Smart as a Bug: A Computational Model of Learning in the Moth Olfactory Network, with Applications to Neural Nets

Charles Boise Delahunt

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

J. Nathan Kutz, Chair

Eve Riskin, Chair

Jeffrey A. Riffell

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Smart as a Bug: A Computational Model of Learning in the Moth Olfactory Network, with Applications to Neural Nets

Charles Boise Delahunt

Co-Chairs of the Supervisory Committee:
Professor J. Nathan Kutz
Applied Mathematics

Professor Eve Riskin
Electrical Engineering

The moth olfactory network is one of the simplest biological neural systems capable of Learning. It is thus ideal for exploring how learning occurs. The network, which includes the antenna lobe, mushroom body, and ancillary structures, contains several key structural motifs widespread in biological neural systems and of great interest. These include cascading networks, large dimension shifts from stage to stage, high-dimensional sparse codings of data, randomness, Hebbian ("fire together, wire together") plasticity, and octopamine stimulation as a vital part of the learning mechanism. While these components are widespread in natural neural systems, they are largely absent from the engineered neural nets of machine learning.

This thesis has three goals: To characterize the various components of the moth's olfactory system and how they enable it to learn; to port the moth's "bag of tricks" to machine learning contexts; and to examine learning as an injury mitigation mechanism.

Our approach is to build a full computational model of the moth olfactory system with the following properties: Its structure and mechanics are tightly tethered to current knowledge of the moth olfactory system; its behavior statistically matches experimental data; and it is able to robustly learn new odors. To our knowledge this is the first full neural network

model that is tightly tied in structure and behavior to a real biological system, and that also demonstrates learning behavior.

From a Biology perspective, the model is a valuable platform to examine how key structural features enable learning in nature. For example, we analyse the role of octopamine stimulation and the functions of high-dimensional sparse network stages in learning. The model also allows predictions about structural details of the olfactory system that are not currently well-characterized. In addition, we explore the role of learning, and other structural features, as injury mitigation mechanisms.

From a Machine Learning perspective, the model allows us to identify promising structures and tools that can be ported to ML systems. For example, it offers bio-mimetic solutions to two open concerns in human-built Neural Nets: It uses a biologically-plausible optimization method to train the network, potentially bridging a long-standing gap between natural and human-built neural nets; and it requires few training samples, offering a potential means to address a current bottleneck in use of neural nets, viz their vast appetite for training data.

TABLE OF CONTENTS

Page

List of Figures ii

Chapter 1: Thesis Introduction 1

Chapter 2: Learning mechanisms in the moth olfactory network 3
2.1 Introduction 3
2.2 Methods 4
2.3 Results 24
2.4 Discussion 37

Chapter 3: Putting a bug in machine learning 42
3.1 Introduction 42
3.2 Methods 43
3.3 Results 49
3.4 Discussion 59

Chapter 4: Built to Last: Injury mitigation mechanisms in the MON 63
4.1 Introduction 63
4.2 Methods 65
4.3 Results 73
4.4 Discussion 81

Chapter 5: Supplementary Information 88

Bibliography 95

LIST OF FIGURES

Figure Number Page

2.1 **AL-MB overview. A:** System schematic: Chemical sensors (RNs) excite a noisy pre-amp network (AL), which feeds forward to a plastic sparse memory layer (MB), which excites readout (decision) neurons (ENs). Green lines show excitatory connections, red lines show inhibitory connections (LH inhibition of the MB is global). Light blue ovals show plastic synaptic connections into and out of the MB. **B:** Neuron timecourse outputs from each network (typical simulation) with time axes aligned vertically. Timecourses are aligned horizontally with their regions-of-origin in the schematic. The AL timecourse shows all responses within ± 2.5 std dev of mean spontaneous rate as medium blue. Responses outside this envelope are yellow-red (excited) or dark blue (inhibited). MB responses are shown as binary (active/silent). Timecourse events are as follows: (1) A period of no stimulus. All regions are silent. (2) Two odor stimuli are delivered, 3 puffs each. AL, MB, and ENs display odor-specific responses. (3) A period of control octopamine, ie without odor or Hebbian training. AL response is varied, MB and EN are silent. (4) The system is trained (octopamine injected) on the first odor. All regions respond strongly. (5) A period of no stimulus. All regions are silent, as in (1). (6) The stimuli are re-applied. The AL returns to its pre-trained activity (it is not plastic). In contrast, the MB and EN are now more responsive to the trained odor, (response to the untrained odor is unchanged). Green dotted line in the EN represents a hypothetical "action" threshold. The moth has learned to respond to the trained odor. 6

2.2 **Schematic of an AL glomerulus.** Detail of neural connections within a glomerulus. Red = inhibitory, green = excitatory, blue = increases responsiveness. RNs enter from the antennae. LNs enter from other glomeruli; one full LN is shown. It is not known if octopamine modulates LNs and PNs (see section 2.4.1). 10

2.3 **Time series of PN firing rates from wet-lab.** x-axis = time, y-axis = FR. Blue lines = mean spontaneous rate, shaded regions = ±1 and 2 std. Red dots are odor responses. Green dots are response to control (mineral oil).
A: PN response, given odor plus coincident sugar reward, ie plus octopamine (time series for PNs with odor only are similar, but with less strong odor responses). Top row: unresponsive to odor. Middle row: excited response to odor. Bottom row: inhibited response to odor.
B: PNs with octopamine wash added in mid-experiment, then rinsed away (duration shown by black line). Octopamine can alter (up, down, or not at all) the spontaneous FR and/or the odor response, so there are 9 possible modulation regimes. This grid of timecourses shows a typical PN from each regime. Top row: spontaneous FR in unaffected. Middle row: spontaneous FR is boosted. Bottom row: spontaneous FR is inhibited. First column: odor response is unaffected. Second column: odor response is boosted. Third column: odor response is inhibited. . . 23

2.4 **Wet-lab data and Model Calibration:** Comparison of PN firing rate activity from wet-lab data and simulations. **Panel A**: Histograms and CDFs of wet-lab data and simulations. **Col a**: mean spontaneous FRs μ_s. **Col b**: σ_s/μ_s of spontaneous FRs, a measure of noisiness of a PN. **Col c**: odor response, measured as distance from μ_s in σ_s units. Distance $> 2\sigma_s$ implies a strong activation/inhibition. **Col d**: odor response during octopamine, in σ_s units distance from μ_s. Note that PN responses are broadened (i.e. more PNs are strongly activated or inhibited). The dotted line in the CDF inset is the same as the CDF of the odor response without octopamine, to show the broadening towards both extremes. **Col e**: change in mean spontaneous FRs due to octopamine, measured in σ_s units distance from (non-octopamine) μ_s. Some PNs are excited, some are inhibited. x-axis units: col a = raw spikes/sec; col b = ratio σ_s/μ_s; cols c, d, e = units of σ_s. **Panel B**: Activity of PNs indexed by increasing spontaneous FR. y-axis: Raw spikes/sec FR. Blue lines = mean spontaneous FRs μ_s (cf col a). Shaded regions = $\sigma_s, 2\sigma_s$ envelopes (cf col b). Solid red dots = odor response FRs (cf col c). Hollow red dots = odor response FRs during octopamine (cf col d). Red lines show the change in odor response FRs due to octopamine (cf broadened response). Black stars (*) = spontaneous FRs during octopamine (cf col e). 26

2.5 **KC responses to odor during training:** KCs respond sparsely to odor pre- and post-training, ie absent octopamine (blue and green dots and curves). Octopamine induces transient increased responsivity (red dots and curves). Training results in permanent increases in response to the trained odor, but no increase in response to control odor (green dots and curves).
A: KC response to an odor before, during, and after training. x-axis: indexed KCs (500 shown). y-axis: consistency of response (in %). The plots are for odor 1 as the trained odor (ie same data as panel B). Blue = pre-training (no octopamine). Red = during training (with octopamine); note the heightened transient response. Green = post-training (no octopamine). There is a permanent increase in the number of KCs that respond to the trained odor.
B: Response rate vs. percentage of active KCs for trained and control odors before, during, and after training. x-axis: percentage of KCs responding at the given rate. y-axis: consistency of response (in %). Blue = pre-training. Red = during octopamine (transient). Green = post-training. The LH plot shows odor 1 as the reinforced odor. The scatterplots in (A) correspond to the three curves in this plot. Note that the permanent KC response curve shifts up and to the right (blue→green) in the trained odor, ie more KCs respond to the odor (right shift) and they respond more consistently (upward shift). The RH plot shows odor 2 as a control. The control's permanent KC response curve does not shift.
C: As (B) above, but in this experiment odor 1 is now the control (LH plot), and odor 2 is reinforced (RH plot). In this case, the response curve of odor 2 (reinforced) shifts to the right (blue→green), while the response curve of odor 1 (control) is unchanged. 29

2.6 **Effect of training on EN FRs: A:** Typical timecourse of EN responses from an experiment with a single moth. First, 16 puffs of each odor were delivered, to establish naive odor responses. Note EN response variability due to noise in the system, especially in the AL. Next, the moth was trained on the first (blue) odor trained over 2 sessions (10 puffs), by delivering odor and octopamine concurrently This timecourse corresponds to the {odor, #sessions} pair in the first column in panel B, at index 2 on the x-axis. Octopamine was then withdrawn, and the four odors were again delivered in series of puffs, to establish post-training changes in EN response. The long green line represents a hypothetical trigger threshold, such that EN response > threshold would induce a distinct behavior.
B: EN response changes due to training: Aggregated results with 11 noise realizations for each {odor, #sessions} pair. Each column shows results of training a given odor, color coded: blue, purple, red, green. x-axis = number of training sessions.
First row: The y-axis measures percent change in EN FR. The line shows mean percent change. The error bars show ±1, 2 std devs.
Second row: The y-axis measures percent changes in EN response, relative to the trained odor (ie subtracting the trained odor's change from all odors). This shows how far each control odor lags behind the trained odor. The line shows mean percent lag. The error bars show ±1, 2 std devs. 32

2.7 **Effect of training on EN FRs, given odors with unequal naive response magnitudes.** When odors induced naive EN responses of very different magnitudes, then trained odor response increased much more than control odor responses either in raw magnitude, or as a percentage, or both.
A: Typical timecourse showing magnitudes EN responses before and after training the third (red) odor, indicated by red arrow, over 15 odor puffs. This corresponds to the third column in panels B - D, at index 3 on the x-axis. Note that only the third (red) odor's EN response changes magnitude.
Panels B - D: Changes to ENs during training. x-axis = number of training sessions. Each column shows results of training a given odor, color coded: blue, black, red. y-axis measures raw EN or percent change in EN. 21 trials per data point.
B: Percent change (from pre-training) in ENs, mean ±2 stds.
C: Raw EN FRs, mean ±2 stds.
D: Changes in raw EN FRs, normalized by trained odor (ie subtract the trained odor's changes from all odors), mean ±2 std devs. This shows how far each control odor lagged behind the trained odor.
Note that the trained odor dominates in either raw increase (panels C, D) if naive response to trained odor was large, or in percent increase (panel B) if naive response to trained odor was small. 33

2.8 **Effects of sparsity on learning and EN reliability** Results for a typical experiment on a moth with two odors.
A: EN responses timecourses for two odors, at varying levels of KC activation (a, b: <1%. c, d: 5 to 15%. e, f: 20 to 45%. Order of events: 3 puffs of each odor as baseline, train on first odor (only one session shown), then 3 puffs each post-training. At very sparse levels (a, b) training is focused but odor response is not reliable. At low sparsity levels (e, f) training is unfocused, boosting EN response to control odor and to background noise.
B: Two Figures of Merit (FoMs) plotted against MB sparsity. Low KC activation (high sparsity) correlates with well-focused learning, but low odor response SNR. High KC activation (low sparsity) correlates with poorly-focused learning, but high odor response SNR. The FoMs are each normalized for easier plotting. y-axis:
Blue data: $\frac{\mu(f)}{\sigma(f)}$, a measure of odor EN response SNR, where f = EN odor response.
Red data: $\frac{\mu(f_T)}{\mu(f_C)}$, a measure of learning focus, where $\mu(f_T)$ = mean EN post-training response to reinforced odor; $\mu(f_C)$ = mean EN post-training response to control odor (values are thresholded at 1 for plotting). A high value indicates that increases in EN response due to training were focused on the trained odor; low values indicate that irrelevant signal (F_C) was also boosted by training.
The points are experimental data, the curves are cubic fits. Vertical green lines indicate the 5 - 15% sparsity region, typical in biological neural systems. 36

3.1 **Network schematic.** Green lines show excitatory connections, red lines show inhibitory connections. Light blue ovals show plastic connections into and out of the MB. The glomeruli (processing units) in the AL competitively inhibit each other. Global inhibition from the lateral horn induces sparsity on MB responses. The ENs give the final, actionable readouts of the system's response to a stimulus. 43

3.2 **Downsampled MNIST digits** used in the experiments (random selection). . . 46

3.3 **Pre- and post-training EN time courses** (normalized) for a typical moth trained on 15 samples per class, showing post-training separation. Each timecourse shows EN response to 150 digits (15 ones, then 15 twos, etc). Top left = naive response (all ENs similar). Other subplots show trained ENs (trained class responses framed in red, some confounding class responses in dashed red). 51

3.4 **Pre- and post-training EN response distributions** (normalized) from a typical experiment, showing post-training separation of class response distributions. Dots show mean(μ), bars show mean(σ) averaged over μ and σ from 13 moths. Mean accuracy for this template was 76%, range 71-83%. Top left = naive response (all ENs similar). Other subplots show trained EN responses in blue. 10 training samples/class. 52

3.5 **Effects of sparsity in the MB:** Optimal accuracy (blue domed curve, $\mu \pm \sigma$) occurred at 5-20%, a compromise between learning focus and high intra-class signal-to-noise ratio (SNR). Descending red curve = mean separation of trained vs control (learning focus). Black ascending curve = mean intra-class SNR. Learning focus and SNR are scaled for plotting. 17 moths per sparsity level. 55

3.6 **Growth rates and sniffing. A:** Effects of growth rate. Solid horizontal blue curve = very fast learner. Solid ascending red curve = slow learner. The fast learner attained 75% accuracy in one-shot, but with no further gains. The slow learner ultimately attained higher accuracy. Dotted lines show the same effect in accuracy using threshold classifier. **B:** Effects of sniffing and noise on one-shot learning, in a "slow-learner" moth template. Multiple sniffs greatly improved one-shot accuracy. Noise in the AL: The clusters of $\mu \pm \sigma$ bars represent varying levels of AL noise, with low-to-high noise plotted left-to-right at each x-axis location. AL noise level did not affect accuracy. 13 moths per data point. 58

4.1 **Schematic overview of the Moth Olfactory Network** (MON) **and axonal injury mechanisms. A, B:** The MON is organized as a feedforward cascade of five distinct subnetworks and a reward mechanism. **C:** Focal Axonal Swellings (FAS) are ubiquitous across all severities of traumatic brain injuries and present in other leading brain disorders. They can cause some or all neural spikes in the train to die off in transit, reducing the overall firing rate arriving at the downstream target neuron. Adding FAS effects to the MON are the basis of our damage/injury protocols. Panel A is adapted from [18] and Panel C from [54]. 64

4.2 **Location of injury in experiments.** Red stars: Damage to Antennae (RNs), which reduces input to the AL. Orange stars: Damage to the AL→MB channel, which reduces signals passed by PNs and QNs to the MB. Left-hand diagram is adapted from [18]. 69

4.3 **Typical EN timecourse.** Readouts from the EN in a typical experiment, in which injury attenuated the EN odor response, and training partly restored it. Naive response (20-55), injury (red dot at 60), injured response (100-150), 5 puffs training (high response, with yellow, 170-190), post-training response (240-280). 71

4.4 **Learning as injury compensation mechanism.** Red/orange: Post-injury EN odor response, normalized by naive, healthy odor response. Blue: Post-training EN response, normalized by naive, healthy odor response. Green: Relative increase from post-injury response due to training. $\mu \pm \sigma$. **A:** Injury to RNs: Trained EN responses (blue) fully regained their pre-injury levels (black line) from injured levels (red) if injury was on average $\leq 20\%$. The ability of training to recover lost ground was fairly steady vs injury level (green curve). **B:** Injury to PNs was more traumatic: Post-injury EN response (orange) was lower, and trained responses (blue) fully regained pre-injury levels if injury was on average $\leq 8\%$. Also, the ability of training to recover lost ground decreased as injury level increased (green curve). Each datapoint show the mean and std dev, over 60 moths, of mean EN odor response. 74

4.5 **Effects of parallel inhibitory channels. A:** EN odor response post-injury normalized by naive, healthy odor responses vs injury level. Each curve corresponds to a number of QNs per 5 PNs, from 0 to 7. Higher QN:PN ratios resulted in much lower impact on EN responses for a given level of injury. **B:** EN odor responses post-training normalized by naive, healthy odor responses vs injury level. Each curve corresponds to a number of QNs per 5 PNs, from 0 to 7. Higher QN:PN ratios resulted in stronger recovery. **C:** Ratio of post-training to post-injury EN odor responses vs injury level. Recovery rate dropped off at injury levels $\geq 20\%$ for #QN = 0, but higher numbers of QNs reduced this drop-off, ie ensured better recovery. **D:** Naive ratio of EN odor response to spontaneous EN noise (SSNR), a measure of signal clarity, was much lower in moths with high QN counts. **E:** Raw Signal-to-Noise Ratio (SNR) of naive, healthy EN responses were fairly uniform across #QNs. **F:** Post-injury SNR normalized by pre-injury SNR. High QN counts gave strong protection against injury-induced degradation of SNR. 78

4.6 **Effect of AL noise:** AL noise protects downstream neurons from loss (A, B), but exacts a cost in terms of signal-to-spontaneous noise ratio (C, D) and SNR (E). **A:** Normalized EN odor response post-injury vs injury level. y-axis = post-injury $\frac{\mu(F)+\sigma(F)}{\mu(F_h)+\sigma(F_h)}$, as a proxy for the highest EN responses of a moth to a series of odor puffs. Higher AL noise resulted in stronger top EN responses post-injury, at any level of injury. **B:** Normalized EN odor response post-training vs injury level. y-axis as in (A). Each curve corresponds to a level of AL noise, from 0 to 1.33. Higher AL noise allowed training to give full recovery of top EN responses from larger injuries. Pre-injury response = black line. **C:** Healthy ratio of EN signal-to-spontaneous noise ratio (SSNR) was much lower at high AL noise levels. **D:** Post-injury SSNR, normalized by pre-injury ratios, vs injury level. In high AL noise moths, injury lowered SSNR far more. **E:** Pre-injury SNR $\frac{\mu(F)}{\mu(s)}$ by AL noise level. SNR was much lower in moths with high-noise ALs. 79

4.7 **Ablation does not map to FAS injury.** Ablation and FAS injury effects had highly variable relationships. In theory, 1 unit Ablation ∼1.85 units FAS injury. In practice: **A:** When RNs were injured, ablation induced a ∼50% bigger loss to EN response than expected. **B:** When PNs were injured, ablation induced a ∼50% smaller loss than expected. 80

4.8 **Injury mitigation hypotheses:** In a cascaded network, various architectures can mitigate the effects of injury to upstream neurons by protecting or restoring functionality of downstream units.
A: (Finding 1) Learning itself can compensate for injury. Octopamine temporarily stimulates the damaged neuron, allowing Hebbian growth to strengthen downstream synaptic connections. Though the injured neuron's signal is not restored, the downstream neurons receive an amplified input, cancelling out the injury.
B: (Finding 2) Parallel inhibitory channels can reduce the effect of generalized injury by spreading damage among excitatory and inhibitory signals, so that losses cancel out in terms of inputs to downstream neurons.
C: (Finding 3) Wide noise envelopes on upstream neuron outputs can protect the strongest stimulus responses from injury-induced attenuation δ, to the degree that their std dev $\sigma > \delta$. This allows the injured neuron's strongest responses to still exceed their activation threshold (green line) for downstream neurons, protecting downstream functionality.
D: (Finding 4) Two simple non-linearities that can result in qualititative change in the relative effects of ablation and FAS injury. In an AND gate, ablation can have worse effect than FAS downstream, depending on the gate's input threshold T. In an OR gate, ablation can be harmless, while FAS can have worse effect downstream, depending on T. 83

5.1 **p-values for trained-control odor pairs: A:** p-values for change in raw EN responses. **B:** p-values for percentage change in EN responses. P-values are sometimes high (for one metric or the other) when trained and control odors have highly disparately-scaled naive responses μ_T (= mean raw T) μ_C (= mean raw C). Plots show results given 20 training puffs.
When μ_T is larger (right end of x-axis), the p-value for raw change (A) is consistently very low, but the p-value for percentage change (rB) can be high, since even a small incidental change to a low-intensity odor can be a large percentage change.
When μ_C is larger (left end of x-axis), the p-value for percentage change (B) is consistently very low, but the p-value for raw change (A) can be high, since even a small percentage change to a high-response odor corresponds to a large raw change.
When naive odor responses are roughly matched, eg within 3x (ie 0.33 to 3), p-values for both raw and percentage change are very low. 92

5.2 **Fractions of p-values below 0.01 for trained-control odor pairs** In most cases, the trained odor shows much larger increases in EN response magnitude. **A:** The percentage of trained-control odor pairs with EN response magnitudes within the ratios given on the x-axis. **B:** The percentage of trained-control pairs, with EN response magnitudes within the ratios given on the x-axis, whose training-induced changes in EN responses were distinct with p-value < 0.01. Each curve is for a different number of training puffs. More training increases distinctions, up to 15 puffs. But additional training actually hinders distinctions, as control odor response reinforcement begins to overtake trained odor reinforcement. 93

5.3 **Effect of AL noise on strongest vs average EN responses** High AL noise had a greater protective effect on the top 15% tranche of EN odor responses than on all odor responses. **A:** Normalized post-injury (red, grey) and post-training (blue, grey) EN odor responses $\frac{\mu(F)+\sigma(F)}{\mu(F_h)+\sigma(F_h)}$. **B:** The same data, but plotting the normalized mean of EN responses $\frac{\mu(F)}{\mu(F_h)}$. Injury mitigation was weaker for mean responses. That is, high EN responses received more protection than low or average EN responses. 94

ACKNOWLEDGMENTS

Deep thanks to my advisor Nathan Kutz, to co-advisor Eve Riskin, and to biologist Jeff Riffell. Many thanks also to committee members Hannaneh Hajishirzi, Sreeram Kannan, and Don Percival. Also thanks to IV Lab for funding, to the Dept of Electrical Engineering, and to Pedro Maia. And of course measureless thanks to Jean T. Olson.

DEDICATION

to my lovely, patient sweetie Jean

Chapter 1

THESIS INTRODUCTION

This thesis explores mechanisms of Learning in biological neural networks, in particular the moth olfactory network (MON). It has three goals: First, to study how the architecture of the moth's olfactory system enables it to learn; second, to port the moth's biological "bag of tricks"to machine learning contexts; third, to examine how Learning (and other neural architectures) act as mechanisms to mitigate neural injury.

The MON is one of the simplest biological neural systems capable of Learning, yet it contains several key structural motifs widespread in biological neural systems and of great interest. These include cascading networks, large dimension shifts from stage to stage, high-dimensional sparse codings of data, randomness, and octopamine stimulation plus Hebbian plasticity as vital parts of the learning mechanism.

We build a computational model (hereafter "MothNet") of the moth olfactory network that closely matches known physiology and wet-lab data, then use this model to run a variety of simulations and experiments. The simulations allow us to analyse how learning occurs in a biological neural net (NN). To our knowledge this is the first full, end-to-end neural network model that demonstrates learning behavior while also tightly matching the structure and behavior of a real biological system.

While the motifs listed above are widespread in biological neural systems, they are largely absent from artificial neural nets (ANN). Thus, we seek to characterize a set of biological elements, a "biological toolkit", that can be assembled into NNs that operate on fundamentally different principles from standard ANNs. *In silico* experiments with MothNet allow us to abstract out critical features in the moth's toolkit that allow it to learn. We then demonstrate the viability of this toolkit in the context of Machine Learning (ML), by teach-

ing MothNet to read hand-written digits, a classic ML task. This transfer of biological tools to ANNs is potentially valuable because biological NNs can do things which ANNs cannot, such as effectively learn from only a few training samples.

In addition, we explore the hypothesis that Learning, and other neural architectures, can be viewed as injury mitigation mechanisms which enable a network to maintain function despite damage. Our results offer evidence that robustness to injury is a core design specification in biological neural systems.

Chapter 2

LEARNING MECHANISMS IN THE MOTH OLFACTORY NETWORK

2.1 Introduction

Learning is a vital function of biological neural networks, yet the underlying biomechanical mechanisms responsible for robust and rapid learning are not well understood. The insect olfactory network, and the moth's olfactory network in particular (e.g. the *Manduca sexta* moth), is one of the simplest biological neural networks that is capable of learning [69]. Additionally, the moth's olfactory processing is amenable to interrogation through experimental neural recordings of key, well-understood structural components that include the antenna lobe (AL) and mushroom body (MB). Thus the moth provides an ideal model organism for characterizing the mechanics of learning, especially as the AL-MB contain many structural motifs that are widespread in biological neural systems. These motifs include: (i) the use of octopamine and dopamine in learning, (ii) a cascading network structure, in this case feed-forward, (iii) large changes in dimensionality (i.e. number of neurons) between networks, (iv) sparse encodings of data in high-dimensional networks, (v) random connections, (vi) the presence of noisy signals, and (vii) Hebbian ("fire-together, wire-together") plasticity.

This chapter describes a computational model (MothNet) that is closely tethered to the known biophysics of the AL-MB interaction, includes these key motifs, and also includes the effects of octopamine stimulation. MothNet demonstrates how in combination these components produce robust and rapid learning. A model schematic is given in Fig 2.1. This gives key biological insights into learning and bio-inspired design principles for neuronal networks more broadly.

Bio-inspired design principles suggest that each of the features mentioned in the preceding

paragraph has high value to the olfactory system. The mechanism of octopamine/dopamine release during learning is of particular interest, both biologically and in the context of machine learning (ML), since it is not well-understood how this stimulation promotes the construction of new sparse codes in the MB. The AL-MB interaction with octopamine and Hebbian plasticity operates in a fundamentally different manner than optimization methods of artificial neural nets (ANNs) such as back-propagation, and thus provides a novel approach to training NNs.

Our goals are to study how these features enable learning in a NN and to derive bio-inspired insight into the mathematical framework that enables rapid and robust learning. Our computational model of the AL-MB dynamics integrates known biological constraints and faithfully reconstructs the dynamics and statistics of neuronal recordings to date. This approach has three advantages: (i) we can meaningfully compare model simulation output to experimental data in order to tune our model parameters, (ii) our results can map back to the original biological system in order to render meaningful biological insights, and (iii) we can characterize the relevance of each structural and dynamical feature of the system to the task of learning in a neural net.

2.2 Methods

In this section, we describe the biological moth olfactory network, as well as the MothNet model. We also provide a Glossary, and describe the wet-lab data used for model calibration.

2.2.1 Moth olfactory system overview

The parts of the AL-MB implicated in learning are organized as a feed-forward cascade of five distinct networks, as well as a reward mechanism [56, 41]. Figure 2.1 gives a system schematic along with typical firing rate (FR) timecourses (from simulation) for neurons in each network.

1. Antennae. Roughly 30,000 noisy chemical receptors detect odor and send signals to

the Antenna Lobe [57].

2. Antenna Lobe (AL). Contains roughly 60 units (glomeruli), each focused on a single odor feature [56]. The AL essentially acts as a pre-amp, boosting faint signals and denoising the antennae inputs [8]. AL neurons are noisy [21].

3. Lateral Horn (LH). Though not fully understood, one key function is global inhibition of the Mushroom Body to enforce sparseness [6].

4. Mushroom Body (MB), here synonymous with the Kenyon Cells (KCs). About 4000 KCs are located in the calyx of the Mushroom Body (MB). These fire sparsely and encode odor signatures [10, 35].

5. Extrinsic Neurons (ENs), numbering $\sim$10's, located downstream from the KCs. These are believed to be "readout neurons" that interpret the KC codes and convey actionable messages (such as "fly upwind") [9, 32].

6. Reward Mechanism. A large neuron sprays octopamine globally over the AL and MB in response to reward, such as sugar at the proboscis. Learning does not occur without this octopamine input [27, 29].

7. Inter-network connections: In the AL-MB these are strictly feed-forward, either excitatory or inhibitory. In particular, Antennae$\rightarrow$AL, AL$\rightarrow$LH, KCs$\rightarrow$ENs are all excitatory. LH$\rightarrow$KCs is inhibitory. AL$\rightarrow$KCs have both excitatory and inhibitory channels.

8. Plasticity: The connections into the KCs (AL$\rightarrow$KCs) and out of the KCs (KCs$\rightarrow$ENs) are known to be plastic during learning [14, 57]. The AL is not plastic.

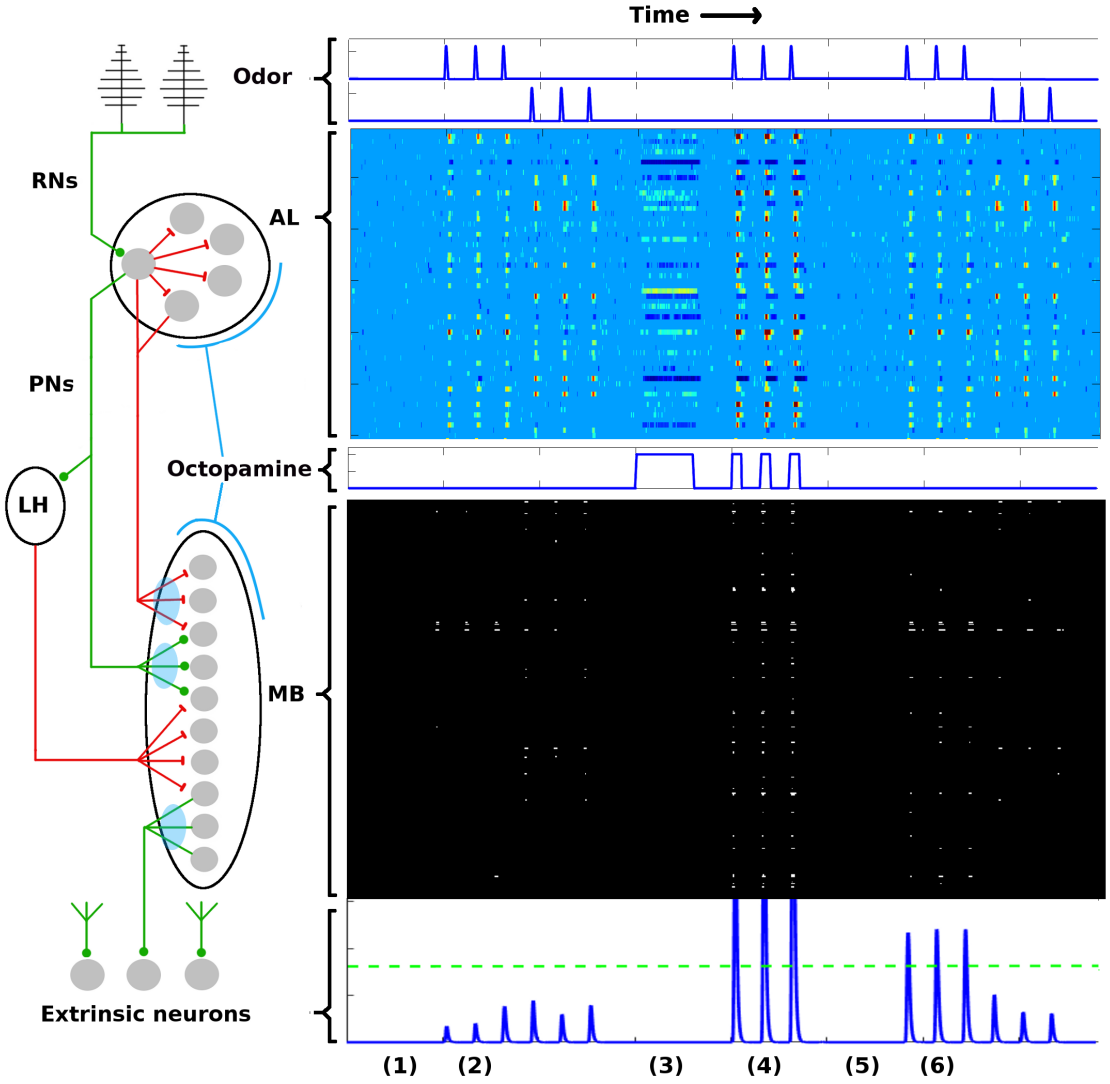

Figure 2.1: **AL-MB overview. A:** System schematic: Chemical sensors (RNs) excite a noisy pre-amp network (AL), which feeds forward to a plastic sparse memory layer (MB), which excites readout (decision) neurons (ENs). Green lines show excitatory connections, red lines show inhibitory connections (LH inhibition of the MB is global). Light blue ovals show plastic synaptic connections into and out of the MB. **B:** Neuron timecourse outputs from each network (typical simulation) with time axes aligned vertically. Timecourses are aligned horizontally with their regions-of-origin in the schematic. The AL timecourse shows all responses within $\pm$ 2.5 std dev of mean spontaneous rate as medium blue. Responses outside this envelope are yellow-red (excited) or dark blue (inhibited). MB responses are shown as binary (active/silent). Timecourse events are as follows: (1) A period of no stimulus. All regions are silent. (2) Two odor stimuli are delivered, 3 puffs each. AL, MB, and ENs display odor-specific responses. (3) A period of control octopamine, ie without odor or Hebbian training. AL response is varied, MB and EN are silent. (4) The system is trained (octopamine injected) on the first odor. All regions respond strongly. (5) A period of no stimulus. All regions are silent, as in (1). (6) The stimuli are re-applied. The AL returns to its pre-trained activity (it is not plastic). In contrast, the MB and EN are now more responsive to the trained odor, (response to the untrained odor is unchanged). Green dotted line in the EN represents a hypothetical "action" threshold. The moth has learned to respond to the trained odor.

2.2.2 Glossary

Antenna lobe (AL): A collection of neurons innervated by odor receptors in the antennae. It sends signals to the mushroom body via projection neurons. Connections in the AL are not plastic.

Mushroom body (MB): A collection of neurons (Kenyon cells - KCs) downstream from the antenna lobe. The MB is believed to store odor codes that serve as a memory, allowing the moth to recognize odors. Connections in the MB are plastic.

Lateral horn (LH): A collection of neurons which receives input from the AL and sends inhibitory output to the MB. One of its roles is to enforce sparse firing in MB neurons.

Receptor neuron (RN): These neurons respond to odors (volatiles) at the antennae and stimulate the antenna lobe. RNs respond to different, distinct odors.

Glomerulus: The antenna lobe is divided into about 60 glomeruli, each of which is a self-contained collection of neurons (projection and lateral), innervated by RNs that respond to particular odors.

Projection neuron (PN): Each glomerulus contains projection neurons, whose output innervates the KCs and also the lateral horn, but not other glomeruli in the AL, ie they are feed-forward only. Most PNs start in one glomerulus and are excitatory. A few PNs arborize in several glomeruli and are inhibitory (we refer to inhibitory PNs as "QNs"). Each glomerulus initiates about five PNs.

Lateral neuron (LN): Each glomerulus contains lateral neurons, which innervate other glomeruli in the AL. LNs are inhibitory. One function is competitive inhibition among glomeruli. An-

other function is gain control, ie boosting low signals and damping high signals.

Kenyon cell (KC): Neurons in the calyx of the MB. These have very low FRs, and tend to respond to particular combinations of PNs. KCs respond sparsely to a given odor. There are about 4000 KCs, ie a two-orders-of-magnitude increase over the number of glomeruli. Each KC synapses with about ten PNs. Connections into and out of KCs are plastic.

Extrinsic neuron (EN): A small number of neurons downstream from the KCs. ENs are thought to be "readout" neurons. They interpret the odor codes of the KCs, deciding to eg "ignore", "approach", or "avoid".

Firing rate (FR): The number of spikes/second at which a neuron fires. Typically FRs are counted using a window (eg 500 mSec). The moth's response to odor puffs is episodic, with FR spikes in FR and rapid return to spontaneous FRs. Neurons respond to relative changes in FR, rather than to raw magnitude changes. A neuron's relative change in FR is scaled by its spontaneous FR (see section 2.2.5 below).

Octopamine: A neuromodulator which stimulates neural firing. The moth spritzes octopamine on both the AL and MB in response to sugar, as a feedback reward mechanism. Dopamine has a similar stimulating effect on both AL and MB, but it reinforces adverse rather than positive events.

2.2.3 Component networks and their MothNet representations

This subsection offers a more detailed discussion of the constituent networks in the biological AL-MB, and details about how they are modeled in MothNet.

Antennae and receptor neurons

The Antennae receptors, activated by chemical molecules in the air, send excitatory signals to Receptor Neurons (RNs) in the AL. Several thousand antennae converge onto 60 units (glomeruli) in the AL [60]. All the receptors for a given atomic volatile converge onto the same glomerulus in the AL, so the glomeruli each have distinct odor response profiles [17]. Since natural odors are a blend of atomic volatiles, a natural odor stimulates several units within the AL [68].

MothNet does not explicitly include antennae. Rather, the first layer of the model consists of the RNs entering the glomeruli. Though ∼500 RNs feed a given glomerulus, the model assumes one RN. The benefit of many RNs converging appears to be noise reduction through averaging [63]. This can be simulated by one RN with a smaller noise envelope. In MothNet, each glomerulus' RN has a spontaneous FR and is excited, according to random weights, by odor stimuli.

Antenna lobe and projection neurons

The AL is fairly well characterized in both structure and dynamics, with a few important gaps. Moths and flies are similar enough that findings in flies (*Drosophila*) can generally be transferred to the moth (in contrast, locusts and honeybees are more complex and findings in these insects do not safely transfer) [70].

The AL contains about 60 glomeruli, each a distinct unit which receives RN input and projects to the KCs via excitatory PNs. The same PN signal also projects to the LH [6]. The AL, unique among the networks, has inhibitory lateral neurons (LNs) [85], the only neurons that are not strictly feed-forward. (There is some evidence of excitatory LNs, eg [62]; MothNet excludes this possibility.) The LNs act as a gain control on the AL, and also allow odors to mask each other by inhibiting other glomeruli's RNs [64, 36]. It is not known whether LNs also inhibit PNs and LNs. Based on calibrations to wet-lab data, in MothNet LNs inhibit all neuron types (cf section 2.4.1). Thus each glomerulus contains dendrites (ie

outputs) for PNs and LNs, and axons (ie inputs) from RNs and LNs, as shown in Figure 2.2.

Each glomerulus does the following: Receives RN input from the antennae receptors upstream; inhibits other glomeruli within the AL via LNs; and sends excitatory signals downstream via Projection Neurons (PNs).

In general, each PN is innervated in a single glomerulus. In moths, there are ∼5 PNs rooted in each glomerulus (60 glomeruli, ∼300 PNs). MothNet assumes all PNs from a given glomerulus carry the same signal (because they share the same glomerulus and therefore inputs, and perhaps also because of ephaptic binding) [76].

Glomeruli also initiate pooled Inhibitory Projection Neurons (QNs) that send inhibitory signals downstream to the KCs.

The AL contains a powerful macro-glomerulal complex (MGC), which processes pheromone. Because pheromone response has fundamentally different dynamics than food odor response [40], MothNet ignores it. Only the glomeruli associated with non-pheromone (food) odors are modeled.

Connections in the AL are not plastic with long-term persistence [15]. While some evidence of short-term plasticity exists, MothNet ignores this option.

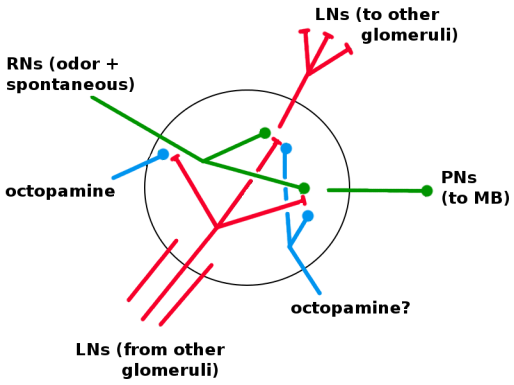

Figure 2.2: **Schematic of an AL glomerulus.** Detail of neural connections within a glomerulus. Red = inhibitory, green = excitatory, blue = increases responsiveness. RNs enter from the antennae. LNs enter from other glomeruli; one full LN is shown. It is not known if octopamine modulates LNs and PNs (see section 2.4.1).

Lateral horn

The LH receives input from the PNs. It then sends an inhibitory signal to the KCs. This inhibition from the LH appears to ensure that the KCs fire very sparsely and thus act as coincidence detectors for signals from the AL [76, 48].

The LH is also suspected of containing a parallel system for processing certain intrinsically-known odors in short-cut fashion (labeled lines) [50]. Since this parallel system is (by definition) not involved with learning, MothNet ignores it. The LH is modeled solely as a simple sparsifying inhibition on the KCs.

(Note: The locust and honeybee, which have more complex olfactory systems and different use-cases in terms of odor processing, have a time-oscillating depolarization mechanism (local potential fields, LPF) which serves a similar purpose to LH inhibition in the moth. LPF oscillations are absent in the moth [56].)

Mushroom body and Kenyon cells

The KCs ($\sim$4000) in the MB are believed to encode odor memories in a high-dimensional, sparse space [80]. Odors with no meaning to the moth still have non-zero codes in the KCs.

KCs receive excitatory input from the PNs and inhibitory input from QNs, both of which vary greatly between KCs, since each KC is innervated by only $\sim$10 PNs [56]. The connection map appears to be random [13]. The KCs also receive generalized damping inhibition from the LH. (There is some evidence in Drosophila of an MB$\rightarrow$MB global inhibitory neuron [48], with the same essential effect as LH inhibition; MothNet ignores this possibility.) KCs fire very sparsely, generally respond to only a single odor, and are silent absent that odor [35]. KCs are treated as noise-free. Their output is an excitatory signal sent to the extrinsic neurons (ENs) [9].

In addition to olfactory input, the KCs receive input signals from other parts of the moth (eg hearing) [76]. Because MothNet targets olfactory learning, it ignores these other inputs and uses a reduced number of KCs ($\sim$2000 instead of $\sim$4000).

The synaptic connections in the MB (PNs→KCs, QNs→KCs, and KCs→ENs) are plastic, ie they can be modified during training [58]. The generalized inhibition from LH→KCs is modeled as non-plastic (actual physiology is not known).

Extrinsic neurons

Though located in the lobes of the MB, here ENs are not considered part of the MB, which is taken to be synonymous with the KCs. ENs are few in number compared to the KCs (∼10s) [32]. They are believed to be "readout" neurons, that interpret the KC codes as actionable signals (eg "approach", "avoid") [57]. We assume that ENs trigger actions when their output FRs exceed some threshold.

We define Learning as: Permanently boosting EN responses beyond their naive (untrained) level, so that EN responses to reinforced stimuli can consistently exceed an action-triggering threshold. This is tantamount to modifying the moth's behavior.

Octopamine (reward circuit)

A large neuron delivers octopamine to the entire AL and MB, in response to positive stimuli, eg sugar at the proboscis. It acts as a reward feedback to the system. A similar neuron delivers dopamine to the AL and MB in response to negative stimuli, and acts as an aversive feedback signal [15]. Learning does not occur without octopamine (dopamine) [29].

Despite their opposite reward values, both octopamine and dopamine act in the same way when sprayed on a neuron: They increase the neuron's general tendency to fire [69]. In MothNet this effect is modeled as making a neuron more responsive to excitatory inputs (eg from odors and RNs) and less responsive to inhibitory inputs (eg from LNs). Details of octopamine's effects, if any, on various neural types are not well-characterized. In MothNet octopamine directly affects RNs and LNs but not PNs in the AL (cf section 2.4.1); has no direct effect on KCs or ENs (though there are strong indirect effects); and has no effect on the LH inhibitory signal.

It is unclear whether octopamine delivery to both the MB and AL is necessary and sufficient for learning [29, 15]. MothNet assumes that octopamine controls an "on/off" switch for Hebbian growth, ie there is no plasticity in the MB (and therefore no learning) without octopamine.

2.2.4 MothNet model description

This section describes the MothNet model (as used in this chapter) in detail. It covers the firing rate measure used to compare model output to wet-lab data; model dynamics; plasticity and other details; model parameters; and moth generation. All coding was done in Matlab.

2.2.5 Firing rate measure

To compare PN firing rate statistics from wet-lab experiments and MothNet simulations (ie model calibration), we use a measure of firing rate (FR) based on Mahalanobis distance, similar to the measure $\frac{DF}{F}$ common in the literature [9, 36, 80, 75]. The premise is that neurons downstream respond to a +1 std change in FRs equally (modulo different connection weights), independent of the sometimes large (up to 40x) magnitude differences in the raw spontaneous FRs of different neurons. The PN firing rate measure is defined as follows:

1. Each PN has a spontaneous firing rate (FR) with a gaussian noise envelope.

2. PNs with FR < 1 spike/sec are ignored, on the assumption that such PNs represent artifacts of experiment (also, the gaussian noise assumption fails). About 10% of PNs in experimental data fall in this category.

3. Output FR activity of PNs is measured as $M(t)$ = distance from mean spontaneous FR, in units of time-varying std dev of spontaneous FR (ie Mahalanobis distance): Let
   $F(t)$ = raw firing rate (spikes per second).
   $S(t)$ = spontaneous firing rate (no odor).

$\mu S(t)$ = moving average of S (no odor).

$\bar{\mu}S(t)$ = smoothed estimate of the moving average μS, eg a quadratic or spline fit.

$\sigma_S(t)$ = standard deviation of S, calculated using $S - \bar{\mu}S$ values within a moving window centered on t.

$\sigma_S(t)$ and $\mu S(t)$ are typically steady absent octopamine, but are often strongly modulated by octopamine.

Then the measure of FR activity M is:

$$M(t) = \frac{F(t) - \bar{\mu}S(t)}{\sigma_S(t)} \tag{2.1}$$

4. M is related to the measure $\frac{DF}{F}$:
   $\frac{DF}{F} = \frac{\Delta F}{F} = \frac{F(t)-\mu S}{\mu S}$, ie $\frac{DF}{F}$ is change in FR, normalized by spontaneous FR. The key difference between M and $\frac{DF}{F}$ is whether or how σ_S is estimated, due to varying exigencies of experiment. Our experimental data allow reasonable estimates of σ_S and μS. MothNet simulations produce very good estimates, since computers are more amenable to repeated trials than live moths.

2.2.6 Model dynamics

MothNet uses standard integrate-and-fire dynamics [16], evolved as stochastic differential equations [34].

Let $x(t)$ = firing rate (FR) for a neuron. Then

$$\tau\frac{dx}{dt} = -x + s(\Sigma \mathbf{w}_i \mathbf{u}_i) = -x + s(\mathbf{w} \cdot \mathbf{u}), \text{ where} \tag{2.2}$$

$\mathbf{w}$ = connection weights;

$\mathbf{u}$ = upstream neuron FRs;

$s()$ is a sigmoid function or similar.

PN dynamics are given here as an example. Full model dynamics are given in SI. PNs are excitatory, and project forward from AL→MB:

$$\tau \frac{d\mathbf{P}}{dt} = -\mathbf{P} + s(\widetilde{\mathbf{P}}) + \mathbf{dW}^P \text{ where} \tag{2.3}$$

$\mathbf{W}(t)$ = brownian motion process;
$\widetilde{\mathbf{P}} = -(1 - \gamma o(t) M^{O,P}) * M^{L,P} * \mathbf{u}^L + (1 + o(t) M^{O,P}) * M^{R,P} * \mathbf{u}^R$;
$M^{O,P}$ = octopamine→PN weight matrix (diagonal $nG \times nG$);
$M^{L,P}$ = LN→PN weight matrix ($nG \times nG$ with $trM^{L,P} = 0$);
$M^{R,P}$ = RN→PN weight matrix (diagonal $nG \times nG$);
$o(t)$ indicates if octopamine is active ($o(t) = 1$ during training, 0 otherwise).
$\mathbf{u}^L$ = LN FRs, vector $nG \times 1$;
$\mathbf{u}^R$ = RN FRs ($nG \times 1$);
γ = scaling factor for octopamine effects on inhibition.

Discretization

The discretization uses Euler-Maruyama, a standard step-forward method for SDEs [34].
Euler (ie noise-free): $x_{n+1} = x_n + \Delta t f(x_n)$
Euler-Maruyama: $x_{n+1} = x_n + \Delta t f(x_n) + \epsilon \text{ randn}(0,1)\sqrt{\Delta t}$, where ϵ controls the noise intensity.

Convergence

Timestep Δt was chosen such that noise-free E-M evolution gives the same timecourses as Runge-Kutta (4th order), via Matlab's ode45 function. $\Delta t = 10$ mSec suffices to match E-M evolution to R-K in noise-free moths. Values of $\Delta t \leq 20$ mSec gives equivalent simulations in moths with AL noise calibrated to match wet-lab data. Values of $\Delta t \geq 40$ mSec show differences in evolution outcomes given AL noise.

Plasticity

The model assumes a Hebbian mechanism for growth in synaptic connection weights [30, 14]. That is, the synaptic weight w_{ab} between two neurons a and b increases proportionally to the product of their firing rates ("fire together, wire together"): $\Delta w_{ab}(t) \propto f_a(t) f_b(t)$.
Thus, synaptic plasticity is defined by:

$$\Delta w_{ab}(t) = \gamma f_a(t) f_b(t), \text{ where } \gamma \text{ is a growth parameter.} \tag{2.4}$$

There are two layers of plastic synaptic weights, pre- and post-MB: AL→MB ($M^{P,K}$, $M^{Q,K}$), and MB→ENs ($M^{K,E}$) . Learning rate parameters of MothNet were calibrated to match experimental effects of octopamine on PN firing rates and known moth learning speed (eg 5 - 10 trials to induce behavior modification) [69]. MothNet does not decay unused synaptic weights. Training does not alter octopamine delivery strength matrices ($M^{O,*}$). That is, the neuromodulator channels are not plastic (unlike, for example, the case in [23]).

Odor and octopamine injections

Odors and octopamine are modeled as Hamming windows. The smooth leading and trailing edges ensures low stiffness of the dynamic ODEs, and allows a 10 mSec timestep to give accurate evolution of the SDEs in simulations.

Training

Training on an odor consists of simultaneously applying puffs of the odor, injecting octopamine, and "switching on" Hebbian growth. Training with 5 to 10 odor puffs typically produces behavior change in live moths.

2.2.7 Model parameters

There is a risk, when modeling a system, of adding too many free parameters in an effort to fit the system. Fewer free parameters are better, for the sake of generality and to avoid

overfitting. Conversely, we wish to reasonably match the physiological realities of the system. Because the key goal of this paper was to demonstrate that a simple model, in terms of parameters and structure, can reproduce the learning behavior of the AL-MB, we made efforts to minimize the number of free parameters. For example, neuron-to-neuron connections in the model are defined by their distributions, ie two parameters each. These are (usually) distinct for different source-to-target pairs (eg LN→RN, LN→LN, etc).
Some mean and std dev parameters for distributions are shared among different neuron types.

Parameter list:

1. Structure: 5 (eg number of neurons in each network)
2. Dynamics: 12 (noise: 2. decay tau, and sigmoid: 3. Hebbian growth: 6. misc: 1).
3. Spontaneous RN FRs: 3.
4. Connection matrices: 27 (to control non-zero connection ratios, 5; synaptic weights (eg $M^{P,K}, M^{R,P}$) means, 12, std devs, 4; octopamine weights (eg $M^{O,R}, M^{O,P}$) means, 6, std devs, 2).

Total free params: 47

Dynamics parameters

The differential equations of all neuron types share the same decay rate, set to allow return to equilibrium in ∼1 second, consistent with wet-lab data. Neurons also share parameters of the sigmoid function within the differential equation. Noise added via the SDE model is controlled by a single parameter ϵ, the same for all neuron types. It is determined by empirical constraint on $\frac{\sigma_S}{\mu S}$, as shown in column 2 of Figure 2.4.

Connection matrix generation

Connection weight matrices (eg $M^{P,K}$ etc) are generated in a standard way, from Gaussian distributions with std dev σ defined proportional to the mean μ, using a scaling factor v:

$M^{*,*} \sim N(\mu_c, \sigma_c^2)$ where μ_c depends on the neuron types being connected, and $\sigma_c = v\mu_c$. Many connection types typically share the same v.

A special feature of the AL is that all the neurons in a given glomerulus share a common environment. For example, all the neurons, of whatever type, in glomerulus A will share the same strong (or weak) LN axon from glomerulus B. Thus, the RN, LN, and PNs in a given glomerulus are all correlated. In addition, neuron types are correlated. To model this dual set of correlations, connection matrices in the AL are generated as follows. As an example, consider LN connection matrices in the AL:

1. A glomerulus-glomerulus connection matrix $M^{L,G}$ is created, which defines LN arborization at the glomerular level.

2. This connection matrix is multiplied by a neural type-specific value to give $M^{L,P}$,$M^{L,L}$, and $M^{L,R}$ connection matrices. This is particularly important when tuning the various inhibitory effects of LNs on RNs, PNs (QNs), and LNs.

3. Sensitivity to GABA: A separate variance factor determines glomerular sensitivity to GABA (ie sensitivity to inhibition). This is tuned to match data in the literature [36], and applies to LN-to-PN(QN) (ie $M^{L,P}$) connections only.

The goal of this two-stage approach is to enforce two types of similarity found in the AL: (i) Connections to all neurons within a single glomerulus are correlated; and (ii) connections to all neurons of a certain type (LN, PN, RN) are correlated.

Due to constraints of the biological architecture there are many zero connections. For example, about 85% of entries in the AL→MB weight matrix are zero because MB neurons connect to only ∼10 projection neurons [12]. All MB→EN weights are set equal at the start of training. Training leads rapidly to non-uniform distributions.

RN spontaneous firing rates

RNs in the glomeruli of the AL have noisy spontaneous firing rates [8]. MothNet simulates this by assigning spontaneous firing rates to RNs. These spontaneous firing rates are drawn from a gamma distribution plus a bias:
$\gamma(x|\alpha,\beta,b) = b + \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$, where α,β are shape and rate parameters, and $\Gamma(\cdot)$ is the Gamma function.
This can be thought of as a source of energy injected into the system, at the furthest upstream point (absent odor). Other energy sources are odor signals and octopamine. The spontaneous firing rates of all other neurons in MothNet are the result of their integrate-and-fire dynamics responding as RN spontaneous FRs propagate through the system.

2.2.8 Discrepancies between biology and model

There are some known discrepancies between the MothNet model and the moth AL-MB. These are listed below.

Connection weight distributions

This model version uses gaussian distributions to generate initial connection weights. However, moths used in live experiments are older and thus presumably have modified PN→KC and KC→EN connection weights. If this modification was strong, we might expect the connection weight distributions to tend towards a scale-free rather than gaussian distribution [5]. This represents an unknown discrepancy between structure parameters of the live moths used in experiments vs the model.

Hebbian pruning

This chapter's MothNet version contains no pruning mechanism to offset, via decay, the Hebbian growth mechanism. Such pruning mechanisms are common in nature, so it is reasonable to suppose that one might exist in the AL-MB. The moth has inhibitory as well

as excitatory feed-forward connections from AL to MB. In this version of MothNet, pruning is functionally replaced by Hebbian growth of QN→KC inhibitory connections, which act to inhibit KCs and thus offset the growth of excitatory PN→KC connections (this does not directly offset KC→EN Hebbian growth). Thus omitting a separate Hebbian decay mechanism is a matter of convenience rather than a match to known biology. (Weight decay is introduced in chapter 2).

Non-olfactory input to KCs:

In addition to olfactory input, the KCs receive signals from other parts of the moth, eg hearing. Because this model targets only olfactory learning, it ignores these other inputs to the KCs, and reduces the total number of KCs (from ∼4000 to ∼2000).

Number of QNs

There are believed to be about 3-6 QNs projecting from the AL to the MB. This model sets their number at about 15. The reason is that, absent a Hebbian pruning system in the model, the QNs function as the brake on runaway increases in KC responses due to Hebbian growth. So the increased number of QNs is a compensation for the lack of a weight-decay system.

Number of ENs

This model version has only one EN, since its goal is to demonstrate simple learning. The moth itself possesses multiple ENs.

LH inhibition

The LH→KC inhibitory mechanism used in this chapter is modeled as a time-invariant global signal, delivered equally to all KCs. This simplifies the model parameter space while retaining the essential functionality of the LH. A more refined version of LH→KC inhibition

might vary in strength according to PN output, since the same PN signals that excite the KCs also excite the LH. The actual dynamics of the AL→LH→KC linkage are not known, beyond the principle that inhibition from the LH sparsifies the KC codes and makes the individual KCs act as coincidence detectors.

2.2.9 Wet-lab moth electrode data

Model parameters were calibrated by matching MothNet performance to electrode readings from the ALs of live moths. The various performance metrics are described in section 2.3.1.

Electrode data was collected by the lab of Prof Jeff Riffell (Dept of Biology, UW). It consists of timecourses of PN firing rates measured via electrode in the AL of live moths, during a variety of regimes including:

1. Series of 0.2 sec odor puffs delivered without octopamine. These experiments gave data re PN response to odor relative to PN spontaneous (baseline) FRs, absent octopamine.

2. Series of 0.2 sec odor puffs delivered coincident with sugar reward (which delivers octopamine). This gave data re how PN odor response is modulated by octopamine, relative to octopamine-free spontaneous FR. See Figure 2.3 panel A.

3. Series of 0.2 sec odor puffs, delivered first without and then coincident with an octopamine wash applied to the AL. This gave data re how PN spontaneous FR and PN odor response are modulated by octopamine. See Figure 2.3 panel B.

The applied odor consisted of a collection of 5 volatiles, which taken together stimulate many glomeruli in the AL. It was selected to ensure sufficient odor-responsive PNs, such that inserted electrodes would detect interesting (ie responsive) PNs. Further details re wet-lab data collection can be found in [74]. Example timecourses are shown in Figure 2.3.

2.2.10 Simulation setup

For learning experiments, the time sequence of events for simulations, shown in Fig 2.1, is as follows:

1. A period of no stimulus, to assess baseline spontaneous behavior.

2. Four odor stimuli are delivered, 16 puffs each.

3. A period of control octopamine, ie without odor or Hebbian training.

4. The system is trained (odor + octopamine + Hebbian mechanism) on one of the odors.

5. A period of no stimulus, to assess post-training spontaneous behavior.

6. The odors are re-applied (16 puffs each), without octopamine, to assess effects of training on odor response.

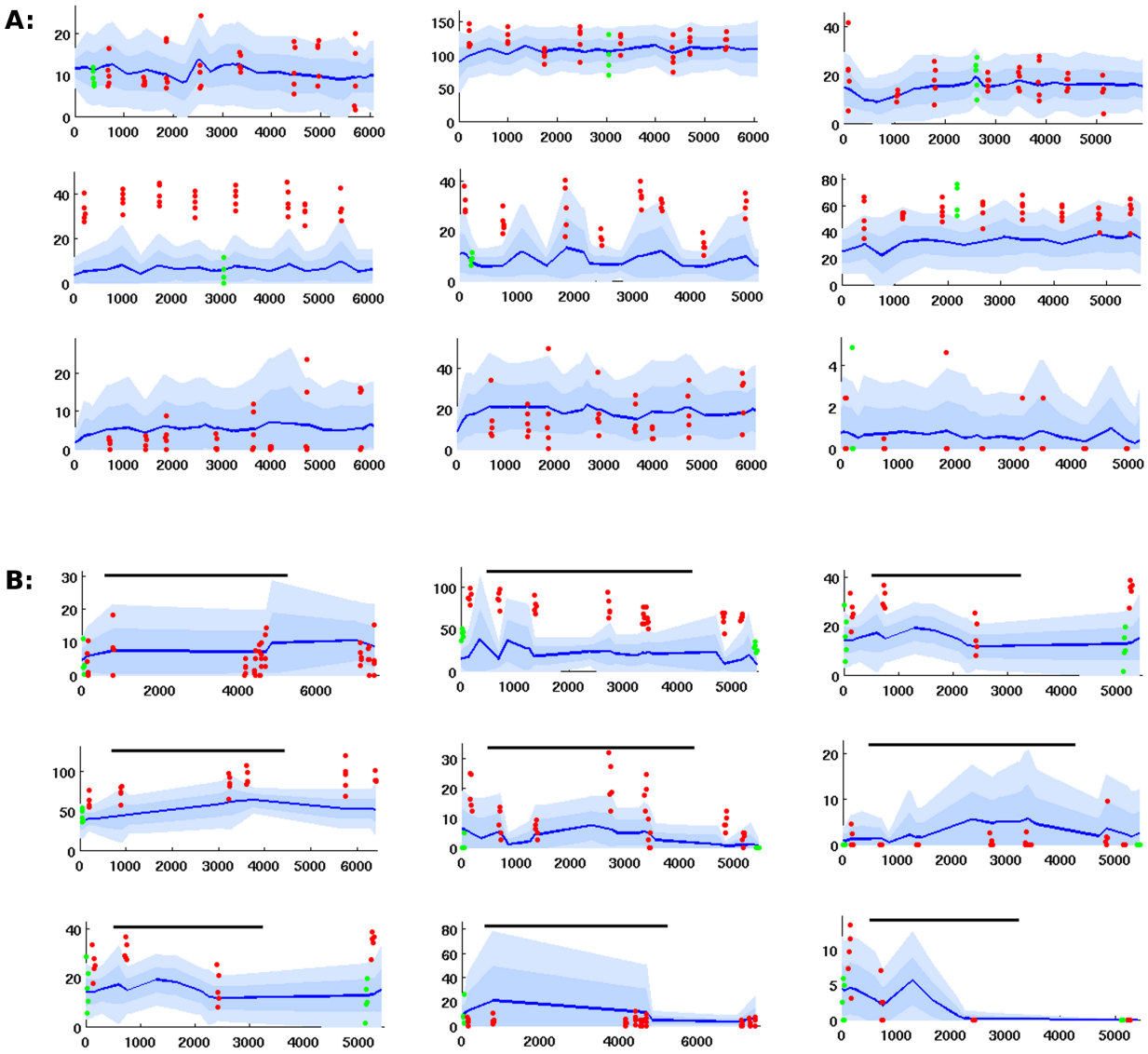

Figure 2.3: **Time series of PN firing rates from wet-lab.** x-axis = time, y-axis = FR. Blue lines = mean spontaneous rate, shaded regions = ±1 and 2 std. Red dots are odor responses. Green dots are response to control (mineral oil).

A: PN response, given odor plus coincident sugar reward, ie plus octopamine (time series for PNs with odor only are similar, but with less strong odor responses). Top row: unresponsive to odor. Middle row: excited response to odor. Bottom row: inhibited response to odor.

B: PNs with octopamine wash added in mid-experiment, then rinsed away (duration shown by black line). Octopamine can alter (up, down, or not at all) the spontaneous FR and/or the odor response, so there are 9 possible modulation regimes. This grid of timecourses shows a typical PN from each regime. Top row: spontaneous FR in unaffected. Middle row: spontaneous FR is boosted. Bottom row: spontaneous FR is inhibited. First column: odor response is unaffected. Second column: odor response is boosted. Third column: odor response is inhibited.

2.3 Results

This section first describes calibration between MothNet simulation outputs and wet-lab data. It then reports learning experiments, giving evidence of robust learning behavior from MothNet experiments, in both KC and EN behaviors. Lastly, it reports experiments on the effects of MB sparsity.

2.3.1 Calibration: MothNet outputs vs wet-lab data

PN behavior

MothNet's performance was calibrated to behave in a statistically similar way to wet-lab PN firing rate data. PN counts from wet-lab experiments were as follows:

1. For behavior with odor but without octopamine: 129 units with FR >1 spike/sec.
2. For behavior with odor, always with octopamine: 180 units with FR >1 spike/sec.
3. For behavior with odor, with and without octopamine: 52 units with FR >1 spike/sec.

Let

μ_s = mean of spontaneous FRs;

σ_s = std dev of PN spontaneous FRs, absent both odor and octopamine.

μ_s and σ_s are calculated for each PN.

Behavior of PNs in simulations and experiment (in terms of response to odor, octopamine, both, or neither) were compared using the PDFs (over all PNs) of the following:

1. μ_s
2. The ratio $\frac{\sigma_s}{\mu_s}$ (a measure of noisiness).
3. Odor responses absent octopamine, as distance from μ_s in units of σ_s (ie Mahalanobis distance).
4. Odor responses with octopamine, as distance from μ_s in units of σ_s.
5. Change in mean spontaneous FR due to octopamine, as distance from μ_s in units of σ_s.

Due to the limited number of PNs with wet-lab FRs, only qualitative comparisons of model and experiment were considered relevant. That is, excessive tuning of the model

would have only chased a specific instantiation of the true moth meta-distributions rather than matching either those meta-distributions or, more importantly, the general learning behavior of the moth.

Figure 2.4 shows a strong match between model behavior and experiment, for PN behavior. One point of difference is that the range of $\frac{\sigma_s}{\mu_s}$ values are generally narrower for the model than for experiment. That is, the PNs of the model are less varied in their level of noisiness than those of live moths.

KC, EN behavior

There is less experimental data about KC firing rates in general, and no data to our knowledge measuring KC firing rates in response to octopamine stimulation. KC behavior in simulations matches experimental statistics in the literature [80].

There are no bulk data, to our knowledge, measuring EN firing rates in response to odors and/or octopamine. However, calibrating EN response is not necessary to demonstrate an ability to learn. The key marker is post-training increase in EN response.

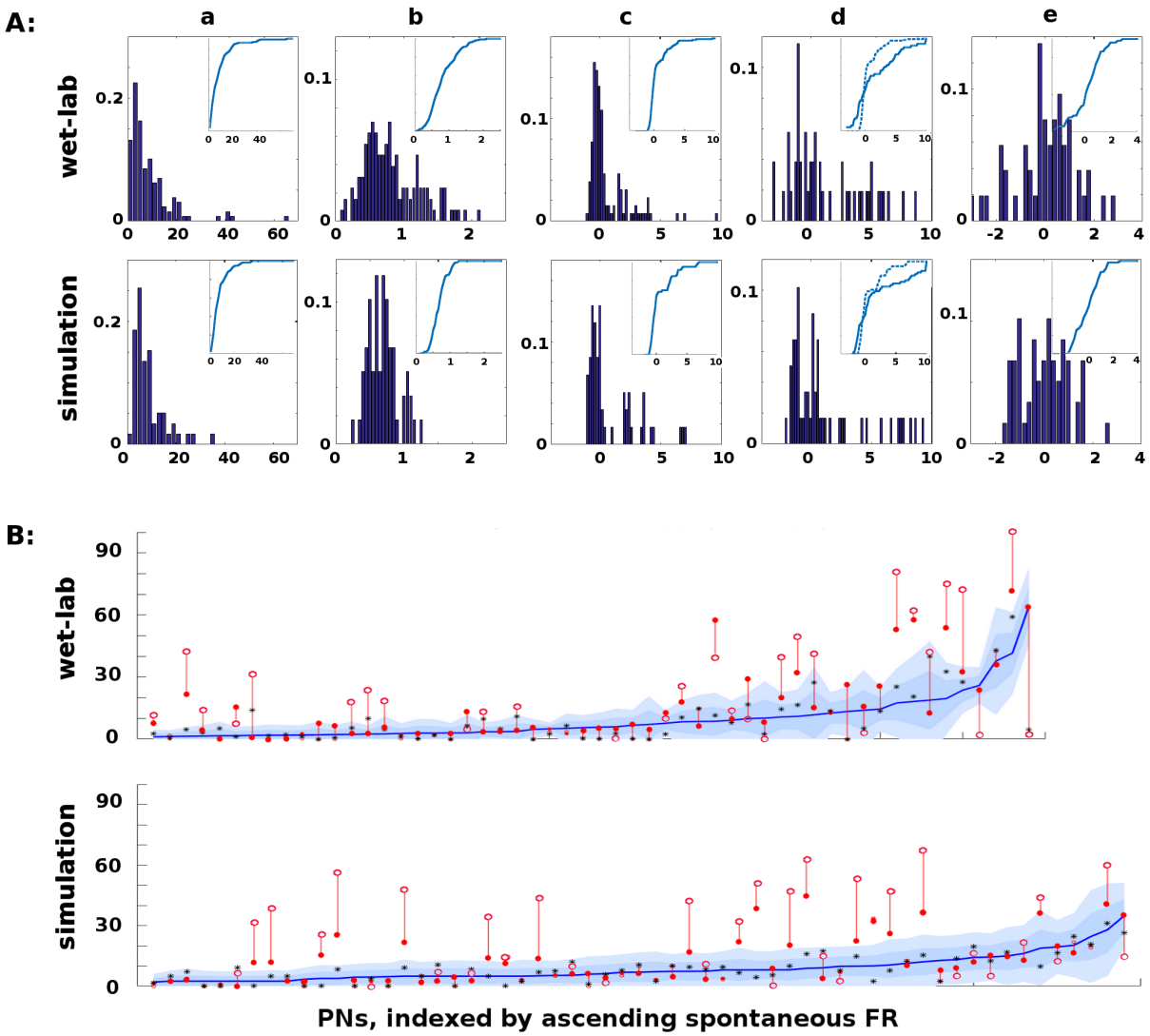

Figure 2.4: **Wet-lab data and Model Calibration:** Comparison of PN firing rate activity from wet-lab data and simulations. **Panel A**: Histograms and CDFs of wet-lab data and simulations. **Col a**: mean spontaneous FRs μ_s. **Col b**: σ_s/μ_s of spontaneous FRs, a measure of noisiness of a PN. **Col c**: odor response, measured as distance from μ_s in σ_s units. Distance $> 2\sigma_s$ implies a strong activation/inhibition. **Col d**: odor response during octopamine, in σ_s units distance from μ_s. Note that PN responses are broadened (i.e. more PNs are strongly activated or inhibited). The dotted line in the CDF inset is the same as the CDF of the odor response without octopamine, to show the broadening towards both extremes. **Col e**: change in mean spontaneous FRs due to octopamine, measured in σ_s units distance from (non-octopamine) μ_s. Some PNs are excited, some are inhibited. x-axis units: col a = raw spikes/sec; col b = ratio σ_s/μ_s; cols c, d, e = units of σ_s. **Panel B**: Activity of PNs indexed by increasing spontaneous FR. y-axis: Raw spikes/sec FR. Blue lines = mean spontaneous FRs μ_s (cf col a). Shaded regions = $\sigma_s, 2\sigma_s$ envelopes (cf col b). Solid red dots = odor response FRs (cf col c). Hollow red dots = odor response FRs during octopamine (cf col d). Red lines show the change in odor response FRs due to octopamine (cf broadened response). Black stars (*) = spontaneous FRs during octopamine (cf col e).

2.3.2 Learning experiments: Behavior of AL, KCs

This section describes behavior of various neural types in MothNet simulations as described in section 2.2.10. In particular, it describes evidence of robust learning behavior in both KCs and ENs.

AL behavior

Fig 2.1 shows MothNet PN firing rates for a typical simulation with two odors, one of which was reinforced with octopamine. The two odors excited (and inhibited) distinct PNs. Octopamine (without odor) resulted in more PNs being excited beyond their usual noise envelopes, and also some PNs being inhibited beyond their usual envelopes. Octopamine and odor, applied together, resulted in broader, supra-additive excitation of PNs. Behavior returned to baseline after octopamine was withdrawn, since AL connection weights are not plastic.

KC behavior

KC behavior for a typical simulation with two odors, one of which gets reinforced with octopamine, is shown in Figure 2.1.

Odor-only response: KC spontaneous baseline response was essentially zero. KC response to individual untrained odors, absent octopamine, was very sparse (2% to 4%). Different odors excited distinct sets of KCs. KC odor responses had high trial-to-trial consistency, as shown in Figure 2.5. These results were consistent with wet-lab experiments from the literature [80].

Octopamine-only response: Given octopamine but no odor, KCs were unresponsive, similar to baseline. This was likely due to LH inhibition offsetting the increased spontaneous PN activity due to octopamine. Note that moths in the wild do not experience this regime

(since reward is always coupled with an odor). No wet-lab data for KCs with octopamine were available for comparison.

Odor + octopamine response: KCs transient response to odor + octopamine was much broader (typically 2x to 3x), with high trial-to-trial consistency. See Fig 2.5 panels B, C. No wet-lab data were available.

Post-training response: After an odor had been trained, the odor-only response was permanently broadened (typically by 50% to 100%, eg from 2% to 3% or 4%). Since AL odor responses revert to baseline levels when octopamine is withdrawn, the increase in KC response to trained odors must be due to increased synaptic weights PN→KC. Spontaneous baseline and octopamine-only responses remained essentially zero.

We note that KC response to octopamine, or to odor plus octopamine, is not an artifact of freely-tuned parameters. Rather, it follows from the effects of tuning the AL to match wet-lab data. Octopamine's effects on PN firing rates are fully determined by calibration of MothNet to wet-lab data. KCs respond only to PN (and QN) behavior (and LH inhibition). Feed-forward connections $M^{P,K}$ are determined by matching KC baseline responses, spontaneous and to odor, with [80]. Then KC behavior is determined entirely by PN behavior, plus the assumption that octopamine has no direct effect on KC firing rates. Thus KC behavior with octopamine is fully determined once the model is tuned to PN data.

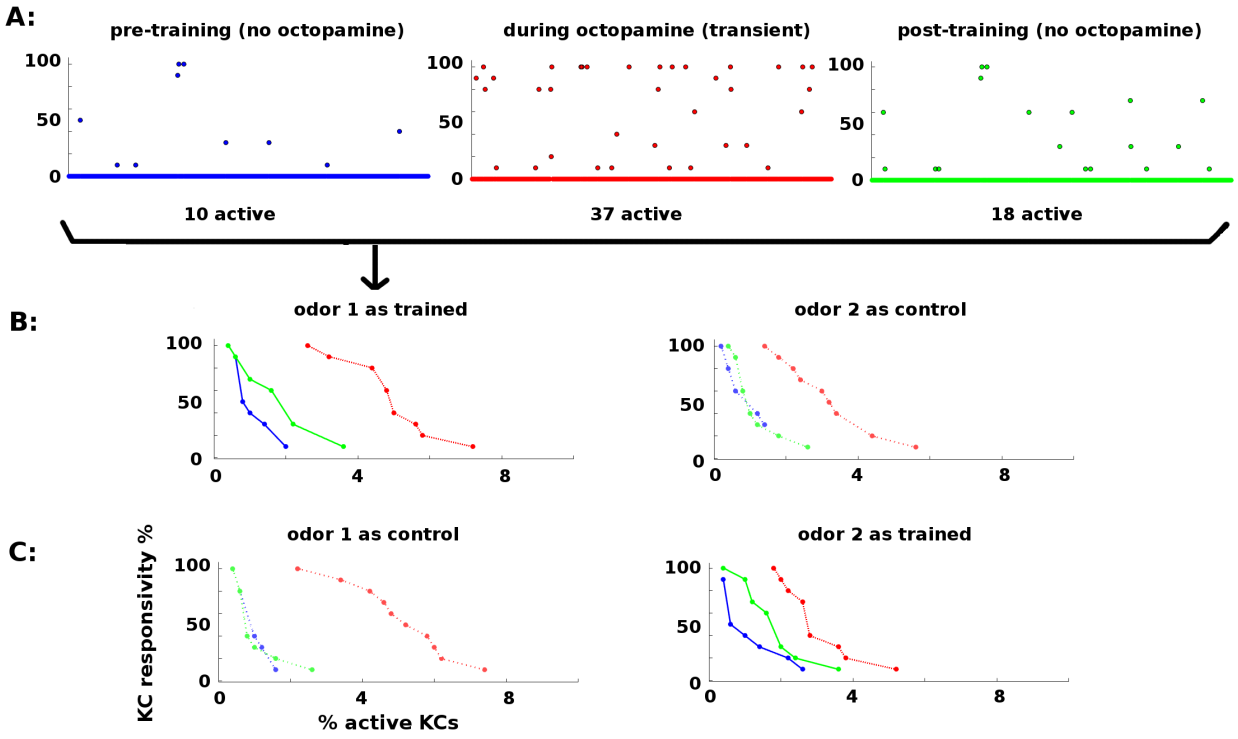

Figure 2.5: **KC responses to odor during training:** KCs respond sparsely to odor pre- and post-training, ie absent octopamine (blue and green dots and curves). Octopamine induces transient increased responsivity (red dots and curves). Training results in permanent increases in response to the trained odor, but no increase in response to control odor (green dots and curves).
A: KC response to an odor before, during, and after training. x-axis: indexed KCs (500 shown). y-axis: consistency of response (in %). The plots are for odor 1 as the trained odor (ie same data as panel B). Blue = pre-training (no octopamine). Red = during training (with octopamine); note the heightened transient response. Green = post-training (no octopamine). There is a permanent increase in the number of KCs that respond to the trained odor.
B: Response rate vs. percentage of active KCs for trained and control odors before, during, and after training. x-axis: percentage of KCs responding at the given rate. y-axis: consistency of response (in %). Blue = pre-training. Red = during octopamine (transient). Green = post-training. The LH plot shows odor 1 as the reinforced odor. The scatterplots in (A) correspond to the three curves in this plot. Note that the permanent KC response curve shifts up and to the right (blue→green) in the trained odor, ie more KCs respond to the odor (right shift) and they respond more consistently (upward shift). The RH plot shows odor 2 as a control. The control's permanent KC response curve does not shift.
C: As (B) above, but in this experiment odor 1 is now the control (LH plot), and odor 2 is reinforced (RH plot). In this case, the response curve of odor 2 (reinforced) shifts to the right (blue→green), while the response curve of odor 1 (control) is unchanged.

2.3.3 Learning experiments: EN behavior

A key finding is that ENs, which are the actionable output of the system, demonstrate robust learning behavior in MothNet simulations. Learning is defined as rewiring the system so that EN responses to trained odors are significantly stronger than to control odors, such that the system can robustly distinguish a trained odor from control odors via thresholding. Typical EN response timecourses, in simulations with multiple odors where one odor is trained on 15 odor puffs, are shown in Figures 2.6A and 2.7A.

Because the EN response is driven solely by feed-forward signals from KCs, in the absence of odor ENs had response ≈ 0, with or without octopamine, as expected. In particular, training did not materially increase EN spontaneous (ie no-odor) response.

EN response to odor + octopamine was always very strong. The functional value of this is discussed in section 2.4.

Training consistently increased the EN response to the reinforced odor much more than response to control odors, measured as percentage increase over naive odor response.

For orthogonal odors, the contrast is stark. Figure 2.6 shows a typical case when the odors' projections onto the moth's AL glomeruli were orthogonal, and when the naive EN responses to the different odors were similar in magnitude. For overlapping odors with widely disparate naive EN responses (Figure 2.7), the effect had some nuance: Differential learning was expressed by substantially higher increase in EN response to reinforced vs unreinforced odors, because MothNet lacks a synaptic weight decay mechanism.

Figure 2.7 shows a typical case where the odors' projections onto the moth's AL glomeruli overlapped, and where the naive EN responses to the different odors varied widely in magnitude.

The differential effect on post-training EN response to overlapping trained vs control odors (ie reinforced vs unreinforced) was routinely significant with $P < 0.01$ when the two odors had roughly similar naive EN responses (within a factor of 3).

Exceptions sometimes occurred when the two odors had highly disparate raw naive EN

responses (greater than a factor of 3 either way). In these cases, one of two things happened: Either the differential raw change in post-training EN responses was significant with $P < 0.01$ (when naive EN response to trained odor $>$ 3*naive EN response to control odor); or the differential percentage change post-training EN responses was significant with $P < 0.01$ (when naive EN response to trained odor $<$ 0.33*naive EN response to control odor). Details of ANOVA treatment can be found in SI.

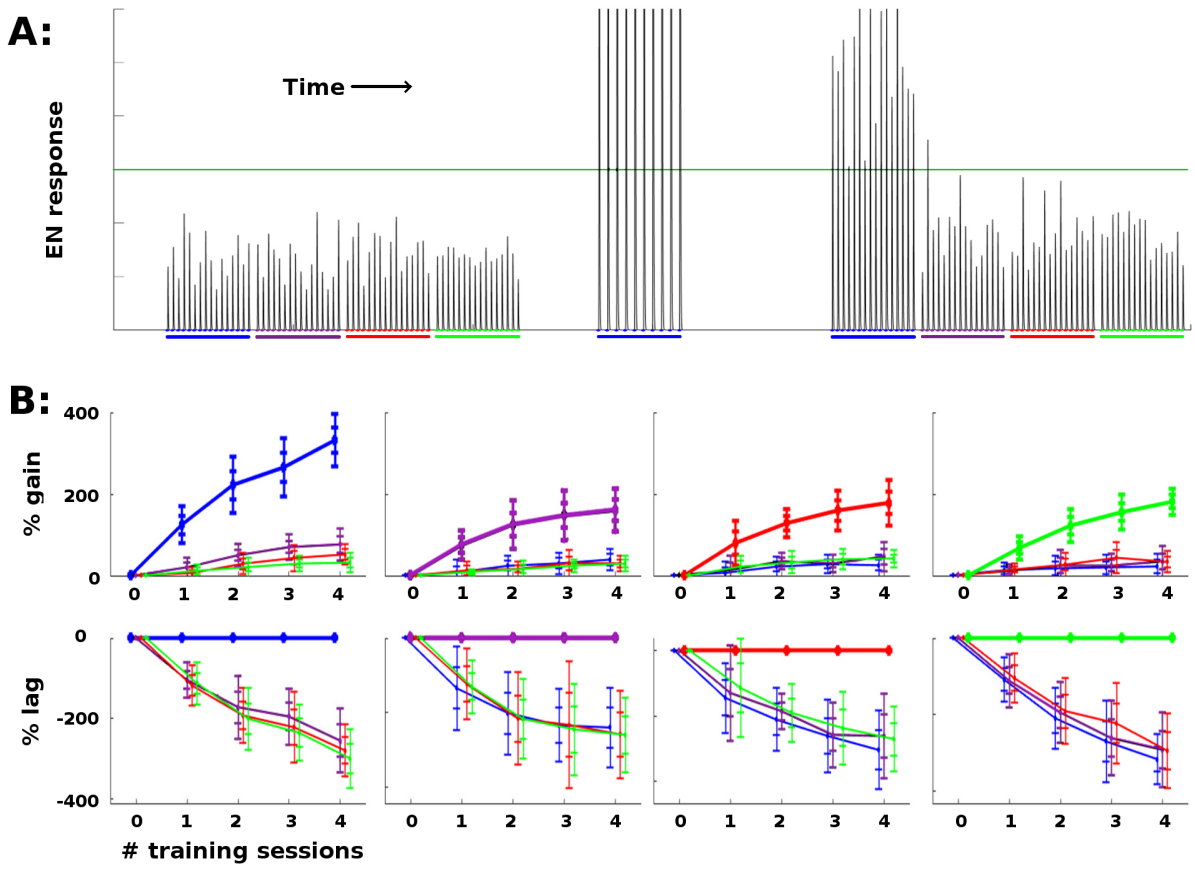

Figure 2.6: **Effect of training on EN FRs: A:** Typical timecourse of EN responses from an experiment with a single moth. First, 16 puffs of each odor were delivered, to establish naive odor responses. Note EN response variability due to noise in the system, especially in the AL. Next, the moth was trained on the first (blue) odor trained over 2 sessions (10 puffs), by delivering odor and octopamine concurrently This timecourse corresponds to the {odor, #sessions} pair in the first column in panel B, at index 2 on the x-axis. Octopamine was then withdrawn, and the four odors were again delivered in series of puffs, to establish post-training changes in EN response. The long green line represents a hypothetical trigger threshold, such that EN response $>$ threshold would induce a distinct behavior.
B: EN response changes due to training: Aggregated results with 11 noise realizations for each {odor, #sessions} pair. Each column shows results of training a given odor, color coded: blue, purple, red, green. x-axis = number of training sessions.
First row: The y-axis measures percent change in EN FR. The line shows mean percent change. The error bars show ± 1, 2 std devs.
Second row: The y-axis measures percent changes in EN response, relative to the trained odor (ie subtracting the trained odor's change from all odors). This shows how far each control odor lags behind the trained odor. The line shows mean percent lag. The error bars show ± 1, 2 std devs.

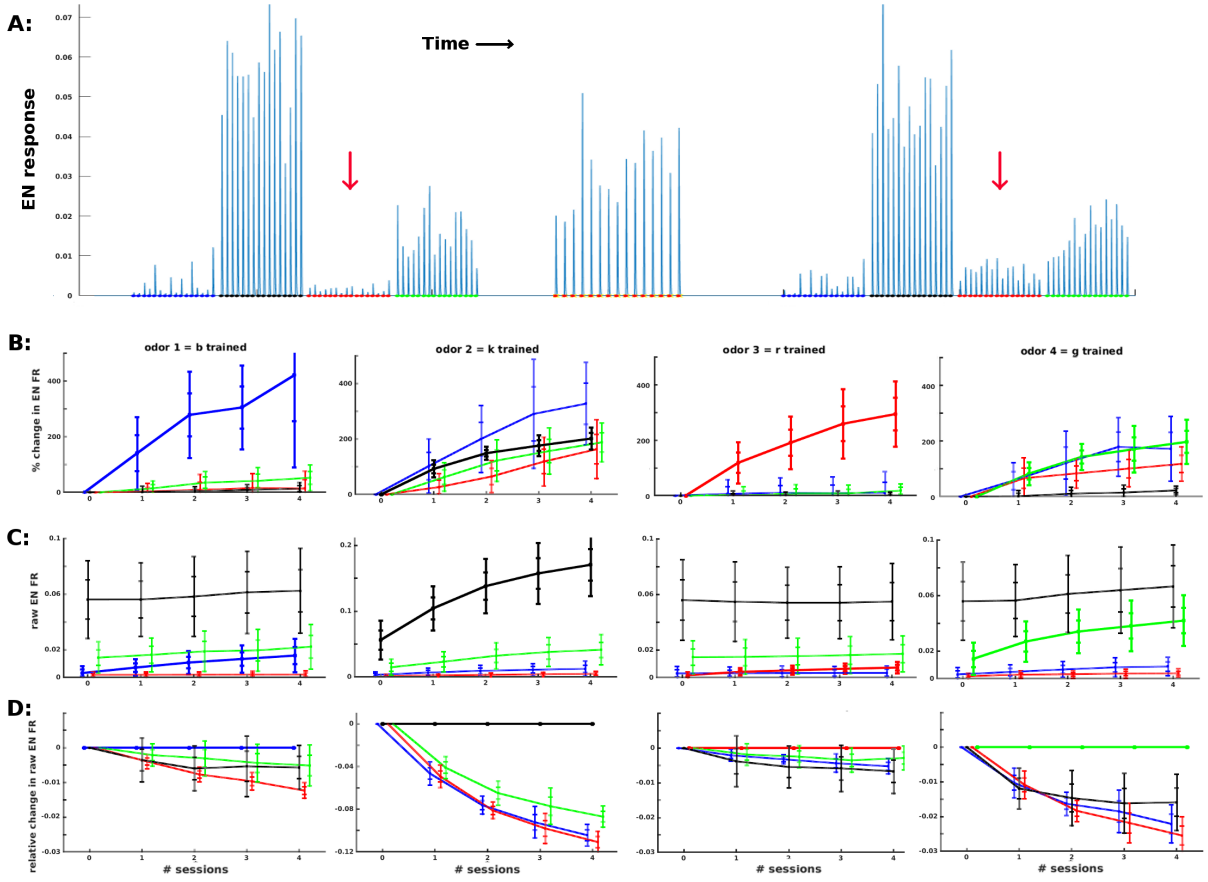

Figure 2.7: **Effect of training on EN FRs, given odors with unequal naive response magnitudes.** When odors induced naive EN responses of very different magnitudes, then trained odor response increased much more than control odor responses either in raw magnitude, or as a percentage, or both.

A: Typical timecourse showing magnitudes EN responses before and after training the third (red) odor, indicated by red arrow, over 15 odor puffs. This corresponds to the third column in panels B - D, at index 3 on the x-axis. Note that only the third (red) odor's EN response changes magnitude.

Panels B - D: Changes to ENs during training. x-axis = number of training sessions. Each column shows results of training a given odor, color coded: blue, black, red. y-axis measures raw EN or percent change in EN. 21 trials per data point.

B: Percent change (from pre-training) in ENs, mean ±2 stds.

C: Raw EN FRs, mean ±2 stds.

D: Changes in raw EN FRs, normalized by trained odor (ie subtract the trained odor's changes from all odors), mean ±2 std devs. This shows how far each control odor lagged behind the trained odor.

Note that the trained odor dominates in either raw increase (panels C, D) if naive response to trained odor was large, or in percent increase (panel B) if naive response to trained odor was small.

2.3.4 MB sparsity experiments

To assess the role of MB sparsity during learning, we ran MothNet experiments that varied the level of generalized inhibition imposed on the MB by the lateral horn, which controls MB sparsity level. Each experiment set a certain level of LH inhibition then ran simulations as in section 2.2.10, training one odor with 15 odor puffs, and leaving one control odor untrained. EN responses to both trained and control odors were recorded, as well as MB sparsity levels, ie percentage of KCs active in response to odor. These simulations indicated that generalized inhibition from the LH onto the MB serves to maintain an optimal level of sparseness in the KCs, in order to ensure effective EN behavior, especially during training. This requires balancing opposed demands, viz for reliable odor response and for well-targeted learning.

Too little damping resulted in a high percentage of KCs being active. This yielded consistent EN responses to odor. But it caused EN responses to control odors and to noise to increase significantly during training. This reduced contrast between EN responses to trained and untrained odors and also increased spontaneous EN noise.

Too much damping resulted in a very low percentage of KCs being active. This ensured that training gains were focused on the trained odor, while EN response to control odors or noise were not boosted. However, in this regime EN responses to odors, both pre- and post-training, were generally inconsistent because too few KCs were activated.

Thus, projection to the sparse high-dimensional MB fulfilled a vital role in MothNet's learning system. LH inhibition of the MB had an optimal regime, where EN odor responses were reliable and training gains were focused on the trained odor only. Timecourses illustrating this effect are seen in Fig 2.8A. Fig 2.8B shows how this trade-off varied with MB sparsity, by plotting two figures-of-merit:

$$\text{Signal-to-Noise Ratio (SNR)} = \frac{\mu(f)}{\sigma(f)} \text{ where } f = \text{ EN odor response;} \tag{2.5}$$

and

$$\text{“Learning Focus”} = \frac{\mu(f_T)}{\mu(f_C)}, \text{ where} \tag{2.6}$$

$\mu(f_T)$ = mean post-training EN response to trained odor, $\mu(f_C)$ = mean post-training EN response to control odor.

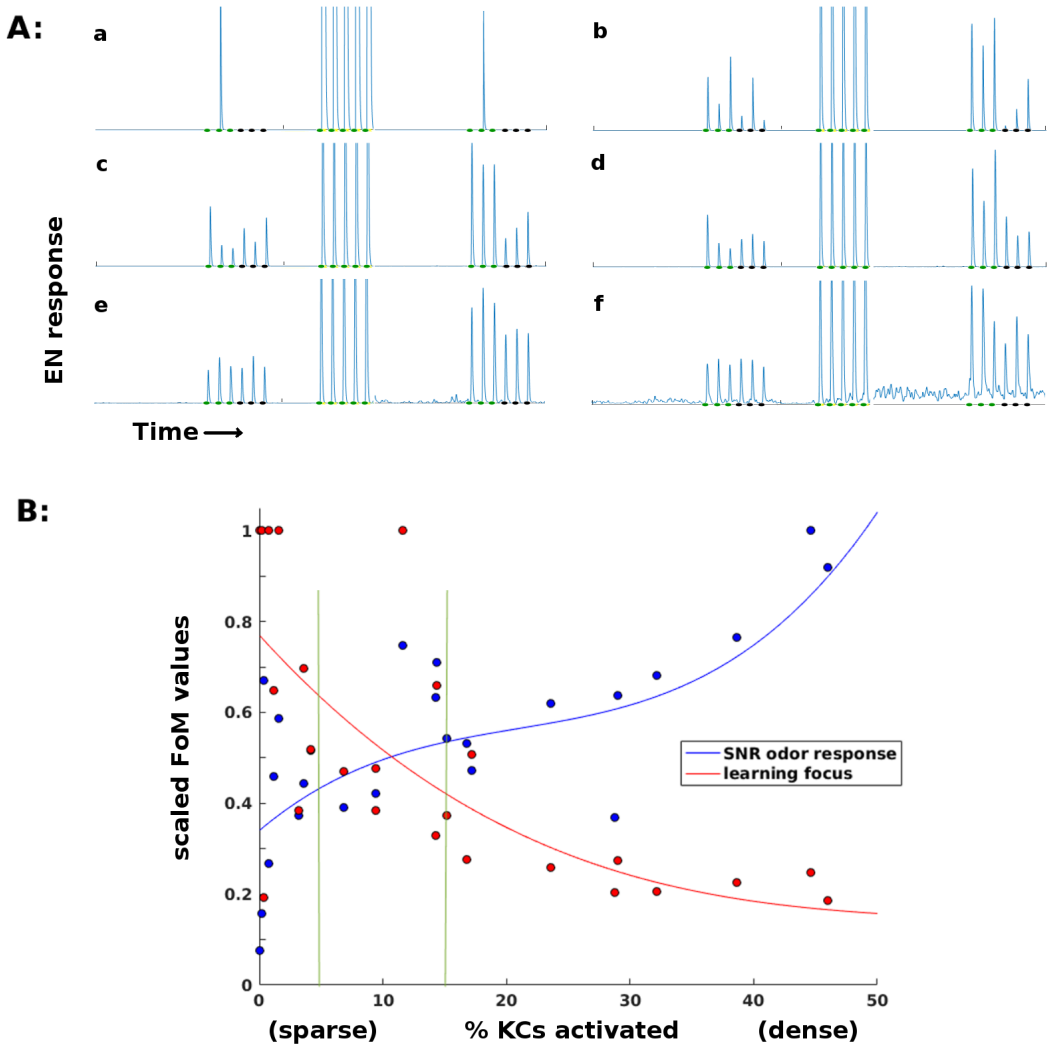

Figure 2.8: **Effects of sparsity on learning and EN reliability** Results for a typical experiment on a moth with two odors.

A: EN responses timecourses for two odors, at varying levels of KC activation (a, b: <1%. c, d: 5 to 15%. e, f: 20 to 45%. Order of events: 3 puffs of each odor as baseline, train on first odor (only one session shown), then 3 puffs each post-training. At very sparse levels (a, b) training is focused but odor response is not reliable. At low sparsity levels (e, f) training is unfocused, boosting EN response to control odor and to background noise.

B: Two Figures of Merit (FoMs) plotted against MB sparsity. Low KC activation (high sparsity) correlates with well-focused learning, but low odor response SNR. High KC activation (low sparsity) correlates with poorly-focused learning, but high odor response SNR. The FoMs are each normalized for easier plotting. y-axis:

Blue data: $\frac{\mu(f)}{\sigma(f)}$, a measure of odor EN response SNR, where f = EN odor response.

Red data: $\frac{\mu(f_T)}{\mu(f_C)}$, a measure of learning focus, where $\mu(f_T)$ = mean EN post-training response to reinforced odor; $\mu(f_C)$ = mean EN post-training response to control odor (values are thresholded at 1 for plotting). A high value indicates that increases in EN response due to training were focused on the trained odor; low values indicate that irrelevant signal (F_C) was also boosted by training. The points are experimental data, the curves are cubic fits. Vertical green lines indicate the 5 - 15% sparsity region, typical in biological neural systems.

2.4 Discussion

Experiments with MothNet offer a variety of insights into the moth olfactory network and how it learns, including: Predictions about aspects of the network that are still unclear in the literature; the role of sparse layers; the role of octopamine; and the value of randomness. These items are discussed below.

2.4.1 Predictions re details of AL-MB structure

Model simulations enable us to make predictions about some unresolved aspects of the moth's AL-MB system. Some examples:

Do LNs inhibit PNs and LNs as well as RNs?

In the AL, LNs are believed to inhibit RNs. It is not currently known whether LNs also inhibit PNs and LNs. Efforts to calibrate MothNet to wet-lab data indicate that LNs need to inhibit not just RNs, but also (to a lesser degree) LNs and possibly PNs. The model weight strengths for LN→RN, LN→LN, and LN→PN are in the ratio of 6:2:1. That LNs would inhibit LNs makes sense when the goal is to maximize the PN output of the active glomerulus: By inhibiting the LNs of rival glomeruli, the active glomerulus reduces the amount of inhibition directed at itself. Similarly, that LNs would inhibit PNs makes sense when the goal is to reduce the PN output of rival glomeruli.

Octopamine's effects on different neuron types

Octopamine increases the responsivity of a neuron to incoming signals. It is unclear how or whether octopamine affects various neuron types (ie RNs, PNs, LNs, KCs). Calibration of MothNet's AL behavior, and tuning of KC behavior to enable learning, indicate that octopamine needs to target RNs and LNs, but not PNs, KCs, or ENs. Logical arguments support these findings from MothNet calibration:

RNs: Because RNs initially receive the odor signal, these are logical neurons to stimulate with octopamine, because it sharpens their response to the exact signature being trained, which in turn sharpen the AL's output code for that odor.

LNs: LNs have the dual roles of inhibiting rival glomeruli and limiting overall PN output in the AL. For the first role, increased LN response to RNs will tend to sharpen AL response to the trained odor, by accentuating inhibition of rival glomeruli PNs. For the second role, increased LN activity mitigates the risk that increased RN activity (due to octopamine) might blow up the overall PN output of the AL.

PNs: MothNet simulations suggest that PNs should receive little or no octopamine stimulation. While increasing PN responsivity would benefit RN-induced sharpening of the trained odor's signature, there are three downsides. First, RN input to PNs is intrinsically noisy, so higher PN responsivity amplifies noise as well as signal. Second, since PNs respond to LNs, higher PN activity tends to reduce the impact of LN inhibition, and thus reduces the inhibition-induced sharpening of the AL odor response caused by octopamine. Third, increasing PN responsivity can have an outsize effect on overall PN firing rates, ie it is a very "high-gain" knob and therefore risky.

KCs: KCs might arguably be good targets for octopamine, as this could make them more sensitive to odor codes outputted from the AL. This contradicts MothNet simulation results indicating that direct octopamine stimulation of KCs reduces sparseness in the MB, which can be disastrous to learning (section 2.3.4).

2.4.2 Noise filtering role of the sparse, high-dimensional stage

Projection from a dense, low-dimensional coding space (eg the AL) to a sparse, high-dimensional coding space (eg KCs in the MB) is a widespread motif of biological neural systems, with size shifts are routinely on the order of 20x to 100x [22, 4]. The reasons for this pattern are not well-understood. Some proposed reasons include information capacity, long-range brain communication, and reduced training data needs [22].

MothNet simulations bring to light another, central, role of sparseness: It acts as a robust

noise filter, to protect a Hebbian growth process from amplifying upstream noise to out-of-control levels. Though noise may be useful (or unavoidable) in upstream networks such as the AL, noise that reaches the neurons on both sides of a synaptic connection will be amplified by Hebbian growth during learning, swamping the system's downstream neurons (eg ENs) with noise. However, the "fire together, wire together" principle of Hebbian learning is an AND gate. Thus it suffices to remove noise from just one of the two connected neurons to prevent synapse growth. Sparsity does precisely this.

MothNet simulations show that sufficient sparseness in the MB ensures that noise does not get amplified during training, so that post-training EN spontaneous firing rates and EN odor responses remain unchanged. Conversely, when KC response is not sufficiently sparse, any training leads rapidly to noisy EN spontaneous response levels and amplified EN responses to control odor. This implies that the noise filtering induced by MB sparseness is required for a workable Hebbian learning mechanism.

Beyond the particular demands of Hebbian plasticity, robust noise filtering may be a core function of sparse, high-dimensional stages within any network cascade where noise accumulates due to (beneficial) use in upstream stages.

2.4.3 Roles of octopamine

The levels of octopamine stimulation in MothNet are calibrated to wet-lab data on PN responses to octopamine. Thus the simulations give insights into downstream effects of octopamine, in particular on plasticity, KC responses, EN responses, and Hebbian learning.

Accelerant

Moths can learn to respond to new odors remarkably quickly, in as little as 5 exposures. Simulations indicates that while Hebbian growth can occur without octopamine, it is so slow that actionable learning, ie in terms of amplified EN responses, does not occur.

This implies that octopamine, through its stimulative effect, acts as a powerful accelerant to learning. Perhaps it is a mechanism that allows the moth to work around intrinsic organic

constraints on Hebbian growth of new synapses, constraints which would otherwise restrict the moth to an unacceptably slow learning rate. To the degree that octopamine enables a moth to learn more quickly, with fewer training samples, it would clearly be highly adaptive.

Active learning

Simulations indicate that octopamine strongly stimulates the EN response to even an unfamiliar odor. Since octopamine is delivered as a reward, this has a beneficial effect in the context of the moth as a learning agent. In the paradigm of reinforcement learning [78], an agent (the moth) can in some cases learn more quickly when it has choice as to the sequence of training samples (Active Learning) [73].

In particular, when a certain class of training sample is relatively rare, it benefits the agent to actively seek out more samples of that class [3]. Octopamine enforces high EN response to a reinforced odor, ensuring that ENs will consistently exceed their "take action" threshold during training. If the action is to "approach", the moth is more likely to again encounter the odor, thus reaping the benefits predicted by Active Learning theory. This advantage applies in the context of positively-reinforced odors.

In the case of aversive learning, the high EN responses to unfamiliar but objectionable odors, due to dopamine, would cause the moth to preferentially avoid further examples of the odor. This would slow learning of aversive responses (a drawback), but would also minimize the moth's exposure to bad odors (danger avoidance, a benefit).

2.4.4 Value of randomness

The principle of randomness permeates the moth olfactory network, for example in neural connection maps and in highly variable performance characteristics of chemical receptors in the antennae. A key result from MothNet experiments is that a biologically-based neural net permeated with the random principle can robustly learn. This is in marked contrast to engineered computers, where poorly-spec'ed components are a liability. Particular benefits of randomness include:

Random KC→EN connections: These guarantee that projections onto ENs are incoherent relative to whatever low-dimensional manifold in KC-space, which ensures (by the Johnson-Lindenstrauss lemma) that the EN readouts will preserve distinctions between elements in KC-space [22].

Variable gaba-sensitivity in glomeruli: This increases the range of sensitivity to odor concentration, because some glomeruli simply don't turn off when a different, stronger odor tries to mask them through lateral inhibition (LNs).

Variable sensitivity of antennae receptors: This gives a natural sensitivity to concentration, as progressively stronger odor will progressively activate the less-sensitive receptors, increasing the total RN input to glomeruli.

Resistance to injury: A randomly-connected network is more robust to damage. When exact connection maps and strengths are not required in the first place, damage has less impact on the fundamental nature of the system.

Most importantly, randomness (of connections and component properties) is evolutionarily cheap, easy, and available. So perhaps the core benefit of randomness to the moth olfactory network is that it works at all.

Chapter 3

PUTTING A BUG IN MACHINE LEARNING

3.1 Introduction

Although originally inspired by the biological structure of networks of interacting neurons [38, 20], engineered (artificial) neural networks (ANNs) have since developed sets of tools (such as backprop, maxPool, etc) that are not biologically plausible. These tools, combined into complex and deep neural net (DNN) architectures, have achieved strong success in a wide array of tasks [72]. But they are also known to fail on critical tasks such as learning from few samples. We seek to improve ANN performance on such tasks by exploiting features of the biological structures involved in learning. Key features in biological NNs include: High noise, random connections, Hebbian synaptic growth, high-dimensional sparse layers, large dimension shifts between layers, and generalized stimulation of neurons during learning. In combination, these features allow for rapid and effective learning.

To test whether such a biological NN toolkit can effectively tackle general classification tasks in the machine learning (ML) context, we assigned the MothNet model the task of learning the handwritten digits of the MNIST dataset. The MNIST dataset is a classic learning task that offers some challenge, yet is simple enough to be potentially within the means of an insect brain. MothNet was able to learn to read with high accuracy given very few (or even just one) training samples. This demonstrates that even a simple biological architecture contains novel and effective tools applicable to ML tasks, in particular tasks involving few training samples or the need to add and train new classes without retraining the full NN.

The experiments described in this chapter elucidate mechanisms for fast learning that rely on cascaded networks, sparsity, and Hebbian plasticity. These motifs (though only a

small subset of the diverse repertoire used by biological NNs) offer a novel, alternative toolkit for building DNNs that more closely mimic biological brains.

3.2 Methods

3.2.1 MothNet2 (computational model)

The computational model (hereafter "MothNet2") of the moth olfactory network used for these MNIST experiments is a variant of the MothNet model of chapter 2. Fig 3.1 gives a schematic. MothNet2 has modifications relative to MothNet, listed below.

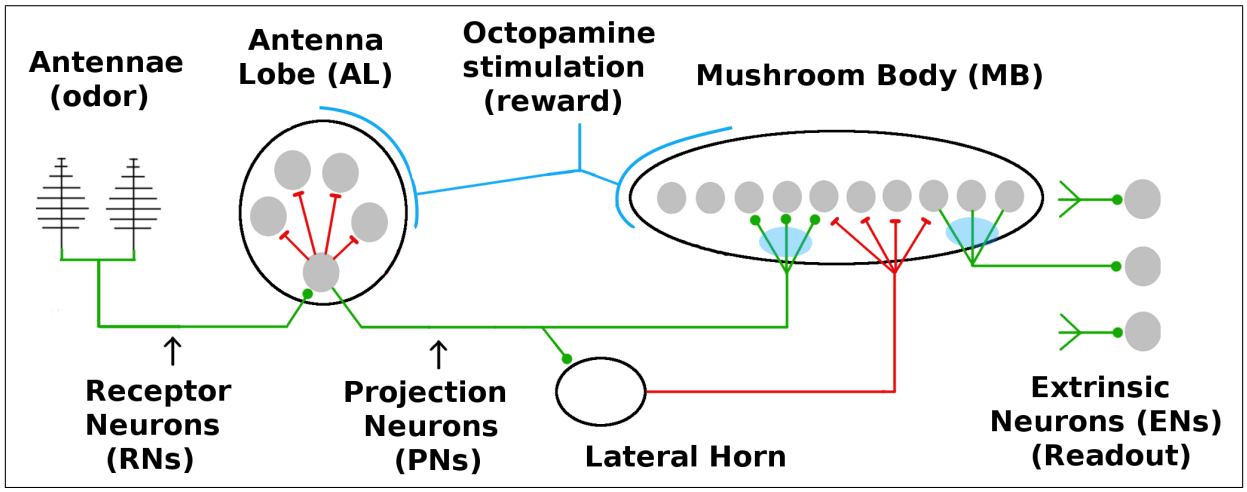

Figure 3.1: **Network schematic.** Green lines show excitatory connections, red lines show inhibitory connections. Light blue ovals show plastic connections into and out of the MB. The glomeruli (processing units) in the AL competitively inhibit each other. Global inhibition from the lateral horn induces sparsity on MB responses. The ENs give the final, actionable readouts of the system's response to a stimulus.

No QNs:

There are no inhibitory feed-forward neurons from AL→MB.

Sparsity mechanism:

As before, the lateral horn is modeled purely as a mechanism to enforce sparsity on the MB (high sparsity corresponds to a low percentage of KCs being active). In the MothNet of chapter 2, MB sparsity was induced by a uniform, time-invariant inhibitory signal to all KCs. This gave higher density of active KCs during training as a side effect of higher AL outputs due to octopamine. In MothNet2, the sparsity percentage is enforced directly. We posit two sparsities, $s_1\%$ in the "default" regime (ie absent octopamine or synaptic growth), and $s_2\%$ in the learning regime (ie during octopamine and synaptic growth), with $s_2 > s_1$. A target sparsity $s\%$ is induced by injecting adaptive global inhibitory input to MB neurons sufficient to damp $(100 - s)\%$ of MB neuron inputs.

Multiple ENs:

MothNet2 has 10 ENs, one for each class (digit) in the MNIST dataset. Each EN begins fully connected to the KCs with uniform weights.

Plasticity

MothNet2 has several modifications to accommodate multiple ENs. Like MothNet, MothNet2 assumes a Hebbian growth mechanism for synaptic weights, $\Delta w_{ab}(t) \propto f_a(t) f_b(t)$, and two layers of plastic synaptic weights: AL→MB, and MB→ENs (ie pre- and post-MB). Hebbian plasticity is assumed to be "switched on" by reward, so it is cosynchronous with octopamine. Thus, plasticity only occurs during training sessions.

In addition, MB→EN weights that are inactive are subject to proportional decay. Thus, two equations govern synaptic plasticity:

$$\Delta w_{ab}(t) = \gamma f_a(t) f_b(t), \text{ where } \gamma \text{ is a growth parameter (same as eqn 2.4); and}$$

$$\Delta w_{ab}(t) = \delta w_{ab}(t), \text{ where } \delta \text{ is a decay parameter, if } f_a(t) f_b(t) = 0. \tag{3.1}$$

A rough balance between growth and decay rates is important for optimal learning.

There are two layers of plastic synaptic weights: AL→MB, and MB→ENs (ie pre- and post-MB). These two layers have distinct influences during training. All ENs are affected to some degree by the AL→MB connection weights, since one PN connects to many MB neurons, each of which can connect to multiple ENs. However, the MB→EN connections are unique to each EN.

The bulk of useful plasticity in MothNet2 occurs in MB→EN weights, because each EN responds to a particular stimulus class. AL→MB weights also grow, though at a much slower rate (by design), and with less apparent impact on learning.

Each of the ten ENs is arbitrarily assigned to a digit in the naive (untrained) moth, so that EN_j trains to recognize class (ie digit) j. Plasticity is activated during training on a sample from class j as follows:

1. AL→MB connections: All weights are updated according to Hebbian growth equation 2.4. None decay, ie eqn 3.1 is not applied.

2. MB→EN connections: Only those connections that feed EN_j are updated. Any connections feeding EN_j that have non-zero input are subject to weight growth according to eqn 2.4. Connections feeding EN_j with zero input are subject to weight decay via eqn 3.1. This EN-specific rule implies that training is supervised, ie the moth knows the class of the training sample. It also enables new classes to be added to the NN simply by inserting and training new ENs, without retraining the rest of the system.

3.2.2 Training data

The MNIST dataset consists of grey-scale thumbnails, 28 x 28 pixels, of handwritten digits 0 - 9. We used pixels of the thumbnails as input features to the MothNet2 classifier. Both biological and engineered systems routinely choose better feature sets (for example, convolutional kernels [44]). However, pixels-as-features provide a good test of whether MothNet can effectively learn to discriminate classes given inputs with inter-class correlations.

MothNet2 (like the moth) feeds one feature to each glomerulus in the AL. Use of full MNIST thumbnails, with pixels-as-features, would imply $28^2 = 784$ glomeruli. To keep the

scale of MothNet2 somewhat close to that of the true moth (60 glomeruli), we preprocessed the thumbnails as follows:

1. Downsample by 2 (linear interpolation).

2. Mean-subtract using 500 random, set-aside digits (50 from each class).

3. Select only the most-active pixels by thresholding the various class averages, ie retain only the most generally active pixels in the thumbnails, while also preserving the most active pixels of each class. These pixels defined the receptive field, and excluded border pixels which supply no information.

Each pixel becomes a feature that feeds into one glomerulus of MothNet2's AL. The experiments described here used 83 pixels (out of $14^2 = 196$ total) to represent the MNIST digits. Examples are shown in Figure 3.2.

4. Optional reinforcement (for training only): Thumbnails can be used as-is for training, or reinforced by adding a class average and/or subtracting a control class average (negative values are then zeroed out). This reinforcement improved the training response slightly, but it is not available in one-shot scenarios and is not so biologically obvious.

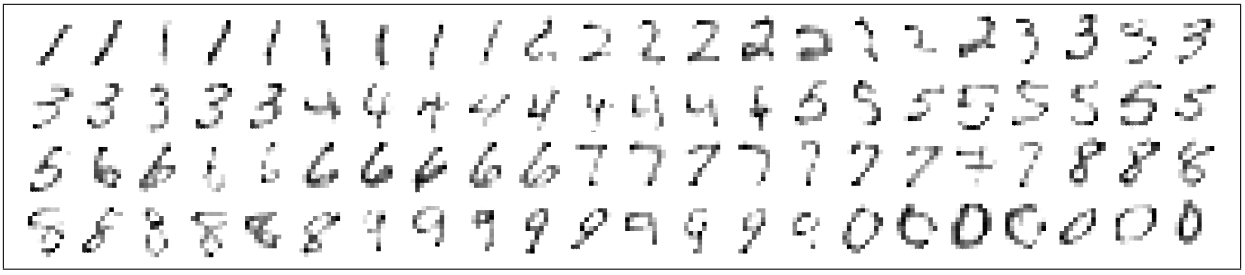

Figure 3.2: **Downsampled MNIST digits** used in the experiments (random selection).

3.2.3 Experiment design

Moth generation: Particular moth behaviors can be modulated by varying one or two template parameters. Each experiment used a fixed basic moth template, modified only by varying the parameter(s)-under-test. Each of these modified templates then randomly

generated many moths, typically 13-17 per data point. The MothNet2 basic templates varied slightly between experiments, eg in some parameter values, or in whether training samples were reinforced with class averages. We found that these slight differences in moth template had minor effect, and that a wide range of parameters and templates delivered effective learning behavior.

Experiment sequence of events

Each experiment contained three stages:

1. Pre-training Baseline (15 digits per class), used to assess naive classifier accuracy;
2. Training (equal numbers of digits per class, randomly ordered);
3. Post-Training Validation (15 digits per class), used to assess post-training classifier accuracy.

All digits were randomly chosen without replacement from non-intersecting pools of digits. Using more than 15 digits in baseline and validation sets did not significantly affect results.

In some experiments, training included multiple "sniffs", ie repeated presentations of each sample. Since the moth olfactory system can learn from few samples [69], we focused on small training sets (1-20 samples/class). The order of training samples did not matter, perhaps because the strongest plasticity was specific to the readout neuron (EN) targeting a given class and was focused on the active pixels of that class.

3.2.4 Classifiers

System readout units are the ENs, downstream from the sparse MB layer and its plastic connections. These ENs are silent absent any input sample, and they consistently spike, more or less strongly, in response to input samples. We classified test digits using a summed log-likelihood over the distributions of responses to each digit class in each EN:

$$\hat{s} = \min_{j \in J} \{ \sum_{i \in J} (\frac{E_i(s) - \mu E_{ij}}{\sigma E_{ij}})^4) \}, \text{ where} \tag{3.2}$$

$\hat{s}$ = predicted class of sample s

$E_i(s)$ = response of the ith EN to s

$\mu E_{ij} = \text{mean}(E_i(s)|s \in V, s \in \text{class } j)$

$\sigma E_{ij} = \text{std dev}(E_i(s)|s \in V, s \in \text{class } j)$

$j \in J$ are the classes (0-9)

V is a reference set (eg a validation set).

Roughly, j is a strong candidate for $\hat{s}$ if each EN's response to s is close to that EN's expected response to class j. The use of the 4^{th} power (vs the usual 2^{nd} power) is a sharpener that penalizes outliers.

This summed log-likelihood is a measure of how well the ENs can separate out the response distributions to various classes, combining information from all ENs and including information about responses to classes not targeted by a particular EN. The goal of this classifier is to assess how much discriminatory information MothNet is able to extract from the training data. We do not wish to imply this is biologically realistic (we don't know). Accuracy of naive (ie untrained) moths was about 15%, slightly higher than random guessing, perhaps because the digit "1" often elicited slightly different naive responses than other digits.

For a given experiment, the post-training classification accuracy was calculated on the validation set, ie the same set used to estimate the post-training EN response distribution parameters μE and σE. Similarly, the baseline (pre-training) classification accuracy was calculated on the same baseline set used to estimate naive EN response distribution parameters. We used the validation set to assess post-training discrimination (rather than a separate holdout set) to shorten simulation time, and also because this was sufficient for the purpose of assessing the increase in discrimination from baseline. Results for true holdout sets were roughly the same.

A more biologically plausible classifier might be simple thresholding: Label the test sample as the class of the most responsive EN_j (by Mahalanobis distance),

$$\hat{s} = \max_{j \in J} \{ \frac{(E_j(s) - \mu E_{jj})}{\sigma E_{jj}} \}, \text{ where} \tag{3.3}$$

$\hat{s}$ = predicted class of sample s,

$E_j(s)$ = response of the jth EN to s,

$\mu E_{jj} = \text{mean}(E_j(t)|t \in V, t \in \text{class } j)$,

$\sigma E_{jj} = \text{st dev}(E_j(t)|t \in V, t \in \text{class } j)$,

$j \in J$ are the classes (1-9, 0),

V is a reference set (eg a validation set).

This uses no cross-class information (E_i response to $s \in$ class $i \neq j$).

Accuracy of classification by simple threshold was in general much lower than by the log-likelihood classifier, likely because thresholding does not leverage the full information about each EN's responses to every class, but only each EN_j's response to its assigned class j. For general ML applications, biological plausibility of the classifier is not required.

3.3 Results

This section presents results of MothNet2 experiments focused on: Learning; MB sparsity; one-shot learning; sniffing; and effects of AL noise.

3.3.1 Learning experiments

We ran training experiments with various moth templates to assess their ability to learn the MNIST digits. In general, a wide range of moth templates responded well to training by differentiating their EN responses to different digits (input classes). In naive moths, all ENs had the same response profile, and responded similarly to all digit classes, as expected given the symmetry of random connection strengths to the various ENs (ie rows of $M^{K,E}$). Training caused EN responses to diverge from baseline and from each other, such that each EN responded most strongly to its assigned digit. Common effects of training included:

1. Most ENs (eg 1, 2, 6, 7, 0) tended to amplify the response to their trained digit very well, relative to control responses. This gave strong separation and accurate classification.

2. A few ENs (eg 5) sometimes poorly separated their trained digit from control digits. These were the digits most often misclassified during validation.

3. Some ENs consistently boosted the responses to certain control digits along with their trained digit (eg, EN_9 boosted 4 and 7, EN_4 boosted 7 and 9). These cases typically reflected visible similarities in the digits, and led to characteristic errors. For example, 9s, if misclassified, were typically misclassified as 4s. However, 9s were not misclassified as 7s, because EN_7 usually strongly separated 7 from 9. That is, outputs of EN_9, EN_7, and EN_4 combined were sufficient to distinguish 9 from 7, but not always 9 from 4.

These behaviors are evident in Fig 3.3, which shows timecourses of EN firing rate responses, pre- and post-training, from a typical experiment. Each subplot shows EN response to 150 digits (15 ones, then 15 two's, etc). The post-training responses are normalized by the EN's mean trained class response for clarity. Training typically increased and/or decreased all class responses of an EN, but to different degrees, resulting in the separations seen in the normalized timecourses. Training stage responses are excised from the timecourses to save space; these responses were consistently much stronger due to the stimulating effect of octopamine injected during training, and would extend past the top of the plot.

Figure 3.4 plots EN response distribution statistics (mean $\pm$ std dev) from a typical experiment, in which 13 moths were generated from a template, then trained on 15 samples per class. Post-training accuracy for moths of this template was 71-83%, starting from 14-18% baseline accuracy. Similar results, both accuracies and limitations (such as confusing 4s and 9s) held for a wide range of moth templates and training regimes.

These learning experiments indicate that the moth olfactory network, with minimal modifications, can rapidly learn to read handwritten digits. This aptitude was qualified by an apparent upper limit of about 85% on accuracy, given the downsampled thumbnails, pixels-as-features, and restricted number of training samples used in the experiments.

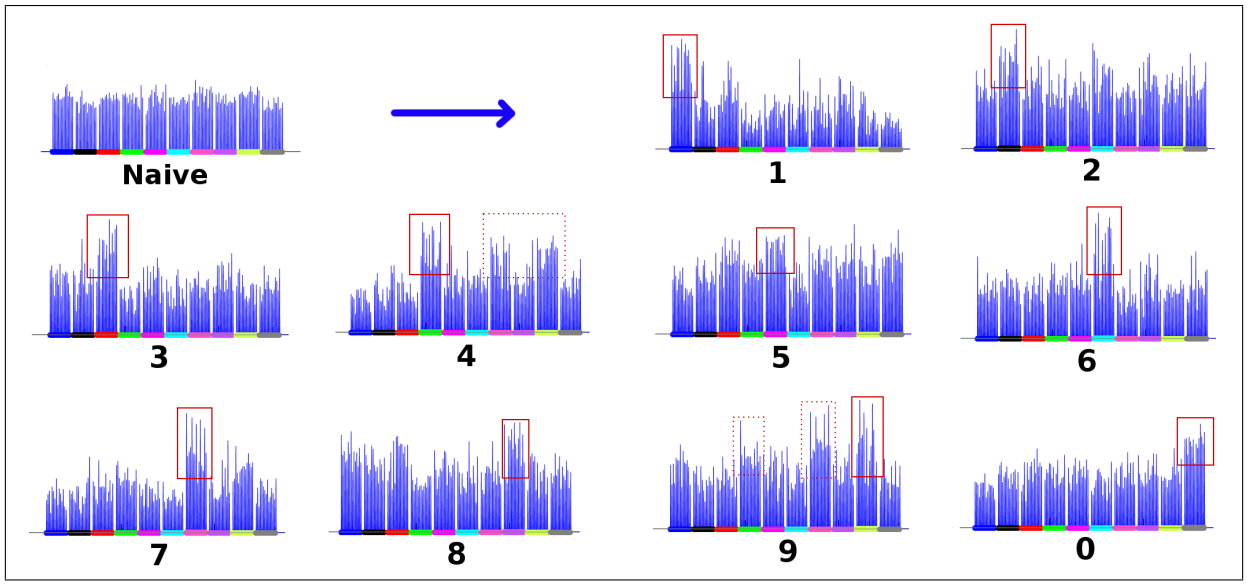

Figure 3.3: **Pre- and post-training EN time courses** (normalized) for a typical moth trained on 15 samples per class, showing post-training separation. Each timecourse shows EN response to 150 digits (15 ones, then 15 twos, etc). Top left = naive response (all ENs similar). Other subplots show trained ENs (trained class responses framed in red, some confounding class responses in dashed red).

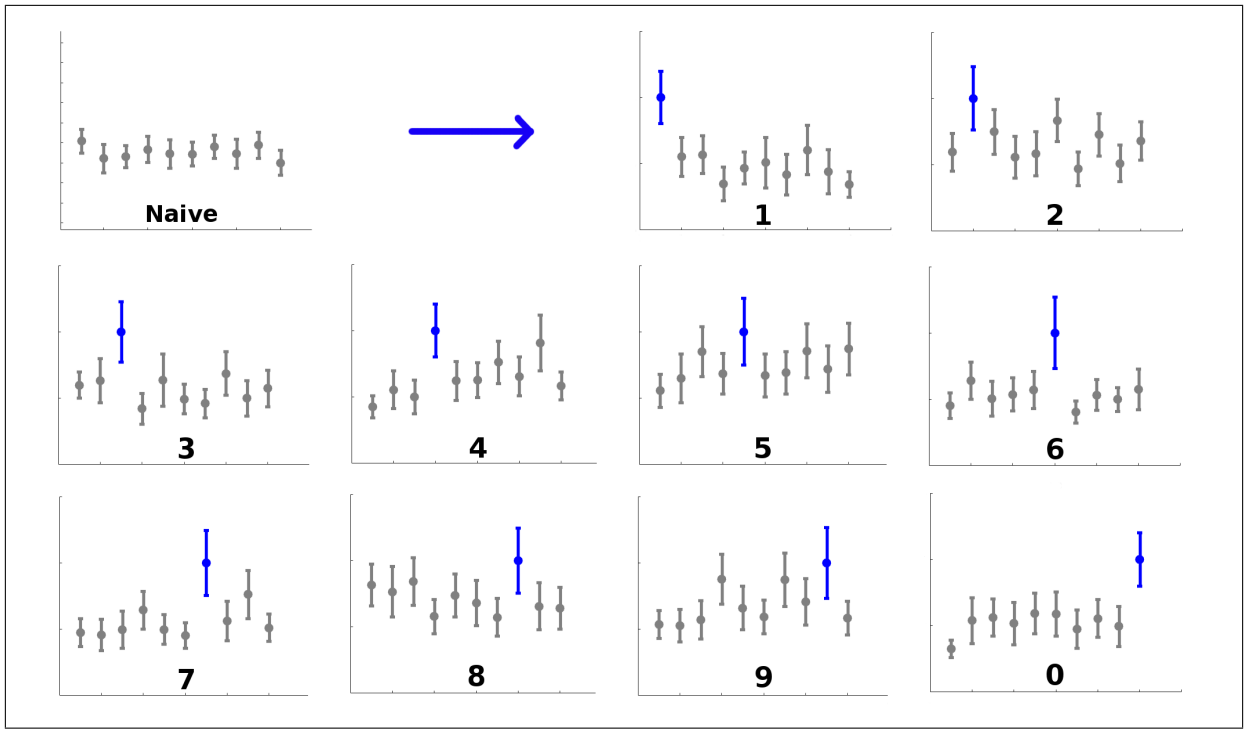

Figure 3.4: **Pre- and post-training EN response distributions** (normalized) from a typical experiment, showing post-training separation of class response distributions. Dots show mean(μ), bars show mean(σ) averaged over μ and σ from 13 moths. Mean accuracy for this template was 76%, range 71-83%. Top left = naive response (all ENs similar). Other subplots show trained EN responses in blue. 10 training samples/class.

3.3.2 Sparsity experiments

High-dimensional, sparse neural layers are a widespread motif in biological NNs [22]. We ran experiments to examine the effects of sparsity in the context of learning the MNIST digits. We trained a reasonably capable moth template, varying only the sparsity level in the MB during training (17 moths per sparsity level). Sparsity here is measured as the fraction of MB neurons that are responsive to stimuli (1% is very sparse, 50% is very dense). In MothNet2, MB sparsity levels (both non-training and training) are parameters. Sparsity levels in the MB affected two crucial behaviors: Intra-class SNR of EN responses; and how well a given EN_j's response to training focused on class j.

Results mirrored sparsity results in chapter 2 section 2.3.4, and showed that the sparse MB layer plays a key role in learning by controlling and focusing Hebbian weight updates.

High MB response fraction (ie high density or low sparsity) correlated strongly with high SNR (ie reliability of intra-class EN responses). But it also resulted in poor post-training classifier accuracy, because Hebbian growth boosted all weights due to the excess of active neurons, and thus weight changes were not restricted to just the most class-relevant MB signals.

Conversely, low MB response (ie high sparsity) resulted in low intra-class SNR, since not enough MB neurons were firing in response to stimulus to reliably activate the ENs. But high sparsity correlated strongly with high "learning focus", ie the tendency for training to focus gains on the correct class, resulting in stronger post-training accuracy.

Post-training discrimination accuracy appeared to represent an optimized trade off between intra-class SNR and learning focus. The tradeoff is shown in Figure 3.5, which plots these effects of different sparsity levels in the MB, as they relate to learning:

1. The descending red curve shows "learning focus", a figure-of-merit defined as the average standardized distance between EN response distributions to trained and control classes,

$$LF = \frac{1}{9}\sum_{i \neq j} \frac{(\mu E_{jj} - \mu E_{ji})}{0.5(\sigma E_{jj} + \sigma E_{ji})}, \text{ notation as in eqn 3.2.} \tag{3.4}$$

This is average Bhattacharyya distance, if EN response distributions are gaussian. It generalizes eqn 2.6. In a very sparse MB, training is strongly focused on the trained class, while in a dense MB, training "raises all boats" and the trained response distributions are very poorly separated. Thus high sparsity focuses learning well.

2. The ascending black dotted curve plots the mean intra-class signal-to-noise ratio (eqn 2.5). This is an opposite situation: A very sparse MB results in high intra-class variance in trained EN responses (low SNR), while a denser MB delivers much more consistent within-class responses (high SNR).

3. The domed blue curve is a fit to classifier accuracies from experiment.

This plot echoes Fig 2.8 in section 2.3.4, on a somewhat more complex model and with the addition of a natural objective function, namely classification accuracy on a test set. Judged by best post-training classification accuracy, optimal MB sparsity level for MothNet2 appears to represent a compromise between delivering sufficient intra-class SNR and sufficient learning focus for inter-class distinctions. This optimal region lies somewhere between 5-20%, consistent with the 5-15% commonly observed in biological NNs.

3.3.3 Growth rate effects, one-shot learning

The speed at which MothNet learns, ie the number of training samples required to reach maximum accuracy, is determined to large degree by the Hebbian growth rate on MB→EN connections and the related decay rate on MB→EN connections. When the connection weights hit the rails of their dynamic range, no further learning is possible. We ran experiments that varied growth rates, to see its effect on accuracy for various training set sizes. One moth template (the "slow learner") had a biologically plausible growth rate (per wet-lab data), while the second moth template (the "fast learner") had a growth rate "turned up to 11". High growth rate moths allowed us to test MothNet2's skill at one-shot learning (ie with just one training sample).

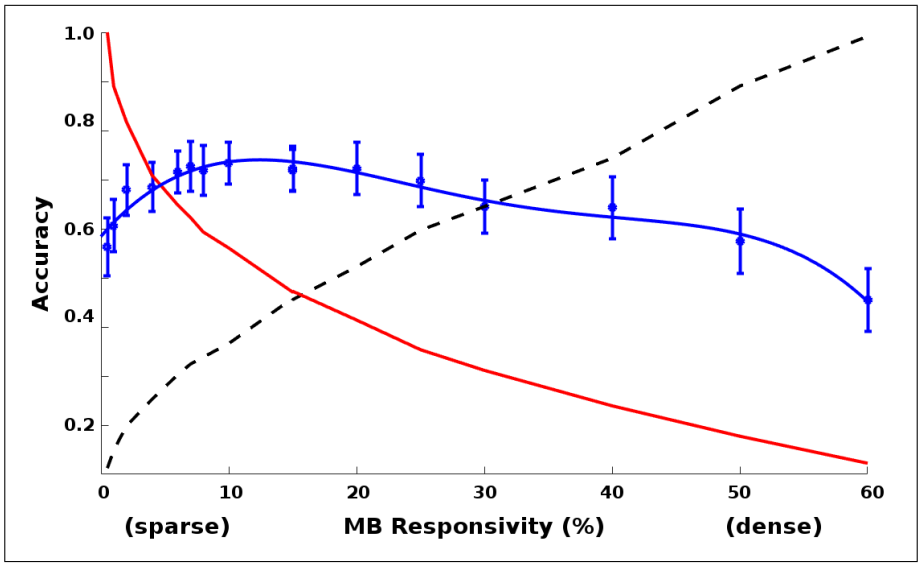

Figure 3.5: **Effects of sparsity in the MB:** Optimal accuracy (blue domed curve, $\mu \pm \sigma$) occurred at 5-20%, a compromise between learning focus and high intra-class signal-to-noise ratio (SNR). Descending red curve = mean separation of trained vs control (learning focus). Black ascending curve = mean intra-class SNR. Learning focus and SNR are scaled for plotting. 17 moths per sparsity level.

Fast learners attained strong immediate accuracy (mean 75%) on just one training sample, but more training gave no further gain. Slow learners took several (∼20) training samples to attain maximun accuracy, but that accuracy was higher (∼80%). When classification was done by simple thresholding (eqn 3.3) rather than the log-likelihood method (eqn 3.2, the long-term advantage of slow over fast learners was more pronounced (43% vs 33%). Figure 3.6 (panel A) shows the trade-off of speed of learning vs maximum attained accuracy. The solid curves show fast- and slow-learner accuracies vs number of training samples. The dashed curves show results from the same experiments, using simple thresholding as an alternative classifier. The effect is similar, but the short-term/long-term contrast is stronger.

Figure 3.6 (panel A) incidentally shows the lower accuracy delivered by using thresholds on EN responses (eqn 3.3 rather than the log-likelihood classifier (eqn 3.2. The difference seen here (80% vs 40%) was typical.

3.3.4 Sniffing, effects of AL noise

Sniffs:

Biological NNs have a remarkable ability to learn from very few training samples. In addition, "sniffing" behavior (repeated sampling of a given odor) is a common biological strategy [84]. We ran experiments to see whether sniffing behavior improves the various stages of learning. Sniffing applied to test samples did not improve test accuracy (experiments not shown).

However, sniffing during training had a large effect, especially in one-shot regimes. When a "slow learner" (in which the template's learning rate was set to a biologically reasonable level) was given a single training sample (one-shot), multiple sniffs raised post-training accuracy from 35% to over 60%. Five sniffs delivered maximal increase, with no further gains from additional sniffs. Figure 3.6 (panel B)) shows this increase in accuracy due to sniffing in the one-shot context. Multiple sniffs (up to 5) during training dramatically improved accuracy. More than 5 sniffs yielded no further benefit.

AL noise:

The moth AL is a very noisy system. That is, neural responses to a given odor stimulus (or absent any odor) have high variance. To see whether this AL noise was beneficial to learning performance, we also varied the AL noise level during the above sniffing experiments, from near-zero through the high end of biologically reasonable (per wet lab data). We had expected the presence of AL noise to improve performance, given multiple sniffs, by acting as a kind of sample augmentation, analogous to that used in DNN training. In fact, noise levels in the AL had no effect on accuracy. This is seen in Figure 3.6B, where each cluster of mean ± std dev bars show different levels of AL noise for a given number of sniffs.

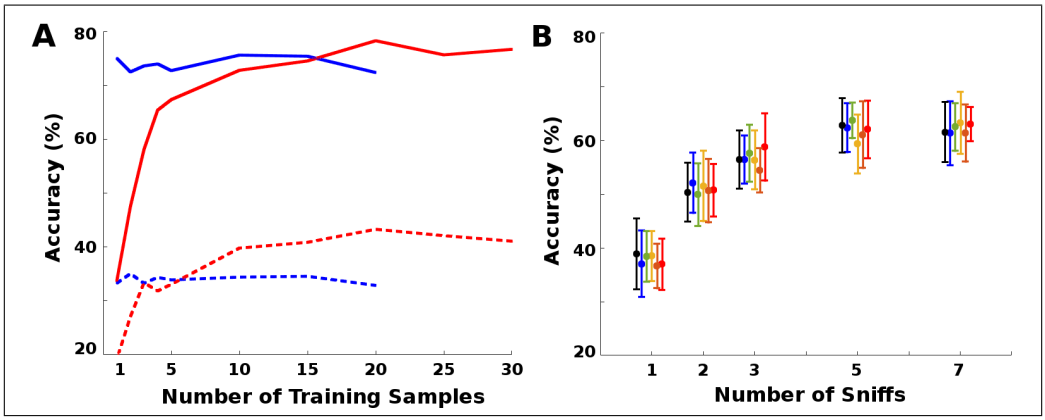

Figure 3.6: **Growth rates and sniffing. A:** Effects of growth rate. Solid horizontal blue curve = very fast learner. Solid ascending red curve = slow learner. The fast learner attained 75% accuracy in one-shot, but with no further gains. The slow learner ultimately attained higher accuracy. Dotted lines show the same effect in accuracy using threshold classifier. **B:** Effects of sniffing and noise on one-shot learning, in a "slow-learner" moth template. Multiple sniffs greatly improved one-shot accuracy. Noise in the AL: The clusters of $\mu \pm \sigma$ bars represent varying levels of AL noise, with low-to-high noise plotted left-to-right at each x-axis location. AL noise level did not affect accuracy. 13 moths per data point.

3.4 Discussion

3.4.1 A biological toolkit for building NNs

In order to learn new odors, the moth olfactory system uses just a few core tools: A noisy pre-amp network with competitive inhibition; Hebbian plasticity controlled by a high-dimensional sparse layer; and generalized (global) stimulation during training.

Our key finding is that NNs built with this toolkit can succeed at a general learning task. MothNet2, a simple combination of these elements structured according to the moth's olfactory architecture, learned to read MNIST digits, increasing accuracy more than 5x (from 15% to 75-85%) given only a few training samples (1-20 per class), almost giving *Manduca sexta* a shot at a post office job. Also, because the key weight updates focus entirely on activity induced by the class being trained (not on activity induced by control classes), new classes can be added and trained without retraining on existing classes.

We note that MothNet2 stayed close to the moth's very simple olfactory architecture, and used a simple feature set (pixels). A free hand with network design, such as using better input features and more complex architectures, might well yield stronger results.

We recognize that other methods, eg an L_2 Nearest-Neighbor algorithm, can learn to classify MNIST digits [45]. However, comparing disparate methods is not our purpose [31]. Our goal is to provide a proof-of-concept as to the abilities of even the simplest of biological NNs, and to clarify how a NN built from biological elements learns a predictive model for a general ML task.

This finding is of interest because the biological elements analyzed here are well-suited to being combined and stacked into larger, deeper neural nets, just as convolutional layers, maxPool, etc are combined to build current DNNs (by contrast, Nearest-Neighbor is not suitable for stacking). The success of actual biological neural nets at a wide range of tasks suggests that these biologically-based NNs (BNNs as it were) may also prove to be effective learners in the ML context. In addition, they offer biological plausibility, the possibility of learning from few samples, and expandability of classes without retraining, all of which are

areas of weakness for current DNNs.

3.4.2 Role of the sparse layer

A second finding, reinforcing results from chapter‘2, emphasizes the key role of sparse layers during learning. Sparse, high-dimensional layers are a widespread motif in biological neural systems, in particular related to memory and plasticity [22]. In the moth, the sparse layer (MB) plays a vital role in learning because the plastic synapses connect into or out of the sparse layer, allowing it to modulate the Hebbian updates to the synaptic connections via the AND gate nature of Hebbian growth. This ensures that learning boosts the important signal (ie the signal associated with a given samples class) and not artifacts. While the sparse MB layer calls to mind the backprop sparse autoencoder [61], the biological role described here has no obvious analogue in backprop sparse encoders, since it is tied to the Hebbian update method. However, biological sparse layers may also perform functions similar to those found in backprop sparse autoencoders, such as reducing noise [81]. and reducing the dimension of the feature space, in an effort to match the complexity of the network to the essential dimension of the classification task [55].

3.4.3 Role of octopamine

Unanswered by these experiments is whether generalized stimulation by octopamine is required in ML systems such as MothNet2, distinct from actual biological systems. In the moth, octopamine stimulation may offer a work-around to avoid biological constraints on Hebbian growth rates and input intensity. That is, octopamine may act primarily as an accelerant. However, engineered NNs can easily crank up growth rates, enforce higher MB activity, and amplify signals during training, all without recourse to the octopamine mechanism but perhaps with the same beneficial effect on learning. In this case, generalized stimulation would not be a necessary part of the biological toolkit described here for application to ML tasks. However, it may be that octopamine stimulation also enables exploration

of the coding solution space not normally activated by stimuli. Alternate ways to replace this functionality in the ML context are not so obvious.

3.4.4 Role of noise

Also unexplained is the role, if any, of high intrinsic noise in the Antennal Lobe during learning. That noise carries no penalty to learning would partly explain its existence in any biological system which found noise desirable for other reasons. However, our experiments found no positive reason to build high intrinsic noise into the AL (or any NN layer) in an ML context. However, a noisy pre-amp (eg the AL) might still be actively beneficial, despite our results, for three reasons. First, when coupled with sniffing, a noisy AL might provide a version of data augmentation (as used in DNNs) by distorting the codes delivered to the MB and readout neurons. This would improve one-shot or few-shot learning, if the distortions induced by the noisy AL somewhat mimicked the within-class sample variation (this condition was likely absent in our MNIST experiments). Second, injecting noise into input layers, or corrupting training samples, can improve NN classification performance [1], suggesting a concrete benefit during training (though perhaps not during classification). Third, injecting noise may be a useful or even necessary way to explore the solution space [7].

3.4.5 Future work

In order to prove out and adapt this biological toolkit to the ML context, some topics invite further study.

A fast substitute for time evolution of ODEs

The ODE time evolutions used in these simulations are computationally expensive. However, both the moth olfactory network and MothNet2 have an episodic response to stimuli: A single stimulus triggers a simple, discrete system response, with rapid return to equilib-

rium and no further time-varying effects such as periodic local field potentials. Thus, the ODE time evolutions used to run MothNet2 are not necessary, if the episodic response and Hebbian updates can be captured by simpler and faster methods. Development of simpler substitutes that retain the net effect are in fact necessary in order to test more complex neural architectures. A downside of this substitution is that it removes the option of new tools based on subtle, less-discrete effects of neural activity as found in other biological NNs.

The value of noise

The ubiquity of high system noise in biological NNs, and findings from autoencoders, suggests that noise at a pre-amp layer has concrete benefits, despite the lack of benefit found in our experiments. Clarifying whether, how, and under what conditions high intrinsic noise improves learning would potentially allow its use as part of a biological toolkit for ML.

More complex architectures

By analogy with DNNs and with biological neural systems, combining the basic biological elements described above into deeper NNs may yield improved performance on ML tasks. This involves experimenting with new architectures, such as parallel AL→MB→EN branches for ensemble averaging, and targeting new ML tasks.

A bigger biological toolkit

Because the moth olfactory network is so simple, and because biological neural systems are so diverse and complex, extracting a toolkit from the moth is like going into a giant hardware store and coming out with a hammer, a screwdriver, and a saw. Countless other useful elements might be abstracted from other biological NNs. This involves accurately modeling these more complex systems and analyzing how they learn, an open-ended task.

Chapter 4

BUILT TO LAST: INJURY MITIGATION MECHANISMS IN THE MON

4.1 Introduction

For biological systems, injury is inevitable. Yet they need to maintain function despite damage or else they perish. For example, progressive damage to a honeybees wings results in shorter foraging trips and less food gathered [33], while bumblebees can adapt to some types of wing injury, resulting in no loss of load lift [71]. This need to maintain function despite injury holds for neural systems in particular. Human-built computing hardware operates in a regime of near-zero tolerance for physical damage. By contrast, biological neural systems necessarily operate in a different regime [67], where lack of tolerance for damage is a ticket to the morgue. However, robustness to injury is often overlooked when analyzing the purpose and function of neural structures. Other design priorities, such as maximum information, high SNR, and low energy consumption are primarily considered. Analyzing neural systems in the context of these design specs is reasonable, but arguably incomplete.

The goal of this chapter is to examine whether and how certain neural mechanisms and architectural structures can be understood as adaptive, built-in systems for robustness to neural injury. That is, we examine how biological neural systems are built to last. Our approach is to run injury experiments on a computational model of the moth olfactory network (MON). This model, hereafter "MothNet3", is a variation of MothNet2, the model developed in chapters 2 and 3. The MON is a simple neural system, fairly well-characterized in the literature, which contains the key architectural elements under test. These elements, widespread in biological neural systems, are (i) the ability to learn (plasticity), (ii) high levels of noise, and (iii) inhibitory feed-forward channels running parallel to excitatory channels.

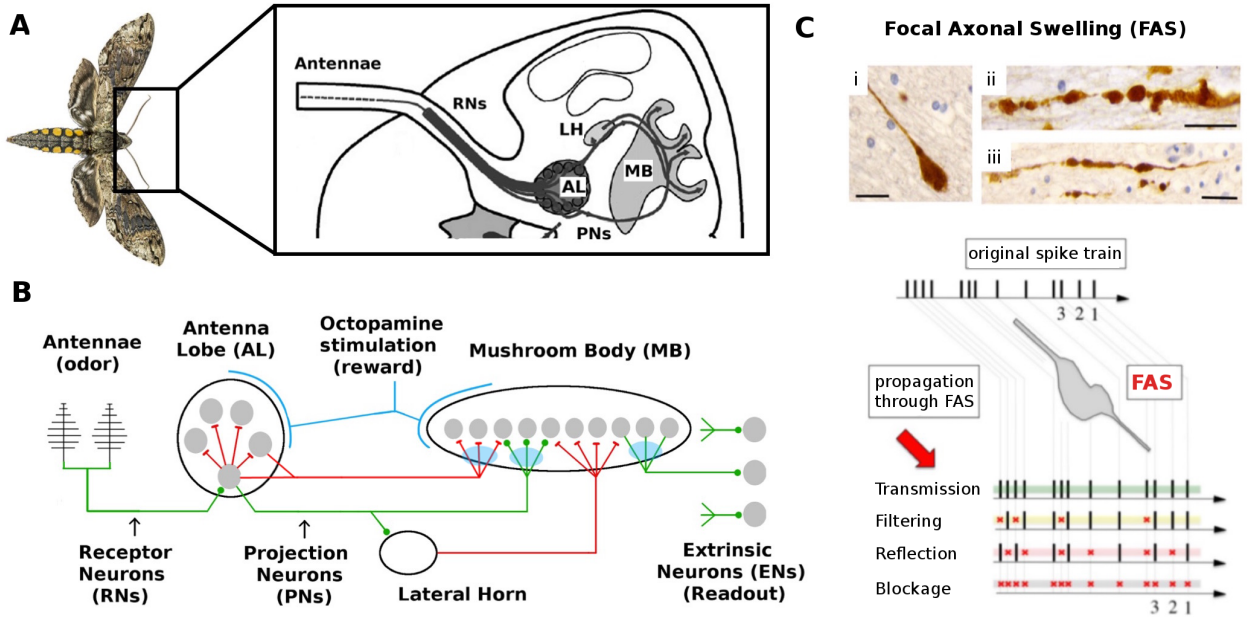

Figure 4.1: **Schematic overview of the Moth Olfactory Network** (MON) **and axonal injury mechanisms. A, B:** The MON is organized as a feedforward cascade of five distinct subnetworks and a reward mechanism. **C:** Focal Axonal Swellings (FAS) are ubiquitous across all severities of traumatic brain injuries and present in other leading brain disorders. They can cause some or all neural spikes in the train to die off in transit, reducing the overall firing rate arriving at the downstream target neuron. Adding FAS effects to the MON are the basis of our damage/injury protocols. Panel A is adapted from [18] and Panel C from [54].

Experiments with MothNet3 led to four main findings concerning injury-mitigation structures in the MON. Schematics for possible mechanisms are given in Fig 4.8.

1. The moth's learning mechanism, based on stimulative neuromodulators (eg octopamine) and Hebbian growth, can compensate for injury by restoring downstream neural responses that have been degraded by injury to upstream neural firing rates (FRs).

2. The existence of inhibitory neurons, parallel to excitatory neurons, that feed-forward from layer to layer, can mitigate the effects of injury by spreading injury among neural types, with excitatory and inhibitory losses cancelling out.

3. High levels of neural noise, ie a broad noise envelope on the FRs, can protect against effects of neural injury by ensuring that even when mean FRs decrease, some stimulus responses are still strong enough to exceed key action-triggering thresholds in downstream neurons if the neuron's FR std dev is large compared to the mean loss in FR due to injury. We note that neurons in the moth antenna lobe (AL) have high levels of noise.

4. Simple ablation injury to an upstream region produces downstream effects that are distinct from those of naturalistic (focal axonal swelling, or FAS) types of injury. The theoretical conversion rate between the two injury types fails entirely, in terms of their effect on downstream readout neurons. In the MON, the actual conversion rate varies according to location of injury. Thus ablation is not a reliable proxy for naturalistic injury.

4.2 Methods

In this section, we first describe MothNet3, the computational model used in these experiments We then describe focal axonal swelling (FAS), a characteristic neural injury which we use as a model of damage, and how it was applied. Lastly, we give some details about the experimental setups involved in our key findings.

4.2.1 MothNet3 architecture

We use a model of the moth olfactory network, MothNet3. This is a slight variant of the models developed in chapters 2 and 3, with architecture modified as needed for each experiment. The relevant structures of the MON, for these experiments, are:

Antennae:

Chemical receptors detect odor and send signals to the Antenna Lobe via receptor neurons (RNs). There are roughly 30k Receptor Neurons (RNs) that deliver odor signals to the AL.

Antennal Lobe (AL):

The AL acts as a pre-amp. It contains ∼60 neural units (glomeruli) which process odors and send excitatory signals via projection neurons (PNs, ∼5 per glomerulus) and inhibitory signals via QNs downstream to the mushroom body (MB) and readout neurons. QNs were structured as follows: Each was innervated in one glomerulus, rather than in several (as in the actual MON). QNs were treated identically to PNs, except that they were inhibitory on KCs. The ratio of QNs:PNs varied from 0 to 1.4 according to the experiment.

Readout Neurons (Extrinsic Neurons, ENs):

Signals from the AL pass to the MB's Kenyon cells, and these in turn feed-forward to the Readout Neurons (ENs), which are assumed to act as decision neurons, with strong EN responses triggering actionable messages (such as "fly upwind").

Although the MB is central to the MON and to learning, its dynamics are not directly relevant to the findings of this chapter. Thus, we focus on the ENs, which represent the final, actionable output of the system. MothNet3 posits one EN, whose output firing rates serve as a measure of the functional effects of upstream injury.

4.2.2 Moth template parameters

In each experiment, several moths (>60 per data point) were randomly generated from a template defining the architecture, eg numbers of neurons, distribution parameters for synaptic connection weights, and how odor projected onto the glomeruli of the AL.

The templates used were realistic in the senses of having (1) PN firing rate behavior matching wet-lab data from live moths and (2) architecture parameters that match what is known from the literature (ie as in chapter 2). Some templates were moved to the boundaries of, or out of, a known realistic regime by varying key parameters-under-test as required by the experiments:

1. The number of QNs per glomerulus varied from 0 to 7, in order to test the injury-

mitigating effect of inhibitory QNs in parallel with PNs feeding-forward AL→MB. PNs were set to 5 per glomerulus, as in live moths. In live moths, each QN is innervated by several glomeruli. In these experiments, each QN was innervated by one glomerulus, like PNs, to provide symmetry to the PN-QN structure and to make QN:PN ratios meaningful. Live moths may have QN:PN ratio of (very) roughly $\leq 20\%$ (ie relatively few QNs), insofar as a ratio can be estimated. Actual values are not known.

2. The level of AL noise, affecting all AL neurons, was varied from 0 to 1.33, where 1.0 represented the noise level of the templates fitted to wet-lab data from live moths. The purpose was to test the injury-mitigating effects of AL noise levels.

3. For experiments testing the effects of learning on injury reduction, a template found to match wet-lab data was used, but with number of QNs = 0 per glomerulus. Setting the number of QNs equal to 2, ie QN:PN ratio = 0.4, gave similar results. These QN numbers bracket the estimated live moth value, and these models remained close to realistic models.

Because some parameters regimes described above deviated from calibrated models, and because moths were randomly generated, sometimes a moth's naive EN responses to odor were dysfunctional in the sense of being untenably noisy. Moths with naive odor:noise ratio (ie EN odor response over EN spontaneous output) < 12 were discarded as being outside a plausibly realistic envelope. These comprised about 12% of moths generated, with the percentage depending on the varied parameters: Templates with high numbers of QNs and/or very high AL noise had more rejected moths; templates with few QNs and "normal" or low AL noise had few rejected moths. Extra moths were generated as needed to fill out numbers.

4.2.3 Focal axonal swelling

FAS is a neural injury associated with traumatic brain injury (TBI), typically caused by physical shock. Examples in current events include blast injuries from recent wars, as well as impact injuries in contact sports. FAS presents as swollen neural axons (the signal delivery pipelines), with diameter changes of up to 30x. This swelling causes signals from the upstream source to be diminished or lost entirely before reaching downstream target neurons [53]. This degradation can be expressed as reduced FRs from upstream neurons, characterized in a computational model by [54], which found that signals traveling down an injured axon are attenuated to greater or lesser degree according to the amount of swelling and the firing rate of the signal. While ablation is a ready and oft-used means to model neural injury, it imposes a binary "all-or-nothing" effect which is not present in FAS injuries. In these experiments we model neural injury according to [54], hereafter"FAS type" or "FAS".

Location of injured regions

Two sites, RNs or PNs/QNs, were targeted for injury. Only one site was injured at a time, the choice determined by the experiment. Locations are shown in Fig 4.2.

RNs: One likely site of FAS injury was posited to be the antennae and RNs, due to their exposure to external impacts.

Roughly 30k RNs are evenly distributed across each antenna. Roughly 500 RNs respond to each of 60 atomic volatiles, and send their inputs to a single glomerulus in the AL where their inputs are averaged to reduce noise. The receptors for a given glomerulus are distributed across the antennae (not concentrated in one spot). Thus we expect injury to an antenna to affect each glomerulus' RN input roughly equally, and to affect a roughly equal percentage of each glomerulus' 500 inputs.

PNs/QNs: There is a channel that carries PN (and QN) axons from AL→MB. We modeled damage to this channel by injuring both PNs and QNs with equal probability.

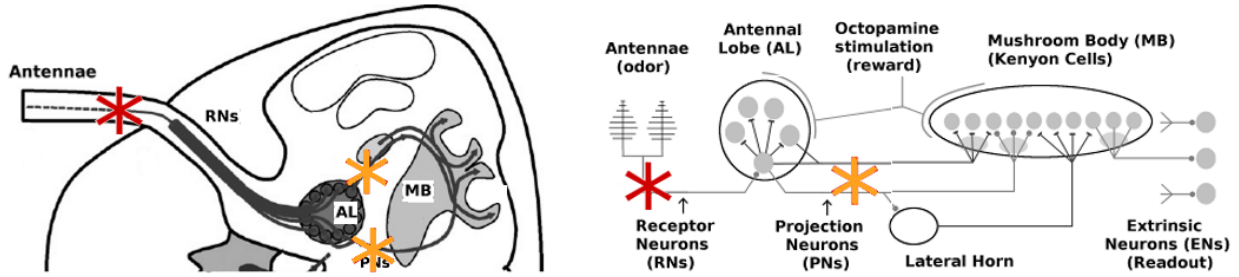

Figure 4.2: **Location of injury in experiments.** Red stars: Damage to Antennae (RNs), which reduces input to the AL. Orange stars: Damage to the AL→MB channel, which reduces signals passed by PNs and QNs to the MB. Left-hand diagram is adapted from [18].

We assumed that Lateral Horn behavior was unaffected by PN and QN injury, ie sparsity percentages enforced on the MB remained stable, since the lateral horn is enervated by the same PN/QN signal that enervates the MB.

Injury methods

FAS due to physical trauma does not affect all neurons in a targeted brain region equally, nor does it operate in an "all or nothing" way [83]. We used the injury regime derived in [54], which calculates the fractions of injured neurons falling into each of four injury types: FR unaffected (transmission); FR cut by half (reflection), FR destroyed (ablation), or FR filtered according to

$f_{injured}(s) = F(f_{healthy}(s))$, where $f_* =$ firing rate, s is a stimulus, and F is a lowpass filter.

Injury fractions were as follows: 15% transmission, 35% reflection, 35% ablation, and 15% low-pass filtering. Neurons in the target group were randomly selected for injury according to the percentage specified in the particular experiment (0 to 60%). Each injured neuron in the target group was then randomly assigned one of the damage types.

Applying this injury regime to populations of PNs and QNs in the model is straightfor-

ward, since these neurons are modeled one-to-one (ie one neuron in the model represents one actual neuron).

Injury to RNs was handled differently than injury to PNs, because in MothNet3 each RN (inputting to one glomerulus in the AL) stands in for ∼500 RNs in the live moth. Injury of RNs worked as follows:

1. The FAS injury level was converted into a theoretically equivalent ablation injury level n, where $100n = \%$ ablation (see formula in section 4.2.4).
2. Each RN's FR was multiplied by $(1 - n)$ (attenuation), since the glomerulus had fewer inputs.
3. The RN noise parameter was multiplied by $\frac{1}{\sqrt{1-n}}$, since the glomerulus was averaging fewer inputs, so there was less noise reduction from averaging.

4.2.4 Simulation protocols

Each experiment consisted of many moths, all generated from the same template with only the parameters-under-test varied. Over 20 moths (trials) were run for each parameter combination (eg "4 QNs, 50% injury, and 10 training odor puffs"), giving over 60 moths for each key parameter pair (eg "4 QNs, 50% injury") with training on 5, 10, or 15 odor puffs. A single trial consisted of first injuring and then training a single moth, in five stages:

1. Pre-injury baseline: Odor without octopamine (15 odor puffs, each 0.2 mSec), to assess naive EN odor response.

2. Injury applied.

3. Post-injury odor: Odor without octopamine (15 odor puffs), to assess the effects of injury on EN odor response.

4. Training: Odor plus octopamine (5, 10, or 15 odor puffs).

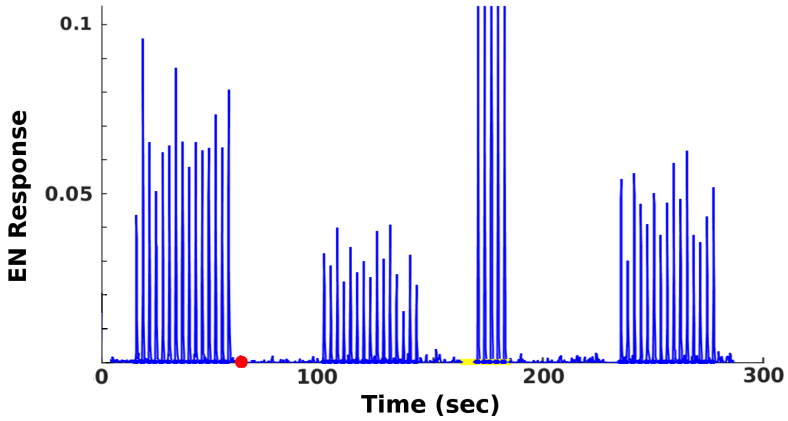

Figure 4.3: **Typical EN timecourse.** Readouts from the EN in a typical experiment, in which injury attenuated the EN odor response, and training partly restored it. Naive response (20-55), injury (red dot at 60), injured response (100-150), 5 puffs training (high response, with yellow, 170-190), post-training response (240-280).

5. Post-Training odor: Odor without octopamine (15 odor puffs), to assess post-training EN odor response.

Firing rate $f(t)$ from a single EN was recorded, to track the actionable effect of injury and training on the system. A timecourse of EN firing rates, from a typical experiment, is shown in Figure 4.3.

4.2.5 Experimental details

Learning compensates for injury (Finding 1)

Moth templates were biologically plausible, in the sense that their AL behavior matched wet-lab data, and they demonstrated learning behavior. Moths in this experiment had 0 QNs, ie no feed-forward inhibitory signals from AL→MB (2 QNs per glomerulus gave very similar results). In two separate experiments, either the RN channel (antennae) or the PN/QN channel (AL→MB) was injured with FAS (0% to 60%).

Parallel inhibitory neurons protect EN responses (Finding 2)

Moths were generated from biologically plausible template but with 0, 2, 4, 5, or 7 QNs per 5 PNs (equivalently, per glomerulus), ie QN:PN ratios were 0%, 40%, 80%, 100%, or 140%. AL noise level was set to 1.0, the value derived from calibration to wet-lab data.

The PN/QN channel was injured with FAS (0% to 60%). PNs and QNs were treated equivalently in terms of injury.

AL noise preserves the highest EN responses (Finding 3)

Moths templates had AL noise level between 0 to 1.33, where 1.0 represents the AL noise level of the model as fitted to wet-lab data, ie live moth behavior. Moths in this experiment had only excitatory PNs (ie 0 QNs). Moth templates with 2 QNs per glomerulus gave similar results.

The RN channel was injured with FAS (0% to 60%).

Effects of ablation vs FAS injury (Finding 4)

Simple ablation of neural channels is a method commonly used to simulate damage to networks. In order to compare the effects of simulating injury with ablation versus with FAS regimes, we ran parallel experiments using ablation instead of FAS injury. Injury levels were set between 0% to 60%, using either FAS type or ablation.FAS injury fractions were 15% transmission, 35% reflection, 35% ablation, and 15% low-pass filtering ($\sim$ 0.9x in most cases). The theoretical conversion rate from FAS type to ablation is as follows:

100 units FAS $\approx$ 15*1 + 35*0.5 + 35*0 + 15*0.9

$= 15 + 17.5 + 0 + 13.5 \approx 46$ units survival

$= 54$ units Ablation, for a conversion rate ≈ 1.85.

For example, 20% ablation nominally converts to 20*1.85 = 37% FAS injury.

For ablation, neurons were randomly selected from the target group according to the injury percentage specified. Ablation to RNs was handled as in section 4.2.3.

4.3 Results

This section gives results of experiments focused on the various findings, F1 through F4.

4.3.1 Learning compensates for injury (F1)

To test whether the moth's learning mechanism itself (ie plasticity via Hebbian updates plus neuromodulatory stimulation) acts to compensate for deficits induced by FAS injury, we conducted two experiments, each injuring a different upstream region of the network (locations shown in Fig 4.2).

In the first experiment, RNs (ie Antennae→AL) were injured (Fig 4.2, red stars). In the second experiment, PNs (ie AL→MB) were injured (Fig 4.2, orange stars). AL noise was set to naturalistic levels (ie realistic per wet-lab data), and #QNs = 0 (#QNs = 2 gave similar results). Injury levels ranged from 0% to 60%, and moths were subsequently trained with 5, 10 or 15 odor puffs. (The vast majority of learning occurred within the first 5 odor puffs.) Mean EN response to a series of odor puffs was recorded, as a measure of the actionable output of the system. A typical timecourse is shown in Fig 4.3. In each experiment, over 60 moths were generated from template and tested for each injury level.

In both experiments, training restored some of the lost EN response due to injury, though the moths were much more robust to RN (antennae) damage. Restoration was complete at injury levels $\leq$20% for RN damage, and $\leq$8% for PN damage. At these injury levels, EN odor responses were reduced to about $\sim$70% of naive baseline and training restored them fully. Fig 4.4 plots these results (left column shows RN injury, right column shows PN injury).

At low injury levels, the system was able to boost EN output by about 140%, a value constrained by limits on the synaptic connection weights (the moths were given sufficient training to saturate). However, at high injury levels learning was not as proportionally effective (green curves in Fig 4.4).

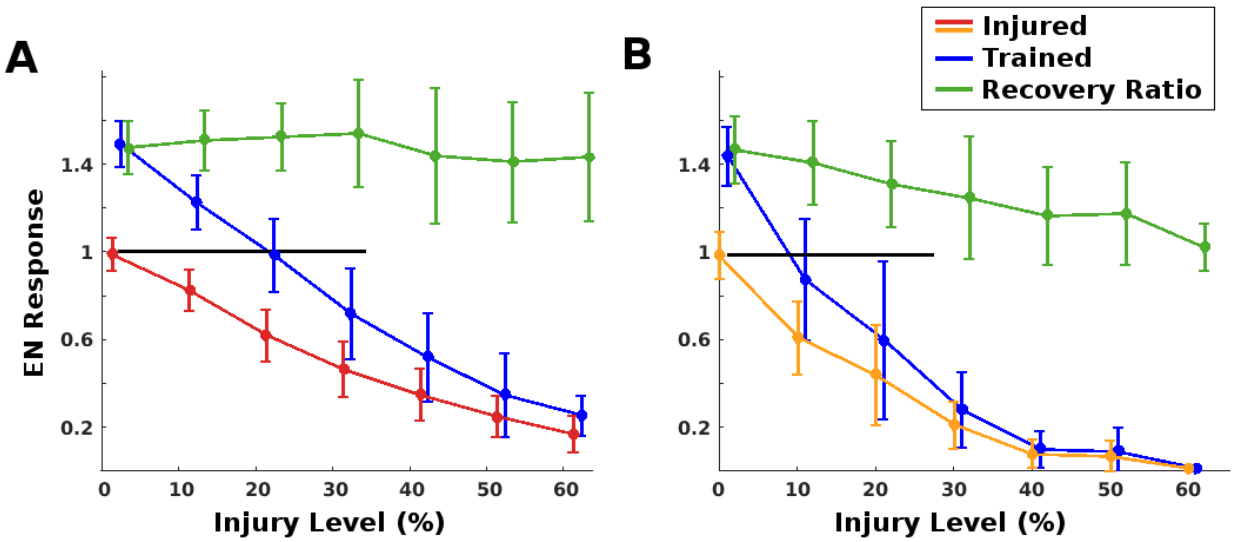

Figure 4.4: **Learning as injury compensation mechanism.** Red/orange: Post-injury EN odor response, normalized by naive, healthy odor response. Blue: Post-training EN response, normalized by naive, healthy odor response. Green: Relative increase from post-injury response due to training. $\mu \pm \sigma$. **A:** Injury to RNs: Trained EN responses (blue) fully regained their pre-injury levels (black line) from injured levels (red) if injury was on average $\leq 20\%$. The ability of training to recover lost ground was fairly steady vs injury level (green curve). **B:** Injury to PNs was more traumatic: Post-injury EN response (orange) was lower, and trained responses (blue) fully regained pre-injury levels if injury was on average $\leq 8\%$. Also, the ability of training to recover lost ground decreased as injury level increased (green curve). Each datapoint show the mean and std dev, over 60 moths, of mean EN odor response.

4.3.2 Parallel inhibitory neurons protect EN responses (F2)

In the moth, five (5) excitatory PNs feed-forward from the AL→MB from each glomerulus. To test whether the presence of parallel inhibitory neurons would mitigate the effect of injury to this AL→MB channel, we posited the existence of QNs that behave identically to PNs, except that they are inhibitory on the MB rather than excitatory. Note that this architecture is slightly different from that of the actual moth, where these parallel QNs are innervated by several glomeruli. We note that live moths have relatively few QNs, perhaps the (rough) equivalent of ≈1 QN per glomerulus (actual values are not known). We tested

moth templates with 0, 2, 4, 5, and 7 QNs per glomerulus. In this experiment, injury was applied to the PN/QN channel (AL→MB) only (orange stars in Fig 4.2). AL noise was set to a natural level, as in experiment F1 above.

As expected, high numbers of QNs correlated strongly with reduced effects from injury. Moths with high QN counts had stronger post-injury EN odor responses. They were also able to fully recover from much higher levels of injury than moths with few or no QNs. For example, full recovery of EN responses occurred at ∼8% when #QNs = 0; at ∼15% when #QNs = 4; at ∼30% when #QNs = 7; Post-injury EN responses are given in Fig 4.5A. Post-training EN responses are given in Fig 4.5B.

High QN counts had another, unexpected advantage. Let signal-to-noise ratio (SNR) = $\frac{\mu(F)}{\sigma(F)}$ where F = the set of discrete EN response FRs to odor puffs (same as eqn 2.5). Naive (uninjured, untrained) SNR was similar for all QN counts (Fig 4.5E). Post-injury, SNR dropped according to severity of injury, but higher QN counts substantially reduced losses to SNR, perhaps because raw EN firing rates were better preserved (Fig 4.5F).

High QNs carried a non-trivial downside, namely much lower EN signal relative to spontaneous FR (noise), ie $\frac{\mu(F)}{\mu(s)}$ where F = discrete EN odor response FRs and s = spontaneous EN FR. This is seen in Fig 4.5D. Note that Fig 4.5D shows stats only from moths that passed the naive low spontaneous noise test (see section 4.2.2). Templates with high QN counts also generated many more moth instances that were rejected due to untenably high naive spontaneous noise.

4.3.3 AL noise preserves the highest EN responses (F3)

If odor-related behavior is triggered when EN responses exceed some threshold, and if the system gets multiple exposures to a given stimulus (by sniffing, or by flying through an odor plume [84, 56]), then to maintain odor-related behavior post-injury it suffices to ensure that at least some EN responses still exceed threshold despite injury-induced attenuation. That is, it suffices to protect the strongest EN responses from effects of upstream injury.

To test whether broad noise envelopes on neural firing rates might protect these strongest

EN responses from injury-induced loss, we applied injury to moth templates with different levels of noise in the AL neurons (ie RNs, PNs, QNs, and lateral inhibitory neurons). This noise level is controlled by a single parameter in MothNet. AL noise was set to be a factor of 0, 0.33, 0.67, 1.0, and 1.33 relative to "true" moth AL noise (ie calibrated to wet-lab data). Injury was applied to RNs in the Antennae→AL channel (Fig 4.2, red stars). Over 60 moths were generated from template for each {AL noise, injury level} datapoint.

We recorded mean + std dev of EN output, $\mu(F) + \sigma(F)$ where F = the set of discrete EN responses to a set of odor puffs. This acts as a proxy for behavior of the strongest (∼top 15%) EN responses.

We normalized this by $\mu(F_h) + \sigma(F_h)$ of healthy EN odor responses:

$$\text{normed } [\mu(F) + \sigma(F)] = \frac{\mu(F) + \sigma(F)}{\mu(F_h) + \sigma(F_h)} \tag{4.1}$$

This measure gives a sense of how injury and AL noise affect the highest-firing tranche of a moth's EN odor responses, relative to healthy behavior.

As expected, AL noise delivered benefit in terms of these top scoring odor responses, reducing the attenuation caused by a given level of injury.

This in turn resulted in higher post-training EN responses. EN output, ie $\mu(F) + \sigma(F)$, fully recovered to pre-injury levels from ≈15% injury when AL noise = 0, and from ≈25% when AL noise = 1 ("true" noise level). No further gains resulted from AL noise at a level greater than naturalistic, suggesting that perhaps the moth's actual AL noise level is a point of maximal return. How AL noise level affected the susceptibility of top EN responses to various levels of injury is shown in Fig 4.6A. Post-training recovery to various injury levels, by AL noise level, is given in Fig 4.6B. Interestingly, this protective effect of AL noise was greater on the set of strongest EN responses than on the set of all responses (Fig 5.3), consistent with our proposed mechanism for the effect (Fig 4.8C).

Experiments also indicated that high AL noise had a (not surprising) functional trade-off: High AL noise decreased signal quality, by two measures. First, the healthy ratio of EN odor response to spontaneous noise $\frac{\mu(F)}{\mu(s)}$ was much lower in moths with high-noise ALs (Fig 4.6C),

and injury degraded this ratio much faster in moths with high-noise ALs (Fig 4.6D). Second, SNR $\frac{\mu(F)}{\sigma(F)}$ in healthy moths varied inversely with AL noise level (Fig 4.6E). Injury degraded SNR in all moths similarly, regardless of AL noise level, at a rate of roughly 1% per 1% injury (results not plotted).

4.3.4 Effects of ablation vs FAS injury (F4)

In theory, ablation injury is roughly 1.85x more harmful than FAS injury ($\sim$54 units ablation $\approx$ 100 units FAS injury, cf calculations in Methods). We wished to test whether functional damage to readout neurons in the MON followed this rule, ie whether ablation is a reliable proxy for naturalistic neural injury when the measured effect is downstream from the location of injury. We ran two experiments, one injuring the RN channel (Fig 4.2A), the other injuring the PN channel (Fig 4.2B). In each experiment, moths were generated from a biologically plausible template, with AL noise at natural levels and #QNs = 0 (#QNs = 2 per glomerulus gave similar results). Half the moths were injured by ablation and half were injured by FAS, with injury levels from 0 to 60%, in order to compare the relative empirical effects on EN outputs. In each experiment, over 60 moths were generated for each injury {type, level} datapoint.

Somewhat surprisingly, the relative effects varied drastically by site of injury. Ablation injury to the RN channel was roughly 50 - 60% *more* harmful than predicted by theory. Conversely, ablation injury to the PN channel was roughly 50% *less* harmful than predicted by theory. The post-injury percent losses to EN response, for FAS injury and for ablation injury, as well as a theoretical post-injury curve for ablation, for each of the two injury regions are given in Fig 4.7 (RNs: A, PNs: B). The curves are fitted sums of two exponentials.

Note that this discrepancy between theoretical and actual effects is not at the site of injury, but at downstream neurons. The impacts of the injuries were transformed nonlinearly as they moved through the system.

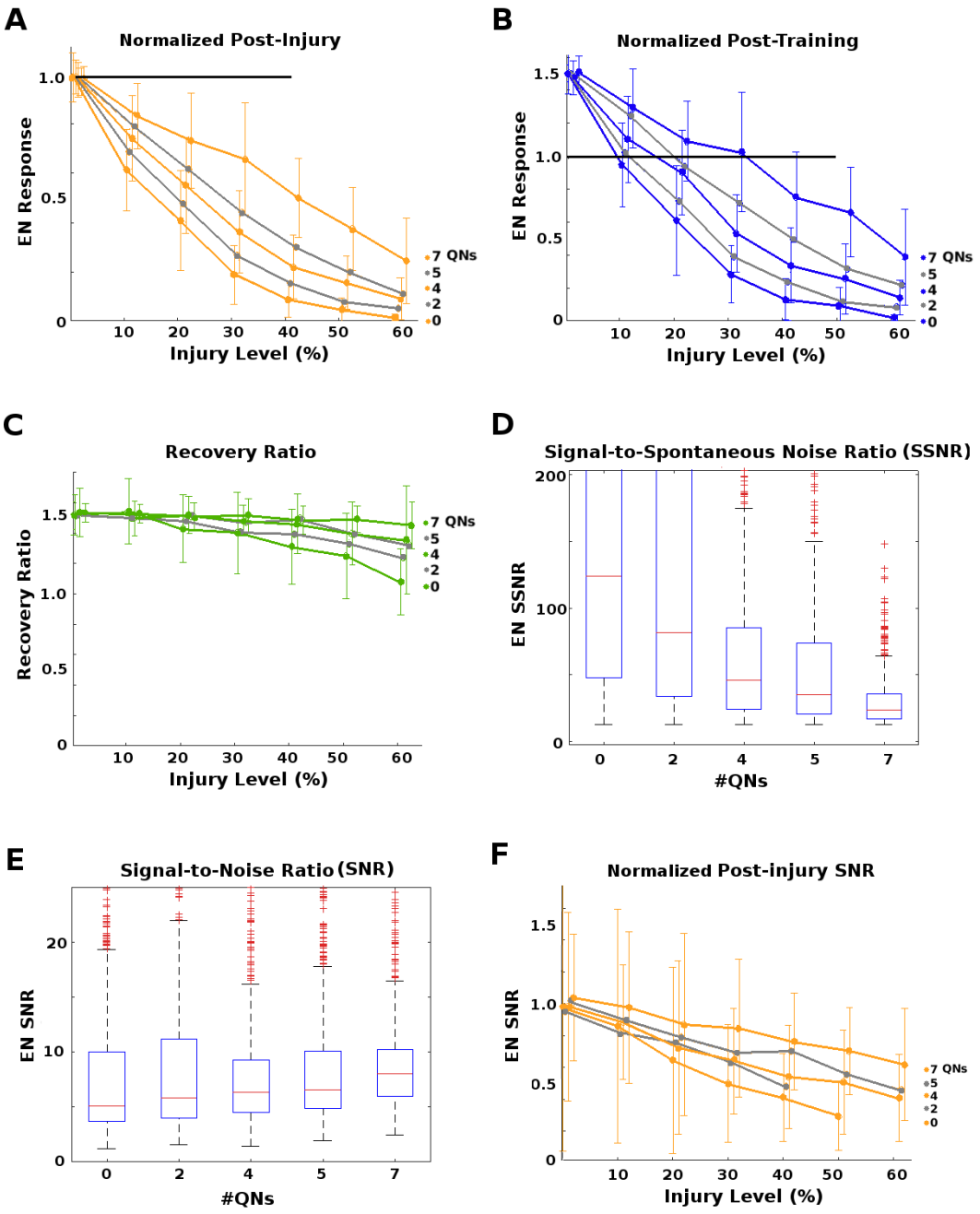

Figure 4.5: **Effects of parallel inhibitory channels. A:** EN odor response post-injury normalized by naive, healthy odor responses vs injury level. Each curve corresponds to a number of QNs per 5 PNs, from 0 to 7. Higher QN:PN ratios resulted in much lower impact on EN responses for a given level of injury. **B:** EN odor responses post-training normalized by naive, healthy odor responses vs injury level. Each curve corresponds to a number of QNs per 5 PNs, from 0 to 7. Higher QN:PN ratios resulted in stronger recovery. **C:** Ratio of post-training to post-injury EN odor responses vs injury level. Recovery rate dropped off at injury levels $\geq 20\%$ for #QN $= 0$, but higher numbers of QNs reduced this drop-off, ie ensured better recovery. **D:** Naive ratio of EN odor response to spontaneous EN noise (SSNR), a measure of signal clarity, was much lower in moths with high QN counts. **E:** Raw Signal-to-Noise Ratio (SNR) of naive, healthy EN responses were fairly uniform across #QNs. **F:** Post-injury SNR normalized by pre-injury SNR. High QN counts gave strong protection against injury-induced degradation of SNR.

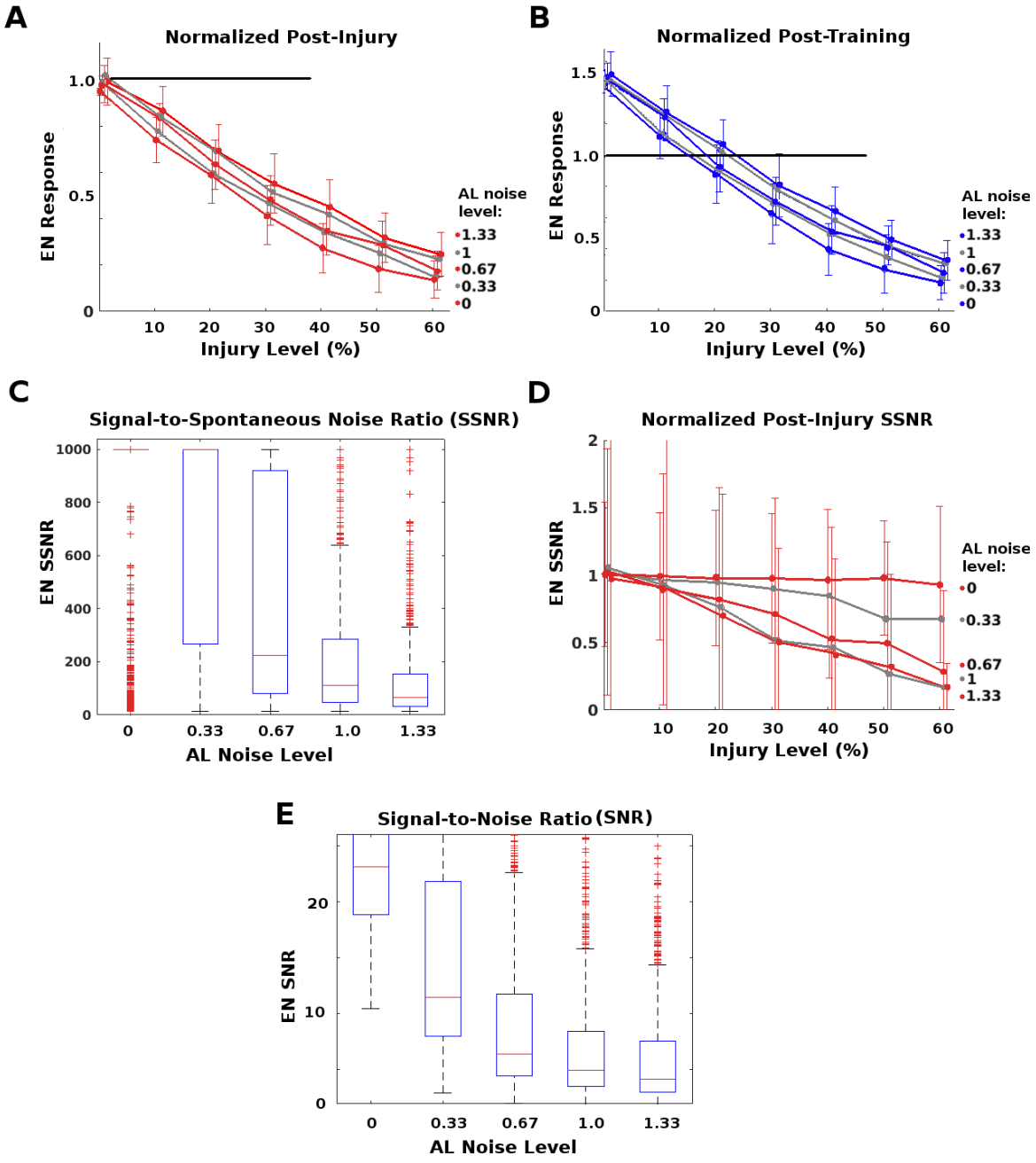

Figure 4.6: **Effect of AL noise:** AL noise protects downstream neurons from loss (A, B), but exacts a cost in terms of signal-to-spontaneous noise ratio (C, D) and SNR (E). **A:** Normalized EN odor response post-injury vs injury level. y-axis = post-injury $\frac{\mu(F)+\sigma(F)}{\mu(F_h)+\sigma(F_h)}$, as a proxy for the highest EN responses of a moth to a series of odor puffs. Higher AL noise resulted in stronger top EN responses post-injury, at any level of injury. **B:** Normalized EN odor response post-training vs injury level. y-axis as in (A). Each curve corresponds to a level of AL noise, from 0 to 1.33. Higher AL noise allowed training to give full recovery of top EN responses from larger injuries. Pre-injury response = black line. **C:** Healthy ratio of EN signal-to-spontaneous noise ratio (SSNR) was much lower at high AL noise levels. **D:** Post-injury SSNR, normalized by pre-injury ratios, vs injury level. In high AL noise moths, injury lowered SSNR far more. **E:** Pre-injury SNR $\frac{\mu(F)}{\mu(s)}$ by AL noise level. SNR was much lower in moths with high-noise ALs.

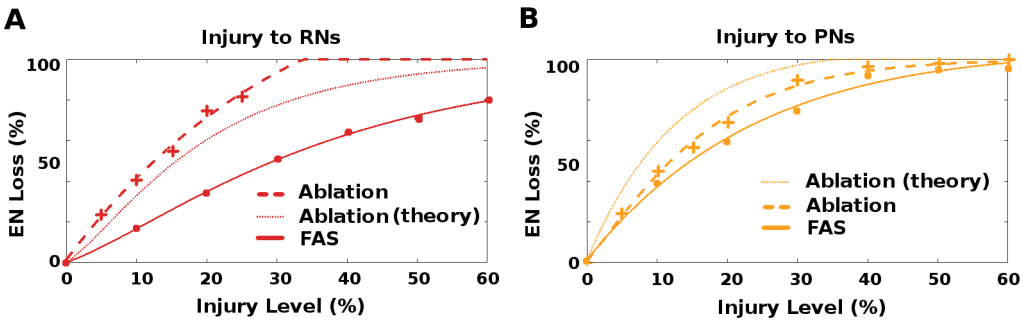

Figure 4.7: **Ablation does not map to FAS injury.** Ablation and FAS injury effects had highly variable relationships. In theory, 1 unit Ablation ∼1.85 units FAS injury. In practice: **A:** When RNs were injured, ablation induced a ∼50% bigger loss to EN response than expected. **B:** When PNs were injured, ablation induced a ∼50% smaller loss than expected.

4.4 Discussion

Our experiments indicate that the neural structures under test have strong injury-mitigation properties, in terms of effects on the downstream readout neurons. In this section we propose mechanisms to explain how these structures protect readout neurons from upstream injury. These mechanisms assume a cascaded network, where system resilience depends on whether downstream units can still deliver key output signals despite upstream injury. Such cascaded networks are endemic among biological neural systems. Mechanisms 1 and 2 also assume integrate-and-fire neural dynamics.

We also discuss the implications of the inconsistent relationship between FAS and ablation injuries, and argue that robustness to injury is a key principle of biological neural design.

4.4.1 How learning compensates for injury (F1)

FAS injury to upstream regions of a network results in dropped spikes from trains and therefore reduced FRs arriving at downstream neurons [83]. When reduced FRs from the damaged region no longer deliver sufficiently strong input to activate downstream neurons, given existing synaptic connection strengths, there is functional loss of information at the output of the downstream neurons. (In these experiments, the upstream neurons are those of the Antennae and/or AL, and the downstream neurons are the ENs and neurons in the MB.)

The learning mechanism in the MON consists of a combination of octopamine stimulation and Hebbian growth. Octopamine stimulation temporarily boosts neural FRs during reinforcement by sugar reward, while Hebbian updates strengthen the synaptic weight w_{ab} between two neurons a and b proportionally to the product of their FRs: $\Delta w_{ab}(t) \propto f_a(t) f_b(t)$ ("wire-together, fire-together").

We propose the following mechanism by which these two mechanisms work together to permanently restore degraded FRs in downstream neurons (schematic in Fig 4.8A):

1. Octopamine causes the injured upstream neurons to temporarily increase their FRs.

2. These transient higher FRs are sufficient to trigger firing in the downstream neurons, given the existing synaptic connection strengths.

3. Because neurons on both sides of the plastic connections are firing, Hebbian growth strengthens the connections.

4. After octopamine is withdrawn, FRs from the injured upstream region return to their reduced rate. Because the synaptic connections are now stronger, these reduced FRs are now sufficient to trigger the downstream neurons. Thus the system's key information (ie downstream response) is restored.

We note that the original injury is not repaired. Rather, synaptic connections downstream are boosted to compensate for the injury-impaired upstream firing rates.

The maximal injury level that can be fully compensated for by training depends, in MothNet, on parameters in the moth template, in particular the parameter controlling maximum synaptic strength (saturation of learning). Therefore, the take-away from our experiments is not any particular threshold value below which damage is irreparable, but that Learning itself functions as an effective injury compensation mechanism, one that allows the system to maintain full performance (according to the key metric, readout response to stimuli) despite substantial injury.

4.4.2 How parallel inhibitory channels reduce downstream effects of injury (F2)

The MON has both excitatory (PN) and parallel inhibitory (QN) projection neurons that feed-forward from AL→MB (Fig 4.1B). We propose the following mechanism to explain how this architecture could protect downstream neurons from damage to this AL→MB channel:

Given an integrate-and-fire neuron model, downstream neurons' behavior depends on the summed input from upstream neurons:

$(\mathbf{w} \cdot \mathbf{u}) = \mathbf{w}^+ \cdot \mathbf{u}^+ - \mathbf{w}^- \cdot \mathbf{u}^-$, where

$\mathbf{w}^+$ = connection weights from excitatory neurons.

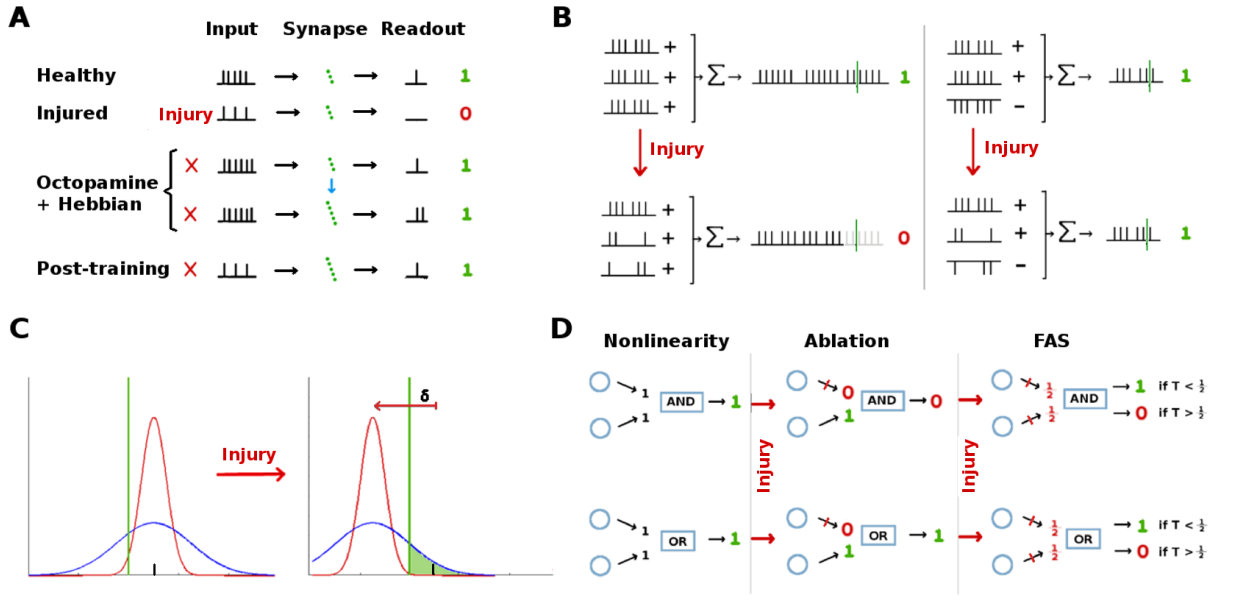

Figure 4.8: **Injury mitigation hypotheses:** In a cascaded network, various architectures can mitigate the effects of injury to upstream neurons by protecting or restoring functionality of downstream units.
A: (Finding 1) Learning itself can compensate for injury. Octopamine temporarily stimulates the damaged neuron, allowing Hebbian growth to strengthen downstream synaptic connections. Though the injured neuron's signal is not restored, the downstream neurons receive an amplified input, cancelling out the injury.
B: (Finding 2) Parallel inhibitory channels can reduce the effect of generalized injury by spreading damage among excitatory and inhibitory signals, so that losses cancel out in terms of inputs to downstream neurons.
C: (Finding 3) Wide noise envelopes on upstream neuron outputs can protect the strongest stimulus responses from injury-induced attenuation δ, to the degree that their std dev $\sigma > \delta$. This allows the injured neuron's strongest responses to still exceed their activation threshold (green line) for downstream neurons, protecting downstream functionality.
D: (Finding 4) Two simple non-linearities that can result in qualititative change in the relative effects of ablation and FAS injury. In an AND gate, ablation can have worse effect than FAS downstream, depending on the gate's input threshold T. In an OR gate, ablation can be harmless, while FAS can have worse effect downstream, depending on T.

$\mathbf{u}^+$ = upstream excitatory neuron FRs.

$\mathbf{w}^-$ = connection weights from inhibitory neurons.

$\mathbf{u}^-$ = upstream inhibitory neuron FRs.

Given FAS injury to this feed-forward channel (eg from physical shock) the net effect on the summed signal reaching downstream target neurons will vary according to the proportion of QNs to PNs ($\mathbf{u}^- : \mathbf{u}^+$) assuming uniform weights $\mathbf{w}$. If all feed-forward signals are excitatory (ie number of QNs $= 0$, so $\mathbf{u}^- = 0$), injury will always lower the summed input reaching the downstream neurons. However, if QNs exist then injury-induced reduction in FRs should be mitigated via a "cancelling out" mechanism, since both excitatory ($\mathbf{u}^+$) and inhibitory ($\mathbf{u}^-$) signals are degraded. If PNs and QNs have equal numbers and synaptic strengths, then after injury their summed total input to the downstream neurons should remain roughly unchanged, ie the downstream neurons are fully robust to injury (in the summed inputs sense). Fig 4.8B shows a schematic.

However, it is clear from our experiments that although high QN:PN ratios provide significant downstream protection from injury, the effect is not as clear-cut as this model suggests: Moths with equal numbers of PNs and QNs still suffered substantial effects from injury.

In addition, our experiments found that high QN counts deliver stronger injury resistance and protect SNR, but at a cost of higher spontaneous EN noise relative to odor response. Presumably, biological networks have QN counts which optimally balance the benefits of injury mitigation on one hand versus the need for high SNR, as well as other concerns such as the energy cost to the organism, on the other. The QN counts in the MON are quite low (QN:PN $< 20\%$). This suggests that the substantial injury mitigation benefits seen in our experiments are not as valuable as they appear, relative to other more pressing architectural constraints.

Learning is not necessary to this mechanism.

4.4.3 How upstream noise protects downstream neurons' triggering behavior (F3)

We propose the following mechanism whereby high noise in upstream networks might protect downstream neurons' functionality from the effects of damage to the upstream networks. We define "preserving functionality" as ensuring that at least a subset of stimuli elicit down-

stream neuron responses that exceed key thresholds (eg ENs exceeding action-triggering thresholds in the MON).

Suppose an upstream neuron FR responds to stimuli as $N(\mu, \sigma)$, and that it needs to exceed some threshold T to activate downstream neurons. Let neural damage reduce this FR by δ on average, so that the new mean $= \mu - \delta$. Then a large noise envelope (large σ) will ensure that some post-injury responses still exceed threshold, to the degree that δ is small compared to σ, ie that $\mu - \delta + \sigma \geq T$.

The idea is sketched in Fig 4.8C, for two FRs characterized by $N(\mu, \sigma_1)$ and $N(\mu, \sigma_2)$ with $\sigma_1 > \sigma_2$. Suppose injury reduces their FRs such that each distribution shifts downwards by the same amount δ. Then samples drawn from the shifted wide gaussian are more likely to be above the original trigger threshold T than samples drawn from the shifted narrow gaussian, according to σ_i/δ.

Our experiments indicate that AL noise does enable the highest EN responses to exceed threshold after injury, even as the average EN response drops. However, increased systemic upstream noise impacts the effectiveness of a healthy system, for example by reducing SNR (as in our experiments). Noise levels in biological networks (such as noise in the AL) may represent an evolved trade-off between injury mitigation effects and negative side-effects such as reduced SNR.

Learning is not necessary to this mechanism.

4.4.4 Unpredictability of ablation vs FAS injury effects (F4)

Simple ablation is a standard and convenient method of applying injury to neural networks. However, our results indicate that ablation is a poor proxy, even when scaled, for naturalistic FAS type injuries, when effects are measured downstream from the injury site. Ablation's effects were sometimes much larger, sometimes much smaller than would be predicted for ablation as applied to a large population of homogeneous neurons and measuring local effects. When the key behaviors that define system functionality are not local to the injured region, but downstream, there are multiple possible reasons for a mis-match of effects.

If the number of neurons N is small, eliminating the signal from one important neuron (as in ablation) is likely to have an effect distinct from merely reducing that signal (as in FAS). Especially when all neurons are excitatory, ablation may have outsized effects in terms of attenuating the signal arriving downstream.

Also, as reduced firing rates in upstream neurons travel through a network (such as the cascade of the MON), nonlinearities due to network properties can distort the effect on downstream regions. Nonlinearities are endemic in NNs (induced for example by the sigmoids within integrate-and-fire neurons, by inhibitory channels, or by neuromodulators), so the unpredictability of ablation's and FAS's relative downstream effects is not surprising. Simple examples of the effects of non-linearities (AND and OR gates) are given in schematic in Fig 4.8D.

In cases where system function is defined by behavior of regions downstream from the injured region, our results suggest that ablation's effects do not consistently map to FAS effects. Thus, for naturalistic simulation of injury, FAS injury regimes are preferable to ablation proxies.

4.4.5 Limitations

These experiments assumed only one readout neuron, and only one broadly activating odor. A more realistic assessment of injury and mitigation might involve several readout neurons, to allow for the case that injury had disparate effects on various readouts.

It might also be more realistic to consider several narrowly-focused odors, because such odors might be more susceptible to catastrophic attenuation if critical processing neurons were injured. For example, this might yield a larger difference between FAS and ablation.

Our choice of injured regions may not have been realistic (in particular the PN/QN channel, chosen to test a particular hypothesis, F2). The choice of injured region is critical to readout function, because of non-linear effects induced as injury-reduced FRs propagate through the system. We also note that there exist many other combinations of {structure-under-test + injured region(s)} which we did not examine.

Finally, the higher levels of injury, eg $\geq 40\%$, might be irrelevant in practice. If high neural injury correlated with major damage to other parts of the organism, then maintaining neural network function would not matter.

4.4.6 Robustness to injury as a central design principle

Each of the architectures tested in these experiments might be justified purely on the basis of their anti-injury benefits. There are also other plausible functions for these architectures. For example, high-noise signals may allow downstream neurons to do Bayesian inference [51]. However, injury mitigation may still be an important or even a primary reason the structures exist, or first evolved.

Indeed, learning itself may be a case of exaption, or borrowed function: It may have originally evolved as a repair mechanism to offset neural injury and maintain function, and was only later ported to the task of developing responses to new information. If this is the case, then the gift of learning is due originally to the exigencies of brain damage.

Our results also show that these architectures can in fact cause worse performance by some other performance metrics, eg SNR. Thus, trying to explain them from the point-of-view of, for example, information theory risks running against the fact that the architectures are actually suboptimal, and therefore will not make sense, relative to that particular lens. In such cases, a neural architecture can be understood only if its injury mitigation function, and the trade-offs between this and other desired functions, are considered.

Chapter 5

SUPPLEMENTARY INFORMATION

5.0.7 Moth AL neural recording datasets

Detailed list of data sets from the lab of Prof. Jeff Riffell, UW:

1. AL, odor only: PNs, one odor, no octopamine. 7 preps with 8 - 16 PNs each.

2. AL, odor + octopamine. PNs, one odor, sugar reward. 10 preps with 9 - 21 PNs each.

3. AL + MB, odor + octopamine. PNs and KCs, one odor, sugar reward. 1 prep, with 7 PNs and 12 KCs.

4. AL, odor + octo wash: PNs, one odor, octopamine directly applied to AL. 7 preps: 6 preps with 8 - 13 PNs each; 1 prep with one pheromone-responsive neuron.

5. AL, odors only (BEA): PNs, several odors and concentrations. 12 preps with 14 - 17 PNs each.

6. AL, odors only (ESO): PNs, several odors and concentrations. 4 preps with 12 - 14 PNs each.

5.0.8 Full equations of model dynamics

$$\tau_{\mathrm{R}} \cdot d\mathbf{u}^{\mathrm{R}} = f_{\mathrm{R}}\big(\mathbf{u}^{\mathrm{R}}, \mathbf{u}^{\mathrm{L}}, \mathbf{u}^{\mathrm{S}}, M^{\mathrm{L,R}}, M^{\mathrm{S,R}}, M^{\mathrm{O,R}}, o(t)\big) \; + \; d\mathbf{W}^{\mathrm{R}} \tag{5.1}$$

$$\tau_{\mathrm{P}} \cdot d\mathbf{u}^{\mathrm{P}} = f_{\mathrm{P}}\big(\mathbf{u}^{\mathrm{R}}, \mathbf{u}^{\mathrm{P}}, \mathbf{u}^{\mathrm{L}}, M^{\mathrm{L,P}}, M^{\mathrm{R,P}}, M^{\mathrm{O,P}}, o(t)\big) \; + \; d\mathbf{W}^{\mathrm{P}} \tag{5.2}$$

$$\tau_{\mathrm{Q}} \cdot d\mathbf{u}^{\mathrm{Q}} = f_{\mathrm{Q}}\big(\mathbf{u}^{\mathrm{R}}, \mathbf{u}^{\mathrm{Q}}, \mathbf{u}^{\mathrm{L}}, M^{\mathrm{L,Q}}, M^{\mathrm{R,Q}}, M^{\mathrm{O,Q}}, o(t)\big) \; + \; d\mathbf{W}^{\mathrm{Q}} \tag{5.3}$$

$$\tau_{\mathrm{L}} \cdot d\mathbf{u}^{\mathrm{L}} = f_{\mathrm{L}}\big(\mathbf{u}^{\mathrm{R}}, \mathbf{u}^{\mathrm{L}}, M^{\mathrm{L,L}}, M^{\mathrm{R,L}}, M^{\mathrm{O,L}}, o(t)\big) \; + \; d\mathbf{W}^{\mathrm{L}} \tag{5.4}$$

$$\tau_{\mathrm{K}} \cdot d\mathbf{u}^{\mathrm{K}} = f_{\mathrm{K}}\big(\mathbf{u}^{\mathrm{P}}, \mathbf{u}^{\mathrm{Q}}, \mathbf{u}^{\mathrm{LH}}, M^{\mathrm{P,K}}, M^{\mathrm{Q,K}}\big) \; + \; d\mathbf{W}^{\mathrm{K}} \tag{5.5}$$

$$\tau_{\mathrm{E}} \cdot d\mathbf{u}^{\mathrm{E}} = f_{\mathrm{E}}\big(\mathbf{u}^{\mathrm{K}}, \mathbf{u}^{\mathrm{E}}, M^{\mathrm{K,E}}\big) \tag{5.6}$$

where

$$
\begin{cases}
f_{\mathrm{R}} = -\mathbf{u}^{\mathrm{R}} + \mathrm{sigmoid}\big[-\big(I - \gamma \cdot o(t) \cdot M^{\mathrm{O,R}}\big) M^{\mathrm{L,R}}\, \mathbf{u}^{\mathrm{L}} \;+\; \big(I + o(t) \cdot M^{\mathrm{O,R}}\big) M^{\mathrm{S,R}}\, \mathbf{u}^{\mathrm{S}} \big] \\
f_{\mathrm{P}} = -\mathbf{u}^{\mathrm{P}} + \mathrm{sigmoid}\big[-\big(I - \gamma \cdot o(t) \cdot M^{\mathrm{O,P}}\big) M^{\mathrm{L,P}}\, \mathbf{u}^{\mathrm{L}} \;+\; \big(I + o(t) \cdot M^{\mathrm{O,P}}\big) M^{\mathrm{R,P}}\, \mathbf{u}^{\mathrm{R}} \big] \\
f_{\mathrm{Q}} = -\mathbf{u}^{\mathrm{Q}} + \mathrm{sigmoid}\big[-\big(I - \gamma \cdot o(t) \cdot M^{\mathrm{O,Q}}\big) M^{\mathrm{L,Q}}\, \mathbf{u}^{\mathrm{L}} \;+\; \big(I + o(t) \cdot M^{\mathrm{O,Q}}\big) M^{\mathrm{R,Q}}\, \mathbf{u}^{\mathrm{R}} \big] \\
f_{\mathrm{L}} = -\mathbf{u}^{\mathrm{L}} + \mathrm{sigmoid}\big[-\big(I - \gamma \cdot o(t) \cdot M^{\mathrm{O,L}}\big) M^{\mathrm{L,L}}\, \mathbf{u}^{\mathrm{L}} \;+\; \big(I + o(t) \cdot M^{\mathrm{O,L}}\big) M^{\mathrm{R,L}}\, \mathbf{u}^{\mathrm{R}} \big] \\
f_{\mathrm{K}} = -\mathbf{u}^{\mathrm{K}} + \mathrm{sigmoid}\big[-\big(\mathbf{u}^{\mathrm{LH}} + M^{\mathrm{Q,K}}\, \mathbf{u}^{\mathrm{Q}}\big) \;+\; M^{\mathrm{P,K}}\, \mathbf{u}^{\mathrm{P}} \big] \\
f_{\mathrm{E}} = -\mathbf{u}^{\mathrm{E}} + M^{\mathrm{K,E}}\, \mathbf{u}^{\mathrm{K}}
\end{cases}
$$

Table 5.1: Variables and parameters for neuronal network model

Symbol	Type	Size/Value	Description and Remarks
R	superscript		Refers to the *receptor neurons* subpopulation.
P	superscript		Refers to the *excitatory projection neurons* subpopulation.
Q	superscript		Refers to the *inhibitory projection neurons* subpopulation.
L	superscript		Refers to the *lateral neurons* subpopulation.
K	superscript		Refers to the *kenyon cells* subpopulation.
E	superscript		Refers to the readout *extrinsic neurons* subpopulation.
O	superscript		Refers to the *octopamine* neurotransmitter.
nG	scalar	60	Number of glomeruli in the antenna lobe. *
nS	scalar	2-4	Number of different stimuli (odors).
nQ	scalar		Number of inhibitory projection neurons.
nK	scalar	2000	Number of kenyon cells.
nE	scalar	1	Number of extrinsic neurons.
$\mathbf{u}^{\mathrm{R}}$	vector	$nG \times 1$	FRs of the receptor neurons subpopulation.
$\mathbf{u}^{\mathrm{P}}$	vector	$nG \times 1$	FRs of the exc. projection neurons subpopulation.
$\mathbf{u}^{\mathrm{Q}}$	vector	$nQ \times 1$	FRs of the inh. projection neurons subpopulation.
$\mathbf{u}^{\mathrm{L}}$	vector	$nG \times 1$	FRs of the lateral neurons subpopulation.
$\mathbf{u}^{\mathrm{K}}$	vector	$nK \times 1$	FRs of the kenyon cells subpopulation. Sparse.
$\mathbf{u}^{\mathrm{E}}$	vector	$nE \times 1$	FRs of the extrinsic neurons subpopulation.
$\mathbf{u}^{\mathrm{S}}$	vector		
$\mathbf{u}^{\mathrm{LH}}$	vector		Inhibition from the LH, identical for all KCs.
$M^{\mathrm{S,R}}$	matrix	$nG \times nS$	Stimulus $\rightarrow \mathbf{u}^{\mathrm{R}}$ connections.
$M^{\mathrm{O,R}}$	matrix	$nG \times nG$	Octopamine $\rightarrow \mathbf{u}^{\mathrm{R}}$ connections. Diagonal matrix.
$M^{\mathrm{O,L}}$	matrix	$nG \times nG$	Octopamine $\rightarrow \mathbf{u}^{\mathrm{L}}$ connections. Diagonal matrix.
$M^{\mathrm{R,L}}$	matrix	$nG \times nG$	Connection weights $\mathbf{u}^{\mathrm{R}} \rightarrow \mathbf{u}^{\mathrm{L}}$.
$M^{\mathrm{R,P}}$	matrix	$nG \times nG$	Connection weights $\mathbf{u}^{\mathrm{R}} \rightarrow \mathbf{u}^{\mathrm{P}}$. Diagonal matrix.
$M^{\mathrm{R,Q}}$	matrix	$nQ \times nG$	Connection weights $\mathbf{u}^{\mathrm{R}} \rightarrow \mathbf{u}^{\mathrm{Q}}$.
$M^{\mathrm{P,K}}$	matrix	$nK \times nG$	Connection weights $\mathbf{u}^{\mathrm{P}} \rightarrow \mathbf{u}^{\mathrm{K}}$.
$M^{\mathrm{Q,K}}$	matrix	$nK \times nQ$	Connection weights $\mathbf{u}^{\mathrm{Q}} \rightarrow \mathbf{u}^{\mathrm{K}}$.
$M^{\mathrm{L,R}}$	matrix	$nG \times nG$	Connection weights $\mathbf{u}^{\mathrm{L}} \rightarrow \mathbf{u}^{\mathrm{R}}$.
$M^{\mathrm{L,P}}$	matrix	$nG \times nG$	Connection weights $\mathbf{u}^{\mathrm{L}} \rightarrow \mathbf{u}^{\mathrm{P}}$.
$M^{\mathrm{L,Q}}$	matrix	$nQ \times nG$	Connection weights $\mathbf{u}^{\mathrm{L}} \rightarrow \mathbf{u}^{\mathrm{Q}}$.
$M^{\mathrm{L,L}}$	matrix	$nG \times nG$	Connection weights $\mathbf{u}^{\mathrm{L}} \rightarrow \mathbf{u}^{\mathrm{L}}$.
$M^{\mathrm{K,E}}$	matrix	$nE \times nK$	Connection weights $\mathbf{u}^{\mathrm{K}} \rightarrow \mathbf{u}^{\mathrm{E}}$.
$o(t)$	function	0 or 1	Flags when octopamine is active (typically during training).
γ	scalar	0.5	Scaling factor for octopamine's effects on inhibition. *

5.0.9 ANOVA analysis of MothNet learning

The differential increase in EN response to trained vs control odors was almost always significant to $p < 0.01$. When odors' naive EN response magnitudes differed by > 3, either raw increases or percentage increases (not both) sometimes did not attain this level of significance, while the other metric did. Fig 5.1 plots the p-values of 336 trained odor/control odor pairs against the ratio of their mean naive responses $\frac{\mu_T}{\mu_C}$, for 28 moths randomly generated from a template, with three control odors and one trained odor. Each p-value is for the trained odor vs one control odor (so there are 12 data points per moth). Column 1 shows p-values for change in raw EN response (as in Fig 2.7C), trained vs control. Trained odors with very low-magnitude naive response often did not have raw increases larger than high-magnitude control odors. Column 2 shows p-values for percentage change in EN response (as in Fig 2.7B), trained vs control.

Unless the naive EN responses for the two odors were highly disparate (eg by factor of >3x), the differential increase in EN response of the trained vs control odors is almost always significant, measured both as raw and as percentage. Fig 5.2 plots the percentage of 336 trained-control pairs that had p-values for both measures of EN response increase (ie as raw and as percentage) below the listed threshold (eg $p = 0.01$), for 336 trained-control pairs whose ratio ($\frac{\mu_T}{\mu_C}$ or $\frac{\mu_C}{\mu_T}$) is within the bound given on the x-axis. Fig 5.2 shows how many moths, generated from template with no constraint on unbalanced naive odor EN responses, had differential post-training EN responses with significance $p < 0.01$ for both measures (as raw and as a percentage).

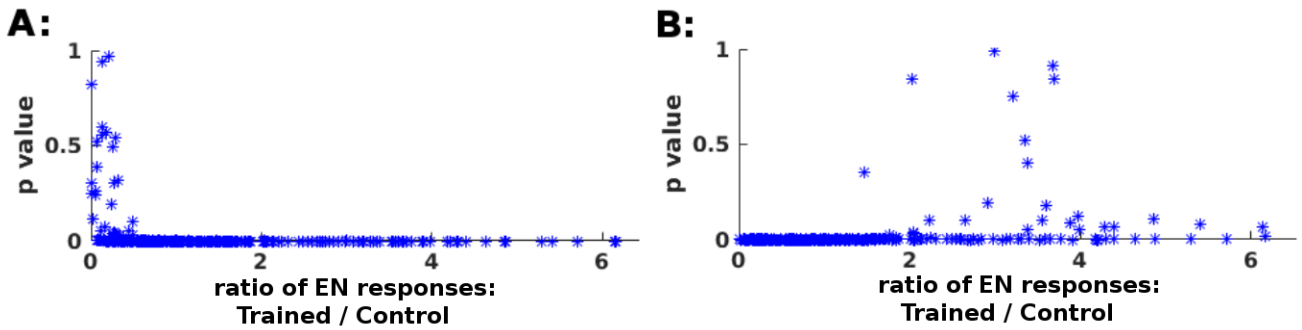

Figure 5.1: **p-values for trained-control odor pairs: A:** p-values for change in raw EN responses. **B:** p-values for percentage change in EN responses. P-values are sometimes high (for one metric or the other) when trained and control odors have highly disparately-scaled naive responses μ_T (= mean raw T) μ_C (= mean raw C). Plots show results given 20 training puffs.
When μ_T is larger (right end of x-axis), the p-value for raw change (A) is consistently very low, but the p-value for percentage change (rB) can be high, since even a small incidental change to a low-intensity odor can be a large percentage change.
When μ_C is larger (left end of x-axis), the p-value for percentage change (B) is consistently very low, but the p-value for raw change (A) can be high, since even a small percentage change to a high-response odor corresponds to a large raw change.
When naive odor responses are roughly matched, eg within 3x (ie 0.33 to 3), p-values for both raw and percentage change are very low.

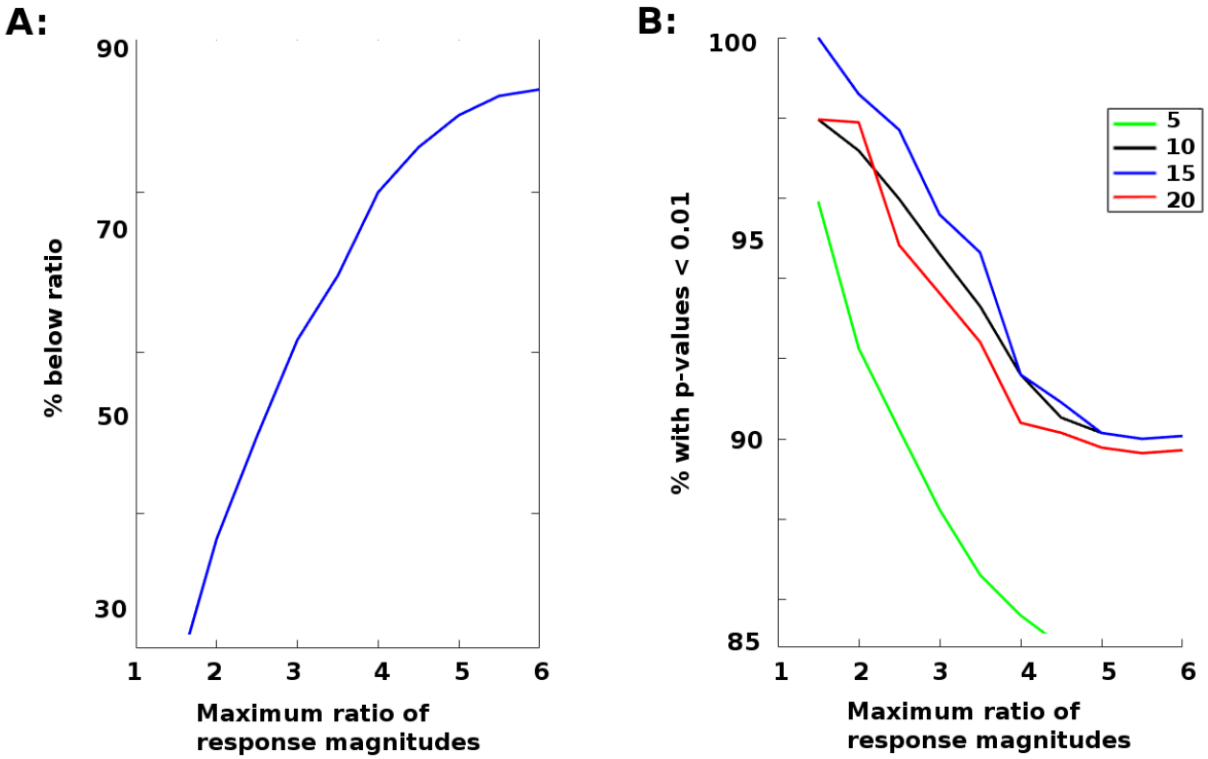

Figure 5.2: **Fractions of p-values below 0.01 for trained-control odor pairs** In most cases, the trained odor shows much larger increases in EN response magnitude. **A:** The percentage of trained-control odor pairs with EN response magnitudes within the ratios given on the x-axis. **B:** The percentage of trained-control pairs, with EN response magnitudes within the ratios given on the x-axis, whose training-induced changes in EN responses were distinct with p-value < 0.01. Each curve is for a different number of training puffs. More training increases distinctions, up to 15 puffs. But additional training actually hinders distinctions, as control odor response reinforcement begins to overtake trained odor reinforcement.

5.0.10 *AL noise effects on strongest vs average EN responses*

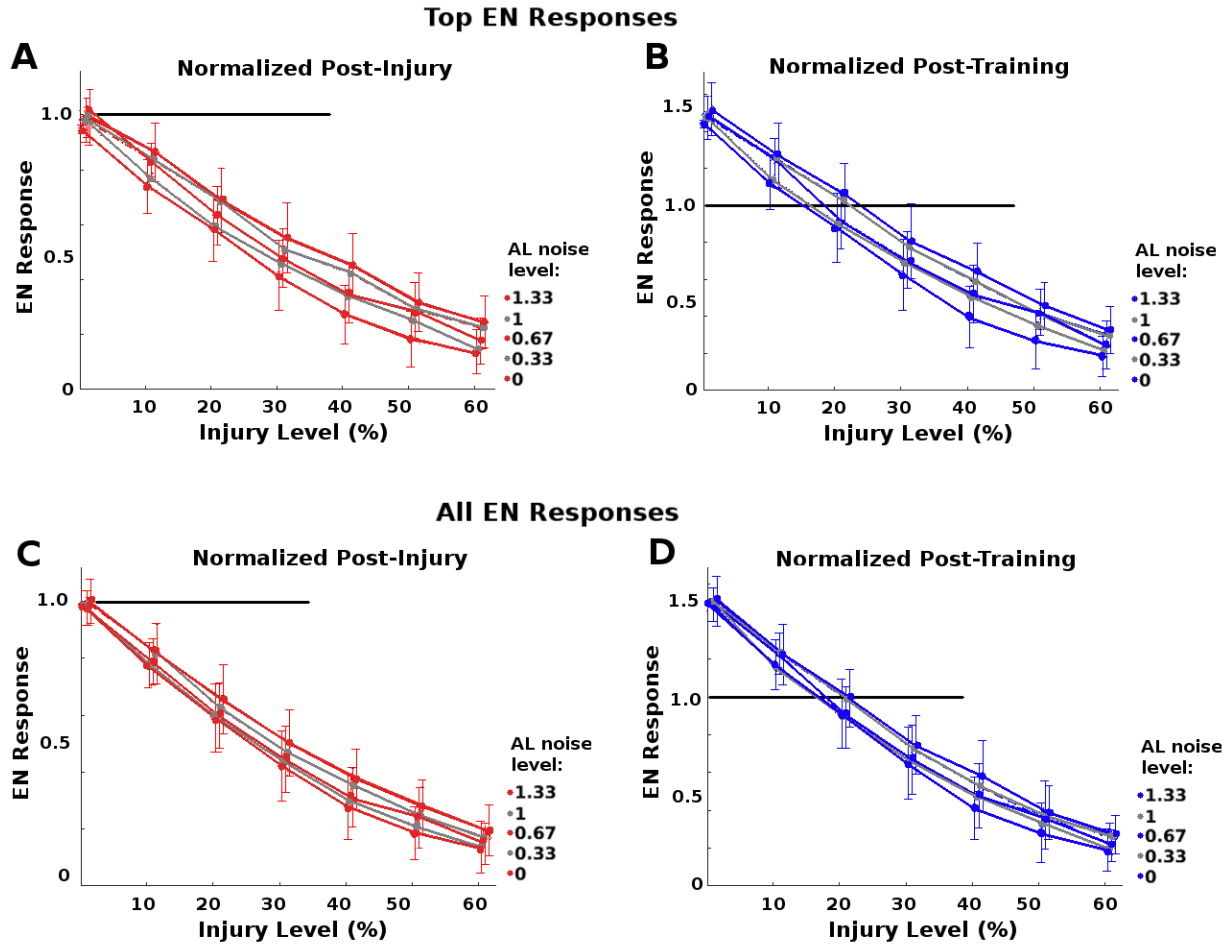

Figure 5.3: **Effect of AL noise on strongest vs average EN responses** High AL noise had a greater protective effect on the top 15% tranche of EN odor responses than on all odor responses. **A:** Normalized post-injury (red, grey) and post-training (blue, grey) EN odor responses $\frac{\mu(F)+\sigma(F)}{\mu(F_h)+\sigma(F_h)}$. **B:** The same data, but plotting the normalized mean of EN responses $\frac{\mu(F)}{\mu(F_h)}$. Injury mitigation was weaker for mean responses. That is, high EN responses received more protection than low or average EN responses.

BIBLIOGRAPHY

[1] Guozhong An. The effects of adding noise during backpropagation training on a generalization performance. *Neural Comput.*, 8(3):643–674, April 1996.

[2] Kenta Asahina, Matthieu Louis, Silvia Piccinotti, and Leslie B. Vosshall. A circuit supporting concentration-invariant odor perception in Drosophila. *Journal of Biology*, 8(1):9, Jan 2009.

[3] Josh Attenberg and Foster Provost. Why label when you can search?: Alternatives to active learning for applying human resources to build classification models under extreme class imbalance. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 423–432, New York, NY, USA, 2010. ACM.

[4] Baktash Babadi and Haim Sompolinsky. Sparseness and expansion in sensory representations. *Neuron*, 83(5):1213 – 1226, 2014.

[5] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[6] Maxim Bazhenov and Mark Stopfer. Forward and back: Motifs of inhibition in olfactory processing. *Neuron*, 67(3):357 – 358, 2010.

[7] Yoshua Bengio and Asja Fischer. Early inference in energy-based models approximates back-propagation. *arXiv e-prints*, abs/1510.02777, October 2015.

[8] Vikas Bhandawat, Shawn R Olsen, Nathan W Gouwens, Michelle L Schlief, and Rachel I Wilson. Sensory processing in the Drosophila antennal lobe increases reliability and separability of ensemble odor representations. *Nature Neuroscience*, 10:1474–1482, 2007.

[9] RAA Campbell, KS Honegger, H Qin, W Li, E Demir, and GC Turner. Imaging a population code for odor identity in the Drosophila mushroom body. *Journal of Neuroscience*, 33(25):10568–81, 2013.

[10] Robert A.A. Campbell and Glenn C. Turner. The mushroom body. *Current Biology*, 20(1):R11 – R12, 2010.

[11] Mikael A. Carlsson, Kwok Ying Chong, Wiltrud Daniels, Bill S. Hansson, and Tim C. Pearce. Component information is preserved in glomerular responses to binary odor mixtures in the moth Spodoptera littoralis. *Chemical Senses*, 32(5):433, 2007.

[12] SJ Caron, V Ruta, LF Abbott, and R Axel. Random convergence of olfactory inputs in the Drosophila mushroom body. *Nature*, 497(5):113–7, 2013.

[13] Sophie J. C. Caron. Brains don't play dice—or do they? *Science*, 342(6158):574–574, 2013.

[14] Stijn Cassenaer and Gilles Laurent. Hebbian stdp in mushroom bodies facilitates the synchronous flow of olfactory information in locusts. *Nature*, 448:709 EP –, Jun 2007.

[15] Andrew M. Dacks, Jeffrey A. Riffell, Joshua P. Martin, Stephanie L. Gage, and Alan J. Nighorn. Olfactory modulation by dopamine in the context of aversive learning. *Journal of Neurophysiology*, 108(2):539–550, 7 2012.

[16] Peter Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2005.

[17] Nina Deisig, Martin Giurfa, Harald Lachnit, and Jean-Christophe Sandoz. Neural representation of olfactory mixtures in the honeybee antennal lobe. *European Journal of Neuroscience*, 24(4):1161–1174, 2006.

[18] Julien Dupuis, Thierry Louis, Monique Gauthier, and Valrie Raymond. Insights from honeybee (Apis mellifera) and fly (Drosophila melanogaster) nicotinic acetylcholine receptors: From genes to behavioral functions. *Neuroscience and Biobehavioral Reviews*, 36(6):1553 – 1564, 2012.

[19] Basil El Jundi, Wolf Huetteroth, Angela E. Kurylas, and Joachim Schachtner. Anisometric brain dimorphism revisited: Implementation of a volumetric 3d standard brain in Manduca sexta. *The Journal of Comparative Neurology*, 517(2):210–225, 2009.

[20] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.

[21] C. Giovanni Galizia. Olfactory coding in the insect brain: data and conjectures. *European Journal of Neuroscience*, 39(11):1784–1795, 2014.

[22] Surya Ganguli and Haim Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual Review of Neuroscience*, 35(1):485–508, 2012. PMID: 22483042.

[23] W. Shane Grant, James Tanner, and Laurent Itti. Biologically plausible learning in neural networks with modulatory feedback. *Neural Networks*, 88(Supplement C):32 – 48, 2017.

[24] Eyal Gruntman and Glenn C. Turner. Integration of the olfactory code across dendritic claws of single mushroom body neurons. *Nature Neuroscience*, 16:1821 EP –, Oct 2013. Article.

[25] Nitin Gupta and Mark Stopfer. Functional analysis of a higher olfactory center, the lateral horn. *Journal of Neuroscience*, 32(24):8138–8148, 2012.

[26] Elissa A. Hallem and John R. Carlson. Coding of odors by a receptor repertoire. *Cell*, 125(1):143–160, April 2006.

[27] M Hammer and R Menzel. Learning and memory in the honeybee. *Journal of Neuroscience*, 15(3):1617–1630, 1995.

[28] Martin Hammer. An identified neuron mediates the unconditioned stimulus in associative olfactory learning in honeybees. *Nature*, 366:59 EP –, Nov 1993.

[29] Martin Hammer and Randolf Menzel. Multiple sites of associative odor learning as revealed by local brain microinjections of octopamine in honeybees. *Learn Mem*, 5(1):146–156, May 1998. 10454379[pmid].

[30] D. O. Hebb. *The organization of behavior : a neuropsychological theory*. Wiley New York, 1949.

[31] Jim Henson. Ernie and Bert: Teaching Bernice to play checkers. *Sesame Street*, 1976.

[32] Toshihide Hige, Yoshinori Aso, Gerald M. Rubin, and Glenn C. Turner. Plasticity-driven individualization of olfactory coding in mushroom body output neurons. *Nature*, 526:258 EP –, Sep 2015.

[33] Andrew D. Higginson, Christopher J. Barnard, Adam Tofilski, Luis Medina, and Francis Ratnieks. Experimental wing damage affects foraging effort and foraging distance in honeybees Apis mellifera. *Psyche*, 2011.

[34] Desmond J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, 43(3):525–546, March 2001.

[35] Kyle S. Honegger, Robert A. A. Campbell, and Glenn C. Turner. Cellular-resolution population imaging reveals robust sparse coding in the Drosophila mushroom body. *Journal of Neuroscience*, 31(33):11772–11785, 2011.

[36] Elizabeth J. Hong and Rachel I. Wilson. Simultaneous encoding of odors by channels with diverse sensitivity to inhibition. *Neuron*, 85(3):573 – 589, 2015.

[37] JJ Hopfield and DW Tank. Computing with neural circuits: a model. *Science*, 233(4764):625–633, 1986.

[38] D. Hubel and T. Wiesel. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

[39] Iori Ito, Rose Chik-ying Ong, Baranidharan Raman, and Mark Stopfer. Sparse odor representation and olfactory learning. *Nature Neuroscience*, 11:1177 EP –, Sep 2008. Article.

[40] Gregory S.X.E. Jefferis, Christopher J. Potter, Alexander M. Chan, Elizabeth C. Marin, Torsten Rohlfing, Calvin R. Maurer Jr., and Liqun Luo. Comprehensive maps of Drosophila higher olfactory centers: Spatially segregated fruit and pheromone representation. *Cell*, 128(6):1187 – 1203, 2007.

[41] Pal Kvello, Bjarte Lofaldli, Jurgen Rybak, Randolf Menzel, and Hanna Mustaparta. Digital, three-dimensional average shaped atlas of the heliothis virescens brain with integrated gustatory and olfactory neurons. *Frontiers in Systems Neuroscience*, 3:14, 2009.

[42] Gilles Laurent. A systems perspective on early olfactory coding. *Science*, 286(5440):723–728, 1999.

[43] Gilles Laurent. Olfactory network dynamics and the coding of multidimensional signals. *Nature Reviews Neuroscience*, 3:884 EP –, Nov 2002. Review Article.

[44] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Comput.*, 1(4):541–551, December 1989.

[45] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Mller, E. Sckinger, P. Simard, and V. Vapnik. Comparison of learning algorithms for handwritten digit recognition. In *INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS*, pages 53–60, 1995.

[46] Yann LeCun. Facebook AI director Yann LeCun on his quest to unleash deep learning and make machines smarter. *IEEE Spectrum*, 2015.

[47] Yann LeCun and Corinna Cortes. MNIST handwritten digit database, http://yann.lecun.com/exdb/mnist/. *Website*, 2010.

[48] Andrew C. Lin, Alexei M. Bygrave, Alix de Calignon, Tzumin Lee, and Gero Miesenböck. Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination. *Nature Neuroscience*, 17:559 EP –, Feb 2014. Article.

[49] Christiane Linster, Silke Sachse, and C. Giovanni Galizia. Computational modeling suggests that response properties rather than spatial position determine connectivity between olfactory glomeruli. *Journal of Neurophysiology*, 93(6):3410–3417, 2005.

[50] Sean X. Luo, Richard Axel, and L. F. Abbott. Generating sparse and selective third-order responses in the olfactory system of the fly. *Proceedings of the National Academy of Sciences*, 107(23):10713–10718, 2010.

[51] Wei Ji Ma, Jeffrey M. Beck, Peter E. Latham, and Alexandre Pouget. Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9:1432 EP –, Oct 2006. Article.

[52] Pedro D. Maia, Matthew A. Hemphill, Brendan Zehnder, Chenfei Zhang, Kevin K. Parker, and J. Nathan Kutz. Diagnostic tools for evaluating the impact of focal axonal swellings arising in neurodegenerative diseases and/or traumatic brain injury. *Journal of Neuroscience Methods*, 253(Supplement C):233 – 243, 2015.

[53] Pedro D. Maia and J. Nathan Kutz. Compromised axonal functionality after neurodegeneration, concussion and/or traumatic brain injury. *Journal of Computational Neuroscience*, 37(2):317–332, Oct 2014.

[54] Pedro D. Maia and J. Nathan Kutz. Reaction time impairments in decision-making networks as a diagnostic marker for traumatic brain injuries and neurological diseases. *Journal of Computational Neuroscience*, 42:323–347, 2017.

[55] Alireza Makhzani and Brendan J. Frey. k-sparse autoencoders. *CoRR*, abs/1312.5663, 2013.

[56] Joshua P. Martin, Aaron Beyerlein, Andrew M. Dacks, Carolina E. Reisenman, Jeffrey A. Riffell, Hong Lei, and John G. Hildebrand. The neurobiology of insect olfaction: Sensory processing in a comparative context. *Progress in Neurobiology*, 95(3):427 – 447, 2011.

[57] Nicolas Y. Masse, Glenn C. Turner, and Gregory S.X.E. Jefferis. Olfactory information processing in Drosophila. *Current Biology*, 19(16):R700 – R713, 2009.

[58] Randolf Menzel and Gisela Manz. Neural plasticity of mushroom body-extrinsic neurons in the honeybee brain. *Journal of Experimental Biology*, 208(22):4317–4332, 2005.

[59] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[60] Katherine I. Nagel and Rachel I. Wilson. Biophysical mechanisms underlying olfactory receptor neuron dynamics. *Nature Neuroscience*, 14:208 EP –, Jan 2011. Article.

[61] Andrew Ng. Sparse autoencoders, https://web.stanford.edu/class/archive/cs/cs294a/cs294a.1104/spar *Webpage*, year?

[62] Shawn R. Olsen, Vikas Bhandawat, and Rachel I. Wilson. Excitatory interactions between olfactory processing channels in the Drosophila antennal lobe. *Neuron*, 54(4):667, 2008.

[63] Shawn R. Olsen, Vikas Bhandawat, and Rachel Irene Wilson. Divisive normalization in olfactory population codes. *Neuron*, 66(2):287–299, Apr 2010. 20435004[pmid].

[64] Shawn R. Olsen and Rachel I. Wilson. Lateral presynaptic inhibition mediates gain control in an olfactory circuit. *Nature*, 452:956 EP –, Mar 2008. Article.

[65] Javier Perez-Orive, Ofer Mazor, Glenn C. Turner, Stijn Cassenaer, Rachel I. Wilson, and Gilles Laurent. Oscillations and sparsening of odor representations in the mushroom body. *Science*, 297(5580):359–365, 2002.

[66] Emmanuel Perisse, Christopher Burke, Wolf Huetteroth, and Scott Waddell. Shocking revelations and saccharin sweetness in the study of Drosophila olfactory memory. *Curr Biol*, 23(17):R752–R763, Sep 2013. S0960-9822(13)00921-4[PII], 24028959[pmid].

[67] Alexandre Pouget and Charvy Narain. A conversation with Alexandre Pouget. *Cold Spring Harb Symp Quant Biol 2014*, 79:285–287, 2014.

[68] Jeffrey A. Riffell, H. Lei, and John G. Hildebrand. Neural correlates of behavior in the moth Manduca sexta in response to complex odors. *Proceedings of the National Academy of Sciences*, 106(46):19219–19226, 2009.

[69] Jeffrey A. Riffell, Hong Lei, Leif Abrell, and John G. Hildebrand. Neural basis of a pollinator's buffet: Olfactory specialization and learning in manduca sexta. *Science*, 2012.

[70] Jeffrey A. Riffell, Hong Lei, Thomas A. Christensen, and John G. Hildebrand. Characterization and coding of behaviorally significant odor mixtures. *Current Biology*, 19(4):335 – 340, 2009.

[71] Jordan C. Roberts and Ralph V. Cartar. Shape of wing wear fails to affect load lifting in common eastern bumble bees (Bombus impatiens) with experimental wing wear. *Canadian Journal of Zoology*, 93(7):531–537, 2015.

[72] Jurgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61(Supplement C):85 – 117, 2015.

[73] B. Settles. *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool, 2012.

[74] Eli Shlizerman, Jeffrey A. Riffell, and J. Nathan Kutz. Data-driven inference of network connectivity for modeling the dynamics of neural codes in the insect antennal lobe. *Frontiers in Computational Neuroscience*, 8:70, 2014.

[75] Ana F. Silbering and C. Giovanni Galizia. Processing of odor mixtures in the Drosophila antennal lobe reveals both global inhibition and glomerulus-specific interactions. *Journal of Neuroscience*, 27(44):11966–11977, 2007.

[76] Marcus Sjoholm. Structure and function of the moth mushroom body. *Swedish Univ of Agricultural Sciences, Alrarp*, 2006. PhD thesis.

[77] Sen Song and L.F. Abbott. Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, 32(2):339 – 350, 2001.

[78] Richard S Sutton and Andrew G Barto. *Reinforcement Learning*. MIT Press, 1998.

[79] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian J. Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *CoRR*, abs/1312.6199, 2013.

[80] Glenn C. Turner, Maxim Bazhenov, and Gilles Laurent. Olfactory representations by drosophila mushroom body neurons. *Journal of Neurophysiology*, 99(2):734–746, 2008.

[81] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 1096–1103, New York, NY, USA, 2008. ACM.

[82] Rebecca L. Vislay-Meltzer and Mark Stopfer. Olfactory coding: A plastic approach to timing precision. *Current Biology*, 17(18):R797 – R799, 2007.

[83] Jiaqiong Wang, Robert J. Hamm, and John T. Povlishock. Traumatic axonal injury in the optic nerve: Evidence for axonal swelling, disconnection, dieback, and reorganization. *J Neurotrauma*, 28(7):1185–1198, Jul 2011. 21506725[pmid].

[84] Rachel I Wilson. Neural and behavioral mechanisms of olfactory perception. *Current Opinion in Neurobiology*, 18(4):408 – 412, 2008. Sensory systems.

[85] Rachel I. Wilson and Gilles Laurent. Role of GABAergic inhibition in shaping odor-evoked spatiotemporal patterns in the Drosophila antennal lobe. *Journal of Neuroscience*, 25(40):9069–9079, 2005.