	MAM	(***Aso et al., 2014a***)
PAM-$\beta 2\beta'$2a ⊸ MBON-$\gamma 5\beta'$2a	f_{av} ⊸ r_{av}	MAM	(***Aso et al., 2014a***)

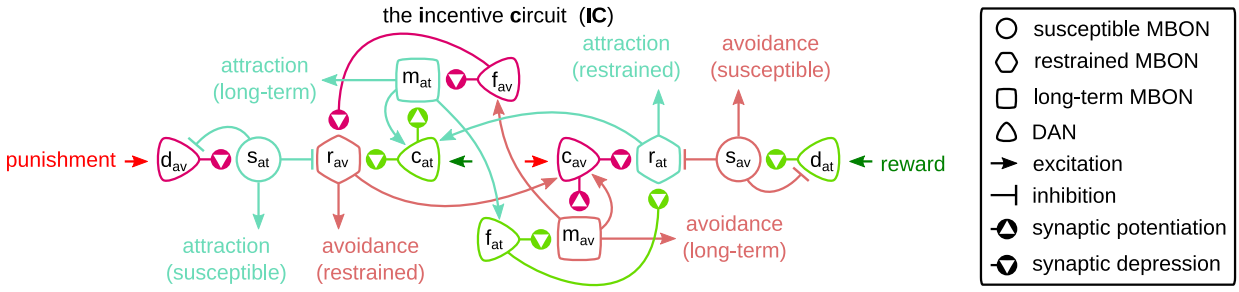


Figure 7. The *incentive circuit* (IC) integrates the different microcircuits of the mushroom body into a unified model allowing the expression of more complicated behaviours and memory dynamics. It unifies the susceptible, restrained, long-term, reciprocal restrained and forgetting memories and the memory assimilation mechanism microcircuits in one circuit that is able to form, consolidate and forget short- and long-term memories.

Memory assimilation mechanism (MAM)

The forgetting DANs allow the developing LTM of one valence to induce forgetting of the LTM of the opposite valence. However the forgetting DANs can also be used for another critical function to maintain flexibility for future learning, which is to erase the memory of the same valence from their respective restrained MBONs. Without this mechanism, the excitation of the charging DANs, c_{at} and c_{av}, by both their respective long-term memories, m_{at} and m_{av}, and restrained MBONs, r_{at} and r_{av}, of the same valence causes a positive feedback loop that could drive the circuit into a state where one LTM MBON will permanently dominate the response. In order to prevent this, we assume that the forgetting DANs, f_{at} and f_{av}, also suppress the KC→ r_{at} and KC→ r_{av} synaptic weights of their respective restrained MBONs, forming the 'memory assimilation mechanism' (MAM) microcircuit (see ***Figure 6***A). This effectively allows memory transfer between the restrained and the *LTM* MBONs, enhancing both the adaptability and the capacity of the circuit. This effect can be observed in ***Figure 6***B and C, where the restrained memory becomes weaker as the LTM becomes stronger, driven by the respective forgetting and charging DANs.

Aso et al. (***2014a***) supports that the forgetting DANs suggested in the RFM microcircuit also depress the KC→restrained MBON synapses of the same valence. More specifically, the avoidance-driving forgetting DAN, f_{av} = PAM-$\beta 2\beta' 2$a, modulates the KC→MBON-$\gamma 4\beta' 2$a synapses, while for the attraction-driving forgetting DAN, f_{at} = PPL1-$\gamma 2\alpha' 1$, modulates the KC→MBON-$\gamma 2\alpha' 1$ synapses, as show in ***Figure 6***D and E respectively. The connections described in all the above microcircuits are summarised in ***Table 1***, while alternative names for these neurons are summarised in Appendix 2 ***Table 1***.

The incentive circuit (IC)

All the microcircuits we have described so far seem to interlock as the unified circuit of 12 neurons shown in ***Figure 7***. For each motivation (attraction or avoidance) there are three types of MBON: susceptible, $s_{\text{at/av}}$, restrained, $r_{\text{at/av}}$, and long-term memory, $m_{\text{at/av}}$; and three types of DAN: discharging, $d_{\text{at/av}}$, charging, $c_{\text{at/av}}$, and forgetting, $f_{\text{at/av}}$ (summarised in ***Table 2***). We call this the 'incentive circuit' (IC), as it motivates the animal to approach or avoid a CS, and it could represent a motif that is repeated in the mushroom body to encode interactions between memories related to opposing motivations.

More specifically, on the outer edges of the model, we have susceptible MBONs, $s_{\text{at/av}}$, which by default respond to all conditional stimuli (CS) restraining the restrained MBONs, $r_{\text{av/at}}$, from expression. The discharging DANs, $d_{\text{av/at}}$, discharge the susceptible MBONs of opposite valence and thus release the restraint of the restrained MBONs of the same valence. The restrained MBONs, $r_{\text{av/at}}$ activate their respective charging DANs, $c_{\text{av/at}}$, which start charging the LTM MBONs of the same valence, $m_{\text{av/at}}$, and at the same time suppress the KC→ $r_{\text{at/av}}$ weights of the restrained MBON of the opposite valence. Similarly, the LTM MBONs, $m_{\text{av/at}}$, enhance the activity of the charging DANs,

$c_{\text{av/at}}$, increasing the momentum of learning (potentiation effect of the dopaminergic learning rule), while simultaneously activating their respective forgetting DANs, $f_{\text{av/at}}$, to decrease the momentum of learning for the opposite LTM MBONs, $m_{\text{at/av}}$, and also suppress the KC→ $r_{\text{av/at}}$ synaptic weights for the same valence restrained MBONs, in order to make space for new memories.

The aversive olfactory conditioning paradigm

So far, we have justified the construction of the incentive circuit from consideration of the connectome in the mushroom body of the adult fruit-fly, coupled with functional suggestions for the role of each neuron in learning. Here we examine whether responses observed in these neurons, collected during the acquisition and reversal phases of an aversive olfactory conditioning paradigm (***McCurdy et al., 2021***), are consistent with the proposed functions. The experimental set-up is summarised in ***Figure 8***A and B.

First, we have a pre-training phase, where the animal is exposed to odour A (OCT) and odour B (MCH) alternately for $5\,\text{sec}$ with $15\,\text{sec}$ breaks (fresh air) in between. This phase runs only once in order to capture the default response of the neuron to each odour and make the animal familiar to it. The acquisition phase follows, where the alternation of the odours continues, while electric shock is also delivered along with odour B (on the 4th second) for $100\,\text{ms}$. This phase is repeated 5 times, followed by half pre-training phase (only exposure to odour A) in order to prepare for the reversal phase. In this phase, the electric shock is delivered with odour A instead of odour B, and it is repeated for only 2 times. ***Figure 8–Figure Supplement 1*** and ***Figure 8–Figure Supplement 2*** summarise the average activity of all the recorded DANs and MBONs of the fly brain respectively, recorded by ***McCurdy et al.*** (***2021***).

Figure 8C shows the recorded activities of the neurons involved in the incentive circuit for the different phases of the task described above. We can see that some general principles described before are visible in these activities. For example, during acquisition, electric shock activates PPL1-$\gamma 1$pedc, which depresses the KC→MBON-$\gamma 1$pedc and release the restraint of MBON-$\gamma 5\beta' 2$a for odour B. At the same time, MBON-$\gamma 5\beta' 2$a excites PPL1-$\gamma 2\alpha' 1$ (in the role of c_{av}) which starts the charging of MBON-$\alpha' 1$ by increasing their KC→MBON synaptic weights. The same DAN causes a decrease in the response of MBON-$\gamma 2\alpha' 1$ and therefore (due to decreases excitation) to of PAM-$\beta' 2$a, which also allows MBON-$\gamma 5\beta' 2$a to increase its activity.

It is harder to see the effects during the reversal phase, as only two trials are available. However, we can clearly see a drop in the activity of MBON-$\gamma 1$pedc and MBON-$\gamma 2\alpha' 1$ for odour A, caused by the increased activity of PPL1-$\gamma 1$pedc and PPL1-$\gamma 2\alpha' 1_2$ respectively due to shock delivery, along with an increase in the activity of MBON-$\gamma 5\beta' 2$a due to the reduced restraint as before. For odour B, the recovery effect of our learning rule restores the synaptic weights of the susceptible and restrained MBON, while its saturation effect keeps the increase or decrease in the synaptic weights of the LTM MBONs stable. Note that the activities shown in ***Figure 8*** are normalised to be between 0 and

Table 2. Map of the neuron types in the incentive circuit to their 'short' and 'code' names. The 'av' or 'at' suffix denote the avoidance- or attraction-driving property of MBON types, or the punishment- or reward-encoding property of DAN types. Note that the forgetting DANs do not encode any reinforcement but instead drive the avoidance- or attraction-specific forgetting.

Neuron type	Short name	Code name	Affiliated motivation
Susceptible MBON	s-MBON	$s_{\text{av/at}}$	avoidance-/attraction-driving
Restrained MBON	r-MBON	$r_{\text{av/at}}$	avoidance-/attraction-driving
Long-term memory (LTM) MBON	m-MBON	$m_{\text{av/at}}$	avoidance-/attraction-driving
Discharging DAN	d-DAN	$d_{\text{av/at}}$	punishment-/reward-encoding
Charging DAN	c-DAN	$c_{\text{av/at}}$	punishment-/reward-encoding
Forgetting DAN	f-DAN	$f_{\text{av/at}}$	avoidance-/attraction-driving

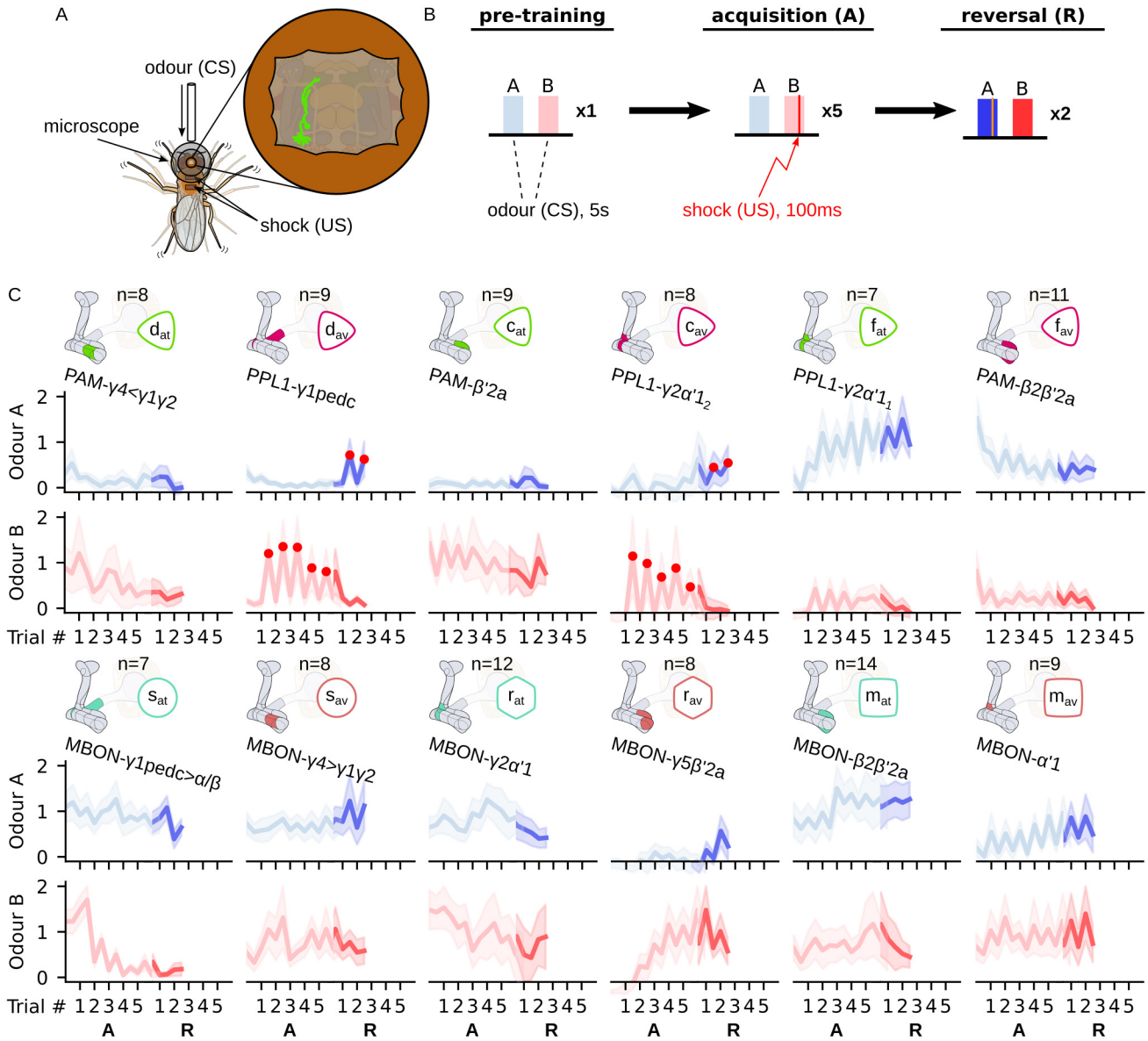


Figure 8. Description of the aversive olfactory conditioning paradigms and responses of neurons in the Drosophila mushroom body affiliated with the incentive circuit. (A) Setup for visualising neural activity via Ca^{2+} imaging during aversive olfactory memory acquisition and reversal. Flies are head-fixed and cuticle dissected for ratiometric imaging of Ca^{2+}-sensitive $GCaMP_6f$ and Ca^{2+}-insensitive tdTomato. (B) The experimental paradigm. 5 sec presentations of odours A (blue) and B (red) continuously alternate, separated by fresh air, while the shock input forms the different phases: 1 repeat of pre-training, where no shock is delivered; 5 repeats of acquisition, where shock is delivered in the last second of odour B; 2 repeats of reversal where shock (thin red line) is paired with odour A. (C) Different panels correspond to different neurons: DANs (top) and MBONs (bottom). At the top of each panel, we show the compartment where the neuron extends its dendrites (for MBONs) or terminates its axon (for DANs) and its name based on this compartment; the sample size, n, used to calculate the means; the neuron from the incentive circuit that we believe it represents. The colour of the neuron in the top row and the one in the compartment denotes its motivation (green – attraction, red – avoidance). Second and third rows in each panel show the mean responses of the neuron in odour A and B respectively, for the acquisition and reversal phases. For each trial there are two response values, corresponding to the first 4 seconds with odour and the final second which (depending on the condition) also includes shock (see Methods). The red dots show the shock-delivery time-steps for the directly affected neurons. The solid lines show the mean activity over different flies and the coloured shades show the standard deviation ($\mu \pm \sigma^2/2$).

Figure 8–Figure supplement 1. Responses of all the recorded DANS in the mushroom body.

Figure 8–Figure supplement 2. Responses of all the recorded MBONs in the mushroom body.

2, so for example the activity of MBON-$\beta 2\beta' 2$a for odour B might be very small and no significant change in its activity happens.

Modelling the neural responses

Many of the activities of the neurons shown in ***Figure 8***C are very noisy, which makes hard to interpret the function of the circuit. Therefore, we computationally modelled the incentive circuit in order to demonstrate all the properties we described before and compare the reconstructed activities to the ones of ***Figure 8***C. We emulated and slightly modified the experimental set-up of ***McCurdy et al.*** (***2021***) and ***Berry et al.*** (***2018***) and compare their findings to the reconstructed responses of our model. More specifically, we run 5 trials of an acquisition phase as described before [or learning phase (***Berry et al., 2018***)], where the animal experiences the pairing of odour B with a reinforcement (electric shock); followed by 5 trials of a *forgetting* phase which could be one of no-shock [or 'extinction' (***Felsenberg et al., 2018***; ***McCurdy et al., 2021***)], unpaired or reversal learning (see ***Figure 9***A). We decided to run 5 trials in order to more clearly demonstrate the behaviour of the neurons. As in the previous section, in the reversal phase, we deliver the shock with odour A instead of odour B, while in the no-shock phase we do not deliver shock to either of the odours. In the unpaired phase we deliver shock during the fresh air break and before delivering odour A. Note that as we are simulating events rather than real time, here the gap between shock and odour is assumed to be too long to express backward conditioning (see Decomposing the dopaminergic factor).

Figure 9B shows the responses produced by the model for the acquisition and the forgetting phases. Starting with the susceptible memory, we see that the reward-encoding discharging DAN, d_{at}, remains silent while the punishment-encoding equivalent, d_{av}, responds to shock (red dots) — neither responds to odour. The susceptible MBONs, s_{at} and s_{av}, initially respond to odour, but the activity of d_{av} caused by shock quickly decreases the response of s_{at} to odour B during the acquisition phase; during reversal the response to odour A decreases, while the response to odour B remains low. For the restrained memory, the continued activity of s_{av} suppresses any response from the attraction-driving restrained MBON, r_{at}, while the response of the avoidance-driving restrained MBON, r_{av} increases as the inhibition from s_{at} decreases, and vice versa.

The reciprocal restrained memory effect can be seen as the increase in r_{av} during acquisition excites, along with the direct input of shock, the punishment-encoding charging DAN, c_{av} (note both charging DANs have some inherent response to odour, but this is due to activation by the restrained and LTM MBONs, not direct input). The activation of charging DAN, c_{av}, will decrease the (already suppressed) responsiveness of the attraction-driving restrained MBON, r_{at}, to odour B during acquisition and odour A during reversal. Note however that there is an observable 'reciprocal' effect that the response of this MBON increases to odour A during acquisition and to odour B during reversal, i.e., the odour paired with 'absence of shock' is becoming more attractive. The overall low response of r_{at} also reduces the activity in the reward-encoding charging DAN, c_{at}, which prevents changes in the odour response of the avoidance-driving restrained MBON, r_{av}.

The long-term memory is also affected by the increased activity of the punishment-encoding charging DAN, c_{av}, during the acquisition phase. This gradually increases the avoidance-driving LTM MBON, m_{av}, response to odour B, setting a positive momentum for increasing weights that continues for the following reversal phase. This in turn activates the Reciprocal forgetting memories (RFM) circuit, specifically the aversive forgetting DAN, f_{av}, becomes more active for odour B inducing a negative momentum for the responsive to this odour of the attraction-driving LTM MBON, m_{at}. Note that again we observe a reciprocal effect in which the opposite momentum for odour A occurs in both LTM MBONs. Reversal weakens the momentum, but it would require a longer reversal phase to change direction.

We can see that most of trends we talked about in the previous section during the acquisition [first 5 trials (A) - faint line] and reversal phases [next 5 trials (R) - dark line] are captured by the computational model. We also show its behaviour during the other two forgetting phases (no-shock

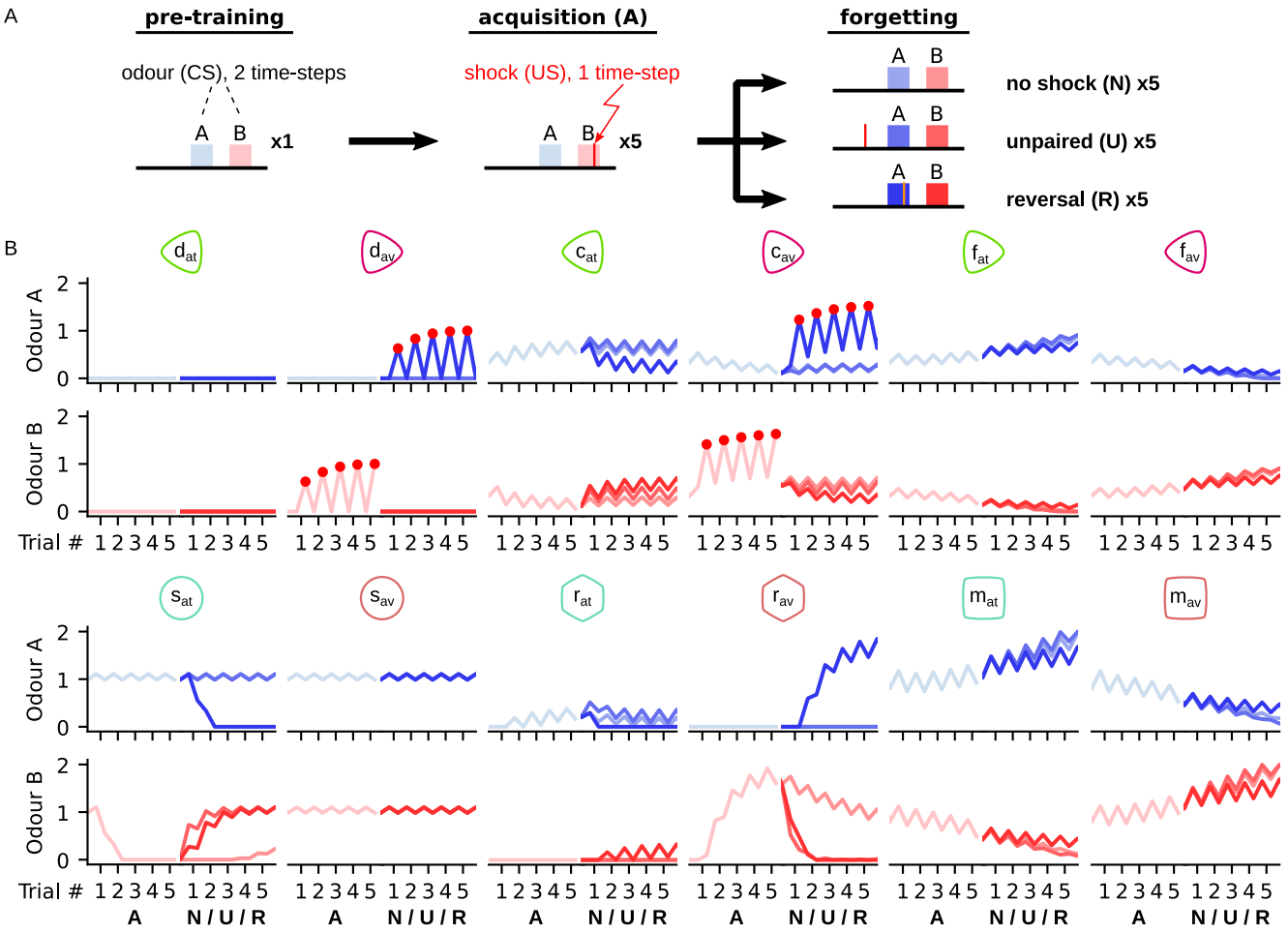


Figure 9. Responses generated by the incentive circuit during the aversive olfactory conditioning paradigms. (A) The experimental paradigms. Odours A and B continuously alternate, while the shock input forms the different phases. First we have 1 repeat of the pre-training phase, where no shock is delivered; then we have 5 repeats of the acquisition phase, where shock is delivered paired with odour B; and finally we have 5 repeats of a forgetting phase (no-shock, unpaired or reversal). The shade and colours of the odours in each experiment is matched to the respective lines of ***Figure 9***B. Blue and red shades denote exposure in odours A and B respectively. The red thin line denotes shock delivery. (B) The responses generated by the incentive circuit for odour A (first and third rows; lines in blue shades) and odour B (second and firth rows; lines in red shades) during the acquisition (A) followed by forgetting, i.e. no shock (N), unpairing (U) or reversal (R) phases (different shades denote the different phases in ***Figure 9***A). Red dots denote shock delivery to the appropriate neurons and time-steps, as in ***Figure 9***A. For each trial we process the odour and shock input in 3 time-steps: (i) no odour and unpaired shock is delivered (unpaired phase for odour A) or none of them (all other phases), (ii) only odour is delivered, and (iii) odour and shock are delivered simultaneously (acquisition phase for odour B and reversal phase for odour A) or only odour (all other phases). Time-step (i) always produces zero response-rate, so only time-steps (ii) and (iii) are visualised in order to make the trends obvious and are equivalent to the 'off-shock' and 'on-shock' values in ***Figure 8***.

Figure 9–Figure supplement 1. KC→MBON synaptic weights for the MBONs of the model during the different phases.

Figure 9–Figure supplement 2. Generated responses when using the prediction-error learning rule.

Figure 9–Figure supplement 3. KC→MBON synaptic weights for the MBONs of the model using the prediction-error learning rule.

Figure 9–Figure supplement 4. Generated responses of the model during the classic unpaired conditioning paradigm.

and unpaired conditioning - lighter lines). Note that the responses in odour B during unpaired forgetting are very similar to the ones in reversal, while the responses in odour A are similar to the ones of the no-shock forgetting. This shows the strong effect of the learning rule, which has the same effects during the no-shock and unpaired phases for odour A, and during the unpaired and reversal for odour B. However similar, the responses are not identical, which is because of the dynamics induced by the circuitry.

We can also notice that some of the trends produced by the model are not consistent with neural behaviours observed in fruit flies. For example, although our model predicts an increase in the response of MBON-$\gamma 1$pedc>α/β for odour B during the reversal and no-shock phases, ***Felsenberg et al.*** (***2018***) and our own observations do not support this prediction, as we have not observed an increase in the responses even after 6 trials. This could be due to inhibitory connections among the susceptible neurons [MBON-$\gamma 4 > \gamma 1 \gamma 2 \dashv$ MBON-$\gamma 1$pedc> α/β (***Aso et al., 2014a***); MBON-$\gamma 1$pedc> $\alpha/\beta \dashv$ MBON-$\gamma 4 > \gamma 1 \gamma 2$ (***Li et al., 2020***)] which are not represented in the model. This would work as a winner-takes-all (WTA) microcircuit for the susceptible MBONs, and allow only the MBON with the highest activity (or a discrete subset of them) to fire while the rest of them remain silent. Similarly to the restrained MBONs, the restraint of their responses does not mean that their synaptic weights do not change; the observation of sharp changes in the responses of these neurons (***Felsenberg et al., 2018***) seems consistent with gradual underlying changes being masked then suddenly revealed by WTA dynamics.

Modelling the behaviour

In terms of the behaviour of fruit flies, we suggest that three MBON types drive attraction and three avoidance. This results in six driving forces per available odour that are combined in order to produce the behaviour of the animal. Three of them (per odour identity) should work as attractive forces (towards odours) and the rest three as repulsive ones (see ***Figure 10***). A simple 'behavioural' readout (used in many previous models) would be to take the sum of all attractive and aversive forces at some time point as a measure of the probability of animals 'choosing' odour A or B, comparable to the standard two-arm maze assay used in many Drosophila studies. However this would not allow us to explore the full dynamics of the circuit as animals simultaneously explore, learn, express learning and forget, while moving through a world with odours. Instead, we simulate the behaviour produced by the incentive circuit in simulated flies in a virtual arena, where they are exposed to two odour gradients, of different strengths, and variously paired with reinforcements. As we have full access to the neural responses, the synaptic weights and the position of the simulated flies for every time-step, this allows us to identify different aspects of the produced behaviour and motivation, including the effect of the LTM on the behaviour and whether 'choice' occurs because the animal is attracted by one odour or repulsed by the other. We can then derive a behavioural preference index based on the location of the simulated flies during relevant time periods (see Calculating the preference Index).

More specifically, we simulate the behaviour of $n = 100$ flies in an arena with two odour sources producing odour gradients, modelled as Gaussian distributions with different standard deviations to produce a mixture of odours of uneven intensity (see ***Figure 11***A). Flies perceive a mixture of both odours where intensity depends on their position with respect to the distributions of the odours – i.e. the closer to a source the more intense the odour stimulation is. The experiments last for $100\,\mathrm{sec}$ each and they are split in 3 phases as shown in ***Figure 11***A. In *pre-training*, the flies are placed in the centre of arena and explore freely for $20\,\mathrm{sec}$. In *training*, either shock or sugar is associated with the region where odour A is stronger, or the region where odour B is stronger, or with both regions (whole arena) for $30\,\mathrm{sec}$. In *post-training*, we remove the reinforcement and let the flies express their learnt behaviour for another $50\,\mathrm{sec}$ creating a no-shock (or a no-sugar) forgetting condition. ***Figure 11***B shows the odour preference of the flies during the different phases for each of the six training conditions, after 1, 5 or 10 repeats of the experiment, when their behaviour is controlled by a combination of the attractive and repulsive forces on the two odours. The actual paths of the

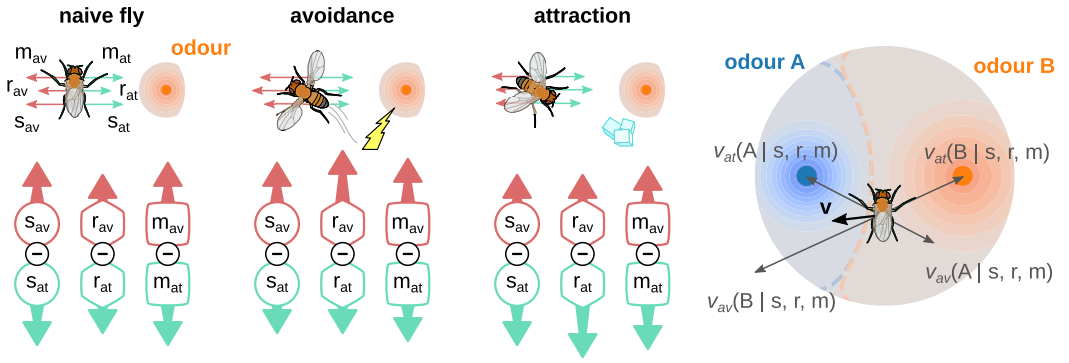


Figure 10. Modelling the behaviour of the simulated flies by a combination attractive and repulsive forces. MBON activity is translated into forces that drive a simulated fly towards or away from odour sources. For naive flies, there is a balance in the forces controlled by the MBONs of the incentive circuit. When electric shock is paired with an odour, the avoidance-driving MBONs have higher activity, which guides the fly away from the odour source. When sugar is paired with an odour, the attraction-driving MBONs have higher activity, guiding the fly towards the paired odour source. Combining the attractive and repulsive forces that drive the fly we conclude in the overall driving force, $\mathbf{v}$, which drives the behaviour of the fly. The mechanism that detects the odour source is not part of this study.

flies for all the 10 repeats are illustrated in ***Figure 11–Figure Supplement 5***. In all cases during the first repeat, the initial (pre-training) behaviour is an unbiased random exploration (blue boxes), but the behaviour differs during the training and post-training phases as well as the following repeats.

During the first repeat of the experiment, the circuit guides the animal towards the most 'beneficial' area, that is, when sugar is presented, we see the flies moving away from the non-rewarded area, and when electric shock is delivered they move towards the non-punishing area, in both cases driven by the susceptible MBONs. Restrained MBONs allow the animal to explore, which explains the not very directed behaviour. The effect of the LTM MBONs is initially weak, hence the behaviour displayed is not clear attraction towards an odour. This is why we do not see the flies concentrating at the source of the odour in ***Figure 11–Figure Supplement 5*** (green lines during the first repeat) but instead spread across the 'beneficial' area, while we see stronger concentration during the post-training phase (especially in the shock-omission case). In post-training where sugar is ommitted, the susceptible MBONs will revert to their original responses (see ***Figure 11–Figure Supplement 1***), also silencing the restrained MBONs, so now the behaviour is guided only by the LTM MBONs, whose effect becomes stronger and drives the animal towards the 'safe' odour source for the punishment cases and away from the least beneficial one for the reward cases. This allows the LTM-charging to continue further increasing the confidence about the selected odour. In the case where the reinforcement is uniformly distributed in the whole arena (odour A/B), flies express a preference to the one or the other odour (depending on their experience) for the punishment case (see bimodal distribution in ***Figure 11***B), while they show no preference to either of them for the reward case, instead occupying the central area. This is because they have actually experienced reinforcement paired with the absence of each odour (as the flies are initially placed in the central area where the stimulation from both odours is weak) and therefore do not acquire attraction to the odours, as might be expected.

By repeating the above experiment 5 times, the flies are exposed to the odours of interest for longer and this builds up the expected behaviour. As odour B is more strongly represented in our set-up, the dominant association becomes for the flies to be attracted by odour B when electric shock in delivered with odour A. When the shock is paired with odour B (in either 'shock+odour B' or 'shock+odour A/B' cases), the aversive association dominates, driving the animals away from odour B almost ignoring odour A. In the sugar-delivery case, the exploration of the rewarding region allows the animals to gain more experience of the paired odour and finally express attraction, driving

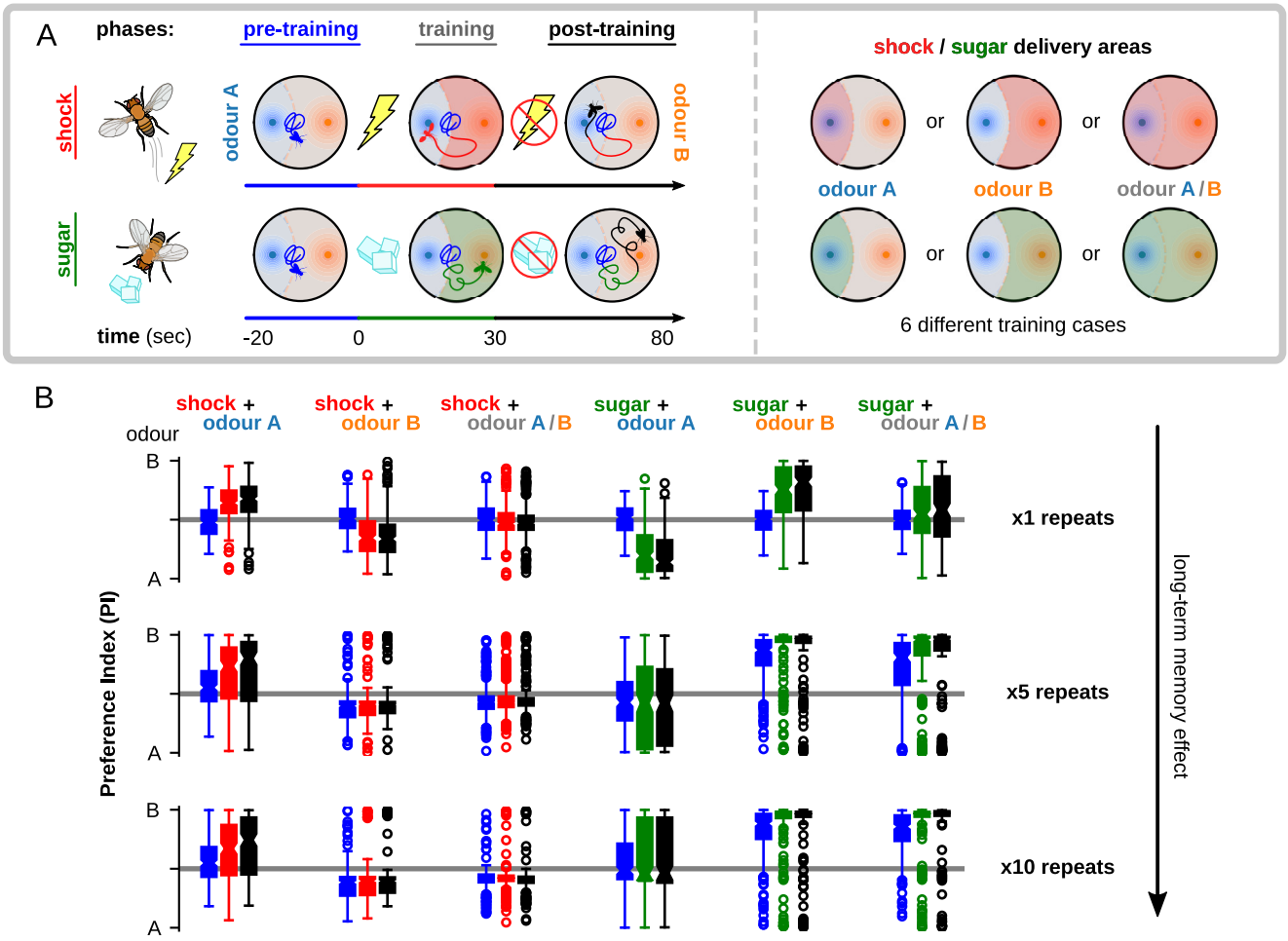


Figure 11. The behaviour of the animal controlled by its neurons during the freely-moving flies simulation. The animal is exposed to a mixture of two odours, whose relative intensity depends on the position of the animal in space. (A) Each experiment lasts for $100\,\text{sec}$ where: the animal is placed at the centre of the arena in time-step $t = -20\,\text{sec}$. During the first $20\,\text{sec}$ (pre-training phase, $t \in [-20, 0)$) the animal explores the arena without any reinforcement (blue tracks/boxes). In the next $30\,\text{sec}$ (training phase, $t \in [0, 30)$) the animal receives reinforcement under one of the six training cases shown on the right: using sugar (green) or shock (red); and reinforcing when the intensity of odour A (shock + odour A) or odour B (shock + odour B) is the highest or irrespective of the odours' intensity (shock + odour A/B). During the last $50\,\text{sec}$ (post-training phase) $t \in [30, 80)$) the animal continues being exposed to the odours without receiving a reinforcement (black tracks/boxes). (B) Distributions of the preference index (PI) extracted from the position of $n = 100$ simulated flies at the end of each phase of the experiments (different colours of box plots) for different training cases (each column). We repeat the experiments for 10 times in order to show the effects of the long-term memory in the behaviour and we plot the PI of the 1st, 5th and 10th repeats in different rows.

Figure 11–Figure supplement 1. The mean KC→MBON synaptic weights over 100 flies for each neuron and time-step and for 1, 5 and 10 repeats.

Figure 11–Figure supplement 2. Summary of behaviour over 100 flies when controlled by the susceptible MBONs only for 1, 5 and 10 repeats.

Figure 11–Figure supplement 3. Summary of behaviour over 100 flies when controlled by the restrained MBONs only for 1, 5 and 10 repeats.

Figure 11–Figure supplement 4. Summary of behaviour over 100 flies when controlled by the LTM MBONs only for 1, 5 and 10 repeats.

Figure 11–Figure supplement 5. Paths of 100 flies when using the dopaminergic learning rule and during all 10 repeats.

Figure 11–Figure supplement 6. Summary of behaviour over 100 flies when using the RPE learning rule for 1, 5 and 10 repeats.

Figure 11–Figure supplement 7. Paths of 100 flies when using the RPE learning rule and during all 10 repeats.

them towards the rewarded odour. When both are rewarded, again, the dominant role of odour B attracts more flies than odour A. This behaviour is now mostly driven by the LTM MBONs (see ***Figure 11–Figure Supplement 4***), while the role of the restrained (see ***Figure 11–Figure Supplement 3***) and susceptible ones (see ***Figure 11–Figure Supplement 2***) becomes to continuously update the LTM trying to shape the behaviour to perfection. This verifies our hypothesis that the roles of the susceptible and restrained MBONs are more important in contexts (i.e. KC activity) naive to the animal, while the long-term memory MBONs take control of the behaviour with experience in the specific context.

Due to the MAM microcircuit, the saturated LTMs constantly erase their respective restrained memories, which allows the restrained memories of the opposite valence to operate. This means that the restrained memories always question the LTMs, trying to learn an alternative behaviour for the animal. This effect becomes stronger when the LTM is less certain, e.g. in the 'sugar+odour A' case, which partially produces avoidance from both odours after 10 repeats of the experiment (see ***Figure 11–Figure Supplement 5***). More confident LTMs lead to more stable behaviours, becoming almost fixed after the same number of repeats.

Dopaminergic learning rule vs reward prediction-error

We have already shown that our novel dopaminergic learning rule (DLR) and the connectome of the incentive circuit build a powerful model for memory dynamics and behavioural control. In order to verify the importance of our DLR in the model, we run the same experiments by replacing it with the reward prediction error (RPE) learning rule (***Rescorla and Wagner, 1972***). Variations of RPE are often used as learning rules in fruit fly models (***Balkenius et al., 2006***; ***Young et al., 2011***), and MBON→DAN feedback connections have recently been interpreted as serving this function (***Zhao et al., 2021***; ***Bennett et al., 2021***; ***Springer and Nawrot, 2021***; ***Eschbach et al., 2020***).

The idea behind RPE is that the fly learns to predict how rewarding or punishing a stimulus is by altering its prediction when this does not match the actual reward or punishment experienced (***Zhao et al., 2021***). This can be adapted to the mushroom body circuit by assuming for a given stimulus represented by KC activation, the MBON output is the prediction, and the KC→MBON synaptic weights should be altered (for the active KC) proportionally to the difference between the MBON output and the actual reinforcement signalled by the DAN.

$$\Delta W_{k2m}^{ij} = \eta k^i(t)[d^j(t) - m^j(t) + w_{\text{rest}}] \tag{5}$$

where $k^i(t)$, $d^j(t)$ and $m^j(t)$ are the activation of the ith KC, the jth DAN and the jth MBON respectively in time-step t, and $\eta = \frac{1}{\tau} > 0$ is the learning rate which works as the inverse of a time constant. Note that this rule allows updates to happen only when the involved KC is active, i.e. $k^i(t) > 0$, implying synaptic plasticity even without DAN activation but not without KC activation, which is in contrast with our DLR and recent findings (***Berry et al., 2018***; ***Hige et al., 2015***) [also in larva (***Schleyer et al., 2018***, ***2020***)].

This effect, i.e. learning when the KC is active even without DAN activation, is visible in ***Figure 9–Figure Supplement 2*** and ***Figure 9–Figure Supplement 3***, where we can see that, for the susceptible MBONs, the synaptic weights recover every time before the shock delivery, when the odour is presented alone, resulting in no meaningful learning and cancelling their susceptible property. Restrained MBONs look less affected (at least in this experimental set-up), while the long-term memory MBONs lose their charging momentum obtained by the saturation effect, resulting in more fragile memories. Furthermore, due to the KC (instead of dopamine) gating of this learning rule, the responses during the unpaired and no-shock conditions look identical in all neurons, while the reversal makes a difference only on the responses to odour A. In general, the responses reproduced using the RPE learning rule have none of the properties of our model that have been shown earlier and also they cannot explain the dynamics of the responses recorded from animals.

In line with what the neural responses reveal, the behaviour of the simulated fly is also affected. As the RPE learning rule essentially eliminates any differences in the MBON properties, it generates

a stiff behaviour that shows no improvement with experience. The fact that the KC activity gates learning and that the prediction error can either be positive or negative, makes the valence of the reinforcement ambiguous allowing it to drive both attraction and avoidance from the associated odour. ***Figure 11–Figure Supplement 7*** and ***Figure 11–Figure Supplement 6*** show that, although the behaviour of the flies looks correct during the first run, it gets confused in the following runs, maintaining a wrong behaviour for the following 9 repeats of the experiment. After contextually pairing an odour with a reinforcement, the odour not associated with the reinforcement becomes more or less attractive guiding the animal towards or away from it, while at the same time the prediction error becomes zero. However, during the post-training phase, the learning will be reversed, forgetting everything and guiding learning to the opposite direction. As we can see in ***Figure 11–Figure Supplement 7***, this leads in confusion during the second run, partially driving learning towards the opposite direction. Note that the intensity of the odour alone (which is related to the distance of the animal from the source) affects both the KC and MBON activity, and as a consequence of the circuitry the DAN activity, the prediction error and the KC→MBON synaptic weights affiliated with this odour. Because of the confusion in the behaviour of the animal during the second run, now the animal starts becoming more attracted towards the most intensive odour (which is odour B) or to none of them, occupying the central area or moving towards odour B in every condition. This behaviour is the result of extreme generalisation of all experiences into a single behavioural tendency to avoid diversity.

Discussion

We have shown that the combination of our novel 'dopaminergic learning rule' (DLR) with the incentive circuit (IC) of the mushroom body is able to generate similar neural responses and behaviours to flies in associative learning and forgetting paradigms. We demonstrate that the commonly used RPE learning rule is not sufficient to explain all the data collected from the animal. Regarding our model, we provide evidence for the existence of all hypothesised connections and suggest that at least 3 types of mushroom body output (susceptible, restrained and long-term memory) and 3 types of dopaminergic neurons (discharging, charging and forgetting) exist in the fruit fly brain, discriminated by their functionality. As we show, this forms a unified system for rapid memory acquisition, transfer from short-term to long-term memory, and exploration/exploitation trade-off in subsequent behaviour.

Advantages of the dopaminergic learning rule

The proposed dopaminergic learning rule (DLR), while remaining very simple, allows the animal to express a variety of behaviours depending on their experience. The rule remains local to the synapse, that is, it depends only on information that is plausibly available in the presynaptic area of the KC axon (***Equation 4***): the activity of the KC, the level of DA, and the deviation of the current 'weight' from a set-point 'resting weight'. We note that it was not possible to obtain good results without this third component to the rule, although the underying biophysical mechanism is unknown; we speculate that it could involve synapsin as it has a direct role in regulating the balance of reserve and release vesicle pools, and is required in the MB for associative learning ***Michels et al.*** (***2011***). The rule also introduces a bidirectional 'dopaminergic factor' based on the results of ***Handler et al.*** (***2019***) who showed the combination of DopR1 and DopR2 receptor activity can result in depression or potentiation of the synapse. In our learning rule, a positive or negative dopaminergic factor combined with active or inactive KCs leads to four possible effects on the synapse: depression, recovery, potentiation and saturation. This allows substantial flexibility in the dynamics of learning in different MB compartments.

In particular, the saturation allows LTM MBONs to consolidate their memories and makes it very hard to forget. This only occurs for consistently experienced associations, which then become strongly embedded. Only a persistent change in the valence of reinforcement experienced with a given stimuli can recover the activity of LTM MBONs through the Reciprocal forgetting memories

(RFM) microcircuit, which equips the circuit with flexibility even in the long-term memories. Further, the fact that the DLR allows short-term (restrained) and long-term memories to interact through the Memory assimilation mechanism (MAM), increases the capacity of the circuit. Whatever the restrained MBONs learn is eventually assimilated by the LTM MBONs, opening up space for the formation of new memories in the restrained MBONs. The DLR plays an important role in this mechanism, as we saw earlier, and the connectivity alone is not enough for it to work properly.

By contrast, the reward prediction-error (RPE) learning rule lacks this flexibility, and fails to maintain useful long-term memories when applied to the same circuit architecture. A literal interpretation of RPE for the MB would require that the difference (error) between the post-synaptic MBON activity and the DA level is somehow calculated in the pre-synaptic KC axon. This seems inconsistent with the observation that learning is generally unaffected by silencing the MBONs during acquisition (***Hige et al., 2015***; ***Krashes et al., 2007***; ***Dubnau et al., 2001***; ***McGuire et al., 2001***). Alternatively (and not directly requiring MBON activity in the KC learning rule) the RPE could be implemented by circuits (***Bennett et al., 2021***; ***Springer and Nawrot, 2021***; ***Eschbach et al., 2020***) in which DANs transmit an error signal computed by their input reinforcement plus the opposing feedback from MBONs (i.e. MBONs inhibit DANs that increase the KC→MBON synaptic weights, or they excite those that suppress the synaptic weights). However, although the evidence for MBON-DAN feedback connections is well-grounded, it is less clear that they are consistently opposing. For example, in the microcircuits we have described, based on neurophysiological evidence, some DANs that depress synaptic weights receive inhibitory feedback from MBONs (***Pavlowsky et al., 2018***) and some DANs that potentiate synaptic weights receive excitatory feedback from DANs (***Ichinose et al., 2015***). As we have shown, the DLR is able to operate with this variety of MBON-DAN connections.

The conditioning effects of the model

During the past decades, a variety of learning effects have been investigated in flies, including forward and backward (relief) conditioning, first- and second-order conditioning and blocking, which we could potentially use to challenge our model. In The dopaminergic learning rule, we demonstrate that our model supports the backward (or relief) conditioning results presented in ***Handler et al.*** (***2019***). Backward conditioning is when the reinforcement is delivered just before the odour presentation and it is based on the time-dependency between the two stimuli. ***Handler et al.*** (***2019***) suggest that the backward conditioning is a mechanism driven by ER-Ca^{2+} and cAMP in a KC→MBON synapse, when a single DAN releases DA on it. In our model, we assume that different time-courses in the response of DopR1 and DopR2 receptors cause the different patterns of ER-Ca^{2+} and cAMP, resulting in the formation of opposite associations for forward and backward conditioning. We note however that in our model, the effect also requires that the target MBON inhibits the respective DAN (as in our susceptible memory microcircuits) altering the time course of neurotransmitter release. This may suggests backward conditioning does not occur in all MB compartments. We believe this mechanism for backward conditioning is better supported than the hypothesised mechanism of post-inhibitory rebound in opposing valence DANs presented in ***Adel and Griffith*** (***2021***), although some role for both mechanisms is possible.

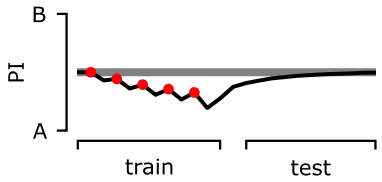


Figure 12. The preference index (PI) of the agent during the classic unpaired conditioning paradigm. During the training phase, we deliver electric shock or odour A alternately. During the test phase, we deliver odour A and odour B alternately. The preference index (PI) is calculated by using the MBON responses for each odour (see ***Figure 9–Figure Supplement 4***).

Backward conditioning can be distinguished from the unpaired conditioning effect; the latter involves the presentation of reinforcement and a specific odour in alternation with less temporal

proximity. It has been observed (***Jacob and Waddell, 2020***; ***Schleyer et al., 2018***) that this procedure will produce a change in response to the odour that is opposite in valence to the reinforcement, e.g. approach to an odour that is 'unpaired' with shock. Note that this effect can be observed both in standard two odour CS+/CS- training paradigms (where an altered response to CS-, in the opposite direction to CS+, is often observed) but also in single odour unpaired paradigms. Surprisingly, our model also produces unpaired conditioning, notably through a different mechanism than backward conditioning. When DANs are activated by a reinforcement without KC activation, the weights of all KCs are potentially altered, e.g., restored towards their resting weight or slightly potentiated. This alteration means that subsequent presentation of odour alone can be accompanied by MBON-driven activation of DANs, resulting in specific alteration of the weights for the presented odour. In the example of ***Figure 12***, odour A starts to self-reinforce its attractive LTM when presented in alternation with shock, and will be preferred to an alternative odour B in subsequent testing. However, repeated presentation of other odours during testing, without further shock, leads to generalisation (equal preference to all experienced odours).

The self-reinforcing property of the positive feedback in the LTM microcircuit can also account for second-order conditioning. If a motivation has been associated to an odour, MBONs related to that motivation will have increased activity when the odour is delivered, even in the absence of reinforcement. In the LTM microcircuit, the positive MBON-DAN connection will consequently activate the charging DAN, so any additional cue (or KC activity) presented alongside the learned odour will also experience increase in the respective KC→MBON weights, creating a similar charging momentum and resulting in a second-order association. Perhaps surprisingly, this predicts that second-order conditioning might happen directly in the long-term memory microcircuit without being filtered by the susceptible and restrained memories first. This would be consistent with the observation that second-order conditioning in flies requires strong induction of the first-order memory and that first-order memory does not appear to be extinguished by the absence of reinforcement during second-order training (***Tabone and Belle, 2011***).

Finally, although we have not tested it explicitly here, it is clear that our learning rule (unlike RPE) would not produce blocking. The blocking effect, as described by ***Kamin*** (***1967***), is when the conditioning to one stimulus subsequently blocks any conditioning to other elements of a mixture including that stimulus. Under RPE learning, this is explained by the first stimulus already correctly predicting the reinforcer, so there is no error to drive a change in the weights. Using the DLR, the update of each synaptic weight does not depend on the responses of the output neurons, so blocking does not happen, which is consistent with the observed behaviour of fruit flies (***Young et al., 2011***; ***Brembs and Heisenberg, 2001***). Although presentation of a learned odour along with a novel odour might, through feedback from the MBONs, alter the DAN responses to the reinforcement, in our circuit this is not generally an opponent feedback so will not cancel the reinforcing effects for the novel odour. This also highlights the difference of our Susceptible memory (SM) and Long-term memory (LTM) microcircuits from the RPE circuits described ***Bennett et al.*** (***2021***); ***Springer and Nawrot*** (***2021***); ***Eschbach et al.*** (***2020***) and ***Zhao et al.*** (***2021***). Nevertheless, as ***Wessnitzer et al.*** (***2012***) and later ***Bennett et al.*** (***2021***) suggest, the fact that blocking has not been observed in fruit flies could also be explained by the way that the mixture of odours is represented by the KCs, i.e., that it might not be simply the superposition of the activity patterns of the individual odours.

Additional mushroom body connections

Our model suggests that only KC→MBON, MBON⊣DAN, MBON→DAN and DAN⊸MBON connections are essential for successful learning in the mushroom bodies. However, there are a number of additional known connections in the mushroom bodies, such as KC→APL, APL⊣KC, DAN→MBON, axoaxonic KC→KC and KC→DAN connections that have been neglected in this model, and need further consideration.

In the larval brain, there are two anterior paired lateral (APL) neurons, one for each mushroom body. They extend their dendrites to the lobes of the mushroom bodies and terminate their axons

in the calyxes releasing the inhibitory GABA neurotransmitter (***Tanaka et al., 2008***). Although there are still two of them, in the adult brain both their dendrites and axons are innervating the calyx and the lobes (***Wu et al., 2013***), suggesting that they function as both global and local inhibitory circuits. Moreover, DAN¬APL (***Liu and Davis, 2009***) and APL¬DAN (***Wu et al., 2012***) connections have been proposed, but there is no clear description of what their function is. Several previous models (***Peng and Chittka, 2017***; ***Delahunt et al., 2018***) have demonstrated that a potential function for this global/local inhibition network is gain control such that the total number of KCs firing to different stimuli remains similar, and indeed that the same effect can be implemented using a flexible threshold for KC firing (***Saumweber et al., 2018***; ***Zhu et al., 2020***; ***Zhao et al., 2020***). In our model we have simplified the KC input, representing just two odours as different patterns across a small number of KCs ($n_k = 10$) with a total activity that sums to 1 for each odour, so the hypothesised gain control function of the APL is not useful here. However it remains an interesting question whether there is learning between the KC and APL in the lobes (***Zhou et al., 2019***), or between the APL and KC in the calyx, and what role this might play in the overall dynamics of memory acquisition.

In addition, ***Eichler et al.*** (***2017***) suggest that most of the KCs input, i.e. $60\,\%$, is from other KCs. We suggest these connections (together with the ones from the APL) might create local winner-takes-all (WTA) networks that force a limited number of KCs per compartment to be active at one time. This predicts that it is possible for the same KC axon to be active in one compartment but inactive in another [consistent with recent data from ***Bilz et al.*** (***2020***)], and that an almost fixed number of KCs might be active at all times, even when no odour is delivered (e.g. fresh air only) enabling acquisition and forgetting at all times. ***Ito et al.*** (***2008***) shows that KCs can be active even in the absence of odours but with no consistent spiking, which is a characteristic of WTA networks when the underlying distribution of spikes across the neurons is almost uniform.

Eichler et al. (***2017***) also observed (from electron microscopy reconstruction in larva) that within a compartment, in a 'canonical microcircuit', KCs make direct synapses to the axons of DANs, and that DAN pre-synapses often simultaneously contact KCs and MBONs. The same connections have been observed in adult Drosophila by (***Takemura et al., 2017***). The extent to which KCs (and thus odour inputs) might be directly exciting DANs remains unclear. ***Cervantes-Sandoval et al.*** (***2017***) show that stimulating KCs result in increased DAN responses and that DANs are activated through the ACh neurotransmitter. However, we note that in our model, such an effect could be explained without assuming a direct connection. For example, in the 'long-term memory' microcircuit, activating the KCs result in increased activity of the LTM MBON which excites the respective charging DAN. The DAN that ***Cervantes-Sandoval et al.*** (***2017***) provide evidence from is PPL1-$\alpha 2\alpha' 2$, which gets excited by MBON-$\alpha 2\alpha' 2$ neurons, as it is characterised by the ACh neurotransmitter (***Aso et al., 2014a***). In our terms, this could be an LTM MBON that excites its respective charging DAN, PPL1-$\alpha 2\alpha' 2$ (***Li et al., 2020***), and provide the source of ACh detected on it. More generally, the altered activity of DANs in response to odours that has been observed during learning can be also observed in our model, without requiring direct KC→DAN connections or their modification. Nevertheless, such connections may possibly play a role in enhancing the specificity of dopamine-induced changes in KC→MBON connectivity. Interestingly, the depression of KC→DAN synapses, in parallel with KC→MBON synapses, could provide an alternative mechanism for implementing RPE learning (***Takemura et al., 2017***).

Takemura et al. (***2017***) demonstrate that the direct synapses observed from DANs to MBONs are functional in altering the MBON post-synaptic current to DAN activation, independently of KCs. This could be a mechanism by which learnt responses to reinforcements are coordinated with the current presence or absence of the reinforcement (***Schleyer et al., 2020***, ***2011***; ***Gerber and Hendel, 2006***). Another possibility is that post-synaptic as well as pre-synaptic changes might be involved in learning at the KC→MBON synapse.

Beyond attraction and aversion

The incentive circuit (IC) consists of 6 MBONs and 6 DANs that link a pair of antagonistic motivations, attraction and avoidance. However, there are ~ 34 MBONs and ~ 130 DANs in the mushroom body of the adult fruit fly brain, within which the IC is an identifiable motif. We suggest the possibility that this motif could be repeated, representing additional opposing motivations, with some neurons having multiple roles depending on the motivational context as proposed by ***Cohn et al.*** (***2015***), working either as restrained MBONs and discharging DANs, or as LTM MBONs and forgetting DANs depending on the reinforcer identity. We have illustrated this concept of a unified system of motivations as The incentive wheel. This could explain how PAM-$\beta 2\beta' 2$a [i.e. MB301B (***May et al., 2020***)] is a sugar-encoding discharging DAN in the appetitive olfactory conditioning context, but it also is an avoidance-driving forgetting DAN in a different context (e.g. aversive olfactory conditioning). In addition, 2 MBONs of the incentive circuit do not interact with the α'/β' KCs of the mushroom body. MBON-$\gamma 4$>$\gamma 1\gamma 2$ and MBON-$\gamma 1$pedc, are part of two autonomous microcircuits, i.e. the susceptible memories (SM), and are working under the context provided by the ~ 675 γ-KCs relative to the task. This makes possible that the KCs from the γ lobe connect to all the susceptible memories of the flies for the ~ 8 available motivations illustrated in the The incentive wheel.

From a functional point of view, the mushroom bodies seem to be involved in the motivation and behaviour of the animal, especially when it comes to behaviours essential for survival. In the mammalian brain, this function is subserved by the limbic system, which is composed of a set of complicated structures, such as the thalamus, hypothalamus, hippocampus and amygdala (***Dalgleish, 2004***; ***Roxo et al., 2011***). According to ***Papez*** (***1937***), sensory (and mostly olfactory) input comes in the limbic system through the thalamus, which connects to both the cingulate cortex (through the sensory cortex) and the hypothalamus (***Roxo et al., 2011***; ***Dalgleish, 2004***). Responses in the cingulate cortex are guiding the emotions, while the ones in the hypothalamus are guiding the behaviour (bodily responses). Finally, the hypothalamus connects with the cingulate cortex through the anterior thalamus (forward) and the hippocampus (backward stream). ***MacLean*** (***1949***) augmented this model, by adding the amygdala and PFC structures that encode primitive emotions (e.g. anger and fear) and connect to the hypothalamus (***Roxo et al., 2011***; ***Dalgleish, 2004***). We suggest that some of the functions we have identified in the MB incentive circuit could be mapped to limbic system structures (see ***Figure 13***).

More specifically, the α'/β'-KCs could have similar role to the neurons in the thalamus, α/β-KCs represent a higher abstraction of the input stimuli and have a similar role to the ones in the sensory cortex, while the γ-KCs represent relatively unprocessed stimuli. This would make the susceptible MBONs a parallel to neurons in the amygdala, creating responses related to primitive motivations and connecting to (inhibiting) the restrained MBONs, which we would compare to the hypothalamus as providing the main control of behaviour. As we suggest that the same MBONs could fulfil a role as LTM or restrained in different circuits (see The incentive wheel), the LTM would also correspond to hypothalamus, with input from the α'/β'-KCs, and thus the RRM, RFM, LTM and MAM microcircuits are assumed to correspond to hypothalamus functions. Following this analogy we predict that the function of the cingulate cortex then is represented by the α/β MBONs, encoding the 'emotions' of the animal towards reinforced stimuli, potentially controlling more sophisticated decision making. This mapping would suggest the connections amongst the restrained/LTM (α'/β') MBONs and the 'emotional' (α/β) MBONs are similar to the hippocampus and anterior thalamus pathways.

While it might seem startling to suggest that a compact circuit of single identified neurons in the insect MB mimics in miniature these far larger and more complex structures in the mammalian brain, the justification comes from the similarity in the behavioural demands common to all animals: surviving and adapting in a changing world.

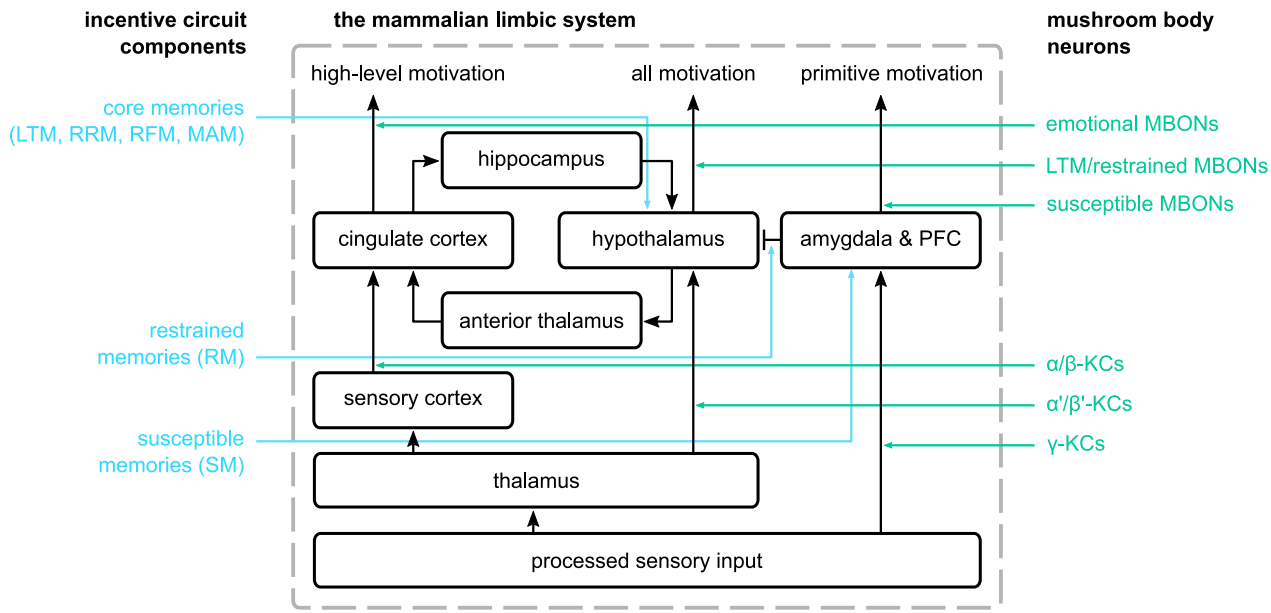


Figure 13. The mammalian limbic system as described by ***Papez*** (***1937***) and ***MacLean*** (***1949***) and the suggested parallels in the proposed incentive circuit. On the left, we show the mushroom body microcircuits that correspond to the different structures in the mammalian limbic system. In the centre, we have the connections among the different structures of the limbic system. On the right, we show the groups of mushroom body neurons we suggest that have similar function to the ones in the limbic system.

Methods

Implementation of the incentive circuit

We represent the connections between neurons by using synaptic-weight matrices and non-linearly transform the information passing from one neuron to another by using an activation function.

Parameters of the model

We assume that the odour identity passes through the projection neurons (PNs) into the mushroom body and its kenyon cells (KCs). It is not in the scope of this work to create a realistic encoding of the odour in the PNs, so we assume that the odour signal is represented by $n_{\mathrm{p}} = 2$ PNs, and that we have $n_{\mathrm{k}} = 10$ KCs that create a sparse representation of the odours (CS). We chose these numbers as 2 PNs is the minimum number of PNs that we can use in order to represent 2 different odours and 10 is a considerably higher number of neurons for the KCs so that 1 PN can connect to 1-6 KCs. Therefore, the vector $\mathbf{p}_{\mathrm{A}} = [1, 0]$ represents the activity of the PNs when odour A is detected, and $\mathbf{p}_{\mathrm{B}} = [0, 1]$ when odour B is detected. The responses of the KCs are calculated by

$$\mathbf{k}(t) = \mathrm{WTA}_{0.6}[\mathbf{p}^T(t) \cdot \mathbf{W}_{\mathrm{p2k}} + \eta] \qquad \text{where } \eta \sim \mathcal{N}(0; 0.001) \tag{6}$$

η is some Gaussian noise, $W_{\mathrm{p2k}} \in \mathbb{R}_{+}^{2\times 10}$ is the weights matrix that allows the transformation of the 2-dimensional odour signal into the 10-dimensional KC responses, and t is the current time-step. The $\mathrm{WTA}_{0.6}[x]$ is an activation function that emulates the KC→KC and APL⊣KC connections, and keeps 60% of KCs active, based on the strength of their activity. This enforces sparsity in the responses of the KCs and makes sure that there will be a fixed number of KCs always active. Note that the number of neurons we are using for PNs and KCs is not very important and we could use any combination of PN and KC populations. However, the bigger the KC population the smaller percentage of them should be active. Using the equation below, the above weights are calculated once at the beginning and they do not change again during the experiments

$$\mathbf{W}_{\mathrm{p2k}} = 2\frac{n_{\mathrm{p}}}{n_{\mathrm{k}}}\mathbf{I}(n_{\mathrm{p}}) \otimes \mathbf{1}(\frac{n_{\mathrm{k}}}{n_{\mathrm{p}}}) = \frac{4}{10}\mathbf{I}(2) \otimes \mathbf{1}(5) = \begin{bmatrix} 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 \end{bmatrix} \tag{7}$$

where $\mathbf{I}(n) \in \{0,1\}^{n\times n}$ is the identity matrix of size n, $\mathbf{1}(n) \in \{1\}^n$ is a vector of n ones and $\otimes$ is the Kronecker product. Note that in order for the above equation to work properly, the fraction $n_\mathsf{k}/n_\mathsf{p}$ must be an integer.

We transform the shock delivery (US), $u(t) \in \{0,1\}$, into an input to the DANs by using the weights matrix $\mathbf{W}_\mathsf{u2d} \in \mathbb{R}_+^{1\times n_\mathsf{d}}$. We represent the activity of the DANs, $\mathbf{d}(t) \in \mathbb{R}^6$, with a 6-dimensional vector, where each dimension represents a different neuron in our model. Specifically,

$$\mathbf{d}(t) = \begin{bmatrix} d_\mathsf{at}(t) & d_\mathsf{av}(t) & c_\mathsf{at}(t) & c_\mathsf{av}(t) & f_\mathsf{at}(t) & f_\mathsf{av}(t) \end{bmatrix} \tag{8}$$

and the shock contribution to the responses of these neurons is given from the weights matrix below

$$\mathbf{W}_\mathsf{u2d} = \begin{bmatrix} 0 & 2 & 0 & 2 & 0 & 0 \end{bmatrix}. \tag{9}$$

In line with the DANs vector representation, we have a similar vector for MBONs, $\mathbf{m}(t) \in \mathbb{R}^6$, where each dimension represents the response of a specific neuron in time t as is shown in the equation below

$$\mathbf{m}(t) = \begin{bmatrix} s_\mathsf{at}(t) & s_\mathsf{av}(t) & r_\mathsf{at}(t) & r_\mathsf{av}(t) & m_\mathsf{at}(t) & m_\mathsf{av}(t) \end{bmatrix}. \tag{10}$$

The weights matrix that encodes the contribution of KCs to the MBON responses, $\mathbf{W}_\mathsf{k2m}(t) \in \mathbb{R}^{10\times 6}$, is initialised as

$$\mathbf{W}_\mathsf{k2m}(t=0) = \mathbf{1}(n_\mathsf{k}, n_\mathsf{m}) = \mathbf{1}(10,6), \tag{11}$$

which effectively is a 10×6 matrix of ones. As these are plastic weights, their value depends on the time-step, and therefore we provide it as a parameter.

There are also connections from MBONs to DANs, $\mathbf{W}_\mathsf{m2d} \in \mathbb{R}^{6\times 6}$, and to other MBONs, $\mathbf{W}_\mathsf{m2m} \in \mathbb{R}^{6\times 6}$, which are given by

$$\mathbf{W}_\mathsf{m2d} = \begin{bmatrix} 0 & -0.3 & 0 & 0 & 0 & 0 \\ -0.3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.05 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.05 & 0 & 0.5 \end{bmatrix} \tag{12}$$

and

$$\mathbf{W}_\mathsf{m2m} = \begin{bmatrix} 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{13}$$

The above matrices summarise the excitatory (positive) and inhibitory (negative) connections between MBONs and DANs or other MBONs. The magnitude of the weights was hand-tuned in order to get the desired result. We also add some bias to the responses of DANs, $\mathbf{b}_\mathsf{d}$, and MBONs, $\mathbf{b}_\mathsf{m}$, which is fixed as

$$\mathbf{b}_\mathsf{d} = \begin{bmatrix} -0.5 & -0.5 & -0.15 & -0.15 & -0.15 & -0.15 \end{bmatrix} \tag{14}$$

$$\mathbf{b}_\mathsf{m} = \begin{bmatrix} -2 & -2 & -0.5 & -0.5 & -0.5 & -0.5 \end{bmatrix}. \tag{15}$$

This bias can be interpreted as the resting value of the neurons, or some external input from other neurons that are not included in our model.

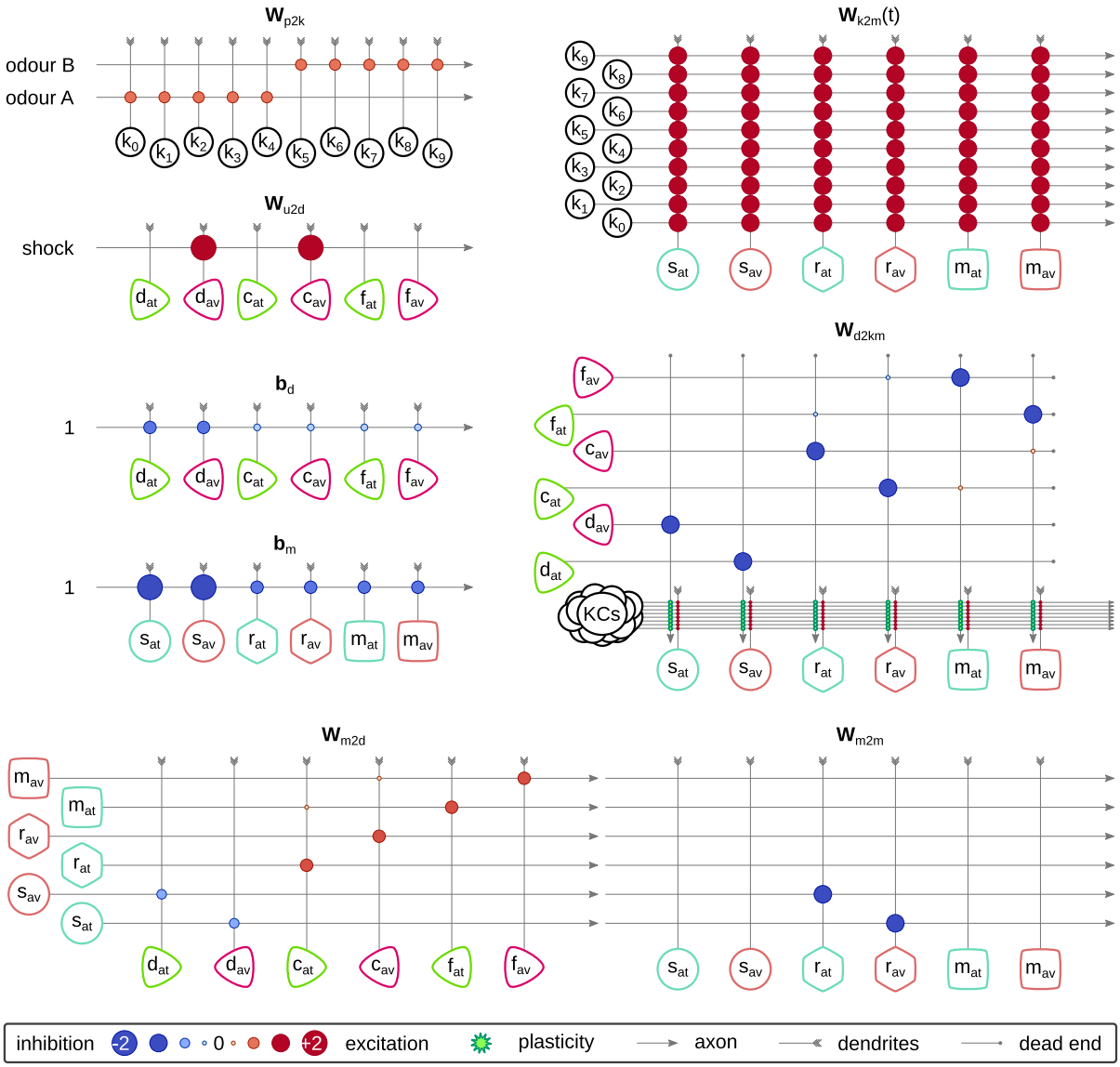


Figure 14. The synaptic weights and connections among the neurons of the incentive circuit (IC). Each panel corresponds to a different synaptic-weights-matrix of the circuit. The size of the circles when the pre-synaptic axon crosses the post-synaptic dendrite shows how strong a connection is, and the colour shows the sign (blue for inhibition, red for excitation). Light-green stars show where the synaptic plasticity takes place and how the DANs modulate the synaptic weights between KCs and MBONs.

Finally, we define the DAN-function matrix, $\mathbf{W}_{\text{d2km}} \in \mathbb{R}^{n_{\text{d}} \times n_{\text{m}}}$, which transforms the responses of the DANs into the dopamine factor that modulates the $\mathbf{W}_{\text{k2m}}$ synaptic weights, and it is given below

$$\mathbf{W}_{\text{d2km}} = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0.05 & 0 \\ 0 & 0 & -1 & 0 & 0 & 005 \\ 0 & 0 & -0.05 & 0 & 0 & -1 \\ 0 & 0 & 0 & -0.05 & -1 & 0 \end{bmatrix}. \tag{16}$$

All the parameters described above are illustrated in ***Figure 14***. The last thing left to describe is the activation function, which is used in order to generate the DAN and MBON responses. This is

$$\sigma(x) = \begin{cases} 2 & \text{if} \quad x \geq 2, \\ x & \text{if} \quad 0 < x < 2, \\ 0 & \text{if} \quad x \leq 0, \end{cases} \tag{17}$$

which is the *rectified linear unit* (ReLU) function, bounded in $\sigma(x) \in [0, 2]$. The reason why we bound the activity is in order to avoid having extremely high values that explode during the charging of the LTM.

Forward propagation

For each time-step, t, we read the environment and propagate the information through the model in order to update the responses of the neurons and the synaptic weights. This process is called *forward propagation* and we repeat it as long as the experiment runs.

First, we read the CS, $\mathbf{p}(t)$, and US, $u(t)$, from the environment and we calculate the KC responses by using ***Equation 6***. In order to calculate the DANs and MBONs update we define the differential equations as follows

$$\tau \frac{d\mathbf{d}}{dt} = u(t) \cdot \mathbf{W}_{\text{u2d}} + \mathbf{m}^T(t) \cdot \mathbf{W}_{\text{m2d}} + \mathbf{b}_{\text{d}} \tag{18}$$

$$\tau \frac{d\mathbf{m}}{dt} = \mathbf{k}^T(t) \cdot \mathbf{W}_{\text{k2m}}(t) + \mathbf{m}^T(t) \cdot \mathbf{W}_{\text{m2m}} + \mathbf{b}_{\text{m}} \tag{19}$$

where $\tau = 3$ is a time-constant that is defined by the number of time-steps associated in each trial and T denotes the transpose operation of the matrix or vector. Using the above differential equations, we calculate the updated responses as

$$\mathbf{d}(t+1) = \sigma[\mathbf{d}(t) + \frac{d\mathbf{d}}{dt}] \tag{20}$$

$$\mathbf{m}(t+1) = \sigma[\mathbf{m}(t) + \frac{d\mathbf{m}}{dt}]. \tag{21}$$

Finally, we calculate the dopaminergic factor, $\delta(t) \in \mathbb{R}^6$, and update the KC→MBON synaptic weights as

$$\delta(t+1) = \mathbf{d}_+^T(t+1) \cdot \mathbf{W}_{\text{d2km}} \tag{22}$$

$$\tau \frac{d\mathbf{W}_{\text{k2m}}}{dt} = \delta(t+1) * [\mathbf{k}^T(t) + \mathbf{W}_{\text{k2m}}(t) - w_{\text{rest}}] \tag{23}$$

$$\mathbf{W}_{\text{k2m}}(t+1) = \max[\mathbf{W}_{\text{k2m}}(t) + \frac{d\mathbf{W}_{\text{k2m}}}{dt}, 0] \tag{24}$$

where $\mathbf{d}_+(t+1)$ is the positive part of the the DANs' responses and '$*$' denotes the element-wise multiplication; $w_{\text{rest}} = 1$ is the resting value of the weights. Note that element-wise multiplication means that each element of the δ vector will be multiplied with each column of the $\mathbf{W}_{\text{k2m}}$ matrix. Also, the element-wise addition of the transposed vector, $\mathbf{k}^T(t)$, to the $\mathbf{W}_{\text{k2m}}(t)$ matrix, means that we add each element of $\mathbf{k}(t)$ to the corresponding row of $\mathbf{W}_{\text{k2m}}(t)$. We repeat the above procedure as many times as it is required in order to complete the running experimental paradigm routine.

Decomposing the dopaminergic factor

In ***Equation 22***, the dopaminergic factor, $\delta(t)$ is derived from the matrix ***Equation 16*** which captures in abstracted and time-independent form the effects of dopamine release. To model these more explicitly, as described in The dopaminergic learning rule, the dopaminergic factor can be decomposed as $\delta(t) = \mathbf{D}_2(t) - \mathbf{D}_1(t)$ where each component has a time dependent form given by the differential equations:

$$\frac{d\mathbf{D}_1}{dt} = \frac{1}{\tau_{\text{long}}}\mathbf{d}_+^T(t)\cdot\mathbf{W}_{d2km}^+ - \frac{1}{\tau_{\text{short}}}\mathbf{d}_+^T(t)\cdot\mathbf{W}_{d2km}^- - \left(1 - \frac{1}{\tau_{\text{short}}} + \frac{1}{\tau_{\text{long}}}\right)\mathbf{D}_1(t-1) \tag{25}$$

$$\frac{d\mathbf{D}_2}{dt} = \frac{1}{\tau_{\text{short}}}\mathbf{d}_+^T(t)\cdot\mathbf{W}_{d2km}^+ - \frac{1}{\tau_{\text{long}}}\mathbf{d}_+^T(t)\cdot\mathbf{W}_{d2km}^- - \left(1 - \frac{1}{\tau_{\text{short}}} + \frac{1}{\tau_{\text{long}}}\right)\mathbf{D}_2(t-1) \tag{26}$$

where W_{d2km}^+ and W_{d2km}^- represent the positive-only (potentiation/saturation) and negative-only (depression/recovery) dopaminergic effects; τ_{short} and τ_{long} are time-constants that define the short (main) and long (secondary) duration of the dopamine effect. The longer the time-constant, the slower the diffusion but also the lower the the peak of the effect. Note the two time-constants must satisfy the constraint $\frac{1}{\tau_{\text{short}}} - \frac{1}{\tau_{\text{long}}} \leq 1$ in order for the above differential equations to work properly.

In ***Figure 2***, where we are interested in more detailed dynamics of the learning rule, and the sampling frequency is high, i.e. $100\,\text{Hz}$, we use $\tau_{\text{short}} = 50$ and $\tau_{\text{long}} = 100$, which essentially means that $\mathbf{D}_1$ will need $0.5\,\text{sec}$ for a full update and $\mathbf{D}_2$ $1\,\text{sec}$. Also note that for the specific (susceptible) type of MBON examined there, the DAN causes depression of the synapse, so there is no positive dopaminergic effect, i.e. $\mathbf{W}_{k2dm}^+ = 0$. By setting $W_{k2dm}^+ = 0$ in equations ***Equation 25*** and ***Equation 26***, we have the fast update with the high peak for $\mathbf{D}_1$ and a slower update with lower peak for $\mathbf{D}_2$, as described in The dopaminergic learning rule.

For the experiments in ***Figure 9*** and ***Figure 11***, we use $\tau_{\text{short}} = 1$ and $\tau_{\text{long}} = 1$, which removes the dynamics induced by the relation between $\mathbf{D}_1$ and $\mathbf{D}_2$, and ***Equation 22*** emerges from

$$\delta(t) = \mathbf{D}_2(t) - \mathbf{D}_1(t) \tag{27}$$

$$= \mathbf{d}_+^T(t)\cdot\mathbf{W}_{d2km}^+ + \mathbf{d}_+^T(t)\cdot\mathbf{W}_{d2km}^-, \qquad \text{for } \tau_{\text{short}} = 1 \text{ and } \tau_{\text{long}} = 1 \tag{28}$$

$$= \mathbf{d}_+^T(t)\cdot\mathbf{W}_{d2km} \tag{29}$$

This essentially means that each update represents a time-step that is longer than the effective period of backward conditioning for the Modelling the neural responses section (where sampling frequency is low, i.e. $\leq 0.5\,\text{Hz}$), while these dynamics seem irrelevant for the Modelling the behaviour section and therefore we use the same time-constants that result in the simplified ***Equation 22***.

Modelling the neural responses

To emulate the acquisition and forgetting paradigms used for flies, we run the simulated circuit in an experiment that consists of $T = 73$ time-steps. Each time-step actually comprises four repeats of the forward propagation update described above, to smooth out any bias due to the order of computations (value vs weights update). After the initialisation time-step at $t = 0$ there are 24 *trials* where each trial consists of 3 *in-trial time-steps*.

Within each trial, the first time-step has no odour, and in the second and third time steps, odour is presented: odour A on even trials and odour B on odd trials. A trial can have no shock (***Figure 15***A), unpaired shock presented in the first time step (***Figure 15***B), or paired shock presented in the third time step (***Figure 15***C).

The first 2 trials compose the 'pre-training phase', where we expose the model to the two odours alternately (i.e. odour A in trial 1 and odour B in trial 2) without shock delivery. Then we have the acquisition phase, where we deliver shock paired with odour B for 10 trials (5 trials per odour; ***Figure 15***D). Before we proceed to the forgetting phases, we leave 2 empty trials (1 per odour), which we call the *resting trials*. The forgetting phases last for another 10 trials (5 trials per odour;

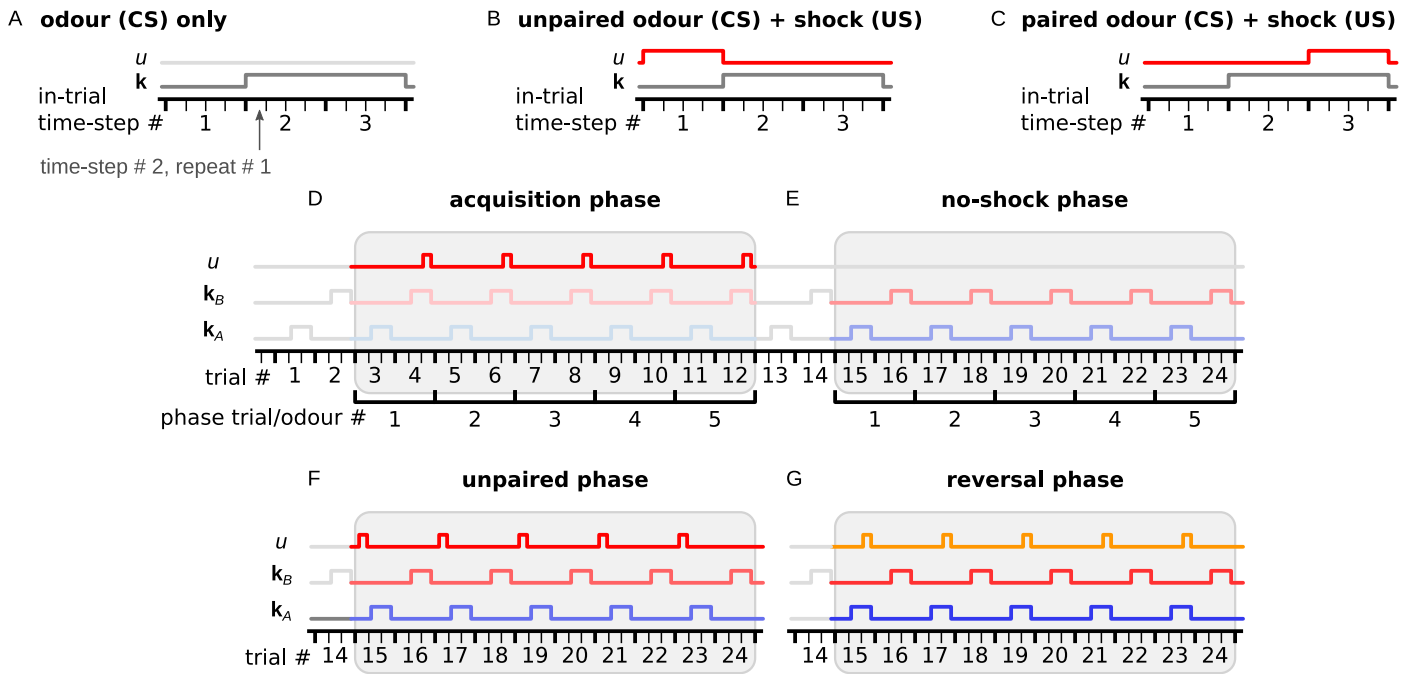


Figure 15. Description of the simulation process from our experiments. A single trial is composed of 3 in-trial time-steps and each time-step by 4 repeats. Odour is provided only during the 2nd and 3rd in-trial time-steps, while shock delivery is optional. (A) In a no-shock trial only odour (CS) is delivered. (B) During an unpaired trial, shock is delivered during in-trail time-step 1. (C) During a paired trial, shock is delivered along with odour delivery and during in-trial time-step 3. (D) The acquisition phase has 5 odour A only trials and 5 paired odour B trials alternating. (E) The no-shock phase has 5 odour A only trials and 5 odour B only trials alternating. (F) The unpaired phase has 5 odour A unpaired trials and 5 odour B only trials alternating. (G) The reversal phase has 5 odour A paired trials and 5 odour B only trials alternating. The colours used in this figure match the ones in ***Figure 9***.

Figure 15E, F and G). During the no-shock phase, no shock is delivered while we continue alternating the odours (see ***Figure 15***E); during the unpaired phase, shock is delivered unpaired from odour A (see ***Figure 15***F); while at the reversal phase shock is paired with odour A (***Figure 15***G).

Modelling the behaviour

In order to generate the results for the Modelling the behaviour and ***Figure 11*** we have created another routine that embeds the simulation of the flies' motion and the environment. We represent the position of the fly, $a(t) \in \mathbb{C}$, and the sources of the odours in the arena, $\mu_A \in \mathbb{C}$ and $\mu_B \in \mathbb{C}$ for odours A and B respectively, in the 2D space as complex numbers in the form $x + iy$. Therefore, the flies are initialised in $a(t = 0) = 0$ and the sources of the odours are placed in $\mu_A = -0.6$ and $\mu_B = 0.6$. The standard deviation of the odour distributions are $\sigma_A = 0.2$ and $\sigma_B = 0.3$ for odours A and B respectively.

We assume that the only odours available are odours A and B, therefore in every time-step, the probability that a fly perceives odour A is calculated by the Bayes rule and it is given by:

$$P_A(t) = \frac{P[\text{odour=A}|a(t)]P[\text{odour=A}]}{P[\text{odour=A}|a(t)]P[\text{odour=A}] + P[\text{odour=B}|a(t)]P[\text{odour=B}]} + \eta \tag{30}$$

where $p[\text{odour=A}|a(t)] = \mathcal{N}(a(t)|\mu_A, \sigma_A)$ is the density value from the Gaussian distribution of odour A for the current position of the fly, $P[\text{odour=B}|a(t)] = \mathcal{N}(a(t)|\mu_B, \sigma_B)$ is the density value from the Gaussian distribution of odour B for the current position of the fly, $P[\text{odour=A}] = 0.5$ is the prior probability of odour A and $P[\text{odour=B}] = 0.5$ is the prior probability of odour B, which we assume that it is drawn from a uniform distribution; $\eta \sim \mathcal{N}(\mu = 0, \sigma = 0.1)$ is the sensors noise drawn from a Gaussian distribution. Finally, as we add a random value to a probability, we also make sure that

$0 \leq P_A(t) \leq 1$. The probability that the fly perceives odour B is then $P_B(t) = 1 - P_A(t)$ as probabilities must sum to one, i.e. $P_A(t) + P_B(t) = 1$.

Therefore the PN activity is calculated using the equation below

$$\mathbf{p}(t) = P_A(t) \cdot \mathbf{p}_A + P_b(t) \cdot \mathbf{p}_B \tag{31}$$

where $\mathbf{p}_A$ and $\mathbf{p}_B$ are the identities of odours A and B respectively in the PNs as described in Parameters of the model. The US is represented by a 2-dimensional vector where the 1st dimension denotes rewarding signal and the 2nd dimension denotes punishment: $\mathbf{u}_{\text{sugar}} = [1, 0]$ and $\mathbf{u}_{\text{shock}} = [0, 1]$; and the contribution of this vector to the responses of the DANs is is given by

$$\mathbf{W}_{\text{u2d}} = \begin{bmatrix} 2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 & 0 \end{bmatrix} \tag{32}$$

which replaces the weights matrix in ***Equation 9*** in our model in order to support 2 identities of reinforcements. Then we run a Forward propagation using the above inputs.

From the updated responses of the MBONs, we calculate the *attraction force*, $\mathbf{v}(t)$, for the mixture of odours which modulates the velocity of the fly. This force is calculated by taking the difference between the responses of the MBONs that drive the behaviour:

$$v_{\text{at}}(\text{CS}|t) = \frac{1}{3}[s_{\text{at}}(t) + r_{\text{at}}(t) + m_{\text{at}}(t)] \cdot \frac{\mu_{\text{CS}} - a(t)}{||\mu_{\text{CS}} - a(t)||} \qquad \text{where CS} \in \{\text{odour A, odour B}\} \tag{33}$$

$$v_{\text{av}}(\text{CS}|t) = \frac{1}{3}[s_{\text{av}}(t) + r_{\text{av}}(t) + m_{\text{av}}(t)] \cdot \frac{\mu_{\text{CS}} - a(t)}{||\mu_{\text{CS}} - a(t)||} \qquad \text{where CS} \in \{\text{odour A, odour B}\} \tag{34}$$

$$\mathbf{v}(t) = \sum_{\text{CS}}^{\{A,B\}} P_{\text{CS}}(t) \cdot v_{\text{at}}(\text{CS}|t) - \sum_{\text{CS}}^{\{A,B\}} P_{\text{CS}}(t) \cdot v_{\text{av}}(\text{CS}|t) \quad . \tag{35}$$

where μ_{CS} is the position of the odour source. The velocity is updated as follows

$$v(t) = v(t-1) + \mathbf{v}(t) + \epsilon_x + i\epsilon_y \qquad \text{where } \epsilon_x, \epsilon_y \sim \mathcal{N}(\mu = 0, \sigma = 0.1) \tag{36}$$

$$\hat{v}(t) = 0.01 \cdot \frac{v(t)}{||v(t)||}. \tag{37}$$

We normalise the velocity so that we keep the direction but replace the step size with $0.01\,\text{m/sec}$. The noise added to the velocity is introduced in order to enable the flies to move in 2 dimensions and not just between the two odour sources. Also, when the attraction force is $\mathbf{v}(t) = 0$, then the noise along with the previous velocity is the one that drives the flies.

We repeat the above process for $T = 500$ time-steps with $5\,\text{Hz}$ (5 time-steps per second), and we provide shock or sugar (when appropriate) between time-steps 100 and 250, otherwise we use a zero-vector as US input to the DANs.

Calculating the preference Index

In ***Figure 11***B, we report the preference index of the flies at the end of each of the 3 phases of the experiment (pre-training, training and post-training), which we measure in times $t = 0\,\text{sec}$, $t = 30\,\text{sec}$ and $t = 80\,\text{sec}$ respectively, using the equation below

$$\text{PI}(t) = \frac{||\mu_{\text{Odour A}} - a(t)|| - ||\mu_{\text{Odour B}} - a(t)||}{||\mu_{\text{Odour A}} - a(t)|| + ||\mu_{\text{Odour B}} - a(t)||} \tag{38}$$

where $a(t)$ is the position of the fly in time t, $\mu_{\text{Odour A}}$ and $\mu_{\text{odour B}}$ are the positions of the odour A and B sources respectively, and $||\cdot||$ denotes the magnitude or length of the complex number inside.

Data collection

In order to verify the plausibility of the incentive circuit, we recorded the neural activity in genetically targeted neurons during aversive olfactory conditioning which is described in more detail in ***McCurdy et al.*** (***2021***). We simultaneously expressed green $GCaMP_6f$ Ca^{2+} indicator and red Ca^{2+}-insensitive tdTomato in neurons of interest in order to visualise the Ca^{2+} changes which reflect the neural activity. We collected data from 357 5- to 8-days-old female flies (2-14 per neuron; 8 flies on average) and for 43 neurons (see ***Figure 8–Figure Supplement 2*** and ***Figure 8–Figure Supplement 1***).

Each fly was head-fixed for simultaneous delivery of odours and electric shock while recording the neural activity. Their proboscis was also glued, while their body and legs were free to move (see ***Figure 8***A). The flies were allowed to recover from the gluing process for $15\,\mathrm{min}$ before placing them under the microscope. We used green ($555\,\mathrm{nm}$) and blue ($470\,\mathrm{nm}$) lights to record GCaMP and Tomato signals. We also used $0.1\,\%$ 3-octanol (OCT) and $0.1\,\%$ 4-methylcyclohexanol (MCH) for odour A and B respectively, and the flow rate was kept constant at 500mL/min for each odour. The flies were allowed to acclimate to the airflow for at least one minute before starting of the experiment.

During the experiments, we alternate trials were $5\,\mathrm{sec}$ of each odour is presented $5\,\mathrm{sec}$ after the (green or red) light is on. We start with 2 pre-training trials (1 per odours) followed by 5 acquisition trials per odour. During acquisition, flies receive alternating $5\,\mathrm{sec}$ pulses of OCT (odour A) and MCH (odour B) paired with electric shock (B), repeated for 5 trials. during reversal, OCT is presented with shock and MCH without, repeated for 2 trials. On trials where electric shock was delivered, it was presented $4\,\mathrm{sec}$ after odour onset for $100\,\mathrm{ms}$ at $120\,\mathrm{V}$.

Calculating off- and on-shock values

From the data collection process described above we get trials of 100 time-steps and at $5\,\mathrm{Hz}$ ($20\,\mathrm{sec}$ each). Odour is delivered between time-steps 25 and 50 (between $5\,\mathrm{sec}$ and $10\,\mathrm{sec}$), and shock is delivered during time-step 45 (at $9\,\mathrm{sec}$). In this work, we report 2 values for each trial: the *off-shock* and *on-shock* values, which represent the average response to the odour before and during the shock delivery respectively, even if shock has not been delivered.

For the off-shock value, from each data-stream of activity from the target neuron, we collect the values from time-steps between 28 ($5.6\,\mathrm{sec}$) and 42 ($8.4\,\mathrm{sec}$). This gives us a matrix of $n_{\mathrm{fly}} \times 15$ values, whose average and standard deviation is the reported off-shock value. Similarly, for the on-shock values, we collect the values in time-steps between 44 ($8.6\,\mathrm{sec}$) and 48 ($9.6\,\mathrm{sec}$), which gives a matrix of $n_{\mathrm{fly}} \times 5$ values, whose average and standard deviation is the on-shock value. We define 'on-shock' as the time window from $8.6\,\mathrm{sec}$ to $9.6\,\mathrm{sec}$, where shock onset occurs at $t = 9\,\mathrm{sec}$.

Acknowledgements

We are grateful to Bertram Gerber for his useful comments on earlier drafts of the manuscript, and to the Insect Robotics group for helpful critique of the figures.

References

Adel M, Griffith LC. The Role of Dopamine in Associative Learning in Drosophila: An Updated Unified Model. Neuroscience Bulletin. 2021; p. 1–22. doi: 10.1007/s12264-021-00665-0.

Ardin P, Peng F, Mangan M, Lagogiannis K, Webb B. Using an Insect Mushroom Body Circuit to Encode Route Memory in Complex Natural Environments. PLOS Computational Biology. 2016; 12(2):e1004683. doi: 10.1371/journal.pcbi.1004683.

Arena P, Patané L, Stornanti V, Termini PS, Zäpf B, Strauss R. Modeling the insect mushroom bodies: Application to a delayed match-to-sample task. Neural Networks. 2013; 41:202–211. doi: 10.1016/j.neunet.2012.11.013.

Aso Y, Hattori D, Yu Y, Johnston RM, Iyer NA, Ngo TT, Dionne H, Abbott L, Axel R, Tanimoto H, Rubin GM. The neuronal architecture of the mushroom body provides a logic for associative learning. eLife. 2014; 3:e04577. doi: 10.7554/elife.04577.

Aso Y, Ray RP, Long X, Bushey D, Cichewicz K, Ngo TT, Sharp B, Christoforou C, Hu A, Lemire AL, Tillberg P, Hirsh J, Litwin-Kumar A, Rubin GM. Nitric oxide acts as a cotransmitter in a subset of dopaminergic neurons to diversify memory dynamics. eLife. 2019; 8:e49257. doi: 10.7554/elife.49257.

Aso Y, Sitaraman D, Ichinose T, Kaun KR, Vogt K, Belliart-Guérin G, Plaçais PY, Robie AA, Yamagata N, Schnaitmann C, Rowell WJ, Johnston RM, Ngo TTB, Chen N, Korff W, Nitabach MN, Heberlein U, Preat T, Branson KM, Tanimoto H, et al. Mushroom body output neurons encode valence and guide memory-based action selection in Drosophila. eLife. 2014; 3:e04580. doi: 10.7554/elife.04580.

Baddeley B, Graham P, Husbands P, Philippides A. A Model of Ant Route Navigation Driven by Scene Familiarity. PLoS Computational Biology. 2012; 8(1):e1002336. doi: 10.1371/journal.pcbi.1002336.

Balkenius A, Kelber A, Balkenius C. From Animals to Animats 9, 9th International Conference on Simulation of Adaptive Behavior, SAB 2006, Rome, Italy, September 25-29, 2006. Proceedings. Lecture Notes in Computer Science. 2006; p. 422–433. doi: 10.1007/11840541_35.

Bazhenov M, Huerta R, Smith BH. A Computational Framework for Understanding Decision Making through Integration of Basic Learning Rules. The Journal of Neuroscience. 2013; 33(13):5686–5697. doi: 10.1523/jneurosci.4145-12.2013.

Bennett JEM, Philippides A, Nowotny T. Learning with reinforcement prediction errors in a model of the Drosophila mushroom body. Nature Communications. 2021; 12(1):2569. doi: 10.1038/s41467-021-22592-4.

Berry JA, Phan A, Davis RL. Dopamine Neurons Mediate Learning and Forgetting through Bidirectional Modulation of a Memory Trace. Cell Reports. 2018; 25(3):651–662.e5. doi: 10.1016/j.celrep.2018.09.051.

Bilz F, Geurten BRH, Hancock CE, Widmann A, Fiala A. Visualization of a Distributed Synaptic Memory Code in the Drosophila Brain. Neuron. 2020; 106(6):963–976.e4. doi: 10.1016/j.neuron.2020.03.010.

Brembs B, Heisenberg M. Conditioning with compound stimuli in Drosophila melanogaster in the flight simulator. The Journal of experimental biology. 2001; 204(Pt 16):2849–59.

Busto GU, Cervantes-Sandoval I, Davis RL. Olfactory Learning in Drosophila. Physiology. 2010; 25(6):338–346. doi: 10.1152/physiol.00026.2010.

Campbell RAA, Honegger KS, Qin H, Li W, Demir E, Turner GC. Imaging a Population Code for Odor Identity in the Drosophila Mushroom Body. The Journal of Neuroscience. 2013; 33(25):10568–10581. doi: 10.1523/jneurosci.0682-12.2013.

Cervantes-Sandoval I, Phan A, Chakraborty M, Davis RL. Reciprocal synapses between mushroom body and dopamine neurons form a positive feedback loop required for learning. eLife. 2017; 6:e23789. doi: 10.7554/elife.23789.

Claridge-Chang A, Roorda RD, Vrontou E, Sjulson L, Li H, Hirsh J, Miesenböck G. Writing Memories with Light-Addressable Reinforcement Circuitry. Cell. 2009; 139(2):405–415. doi: 10.1016/j.cell.2009.08.034.

Cleland TA. Inhibitory glutamate receptor channels. Molecular Neurobiology. 1996; 13(2):97–136. doi: 10.1007/bf02740637.

Cohn R, Morantte I, Ruta V. Coordinated and Compartmentalized Neuromodulation Shapes Sensory Processing in Drosophila. Cell. 2015; 163(7):1742–1755. doi: 10.1016/j.cell.2015.11.019.

Colomb J, Kaiser L, Chabaud MA, Preat T. Parametric and genetic analysis of Drosophila appetitive long-term memory and sugar motivation. Genes, Brain and Behavior. 2009; 8(4):407–415. doi: 10.1111/j.1601-183x.2009.00482.x.

Dalgleish T. The emotional brain. Nature Reviews Neuroscience. 2004; 5(7):583–589. doi: 10.1038/nrn1432.

Davis RL. Mushroom bodies and drosophila learning. Neuron. 1993; 11(1):1–14. doi: 10.1016/0896-6273(93)90266-t.

Delahunt CB, Riffell JA, Kutz JN. Biological Mechanisms for Learning: A Computational Model of Olfactory Learning in the Manduca sexta Moth, With Applications to Neural Nets. Frontiers in Computational Neuroscience. 2018; 12:102. doi: 10.3389/fncom.2018.00102.

Dubnau J, Grady L, Kitamoto T, Tully T. Disruption of neurotransmission in Drosophila mushroom body blocks retrieval but not acquisition of memory. Nature. 2001; 411(6836):476–480. doi: 10.1038/35078077.

Dylla KV, Raiser G, Galizia CG, Szyszka P. Trace Conditioning in Drosophila Induces Associative Plasticity in Mushroom Body Kenyon Cells and Dopaminergic Neurons. Frontiers in Neural Circuits. 2017; 11:42. doi: 10.3389/fncir.2017.00042.

Eichler K, Li F, Litwin-Kumar A, Park Y, Andrade I, Schneider-Mizell CM, Saumweber T, Huser A, Eschbach C, Gerber B, Fetter RD, Truman JW, Priebe CE, Abbott LF, Thum AS, Zlatic M, Cardona A. The complete connectome of a learning and memory centre in an insect brain. Nature. 2017 8; 548(7666):175–182. doi: 10.1038/nature23455.

Eschbach C, Fushiki A, Winding M, Schneider-Mizell CM, Shao M, Arruda R, Eichler K, Valdes-Aleman J, Ohyama T, Thum AS, Gerber B, Fetter RD, Truman JW, Litwin-Kumar A, Cardona A, Zlatic M. Recurrent architecture for adaptive regulation of learning in the insect brain. Nature Neuroscience. 2020; 23(4):544–555. doi: 10.1038/s41593-020-0607-9.

Faghihi F, Moustafa AA, Heinrich R, Wörgötter F. A computational model of conditioning inspired by Drosophila olfactory system. Neural Networks. 2017; 87:96–108. doi: 10.1016/j.neunet.2016.11.002.

Felsenberg J, Jacob PF, Walker T, Barnstedt O, Edmondson-Stait AJ, Pleijzier MW, Otto N, Schlegel P, Sharifi N, Perisse E, Smith CS, Lauritzen JS, Costa M, Jefferis GSXE, Bock DD, Waddell S. Integration of Parallel Opposing Memories Underlies Memory Extinction. Cell. 2018; 175(3):709–722.e15. doi: 10.1016/j.cell.2018.08.021.

Finelli LA, Haney S, Bazhenov M, Stopfer M, Sejnowski TJ. Synaptic Learning Rules and Sparse Coding in a Model Sensory System. PLoS Computational Biology. 2008; 4(4):e1000062. doi: 10.1371/journal.pcbi.1000062.

Gerber B, Stocker RF, Tanimura T, Thum AS. Smelling, Tasting, Learning: Drosophila as a Study Case. In: Meyerhof W, Korsching S, editors. *Chemosensory Systems in Mammals, Fishes, and Insects*, vol. 47 Springer; 2009.p. 139–186. doi: 10.1007/400_2008_9.

Gerber B, Hendel T. Outcome expectations drive learned behaviour in larval Drosophila. Proceedings of the Royal Society B: Biological Sciences. 2006; 273(1604):2965–2968. doi: 10.1098/rspb.2006.3673.

Handler A, Graham TGW, Cohn R, Morantte I, Siliciano AF, Zeng J, Li Y, Ruta V. Distinct Dopamine Receptor Pathways Underlie the Temporal Sensitivity of Associative Learning. Cell. 2019; 178(1):60–75.e19. doi: 10.1016/j.cell.2019.05.040.

Hebb DO. The organization of behavior: A neuropsychological theory. Psychology Press; 2005.

Heisenberg M. Mushroom body memoir: from maps to models. Nature Reviews Neuroscience. 2003; 4(4):266–275. doi: 10.1038/nrn1074.

Hige T, Aso Y, Modi M, Rubin G, Turner G. Heterosynaptic Plasticity Underlies Aversive Olfactory Learning in Drosophila. Neuron. 2015; 88(5):985–998. doi: 10.1016/j.neuron.2015.11.003.

Huerta R, Huerta R, Nowotny T, Nowotny T, García-Sánchez M, Abarbanel HDI, Rabinovich MI. Learning Classification in the Olfactory System of Insects. Neural Computation. 2004; 16(8):1601–1640. doi: 10.1162/089976604774201613.

Ichinose T, Aso Y, Yamagata N, Abe A, Rubin GM, Tanimoto H. Reward signal in a recurrent circuit drives appetitive long-term memory formation. eLife. 2015; 4:e10719. doi: 10.7554/elife.10719.

Ito I, Ong RCy, Raman B, Stopfer M. Sparse odor representation and olfactory learning. Nature Neuroscience. 2008; 11(10):1177–1184. doi: 10.1038/nn.2192.

Jacob PF, Waddell S. Spaced Training Forms Complementary Long-Term Memories of Opposite Valence in Drosophila. Neuron. 2020; 106(6):977–991.e4. doi: 10.1016/j.neuron.2020.03.013.

Kallman BR, Kim H, Scott K. Excitation and inhibition onto central courtship neurons biases Drosophila mate choice. eLife. 2015; 4:e11188. doi: 10.7554/elife.11188.

Kamin LJ. Predictability, surprise, attention and conditioning. Punishment and Aversive Behaviour. 1967; p. 279–296.

Krashes MJ, DasGupta S, Vreede A, White B, Armstrong JD, Waddell S. A Neural Circuit Mechanism Integrating Motivational State with Memory Expression in Drosophila. Cell. 2009; 139(2):416–427. doi: 10.1016/j.cell.2009.08.035.

Krashes MJ, Keene AC, Leung B, Armstrong JD, Waddell S. Sequential Use of Mushroom Body Neuron Subsets during Drosophila Odor Memory Processing. Neuron. 2007; 53(1):103–115. doi: 10.1016/j.neuron.2006.11.021.

Li F, Lindsey JW, Marin EC, Otto N, Dreher M, Dempsey G, Stark I, Bates AS, Pleijzier MW, Schlegel P, Nern A, Takemura Sy, Eckstein N, Yang T, Francis A, Braun A, Parekh R, Costa M, Scheffer LK, Aso Y, et al. The connectome of the adult Drosophila mushroom body provides insights into function. eLife. 2020; 9:e62576. doi: 10.7554/elife.62576.

Lin S, Owald D, Chandra V, Talbot C, Huetteroth W, Waddell S. Neural correlates of water reward in thirsty Drosophila. Nature Neuroscience. 2014; 17(11):1536–1542. doi: 10.1038/nn.3827.

Liu C, Plaçais PY, Yamagata N, Pfeiffer BD, Aso Y, Friedrich AB, Siwanowicz I, Rubin GM, Preat T, Tanimoto H. A subset of dopamine neurons signals reward for odour memory in Drosophila. Nature. 2012; 488(7412):512–516. doi: 10.1038/nature11304.

Liu WW, Wilson RI. Glutamate is an inhibitory neurotransmitter in the Drosophila olfactory system. Proceedings of the National Academy of Sciences. 2013; 110(25):10294–10299. doi: 10.1073/pnas.1220560110.

Liu X, Davis RL. The GABAergic anterior paired lateral neuron suppresses and is suppressed by olfactory learning. Nature Neuroscience. 2009; 12(1):53–59. doi: 10.1038/nn.2235.

MacLean PD. Psychosomatic Disease and the "Visceral Brain". Psychosomatic Medicine. 1949; 11(6):338–353. doi: 10.1097/00006842-194911000-00003.

Mao Z, Davis RL. Eight Different Types of Dopaminergic Neurons Innervate the Drosophila Mushroom Body Neuropil: Anatomical and Physiological Heterogeneity. Frontiers in Neural Circuits. 2009; 3:5. doi: 10.3389/neuro.04.005.2009.

May CE, Rosander J, Gottfried J, Dennis E, Dus M. Dietary sugar inhibits satiation by decreasing the central processing of sweet taste. eLife. 2020; 9:e54530. doi: 10.7554/elife.54530.

McCarthy Ev, Wu Y, deCarvalho T, Brandt C, Cao G, Nitabach MN. Synchronized Bilateral Synaptic Inputs to Drosophila melanogaster Neuropeptidergic Rest/Arousal Neurons. The Journal of Neuroscience. 2011; 31(22):8181–8193. doi: 10.1523/jneurosci.2017-10.2011.

McCurdy LY, Sareen P, Davoudian PA, Nitabach MN. Dopaminergic mechanism underlying reward-encoding of punishment omission during reversal learning in Drosophila. Nature Communications. 2021 2; 12(1):1115. doi: 10.1038/s41467-021-21388-w.

McGuire SE, Le PT, Davis RL. The Role of Drosophila Mushroom Body Signaling in Olfactory Memory. Science. 2001; 293(5533):1330–1333. doi: 10.1126/science.1062622.

Michels B, Chen Yc, Saumweber T, Mishra D, Tanimoto H, Schmid B, Engmann O, Gerber B. Cellular site and molecular mode of synapsin action in associative learning. Learning & Memory. 2011; 18(5):332–344. doi: 10.1101/lm.2101411.

Milyaev N, Osumi-Sutherland D, Reeve S, Burton N, Baldock RA, Armstrong JD. The Virtual Fly Brain browser and query interface. Bioinformatics. 2012; 28(3):411–415. doi: 10.1093/bioinformatics/btr677.

Niv Y. Reinforcement learning in the brain. Journal of Mathematical Psychology. 2009; 53(3):139–154. doi: 10.1016/j.jmp.2008.12.005.

Papez JW. A proposed mechanism of emotion. Archives of Neurology & Psychiatry. 1937; 38(4):725–743. doi: 10.1001/archneurpsyc.1937.02260220069003.

Pavlowsky A, Schor J, Plaçais PY, Preat T. A GABAergic Feedback Shapes Dopaminergic Input on the Drosophila Mushroom Body to Promote Appetitive Long-Term Memory. Current Biology. 2018; 28(11):1783–1793.e4. doi: 10.1016/j.cub.2018.04.040.

Peng F, Chittka L. A Simple Computational Model of the Bee Mushroom Body Can Explain Seemingly Complex Forms of Olfactory Learning and Memory. Current Biology. 2017; 27(2):224–230. doi: 10.1016/j.cub.2016.10.054.

Perisse E, Owald D, Barnstedt O, Talbot C, Huetteroth W, Waddell S. Aversive Learning and Appetitive Motivation Toggle Feed-Forward Inhibition in the Drosophila Mushroom Body. Neuron. 2016; 90(5):1086–1099. doi: 10.1016/j.neuron.2016.04.034.

Plutchik R. The Nature of Emotions. American Scientist. 2001; 89(4):344–350.

Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical conditioning II: current research and theory* New York: Appleton-Century-Crofts; 1972.p. 64–99.

Roxo MR, Franceschini PR, Zubaran C, Kleber FD, Sander JW. The Limbic System Conception and Its Historical Evolution. TheScientificWorldJOURNAL. 2011; 11:2427–2440. doi: 10.1100/2011/157150.

Saumweber T, Rohwedder A, Schleyer M, Eichler K, Chen Yc, Aso Y, Cardona A, Eschbach C, Kobler O, Voigt A, Durairaja A, Mancini N, Zlatic M, Truman JW, Thum AS, Gerber B. Functional architecture of reward learning in mushroom body extrinsic neurons of larval Drosophila. Nature Communications. 2018; 9(1):1104. doi: 10.1038/s41467-018-03130-1.

Schleyer M, Fendt M, Schuller S, Gerber B. Associative Learning of Stimuli Paired and Unpaired With Reinforcement: Evaluating Evidence From Maggots, Flies, Bees, and Rats. Frontiers in Psychology. 2018; 9:1494. doi: 10.3389/fpsyg.2018.01494.

Schleyer M, Saumweber T, Nahrendorf W, Fischer B, Alpen Dv, Pauls D, Thum A, Gerber B. A behavior-based circuit model of how outcome expectations organize learned behavior in larval Drosophila. Learning & Memory. 2011; 18(10):639–653. doi: 10.1101/lm.2163411.

Schleyer M, Weiglein A, Thoener J, Strauch M, Hartenstein V, Weigelt MK, Schuller S, Saumweber T, Eichler K, Rohwedder A, Merhof D, Zlatic M, Thum AS, Gerber B. Identification of dopaminergic neurons that can both establish associative memory and acutely terminate its behavioral expression. Journal of Neuroscience. 2020; 40(31):JN–RM–0290–20. doi: 10.1523/jneurosci.0290-20.2020.

Schroll C, Riemensperger T, Bucher D, Ehmer J, Völler T, Erbguth K, Gerber B, Hendel T, Nagel G, Buchner E, Fiala A. Light-Induced Activation of Distinct Modulatory Neurons Triggers Appetitive or Aversive Learning in Drosophila Larvae. Current Biology. 2006; 16(17):1741–1747. doi: 10.1016/j.cub.2006.07.023.

Schwaerzel M, Monastirioti M, Scholz H, Friggi-Grelin F, Birman S, Heisenberg M. Dopamine and Octopamine Differentiate between Aversive and Appetitive Olfactory Memories in Drosophila. Journal of Neuroscience. 2003; 23(33):10495–10502. doi: 10.1523/jneurosci.23-33-10495.2003.

Senapati B, Tsao CH, Juan YA, Chiu TH, Wu CL, Waddell S, Lin S. A neural mechanism for deprivation state-specific expression of relevant memories in Drosophila. Nature Neuroscience. 2019; 22(12):2029–2039. doi: 10.1038/s41593-019-0515-z.

Smith D, Wessnitzer J, Webb B. A model of associative learning in the mushroom body. Biological Cybernetics. 2008; 99(2):89–103. doi: 10.1007/s00422-008-0241-1.

Springer M, Nawrot MP. A mechanistic model for reward prediction and extinction learning in the fruit fly. eNeuro. 2021; p. ENEURO.0549–20.2021. doi: 10.1523/eneuro.0549-20.2021.

Sten TH, Li R, Otopalik A, Ruta V. An arousal-gated visual circuit controls pursuit during Drosophila courtship. bioRxiv. 2020; p. 2020.08.31.275883. doi: 10.1101/2020.08.31.275883.

Tabone CJ, Belle JSd. Second-order conditioning in Drosophila. Learning & Memory. 2011; 18(4):250–253. doi: 10.1101/lm.2035411.

Takemura Sy, Aso Y, Hige T, Wong A, Lu Z, Xu CS, Rivlin PK, Hess H, Zhao T, Parag T, Berg S, Huang G, Katz W, Olbris DJ, Plaza S, Umayam L, Aniceto R, Chang LA, Lauchie S, Ogundeyi O, et al. A connectome of a learning and memory center in the adult Drosophila brain. eLife. 2017; 6:e26975. doi: 10.7554/elife.26975.

Tanaka NK, Tanimoto H, Ito K. Neuronal assemblies of the Drosophila mushroom body. Journal of Comparative Neurology. 2008; 508(5):711–755. doi: 10.1002/cne.21692.

Turner GC, Bazhenov M, Laurent G. Olfactory Representations by Drosophila Mushroom Body Neurons. Journal of Neurophysiology. 2008; 99(2):734–746. doi: 10.1152/jn.01283.2007.

Waddell S. Dopamine reveals neural circuit mechanisms of fly memory. Trends in Neurosciences. 2010; 33(10):457–464. doi: 10.1016/j.tins.2010.07.001.

Wessnitzer J, Young JM, Armstrong JD, Webb B. A model of non-elemental olfactory learning in Drosophila. Journal of Computational Neuroscience. 2012; 32(2):197–212. doi: 10.1007/s10827-011-0348-6.

Wu CL, Shih MF, Lee PT, Chiang AS. An Octopamine-Mushroom Body Circuit Modulates the Formation of Anesthesia-Resistant Memory in Drosophila. Current Biology. 2013; 23(23):2346–2354. doi: 10.1016/j.cub.2013.09.056.

Wu Y, Ren Q, Li H, Guo A. The GABAergic anterior paired lateral neurons facilitate olfactory reversal learning in Drosophila. Learning & Memory. 2012; 19(10). doi: 10.1101/lm.025726.112.

Wu Z, Guo A. A model study on the circuit mechanism underlying decision-making in Drosophila. Neural Networks. 2011; 24(4):333–344. doi: 10.1016/j.neunet.2011.01.002.

Young JM, Wessnitzer J, Armstrong JD, Webb B. Elemental and non-elemental olfactory learning in Drosophila. Neurobiology of Learning and Memory. 2011; 96(2):339–352. doi: 10.1016/j.nlm.2011.06.009.

Zhang YV, Ni J, Montell C. The Molecular Basis for Attractive Salt-Taste Coding in Drosophila. Science. 2013; 340(6138):1334–1338. doi: 10.1126/science.1234133.

Zhao C, Widmer YF, Diegelmann S, Petrovici MA, Sprecher SG, Senn W. Predictive olfactory learning in Drosophila. Scientific Reports. 2021; 11(1):6795. doi: 10.1038/s41598-021-85841-y.

Zhao F, Zeng Y, Guo A, Su H, Xu B. A neural algorithm for Drosophila linear and nonlinear decision-making. Scientific Reports. 2020; 10(1):18660. doi: 10.1038/s41598-020-75628-y.

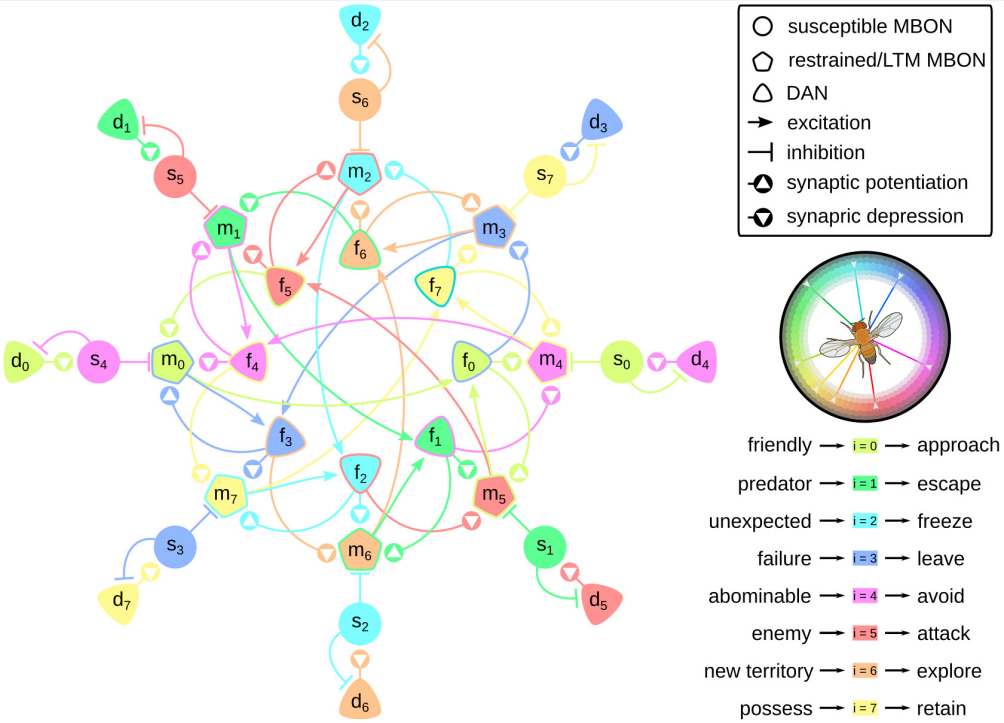
Zhou M, Chen N, Tian J, Zeng J, Zhang Y, Zhang X, Guo J, Sun J, Li Y, Guo A, Li Y. Suppression of GABAergic neurons through D2-like receptor secures efficient conditioning in Drosophila aversive olfactory learning. Proceedings of the National Academy of Sciences. 2019; 116(11):201812342. doi: 10.1073/pnas.1812342116.

Zhu L, Mangan M, Webb B. Biomimetic and Biohybrid Systems, 9th International Conference, Living Machines 2020, Freiburg, Germany, July 28–30, 2020, Proceedings. Lecture Notes in Computer Science. 2020; p. 415–426. doi: 10.1007/978-3-030-64313-3_39.

Appendix 1

The incentive wheel

We have shown that the incentive circuit (IC) is able to explain classical conditioning experiments that have been done with adult fruit flies, and its neurons can replicate the responses of the mushroom bodies in the flies' brain. We have also seen that three types of memories are stored in this model (i.e. susceptible, restrained and long-term) for each of the two represented motivations of the animal (i.e. attraction or avoidance) guided by reinforcements (i.e. reward or punishment). Although this model is sufficient to explain the behaviour of the animals in the laboratory, where the animal is exposed in controlled portions of chemicals and the results are translated into a simple attraction to or avoidance from a source, in the wild there are more than two motivations that modulate the behaviour of the animal either synergistically or opponently.



Appendix 1 Figure 1. The "incentives wheel" model. This model supports that reinforcement is not binary but it draws its values from a spectrum. Different types of reinforcements trigger different dopaminergic neurons (DANs) that enable learning in different parts of the mushroom body. Colours show the variety of motivations that the model can encode associated with humans' 'wheel of emotions' (***Plutchik, 2001***); e.g. light green: trust, green: fear, light blue: surprise, blue: sadness, pink: disgust, red: anger, orange: anticipation, yellow: joy. Neurons of more than one colours are part of multiple circuits that contribute in different motivations.

Real-life experiences are complicated and rich in information. This could produce a whole spectrum of reinforcements and motivations that guide the behaviour of animals. Data show that animals respond differently in different reinforcements, which cannot be represented just by the magnitude of a single variable, e.g. more/less rewarding/punishing. For example, different concentration of salt (***Zhang et al., 2013***), or sugar (***Colomb et al., 2009***) might be combined with the sated state of the animal, activate different subsets of DANs and trigger different behaviours, such as feeding or escaping. When the male fruit fly is exposed to female pheromones, courtship behaviour is triggered through P1 neurons

(***Kallman et al., 2015***; ***Sten et al., 2020***), which can be translated to attraction, but has nothing to do with the appetite of the animal. On the other hand, other male pheromones trigger avoidance which shows that a similar to the incentive circuit could explain this behaviour. The mushroom body has been proved to contribute in many behaviours other than olfactory classical conditioning, including visual navigation, and its output neurons encode richer information that is very close to humans' decision making (***Heisenberg, 2003***).

It is reasonable to think that the ~ 34 output and ~ 130 dopaminergic neurons interacting with the mushroom bodies in the brain of fruit flies are not all used in order to discriminate odours and assign a positive or negative reinforcement to them driving attraction and avoidance. For this reason, we believe that the different MBONs do not represent different odours like it has been proposed before (***Huerta et al., 2004***) neither are split into two groups [e.g. attraction or avoidance, (***Schwaerzel et al., 2003***; ***Schroll et al., 2006***; ***Waddell, 2010***)], but rather represent different motivations of the animal that all together guide its overall behaviour (***Heisenberg, 2003***; ***Krashes et al., 2009***). These motivations are associated to different contexts, which are represented by the responses of the KCs as it has been proposed by ***Cohn et al.*** (***2015***), who showed that the same output neurons in the γ compartment respond differently when a different context is given. This context enables or disables different microcircuits of the mushroom body — similar to the ones described in Mushroom body microcircuits and The incentive circuit (IC) — which result in the activation of a subset of MBONs that represent different motivations, while the overlapping microcircuits result in what we sometimes call "noisy" or "insignificant" changes in the behaviour.

Appendix 1 ***Figure 1*** illustrates such a model, which we call the 'incentive wheel' (IW). In this model, we use 4 identical incentive memories (C0/4, C1/5, C2/6 and C3/7), where the Reciprocal restrained memories (RRM) microcircuit of the one is the Reciprocal forgetting memories (RFM) microcircuit of another. As the structure of the RFM microcircuit is identical to the RRM one, we assume that the RFM of circuit C0/4 is the RRM of circuit C1/5, and the RFM of circuit C1/5 is the RRM of circuit C2/6 and so on. This way, we weave the different circuits into an incentive wheel with opposing motivations. The reinforcements that trigger the DANs in this model are drawn from a spectrum and the output of the MBONs of the model triggers different motivations. The LTMs and restrained memories can both exist in the same neurons of the core of the model, representing different motivations in different contexts. This might cause changes in the behaviour of the circuits that is irrelevant to the associated reinforcement, but relevant to a neighbouring reinforcement of the spectrum.

The 'incentive wheel' is an example of how the incentive circuit can be part of a bigger circuit that can provide a variety of motivations to the animal. An extension of it could have susceptible MBONs, s_i, connecting to other susceptible MBONs from a parallel IW model with higher-order motivations. In Appendix 1 ***Figure 1***, we have associated the different motivations with the primary human emotions from the 'wheel of emotions' (***Plutchik, 2001***). Higher-order motivations could exist by combining primary motivations as if they were emotions and result in more complicated behaviours for the animal.

Appendix 2

Appendix 2 Table 1. Additional information on the mushroom body neurons used for the incentive circuit, sorted by their short name. Information about the neurons has been collected from ***McCurdy et al.*** (***2021***) and ***Aso et al.*** (***2014a***).

Driver	Cluster	Neuron name	#cells	Short name	Alternative names
MB011B	M4/M6	MBON-$\gamma 5\beta' 2$a	1	MBON-01	MB-M6
MB399B	M4/M6	MBON-$\beta 2\beta' 2$a	1	MBON-02	
MB298B	MV2	MBON-$\gamma 4 > \gamma 1\gamma 2$	1	MBON-05	
MB112C		MBON-$\gamma 1$pedc>α/β	1	MBON-11	MB-MVP2
MB077B	V3/V4	MBON-$\gamma 2\alpha' 1$	2	MBON-12	
MB050B	V2	MBON-$\alpha' 1$	3	MBON-15	
MB109B	PAM	PAM-$\beta' 2$a	6-9	PAM-02	
MB301B	PAM	PAM-$\beta 2\beta' 2$a	> 3	PAM-04	subset of MB-M8
MB312B	PAM	PAM-$\gamma 4 < \gamma 1\gamma 2$	13-17	PAM-07	subset of MB-AIM?
MB320C	PPL1	PPL1-$\gamma 1$ped	1	PPL1-01	MB-MP1, MP
MB296B$_1$	PPL1	PPL1-$\gamma 2\alpha' 1_1$	1	PPL1-03	MB-MV1
MB296B$_2$	PPL1	PPL1-$\gamma 2\alpha' 1_2$	1	PPL1-03	MB-MV1

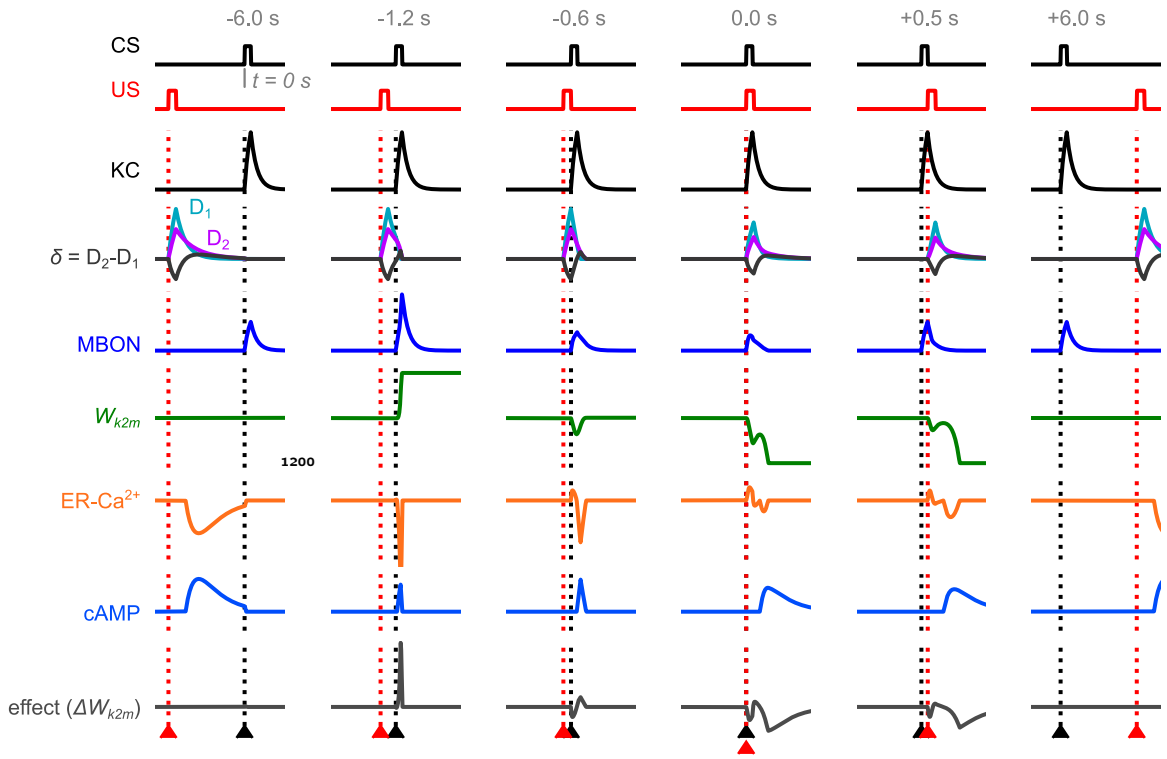


Figure 2–Figure supplement 1. All the chemical levels and neural activities calculated based on the order of the conditional (CS) and unconditional stimuli (US). Black arrowhead marks time of the CS (duration $0.5\,\mathrm{sec}$); red arrowhead marks time of the US (duration $0.6\,\mathrm{sec}$), similar to ***Handler et al.*** (***2019***) - Figure 2D.

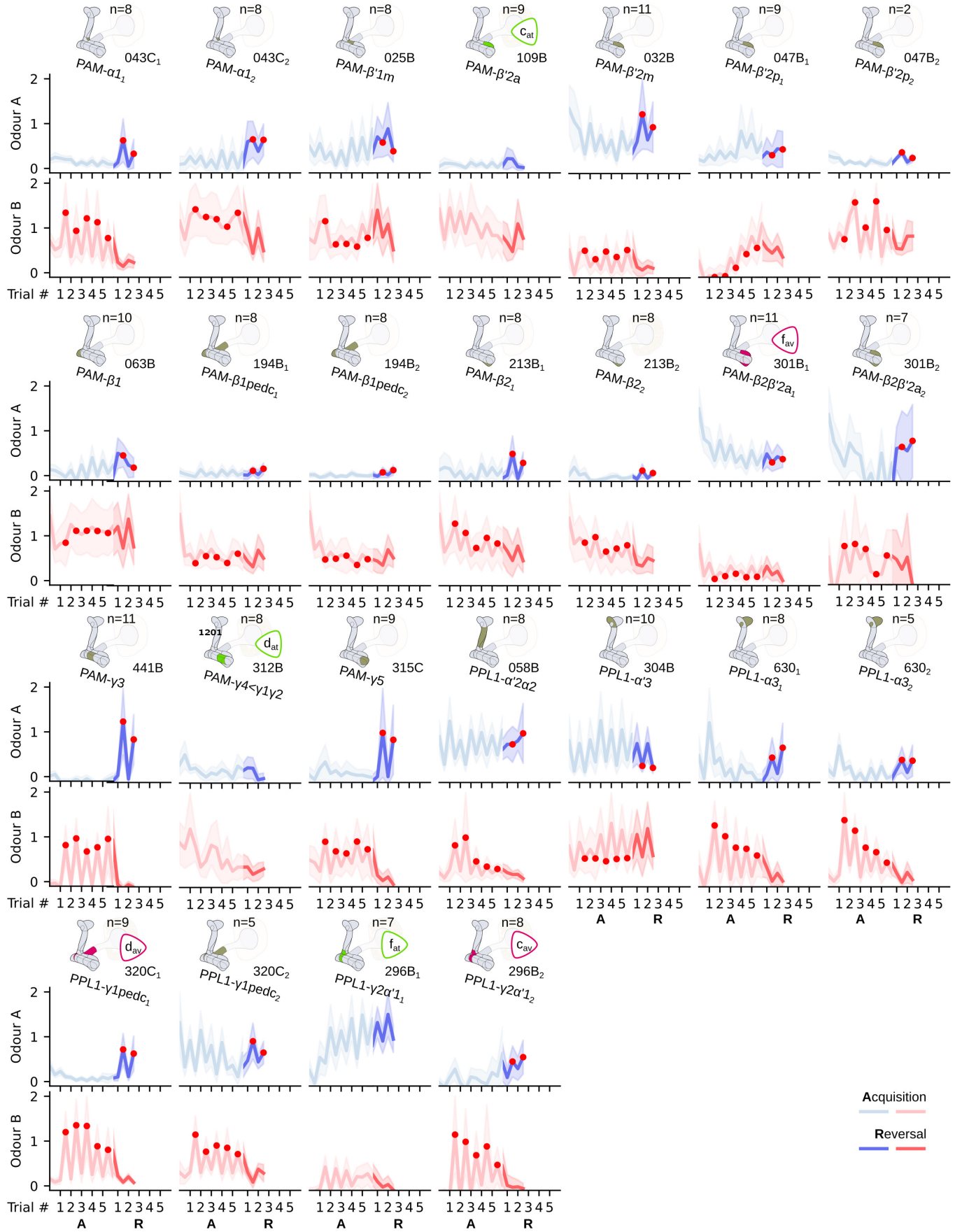


Figure 8–Figure supplement 1. Responses of all the recorded DANs in the mushroom body.

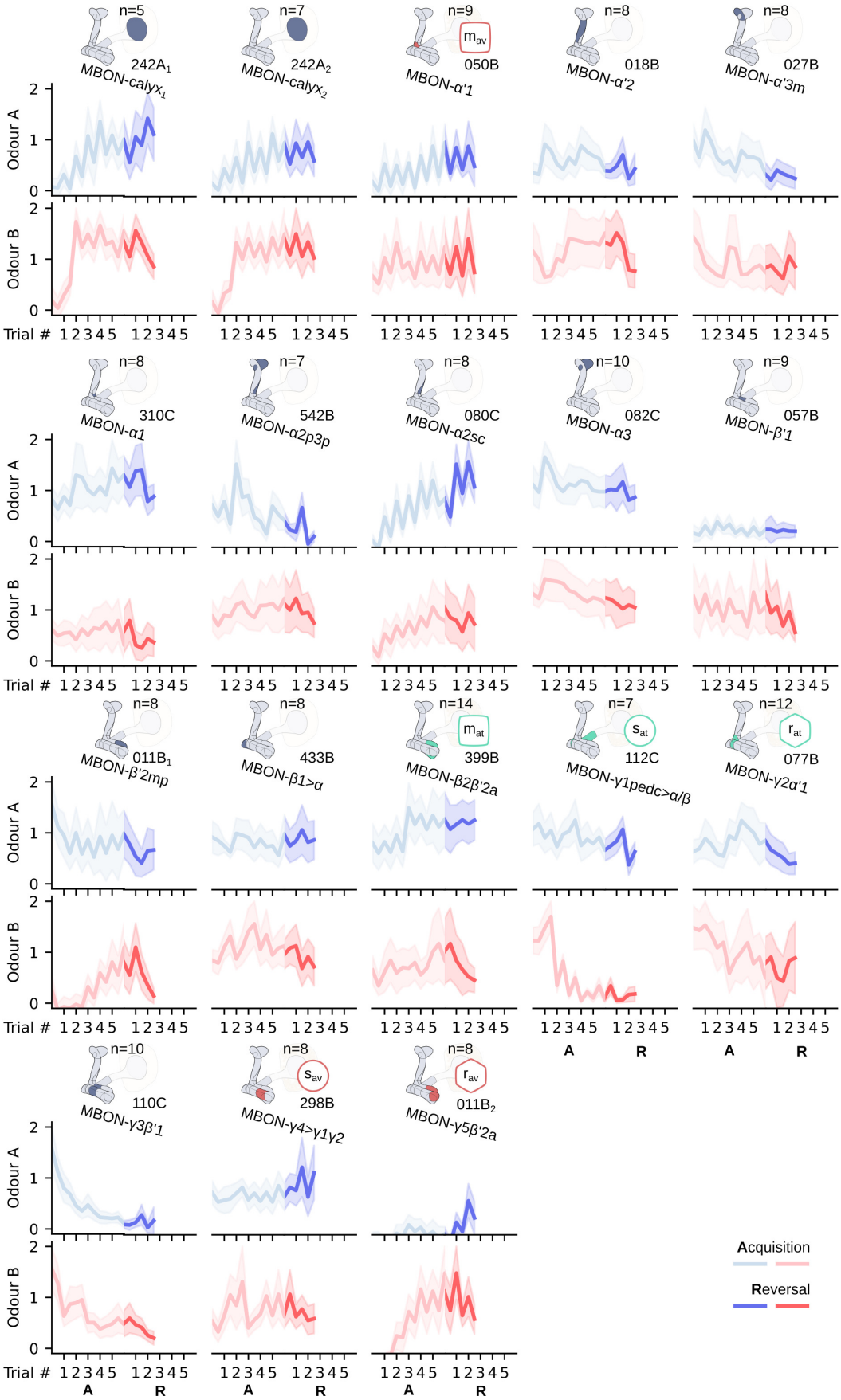


Figure 8–Figure supplement 2. Responses of all the recorded MBONs in the mushroom body.

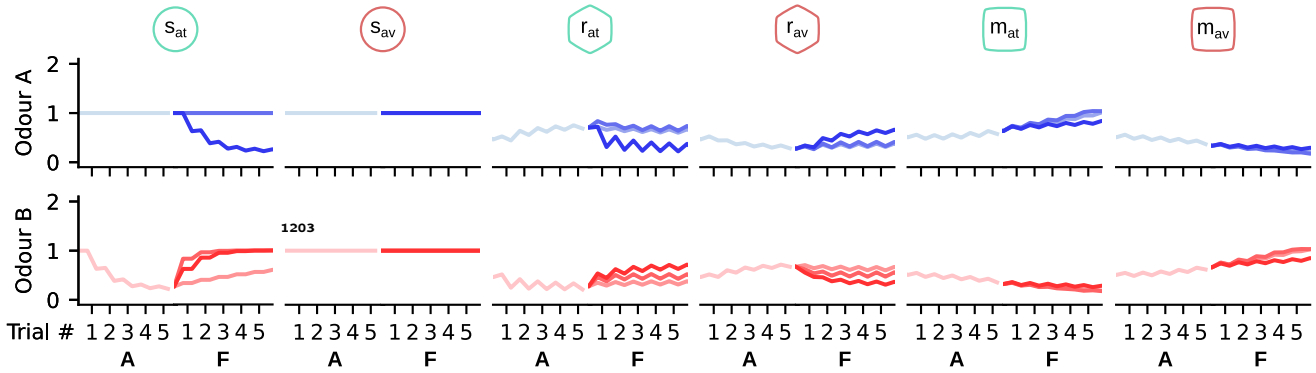


Figure 9–Figure supplement 1. The KC→MBON synaptic weights for the MBONs of the model during the different phases.

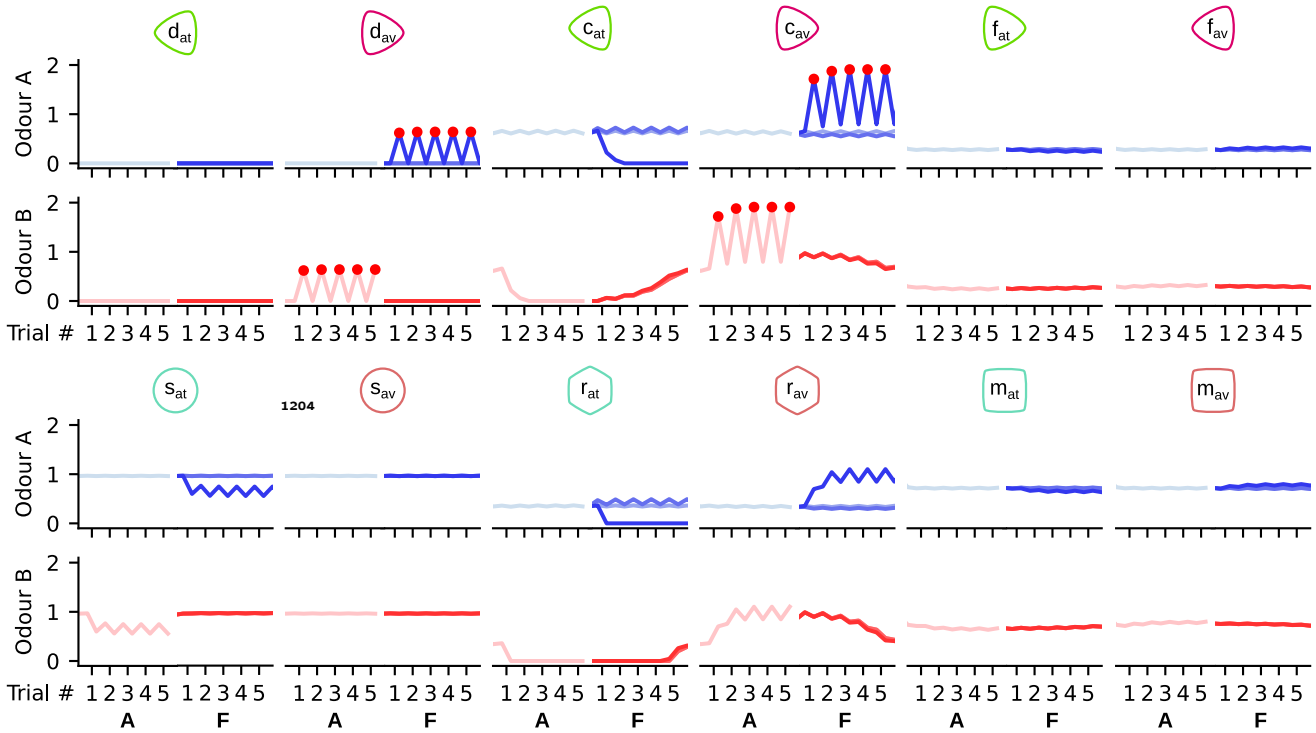


Figure 9–Figure supplement 2. Generated responses when using the prediction-error learning rule.

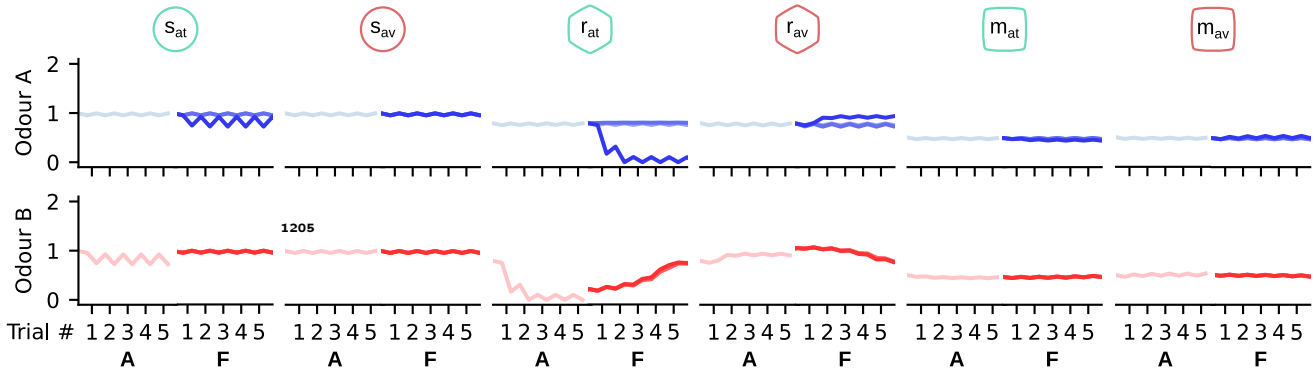


Figure 9–Figure supplement 3. The KC→MBON synaptic weights for the MBONs of the model using the prediction-error learning rule.

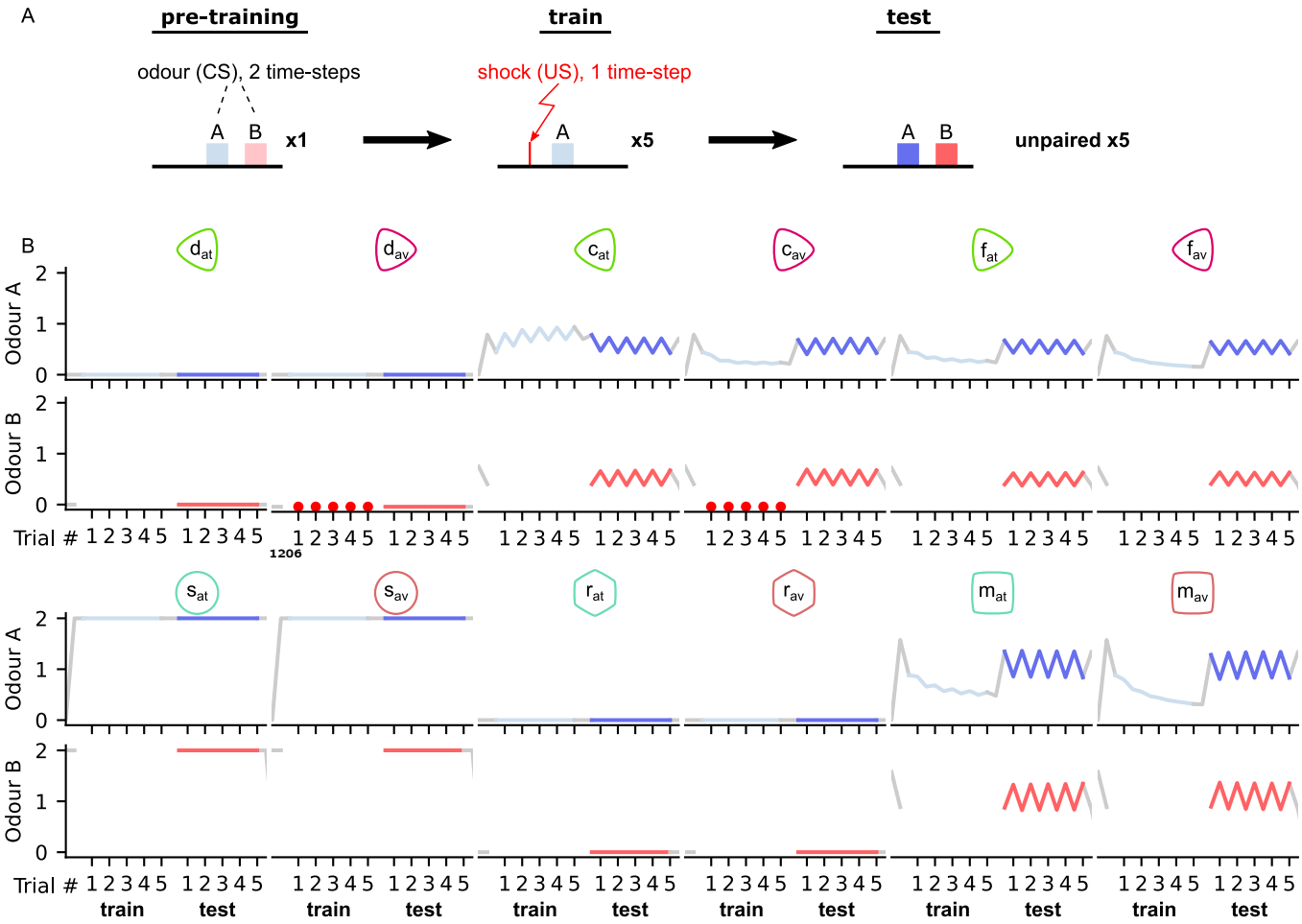


Figure 9–Figure supplement 4. The responses of the model during the classic unpaired conditioning paradigm. (A) During the pre-training, we deliver odour A and then odour B to make the animal familiar to the odours. In the training phase, we deliver electric shock and then odour A alternately for 5 trials. In the testing phase, we replace electric shock with odour B and alternate the odours for 5 trials. (B) The responses produced by the model.

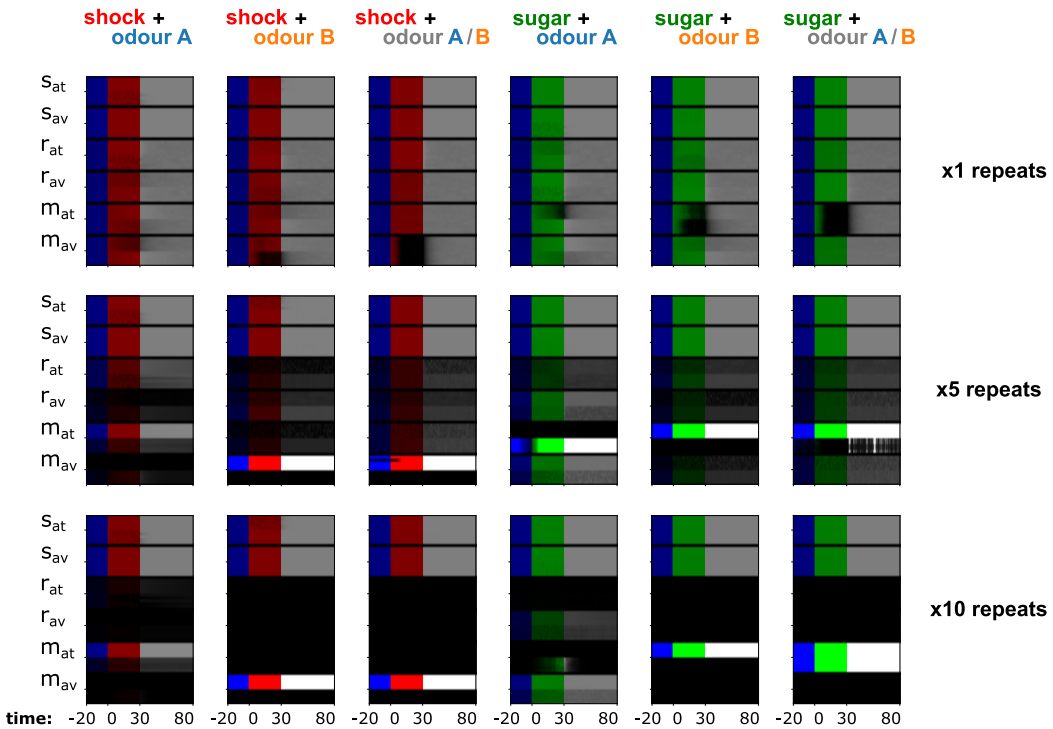


Figure 11–Figure supplement 1. The mean KC→MBON synaptic weights over 100 flies for each neuron and time-step and for 1, 5 and 10 repeats.

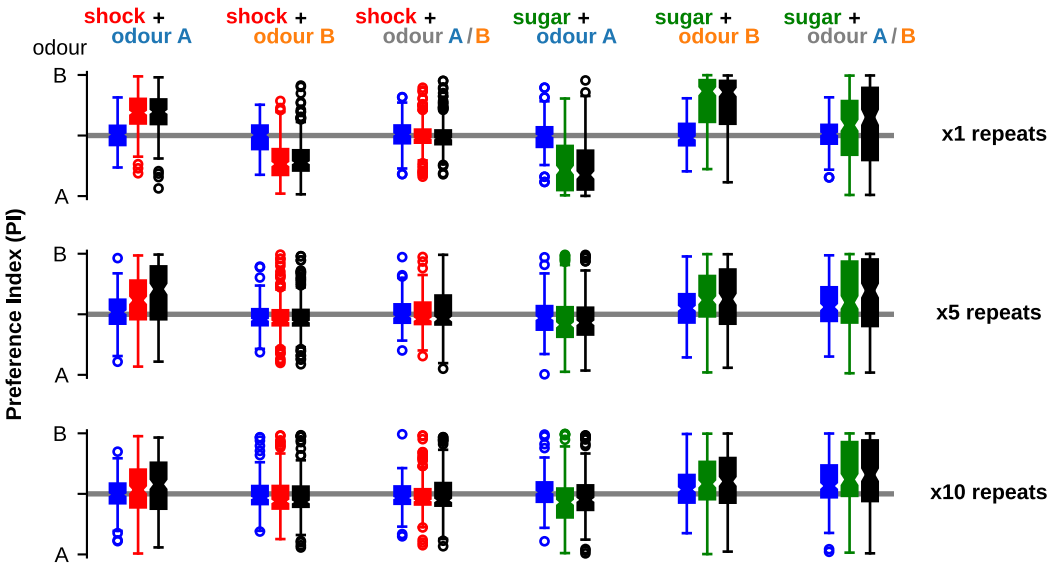
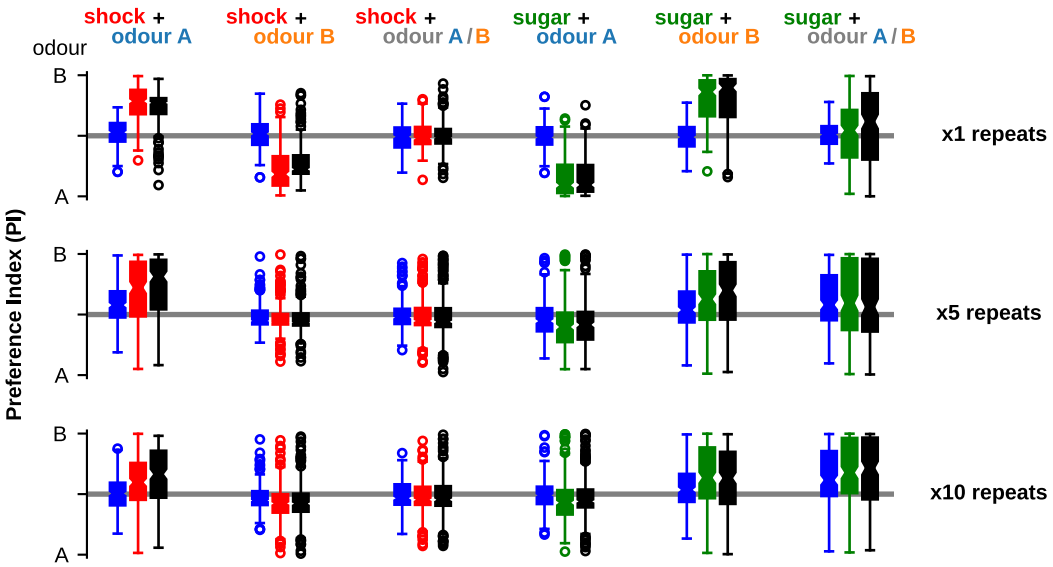
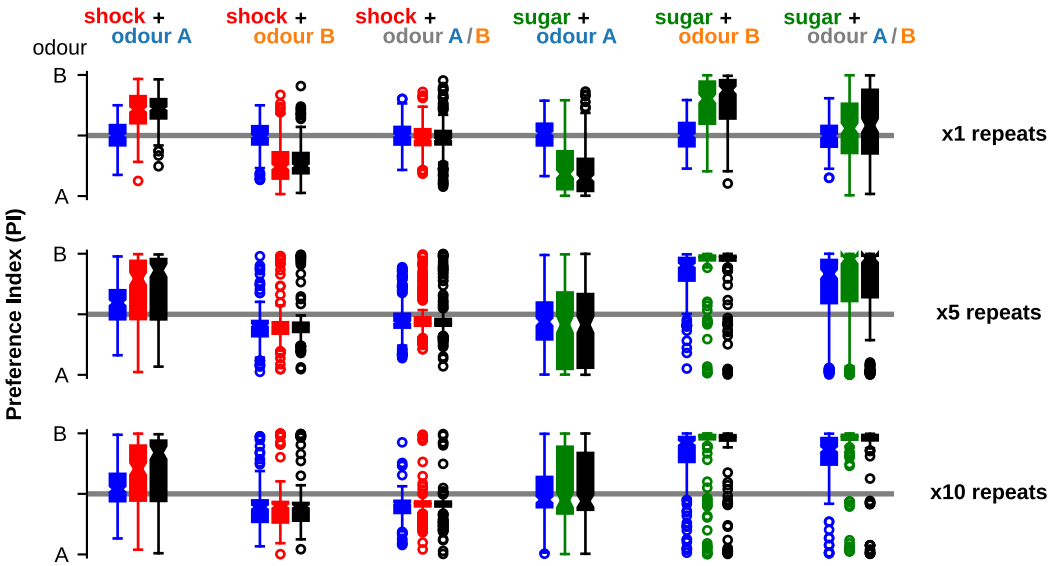


Figure 11–Figure supplement 2. Summary of behaviour over 100 flies when controlled by the susceptible MBONs only for 1, 5 and 10 repeats.



1209

Figure 11–Figure supplement 3. Summary of behaviour over 100 flies when controlled by the restrained MBONs only for 1, 5 and 10 repeats.



1210

Figure 11–Figure supplement 4. Summary of behaviour over 100 flies when controlled by the long-term memory MBONs only for 1, 5 and 10 repeats.

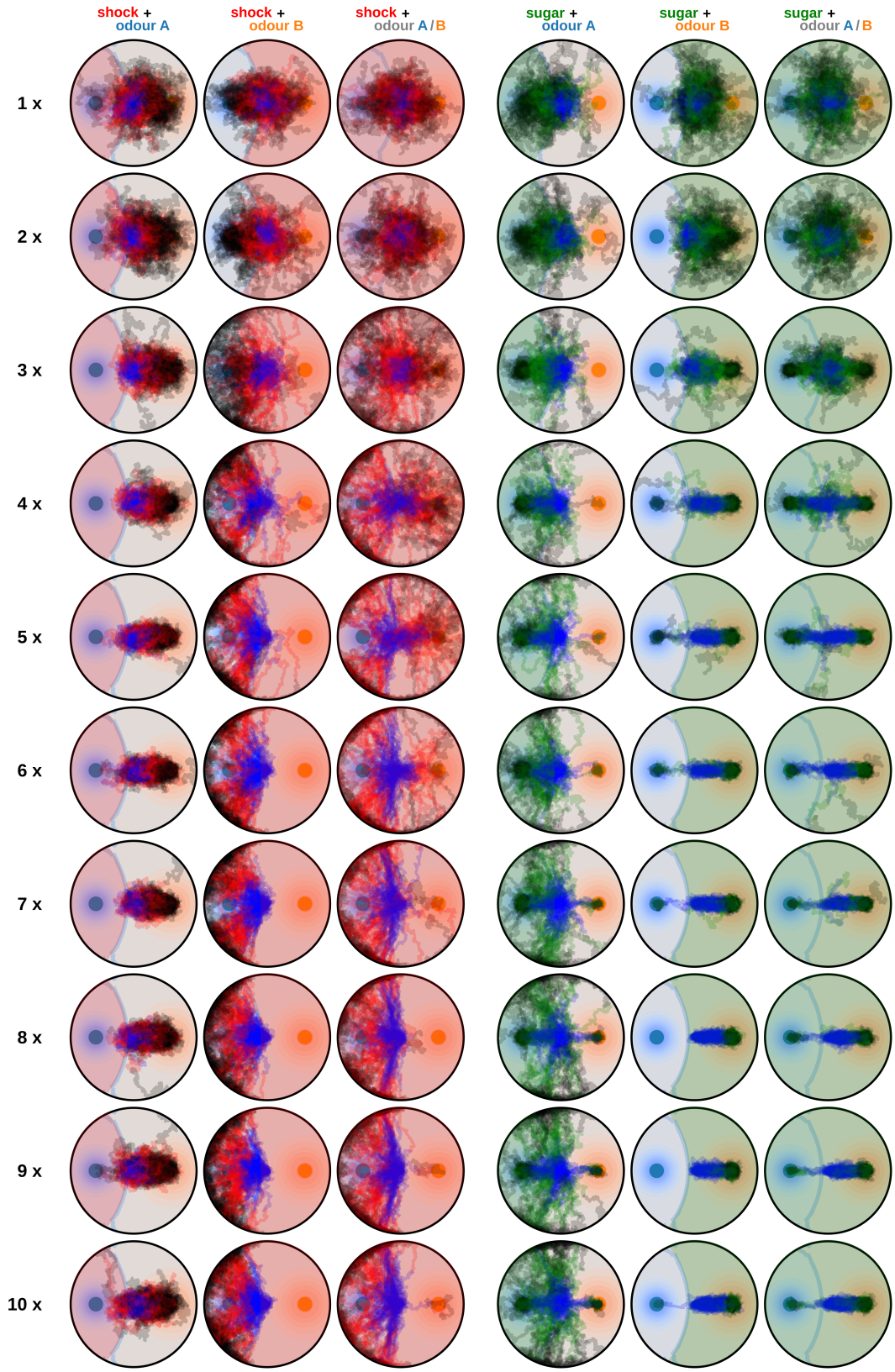


Figure 11–Figure supplement 5. Paths of the flies when using the dopaminergic learning rule and during all 10 repeats of the experiment.

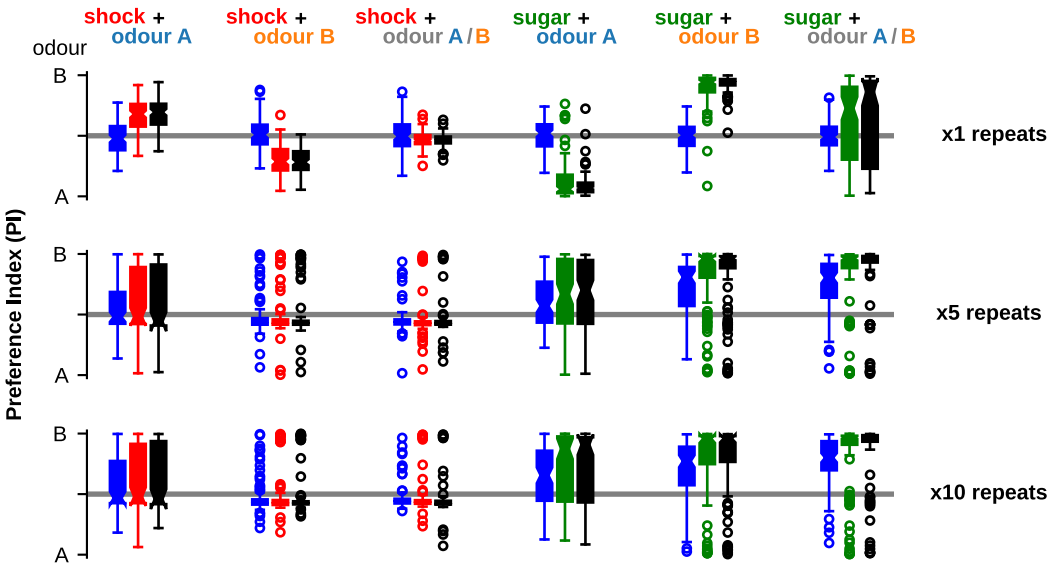


Figure 11–Figure supplement 6. Summary of behaviour over 100 flies when using the reward prediction error learning rule for 1, 5 and 10 repeats.

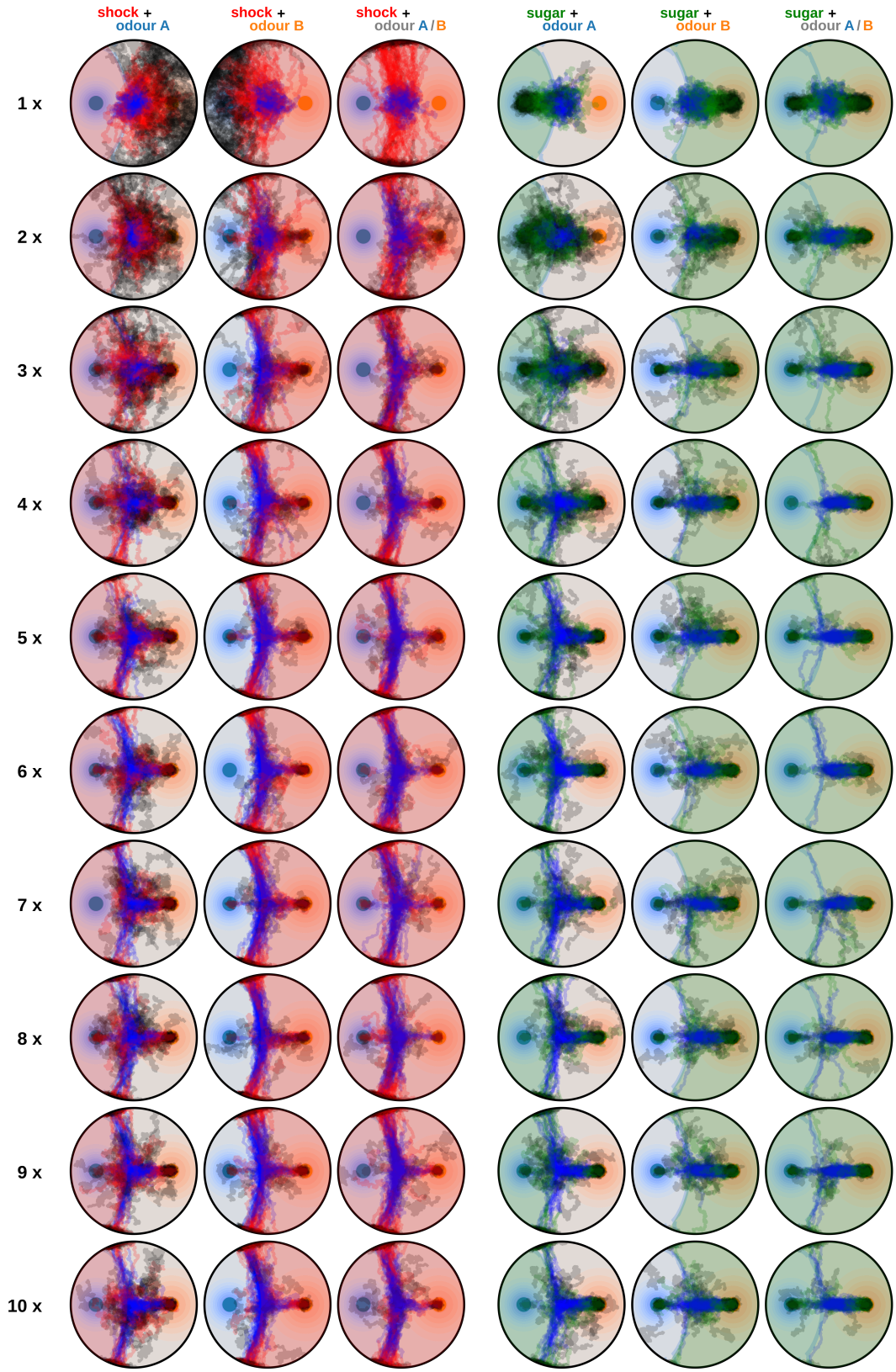


Figure 11–Figure supplement 7. Paths of the flies when using the RPE learning rule and during all 10 repeats of the experiment.