Distribution Agreement

Signature:

________________________________ _______________
Jirui Qiu Date

Context-dependent Encoding of Descending Neurons in *Drosophila*

By

Jirui Qiu
Master of Science
Physics

Gordon Berman, Ph.D.
Advisor

Ilya Nemenman, Ph.D.
Committee Member

Justin Burton, Ph.D.
Committee Member

Sam Sober, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Context-dependent Encoding of Descending Neurons in *Drosophila*

By

Jirui Qiu
B.Sc., Nanjing University, China, 2015

Advisor: Gordon Berman, Ph.D.

An abstract of

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science
in Physics

2017

Abstract

Context-dependent Encoding of Descending Neurons in *Drosophila*

By Jirui Qiu

Animals exhibit a wide variety of behaviors. How these behaviors are initiated, sustained, and modulated, though, remains largely unknown. For a substantial number of animals, including *Drosophila*, descending neurons communicate signals from the brain to local circuits to generate motor outputs. However, little is known about the associations between descending neurons and behavior outputs. Now the combination of optogenetic techniques and an unsupervised measurement of animal behaviors permits us to study the behavioral effects of stimulating individual descending neurons in freely moving *Drosophilae*. Our study finds evidence supporting the theory that these induced behavioral effects are context-dependency. Previous behavioral statuses can affect the behavioral effects induced by activating individual descending neurons.

Our study applies mutual information as a main approach to measuring the correlations between behavioral probability distributions, and to characterizing context-dependency. To further explore the properties of context-dependency, we then apply multiple statistical methods to analyze the experimental results. Based on information theory, we study the general principles of the signal transmission within the neural network and possible theoretical hypotheses for the context-dependency. This thesis aims to provide

insights into how the brain modulates animal behaviors, and pave the way toward a deeper understanding toward the functioning of the neural network.

Context-dependent Encoding of Descending Neurons in *Drosophila*

By

Jirui Qiu
B.Sc., Nanjing University, China, 2015

Advisor: Gordon Berman, Ph.D.

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science
in Physics

2017

Acknowledgements

First, I would like to express my genuine gratitude to my advisor Prof. Gordon Berman for his generous support over the years. Prof. Berman is an extremely knowledgeable and patient mentor, whose guidance helped me not only overcome obstacles in research, but also learn to be a better person. His kindness and humor always motivated me to proceed and face challenges at the times when I got stuck on problems and felt hopeless. Without Prof. Berman's mentoring, I would never be able to deal with all the difficulties and continuously improve myself.

Secondly, I would like to thank the rest of my committee members, Prof. Ilya Nemenman, Prof. Justin Burton and Prof. Sam Sober for their insightful advice. My sincere thanks also go to other faculties and staffs in the physics department at Emory for all their supports and encouragements for me. Special thanks go to Prof. Minsu Kim for serving as my rotation advisor and delivering warm talks during my first year.

Lastly, I would like to thank all the other fellow members working in Prof. Berman's group, Katherine Overman, Josuan Calderon, Itai Pinkovezky and Yating Yang. I always feel extremely lucky to have these bright and talented colleagues, who brought numerous happiness into my life. Especial thanks go to Katherine for her valuable advice on my master thesis and generous help during my years at Emory.

Contents

Table of Figures

Chapter 1

Introduction

1.1 Introduction

To survive in a given environment, animals must adaptively perform behaviors, spanning from the stereotyped ones, like grooming [1] and courtship singing [2], to complex and environment-based ones, like foraging [3], [4] and crossing obstacles [5]. These behaviors result from the interactions between local motor circuits and higher-order processing centers in the brain [6]. Although numerous efforts have been invested in investigating how animals generate, adapt and maintain behaviors [7], [8], little is known about the exact associations between the neural network and the behavior outputs.

Descending neurons in *Drosophila* provide a feasible system for the study of the encoding process within the neural network. Serving as the bottleneck of signal transmissions, descending neurons transmit the signals from the brain to the ventral nerve chord to generate motor outputs. These descending signals play a critical role in the coordination and adaption of behaviors, like navigating complex terrains [9]. Now optogenetic techniques permit precise stimulations of individual descending neurons, and enable us to study the encoding process.

According to our results, depending on the pre-stimulation behavioral states of *Drosophila*, the stimulation of a specialized group of descending neurons can induce varied behavioral effects. Thus, our hypothesis is that the behavioral effects of stimulating descending neurons are context-dependent. Based on information theory, we then study the general principles of the signal transmission within the neural network and the possible theoretical basis of context-dependency. Further studies have been proposed to investigate whether the context-dependency only works for a distinct population of descending neurons or constitutes a general principal in the whole neural system, and to explore its application in modelling neural network.

1.2 Background

1.2.1 *Drosophila*

Drosophila melanogaster, a fly species, has become a popular model organism for behavioral research due to several factors. First, *Drosophilae* are robust and easy to raise in the laboratory environment. In addition, the availability of genetic tools enable the expression of the targeted genes in a tissue-specific manner. Finally, the relatively high transparency of flies' cuticles under red light and the availability of optogenetic techniques at the corresponding wavelengths enable quick and precise control over the activation of neurons.

1.2.2 Descending neurons

In insects, while the majority of sensory processing tasks are conducted by the brain, motor circuits are mostly located in the ventral nerve cord. To generate appropriate behavior outputs, command signals therefore must be transmitted from the brain to the motor circuits in the ventral nerve cord to modulate behaviors. In *Drosophila*, there exist approximately 550 pairs of descending neurons [10] (Figure 1), which extend a single axon from the sensory-processing brain regions to the ventral nerve cord. Each segment of the ventral nerve cord constitutes of a pair of ganglia, which serve as local processors and modulate motor outputs.

Figure 1 Anatomical structure of a descending neuron.

In *Drosophila*, researchers have associated several descending neurons with specific behavior outputs, like courtship singing and altering walking directions [11]. Also, how the descending control manipulates the behavioral choice has been widely discussed, like the competitive mechanism between forward and backward stepping [12].

1.2.3 Optogenetic Techniques

Optogenetics techniques, which use optically-activated proteins to manipulate neuronal function, have been developed rapidly in recent years [13]. In the past, thermogenetic tools were widely used in neuronal activation. However, since this approach relies on a change of temperature to manipulate neural activities, it suffers from a lack of temporal precision and intensity [14]. Optogenetic techniques provide an approach to stimulating or inhibiting a distinct population of neurons in freely moving animals with millisecond precision.

Although a substantial number of light-sensitive proteins have been engineered, the most popular one is channelrhodopsin, which contains the light-isomerizable all-trans-retinal. When the all-trans-retinal absorbs a photon, it undergoes a conformational change to 13-cis-retinal and permits the inflow of ions. However, within milliseconds, 13-cis-retinal would automatically relax back to all-trans-retinal, blocking the inflow of ions. Thus, in order to activate the channelrhodopsins, the fruit flies must be reared with food containing retinal. For different types of channelrhodopsins, the corresponding activation-required light wavelengths may vary. For example, Channelrhodopsin-2 is capable of transducing millisecond long flashes of blue light (475 nm) into spike trains.

Although optogenetic techniques have been widely applied in mammalian behavioral studies, the applications of such techniques in *Drosophila* were relatively limited due to two factors. First, to regulate the neural activities inside the brain tissue, the ideal lights must be capable of penetrating the skin (the cuticle in *Drosophila*) to reach the deep brain structure, so no surgeries would be needed to thin the skin or implant optical fibers. Previous research has found that the penetration of blue light through the cuticle is relatively weaker (~2%) due to intrinsic blue absorption by the retinal chromophore [15], when compared to longer wavelengths such as red light (~7%) [16]. Second, the stimulation light sometimes would reach photosensitive regions in the eyes and induce strong innate behavioral artifacts. Thus, the application of channelrhodopsins in *Drosophila* was only confined to particular cases, where the light could be blocked from reaching the eyes, or the side effects of visual stimulation would not affect the experiments.

To handle these problems, researchers engineered red-shifted variants of channelrhodopsin [17], such as *CsChrimson* (590 nm) [18], whose excitation spectrum is red-shifted 45 nm more than previous channelrhodopsins. *CsChrimson* permits control over neural activity in freely moving *Drosophila*, at wavelengths that are believed not to interfere with normal visual function. For 5-ms impulses of red light (625 nm), the *CsChrimson* can reliably elicit light-driven spikes at frequencies of at least 20 Hz. This discovery of *CsChrimson* permits control over neural activity in freely moving *Drosophila*, at wavelengths that are believed not to interfere the normal visual function. In our data set, the red LED light (627 nm) was turned on throughout the stimulation period at the level of $5\ \mathrm{mW/cm^2}$. Although the channelrhodopsins were kept open all the time, due to the occasional inhibitory effects from other neurons, the targeted descending neuron might not keep firing through the stimulation period.

1.2.4 GAL4-UAS system

To ensure that the genes of the *CsChrimson* only express in a tissue-specific manner and would not interfere with the normal functioning of other parts, the GAL4-UAS system was applied to control the expressions of the *CsChrimson*. Since its development in 1993, the GAL4-UAS system has been wildly used in the studies of genes' expression and functioning within organisms, like *Drosophila* [19].

The GAL4-UAS system incorporates two components, the GAL4 gene and the UAS (Upstream Activation Sequence). Originally serving as the yeast transcription activator protein, GAL4 can bind to the UAS and activate gene expression of the targeted genes, which are next to the UAS. In *Drosophila*, although GAL4 is not normally present, geneticists have engineered a substantial number of its genetic varieties, called GAL4 lines. Under the control of a driver gene, each line can express GAL4 in a tissue specific manner.

Thus, despite the occurrence of the GAL4 gene in each single cell within the organism, GAL4 is only generated in specific cells to trigger the gene expression of the targeted genes.

However, the standard (Generation 1) GAL4 lines usually express in more than one cell type. To achieve cell-type specificity, the split-GAL4 approach [20] was devised to confine the expression to the overlap in the expression patterns of two standard GAL4 lines. In this method, two lines are made to each express half of the GAL4 protein, which is inactive by each line itself. Only in the cells where both promoters are active, the two halves self-assemble by the leucine zipper into functional GAL4 proteins and activate the target genes. Our current data set includes approximately two hundred split GAL4 lines, targeting varied individual descending neurons in *Drosophila*.

1.2.5 Behavioral analysis

As the observable output of the neural system, animal behavior can be measured at varied time-scales of resolution, constituting of location, trajectory, posture and so on. Starting with raw recorded data (often videos), behavioral analysis classifies animal movements to generate detailed behavioral descriptions. Depending on whether the representations of targeted behaviors have been previously established, behavioral analysis can be divided into two groups, supervised behavioral analysis and unsupervised behavioral analysis.

For supervised behavioral analysis, researchers must encode their intuition about behaviors by manually annotating part of the data set, which could later serve as classifiers to automatically label behaviors. This analysis method shows reliable performance when the data set is relatively small and has been successfully applied to multiple areas, like the behavioral sequences of mice [21], [22]. However, supervised behavioral analysis has two serious drawbacks. First, this analysis method required a large amount of training examples. With the data set increasing and more behavioral annotations added, the training process could consume excessive time and energy in order to reach a certain degree of precision. Second, the behavior measurements are only confined to the ones that are believed to be important. Thus, there potentially exists a number of biases and artifacts.

Attempts to circumvent this obstacle lead to the development of unsupervised behavioral analysis, which establishes the representations of behaviors without manual annotation. The capability of identifying the hidden structure of the data set allows the discovery of structures that researchers are not explicitly looking for. Unsupervised behavioral analysis has been successfully applied to the study of various kinds of animals, such as *C.elegans* [23], mice [24] and *Drosophila* [25].

1.3 Methods

1.3.1 Unsupervised Behavioral Analysis

Our study adopts an unsupervised behavior analysis method [25] to map the recorded behaviors of *Drosophila* into the two-dimensional behavioral space (Figure 2). Based on the assumption that a considerable number of animal behaviors consist of stereotyped actions, this analysis method utilizes the underlying structure within the postural movement data to categorize behaviors in a meaningful and discrete manner.

The central concept of this method is to view animal behaviors as a trajectory travelling through the high-dimensional behavioral space constructed by postural dynamics. In this space, stereotyped behaviors correspond to the locations that the trajectory repeatedly hovers around. On the other hand, non-stereotyped behaviors coincide with the intervals between the stereotyped behaviors in the behavioral space that the trajectory quickly travels through. For *Drosophila*, stereotyped behaviors were found to incorporate more than one hundred varied types and occupy approximately fifty percent of the recorded time.

Within the framework of this analysis method, starting from the raw recorded videos, the images of the *Drosophila* are first segmented from the background frame by frame [26]. During this procedure, the images are aligned rotationally and translationally via cross-correlation to a template image, in which the wings and legs of the fruit fly are manually removed. Then the size of the images is adjusted to occupy the same number of pixels. With the application of principal component analysis (PCA), the images are decomposed into thousands of eigenvectors, among which, fifty with the largest eigenvalues are selected to represent the principal postural modes. In order to characterize the inherent dynamics of behaviors, these components undergo a wavelet transformation to generate spectrograms, through which feature vectors are constructed. For the wavelet transformation, there exist 25 frequency channels spaced between 1 and 50 Hz. Thus, timescales of characterized behaviors are between 0.02 second and 1 second. These behaviors whose timescales are larger than 1 second are not considered in our current study. Using t-Distributed Stochastic Neighbor Embedding (t-SNE), the spectrograms are then embedded into the two-dimensional behavioral space. By repeatedly mapping recorded raw data into the 2-dimentional space, a behavioral map containing the description of the postural characteristics of *Drosophila*, is generated. Within this map, each point corresponds to a unique motion and nearby points represent similar motions. Based on the local density maximum, the behavioral space is split into hundreds of regions, each of which represents a particular stereotyped behavior. By manually examining the videos of the behaviors within each region, these behavioral regions can be generally categorized into six groups, representing varied behavior types (Figure 2).

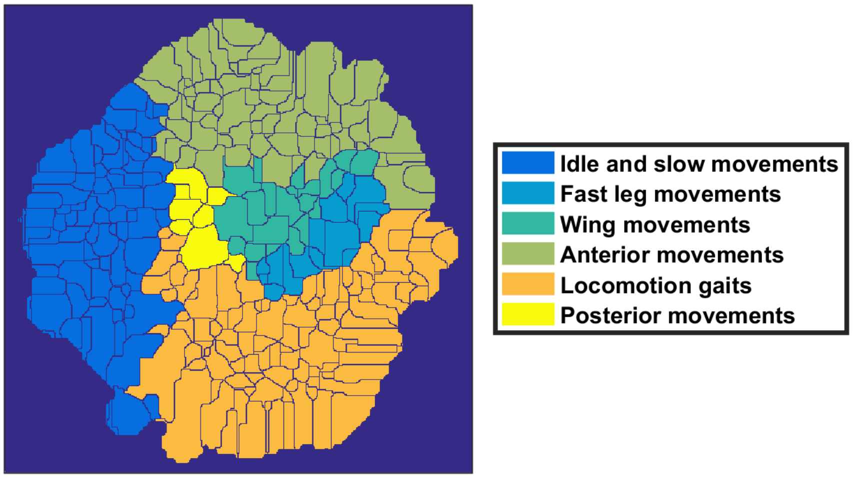

Figure 2 Behavioral space, where regions are clustered into varied groups with different color.

1.3.2 Data Set

Our current data sets of *Drosophila* were contributed by our collaborators at Janelia Research Campus, and consist of approximately two hundred GAL4 lines. Each line corresponds to the activation of a particular population of descending neurons, and consists of twelve individual fruit flies. Although the genes of the twelve flies are exactly the same, only six are reared with food containing retinal. For other six control flies, the channelrhodopsins cannot function without retinal. Each ground-based fly is recorded in a circular arena by 100-Hz high-speed camera for thirty cycles, each of which consists a period of LED-on for 45 seconds and LED-off for 15 seconds, thus yielding 1.8×10^5 movie frames per individual.

1.3.3 Mutual Information

For two random variables X and Y, the mutual information $I(X;Y)$ is the Kullback–Leibler divergence (relative entropy) between the joint probability distribution $p(x,y)$ and the product of marginal distributions $p(x)p(y)$ [27]:

$$I(X;Y) = \sum_{x\in X} \sum_{y\in Y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}. \tag{1}$$

It measures the mutual dependence between the two random variables X and Y. Another common expression of the mutual information is:

$$I(X;Y) = H(X) - H(X|Y). \tag{2}$$

Intuitively, the mutual information $I(X;Y)$ is the deduction in the variability of the variable X given the information of the variable Y.

1.3.4 Estimation of Mutual information

For the calculation of entropy [28], [29], if we randomly draw N sample out of K accessible states, the relation between resulted frequency f_i for state i and the corresponding intrinsic probability p_i can be expressed as $f_i = p_i + \delta f_i$. Thus, the corresponding estimate of the entropy can be written as

$$H_{estimate} = -\sum_i f_i \log_2 f_i = -\sum_i (p_i + \delta f_i) \log_2 (p_i + \delta f_i). \tag{3}$$

With the application of the Taylor expansion, the equation becomes

$$H_{estimate} = -\sum_{i=1}^{K} p_i \log_2 p_i - \sum_{i=1}^{K} [\log_2 p_i + \frac{1}{\ln 2}] \delta f_i - \frac{1}{2} \sum_{i=1}^{K} [\frac{1}{p_i \ln 2}] (\delta f_i)^2 + \cdots. \tag{4}$$

Since it is a random Poisson process, we expect $\langle \delta f_i \rangle = 0$ and $\langle (\delta f_i)^2 \rangle = \frac{p_i}{N}$. Thus, when computing the average of the entropy estimate, we find

$$\langle H_{estimate} \rangle = H_{true} - \frac{K}{(2\ln 2)N} + \cdots. \qquad (5)$$

By substituting equation (5) into the mutual information $I(X;Y) = H(X) + H(Y) - H(X,Y)$, the expression of the mutual information can be written as

$$\langle I(X;Y)_{estimate} \rangle = I(X;Y)_{true} + \frac{K_{X,Y} - K_X - K_Y}{(2\ln 2)N} + \cdots. \qquad (6)$$

Since the number of accessible states $K_{X,Y}$ for the joint distribution $p(x,y)$ is usually much larger than the sum of K_X and K_Y, $\langle I(X;Y)_{estimate} \rangle$ is expected to decrease with the increase of N.

Chapter 2

Results

2.1 Context-dependent Behavioral Transitions

One of the questions we are interested in is how animal behaviors would alter upon the activation of descending neurons. Here, we apply mutual information to study the associations between the pre-stimulation and intra-stimulation behavior probability distributions.

Based on the two-dimensional behavioral space, our pre-stimulation samples correspond to behavior probability distribution within the intervals, 1-5s before the ignition of the LED light. We neglect the period directly adjacent to the initiation of the optogenetic stimulation, because the animal behaviors can be relatively abnormal near the stimulation. To generate the intra-stimulation samples, we then apply the Wilcoxon signed-rank test to identify the significantly upregulated regions in the behavior space in the 0-3s of the stimulation. The criterion is that for each GAL4 line, the median of the intra-stimulation point density in the behavioral space must be significantly not equal to zero. Also, the median of the point density in experimental group must be significantly not equal to the ones in the control group. Among these points, we select the connected upregulated regions and neglect the downregulated ones.

After the establishment of samples, we then must associate the intra-stimulation significantly upregulated regions to the pre-stimulation distributions in the behavioral space. For each stimulation period, we mark the point with highest density in the behavioral map, and measure its distance to each of the regulated regions. In most of the cases, the point would be assigned to the closest region. However, if the distances all surpass the threshold, whose squared value is set to 10 pixels for a square image with length equal to 101 pixels, no assignment would be conducted.

2.2 Shuffling

To study whether these correlations are simply the consequences of the sampling noise, we apply shuffling method to search for GAL4 lines with significantly upregulated mutual information. In the shuffling method, the correspondence between the pre-stimulation distributions and intra-stimulation regions is permutated for each shuffling. Each pre-stimulation distribution is randomly paired to a significant upregulated region. After each shuffling, the mutual information would be recalculated. For each GAL4 line, the shuffling procedure is repeated for one thousand times. Based on the shuffling results, we select the lines whose original mutual information is larger than nighty-five percent of the shuffled ones. Among the 98 GAL4 lines with multiple regions, 16 are found to match this criterion (Figure 3).

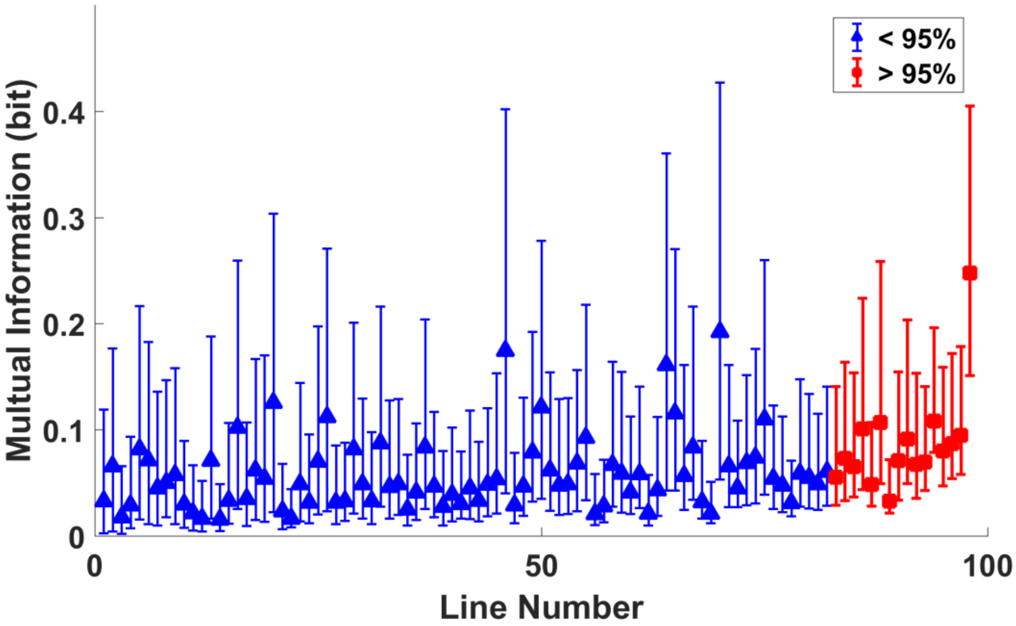

Figure 3 Shuffling results for lines with more than one significant regions. The lines are ranked based on the percentage of shuffling values less than the original ones. Error bars represent the range of the shuffling results. The lines, whose original value of mutual information is larger than ninety-five percent of the shuffling results, are marked as red.

2.3 Estimating Mutual Information

Although the shuffling method enables the detection of the significance level, it is still a relatively conservative and imprecise method. To achieve a more precise assessment of significance, we estimating the true mutual information for all GAL4 lines with two significant regions.

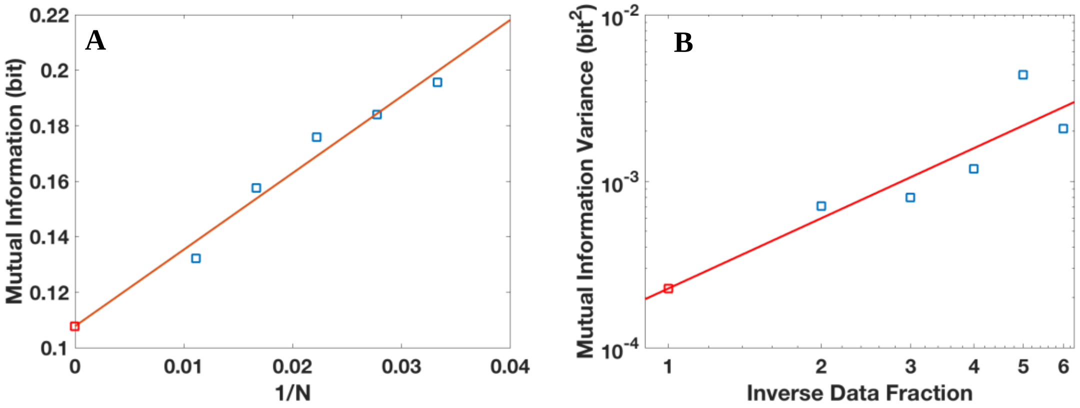

Figure 4 Estimation of the value and variance of mutual information. **A.** Mutual Information vs 1/N. Blue squares represent the mean of the mutual information for the subsamples at each m, where the sample size is $180/m$. Red line is the linear fitted line and the red square represents the estimation of the mutual information. **B.** The estimation of the variance is based on the law $log\,\sigma^2(m) = A + log\,m$. For each m, the variance is calculated for the mean mutual information. This is repeated 10 times to average the variance at each m (blue squares). Then a linear regression is done to estimate the variance of the original data set by setting $m = 1$.

To estimate the mutual information, we divided the N samples in each line nonoverlapping subsamples of size N/m, with the inverse data fraction $m = 2,3, \ldots ,6$. We calculated the mutual information for each subsample and evaluated the variance of the calculated mutual information for each m. We repeated 10 independent partitions at each m to average the mean and the variances. Then, based on equation (6), a linear regression was done to estimate the mutual information for each line (Figure 4A). To estimate the variance, we fit the variance to the 1/(sample size) law $\log \sigma^2(m) = A + \log m$ and found the variance for the original dataset by setting $m = 1$ (Figure 4B).

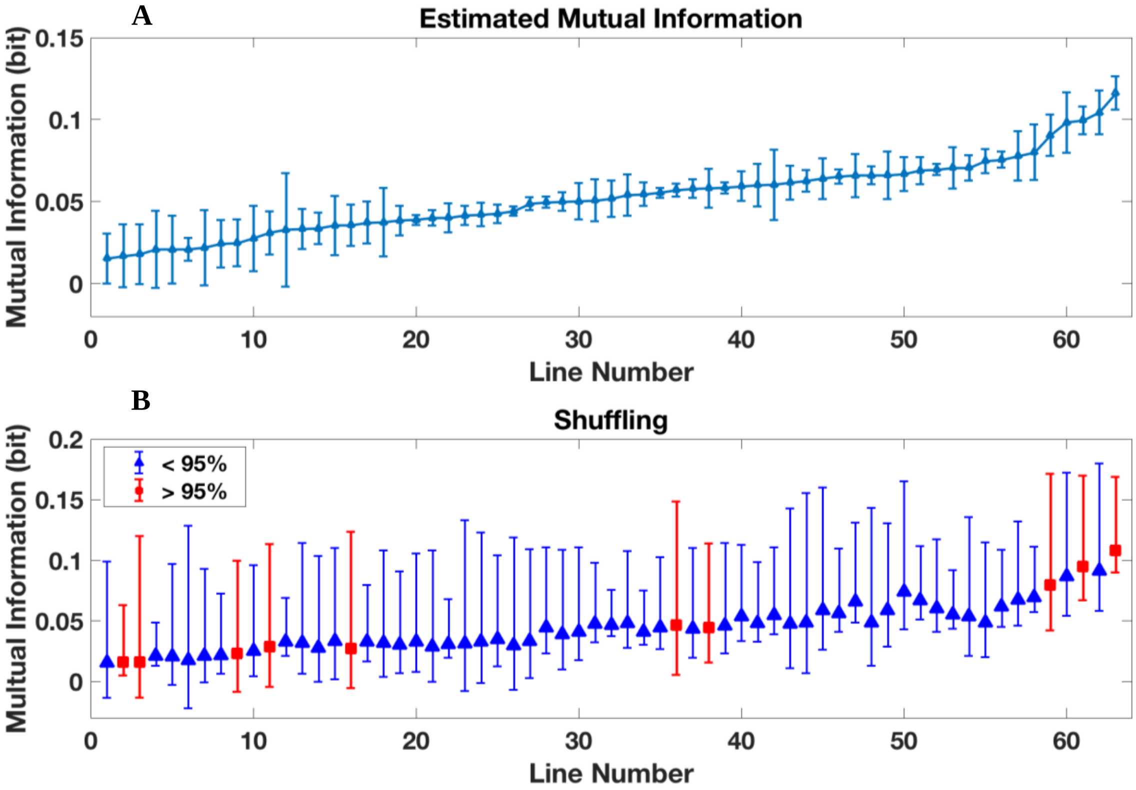

Figure 5 A. The distribution of the estimated mutual information for GAL4 lines with two significant regions. The lines are ranked based on the value of the estimated mutual information triangular dots correspond to the values of the estimated mutual information, and the error bars represent the standard deviation. **B.** The corresponding shuffling results, for the lines in figure A, are ranked in the same order.

Based on the results of estimated mutual information for all 63 GAL4 lines with two significant regions (Figure 5A), we found that largest mutual information turns out to be 0.12 ± 0.01 bit. Given one of the upper bounds of the mutual information for two-region lines is $\ln n_{region} = 1$ bit, this value is relatively large. The exact same line also turns out to be statistically upregulated in the bootstrapping results (Figure 5B).

To compare the results for experimental groups with the ones for control groups, we repeat the same procedure to estimate mutual information for control groups within each line (Figure 6). In 49 of 63 GAL4 lines, mutual information has larger values in experimental groups than control groups. Thus, the activation of descending neurons overall tends to enhance context-dependency in animal behaviors. On the other hand, for the GAL4 lines exhibiting significantly high mutual information for experimental groups, the results of the control group are relatively smaller. It confirms that the behavioral activation of the corresponding descending are significantly context-dependent.

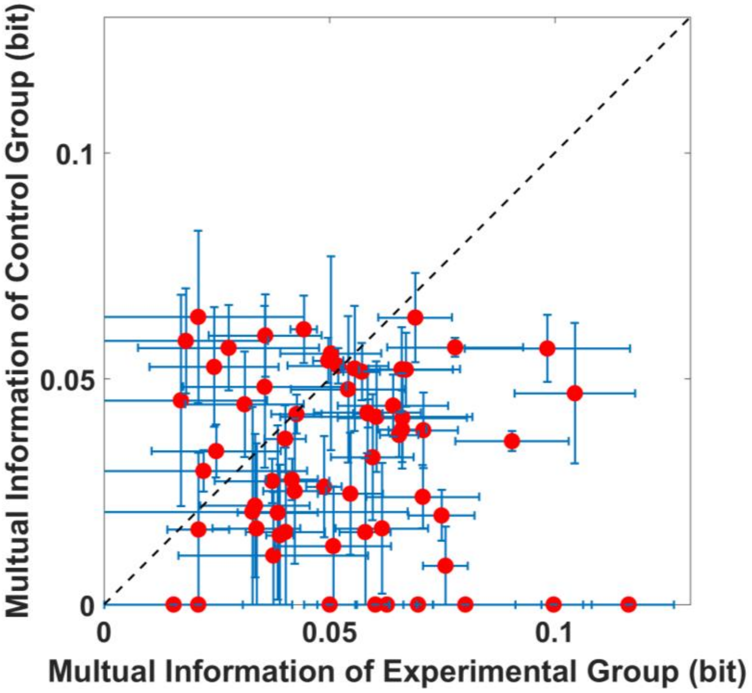

Figure 6 Estimated Mutual information of both experimental and control groups. The red dots represent the results, and error bars correspond to the corresponding standard deviation. The black line denotes the case when data points have equal values for both control groups and experimental groups. In 49 of 63 GAL4 lines, mutual information has larger values in experimental groups than control groups.

To investigate the actual behavior status underlying the context-dependency, we compare pre-stimulation behavior probability distribution and significant intra-stimulation regions in behavioral space. For example, in the GAL4 line with third largest fitted mutual information (Figure 7), although the pre-stimulation behavioral distributions for both regions have a large weight over the anterior-movement area, the distributions show varied weights over wing movements and locomotion gaits for varied upregulated regions. For Region 1, when the pre-stimulation behavior probability distribution has a large weight over the wing movements, the intra-stimulation upregulated regions occupied large area of the idle and slow movements. Although it also covers part of the areas related to the locomotion gaits, these areas are adjacent to the part corresponding to idle and slow movements, thus representing relatively slow locomotion. For Region 2, when the pre-stimulation has a large weight over the locomotion gaits, the intra-stimulation upregulated region also occupies part of the locomotion gaits, corresponding to fast movements. Based on these behavioral transition relations, it seems that depending on the previous behavioral, the activation of the descending neuron triggers the enhancement of varied behaviors. Preferences over slow behaviors induce the enhancement of relatively slow movements, while fast behaviors transit to fast movements.

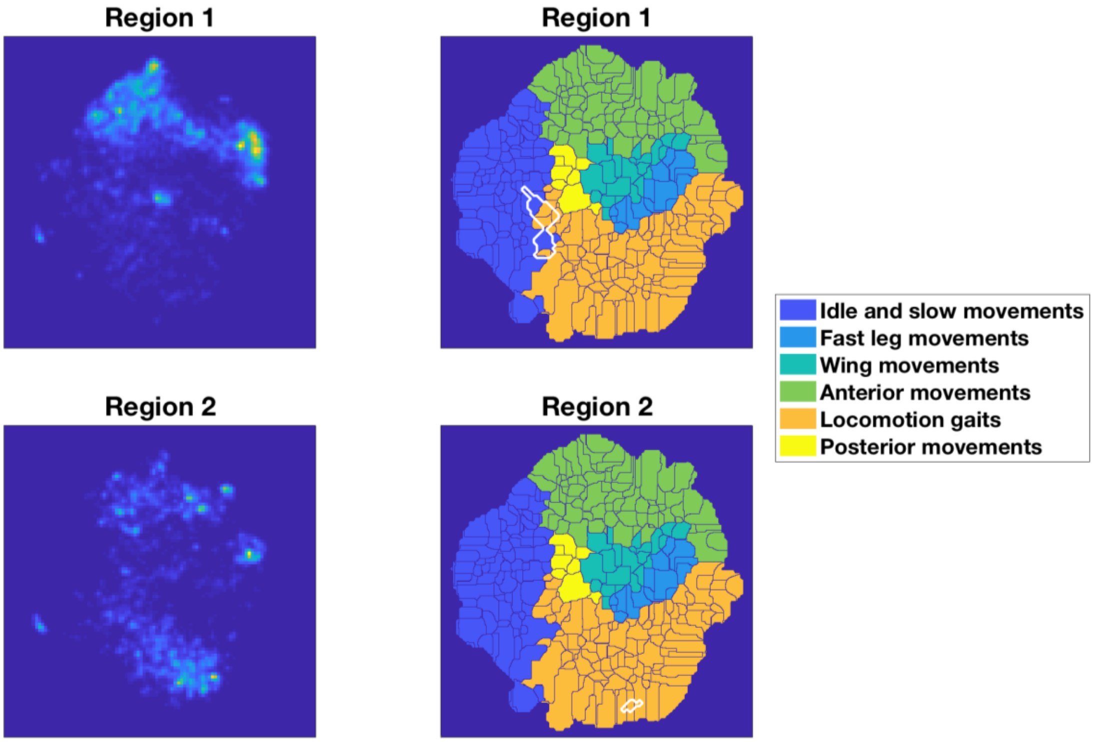

Figure 7 The Pre-stimulation mean behavior probability distribution and the significantly upregulated intra-stimulation regions. The left plots are the pre-stimulation behavior probability distribution on the behavioral space. The left plots show the intra-stimulation statistically upregulated regions, whose contours are marked by white lines.

2.4 Comparison of Behavioral Effects

To investigate the context-dependency on a neural basis, one of the fundamental questions is what factors contribute to the differences between the behavioral effects upon the activation of varied descending neurons. To study the problem, we first found several GAL4 lines, whose corresponding descending neurons descend to the same single segment in the ventral nerve cord from varied brain regions. Here, we aim to investigate the similarities and differences within the behavioral effects, in order to explore the roles of both brains regions and the ventral nerve cord in determining the behavioral effects of neural stimulations.

In the data set, there exists eight lines that innervate the same segment in the ventral nerve cord called the neck neuropil. For each line, the average density distribution on the behavioral map during the first three second of the neural stimulation was calculated. In order to quantify the similarities between these behavioral maps, for each pair of lines, the Jensen-Shannon divergence between their density distributions was calculated. The Jensen-Shannon divergence is a smoothed and symmetrized version of the Kullback-Leibler divergence (relative entropy), $D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$. For discrete probability distribution, the Jensen-Shannon divergence is defined to be

$$D_{JS}(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(Q \parallel M), \text{where } M = \frac{1}{2}(P + Q). \qquad (7)$$

Based on the obtained Jensen-Shannon divergence, we applied the k-medoids algorithm [30] to cluster the data set into two groups (Figure 8). Similar to the k-means algorithm, k-medoids algorithm attempts to minimize the within-cluster sum of squares, $\sum_{i=1}^{k} \sum_{x \in S_i} \|\boldsymbol{x} - \boldsymbol{\mu}_i\|^2$. But instead of calculating the mean of the data points within a cluster to be the center of the observations, k-medoids algorithm adopts one of these data points as the center of the cluster.

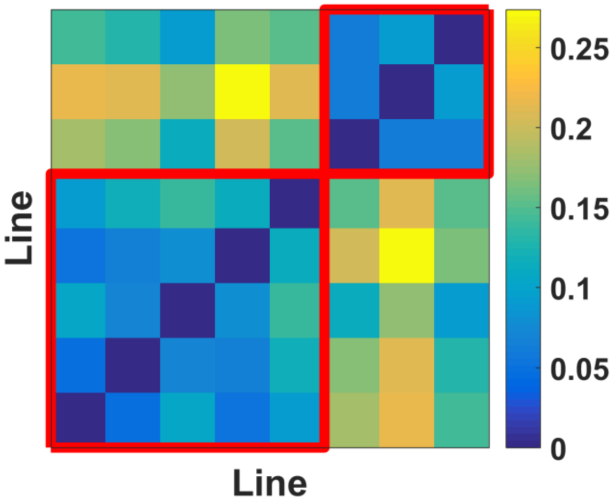

Figure 8 Kullback-Leibler divergence between each pair of lines. k-medoids algorithm was employed to cluster the lines into two groups, whose contour is marked as red. For two distributions, the more similar the distributions are. the smaller Kullback-Leibler divergence becomes.

To examine the results of the clustering, the corresponding intra-stimulation distributions on the behavioral map have been drawn for comparison (Figure 9). As expected, within each cluster, the overall shapes of the distributions look rather similar, while across clusters, the distributions exhibit relatively different patterns. For the first cluster, the distributions occupy large areas corresponding to the fast locomotion gaits, while in the second cluster, the behaviors show preferences over the anterior and slow movements, during which flies usually stand still or move relatively slow. This binary pattern of intra-stimulation behavior probability distribution attracted our attention. One hypothesis is that depending on the type of descending neurons, the effects of activation to the connected segments in the ventral nerve cord could be either excitatory or inhibitory. The amplitude of these effects may also vary across the descending neurons. However, based on the relative small data sets, it is currently impossible to determine the validity of this hypothesis and will be potential directions for the future study.

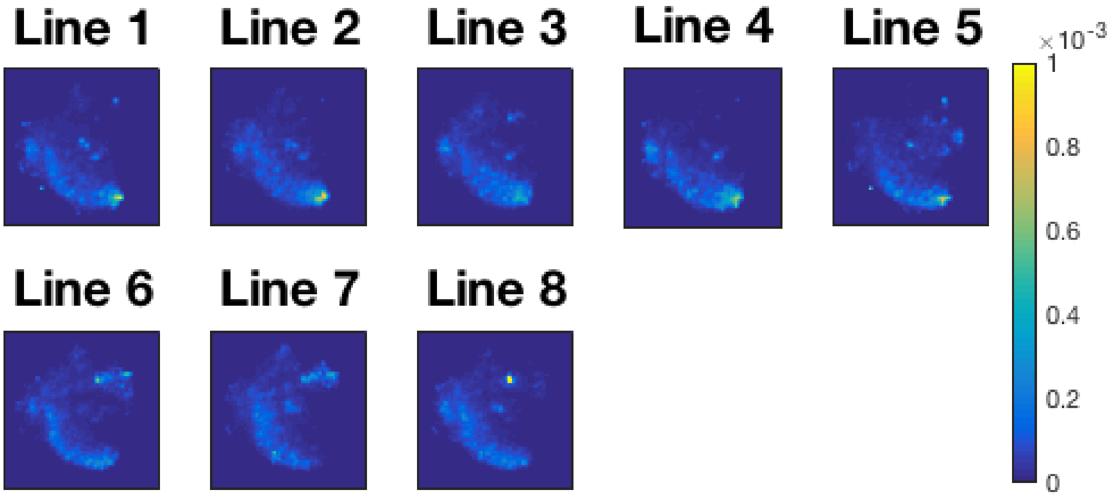

Figure 9 The intra-stimulation behavior probability distribution in the behavior space for the lines within different groups. The first row corresponds to the first group with five lines, while the second row shows the case for the three lines in the second group.

Chapter 3

Discussion

3.1 Introduction

After the discovery of the context-dependency, the following question is why animals choose to conduct behaviors in a context-dependent manner. Here, we mainly focus on the theoretical basis originated from information theory, and briefly discuss the biological basis of this context-dependency.

3.2 Theoretical Basis from Information Theory

For most neurons, the input signals are received by the dendrites and then processed by the cell body. The processed signals are then transmitted through the axon to the dendrites of nearby neurons. These interactions between neurons establish the neural network. Here we apply information theory to investigate the general principles governing the signal transmission within the neural network and the theoretical basis for the context-dependency of behaviors.

3.2.1 Information Transmission within Neural Networks

For the information transmission within organisms, although the links between the input signals and final output can be complex, one interesting question is how the amount of contained information changes throughout this process.

For random variables X, Y, Z, a Markov chain $X \rightarrow Y \rightarrow Z$ can be formed if the joint probability function can be denoted as

$$p(x, y, z) = p(x)p(y|x)p(z|y). \tag{8}$$

Thus, $p(x, z|y) = \frac{p(x,y,z)}{p(y)} = p(x|y)p(z|y)$, which means that given the value of Y, the variables X and Z are conditionally independent, so the corresponding mutual information between variables X and Z given Y is equal to 0, $I(X; Z|Y) = 0$. On the other hand, the mutual information between the variable X and variables Y, Z can be expanded in two ways:

$$I(X; Y, Z) = I(X; Z) + I(X; Y|Z). \tag{9}$$

$$= I(X; Y) + I(X; Z|Y). \tag{10}$$

Due to the nonnegativity of mutual information, we have $I(X; Y|Z) \geq 0$. Based on the conditions $I(X; Z|Y) = 0$ and $I(X; Y|Z) \geq 0$, the induced inequation can be written as

$$I(X; Y) \geq I(X; Z). \tag{11}$$

This useful theorem demonstrates that the further processing of the Y cannot increase the amount of information of X.

For organisms, the more the outside stimuli are processed, the more information is lost through the process. Thus, in order to convey more information towards the outside stimuli, the final outputs are required to be generated within fewer steps. If this number is too large, the organism would become insensitive and unresponsive to the outside stimuli. This insensitivity can restrain the organism's ability to quickly respond to changes and sometimes reduce the chance to survive. This theorem basically sets an upper bound of the number of processing steps within an organism. However, despite the loss of information, most organisms in the nature are equipped with complex neural networks with multiple and complex layers. Thus, the following problem is why a substantial number of organisms chose to develop a brain for information processing through evolution.

3.2.2 Comparisons of Neural Networks with Varied Number of Layers

To study the potential benefits of this multi-layer structure, we first review the previous developments of neural networks [31] and study the limitations induced by the neural networks with only one layer. Here, we examine the performance of the single-layer perceptron neural network, which acts as one of the classic binary classifiers.

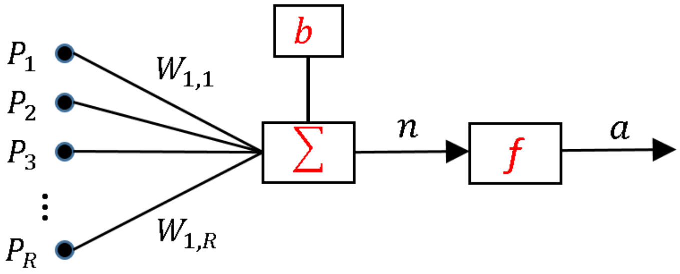

Figure 10 Multiple-input neuron model. The individual inputs p_1, p_2, …p_R are combined based on the weight elements $w_{1,1}$, $w_{1,2}$, …$w_{1,R}$. The bias b is then added to the combine result to generate the net input n, which is then processed by the transfer function f to produce the output a.

First, we start with the single-neuron perceptron network with multiple inputs. For this multiple-input neuron model (Figure 10), the individual inputs p_1, p_2, …p_R are added based on the weight elements $w_{1,1}$, $w_{1,2}$, …$w_{1,R}$ in the weight matrix $\mathbf{W}$. This combined result then adds the bias b to generate the net input n:

$$n = p_1 w_{1,1} + p_2 w_{1,2} + \cdots + p_R w_{1,R} + b. \qquad (12)$$

The bias b is a constant and represents the mean activity level of the neuron without the external inputs. Equation (8) can also be written in the matrix and vector form:

$$n = \mathbf{W}\boldsymbol{p} + b. \qquad (13)$$

Then, this net input n is processed by the transfer function f to generate the output $a = f(n)$. While $\mathbf{W}$ and b can be adjusted through the training algorithm based on the training data set, the transfer function $f(n)$ is usually chosen by the designer. Three most commonly used transfer functions are hard-limit transfer function $f(n) = \begin{cases} 0, n < 0 \\ 1, n \geq 0 \end{cases}$, linear transfer function $f(n) = n$, and log-sigmoid transfer function $f(n) = \frac{1}{1+e^{-n}}$, which transforms the input into the output between 0 and 1. In this section, the hard limit transfer function is applied.

For a perceptron neuron with two inputs, the decision boundary is linear and separates the space into two regions based on the outputs (Figure 11). When more inputs are added, the linearity of the boundary remains. Due to the induced linear decision boundary, the single-neuron perceptron network can only be used to classify the linearly separable data.

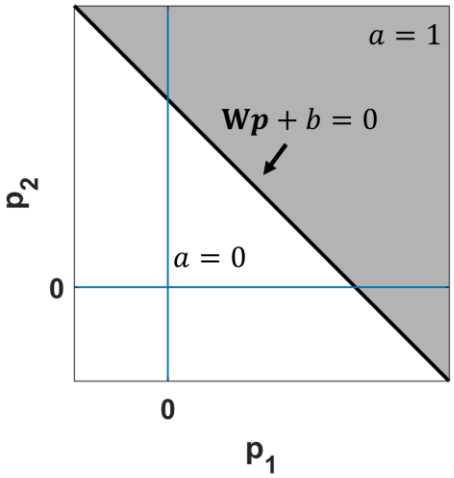

Figure 11 Decision boundary for two-input perceptron neuron model. Based on the values of the inputs p_1 and p_2, the space is linearly divided into two parts, corresponding to different values of the output a.

The following question is whether this limitation can be resolved when more neurons are added. By combining several multiple-input neurons, a single-layer neural network can be established (Figure 12). Like the multiple-input neuron model (Figure 10), R individual inputs are transmitted to S neurons to generate the corresponding outputs a_1, a_2,… a_S. The transfer function within each neuron is not necessarily needed to be the same.

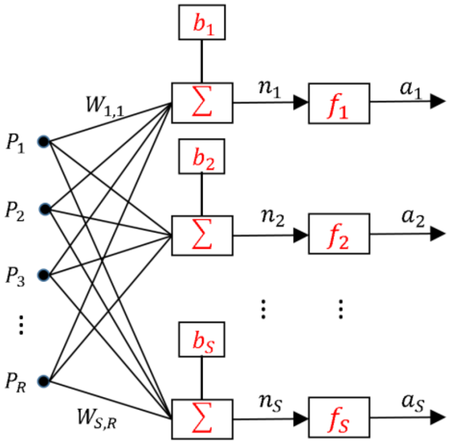

Figure 12 Single-layer neural network. Based on the multiple-input neuron model (Figure 9), a single-layer neural network can be constructed by combining neurons influenced by the same set of inputs.

In this multiple-neuron perceptron network, each neuron classifies the input signals into two categories. Thus, for a single-layer perceptron network with N neurons, there exist 2^N categories. However, although the system simply generates more categories by adding neurons, due to the linearity of the decision boundary, this single-layer perceptron neural network still fails to solve the linearly inseparable problems.

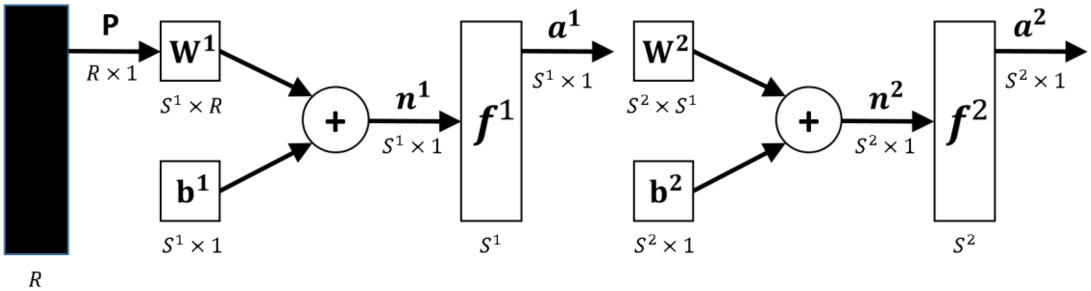

Figure 13 Multi-layer neural network. The plot shows the structure of a two-layer perceptron neural network in vector and matrix notations. The outputs from the first layer $\boldsymbol{a}^1$ serves as the inputs to the second layer to generate the outputs $\boldsymbol{a}^2$.

To overcome the limitations of the single-layer neural network, the multiple-layer perceptron neural network was developed. For multiple-layer perceptron neural network, the outputs from neurons within one layer serve as inputs to neurons in the next layer. By combining several single-layer perceptron neural network together, a multi-layer neural network can be constructed. For example, by using vector and matrix notations, a two-layer perceptron neural network can be constructed (Figure 13).

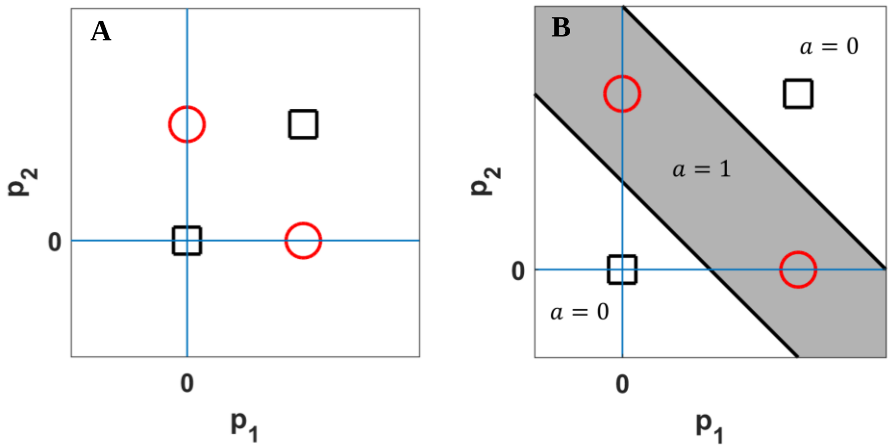

Figure 14 Performance of two-layer perceptron neural network. **A** Four data points belong to two categories denoted by varied shapes (square and round). **B**. In a two-layer perceptron network, the original outputs generated two linear boundaries are further processed by the second layer to complete the classification.

For example, consider the classic exclusive-or (XOR) problem. The four data points belong to two categories denoted by different symbols (Figure 14A) and are not linearly separable. Due to the linearity of boundaries, a single-layer perceptron neural network fails to effectively classify these data points. However, a double-layer network can easily solve the problem. In the first layer, two neurons create two decision boundaries (Figure 14B), whose outputs are then combined by a second layer to generate the final outputs.

Thus, the multi-layer network can overcome the approximation limitation of the single-layer model and achieve better performance. For example, a network, with a sigmoid transfer function on first layer and a linear transfer function on the second layer, exhibits outstanding performance in approximating most functions.

3.2.3 Trade-off between Sensitivity and Complexity

Although the further processing of the input signals cannot increase the amount of contained information, involving multiple layers does equip the system with the ability to perform complex input-output tasks. Thus, the processing of information can be considered as the trade-off between sensitivity and complexity. The reduction of processing steps boosts the sensitivity of the outputs and enables the organisms to quickly generate corresponding behaviors as a respond to the outside stimuli. Sometimes, when it comes to the critical situations, the fast and simple responses enabled by this mechanism can mean the difference between death and survival. For example, when an animal touches an object with a high temperature, it simply needs to pull back that body part quickly. On the other hand, the further processing of the input signals benefits the complexity of the input-output correlations and allows organisms to resolve the complicated tasks.

However, if the loss of information throughout the processing stage is huge, the final outputs would become irrelevant to the input signals and sometimes even seem meaningless. Thus, to compensate the loss of information and maintain an adequate amount of correlations between the outputs and the changing environment, one approach is to receive more input signals at the first stage. On one hand, the number of received input signals can be enlarged by integrating the same type of information from the sensors in multiple body areas. On the other hand, the amount of inputs can also be increased by involving multiple types of sensory signals, like visual, olfactory and auditory information.

3.2.4 Context-dependency – An Approach to Integrating and Compensating the Loss of Information

Apart from increasing the amount of received input signals, another approach to compensating the loss of information is to utilize the previously received information. By combing the currently received signals with the former ones, the whole neural system is enabled to gather enough information for generating corresponding outputs, while maintaining the complexity of the input-output correlations. Thus, the motor outputs of the neural system would become context-dependent and rely on the previous states. On the other hand, due to limitations on the amount of originally received signals imposed by the number of sensors, the utilization of previous information might be simply targeted at integrating more information, instead of compensating the losses.

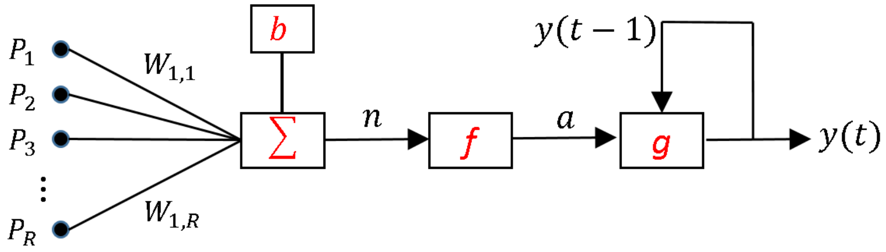

Figure 15 Implementation of context-dependency. The current input-induced effect a and the previous state $y(t-1)$ are further processed to generate the output $y(t)$.

Although there exist various models involving the dependency of the previous states, the exact mechanisms can be complex and is not the focus of the current discussion. One of the most straightforward method (Figure 15) to implement the impacts from the previous states is to add an additional function g to further process the current input-induced effects a with the previous state $y(t-1)$ to generate the output $y(t)$. The mathematical expression can be written as

$$y(t) = g[a + y(t-1)]. \tag{14}$$

3.3 Biological Basis of Context-Dependency

At the biological level, our belief in the context-dependency roots in interactions between neural activities. Although the exact details of how the brain makes decisions and modulates animal behaviors remain largely unknown, consensuses have been reached toward the importance of the complex interactions between neurons. Due to these interactions, the alternation of the activity of a single neuron could potentially exert significant impacts upon the whole neural system. On the other hand, the transmission of the impacts and signals could not be completed instantaneously. Even from the perspective of a single neuron, its activity is regulated by multiple ion channels, whose activation and inactivation consume time. Depending on the timing of exerted outside stimulus and its own activity state, the corresponding impacts upon neural activities may vary. Thus, we assume the behavioral effects of activating individual descending neurons would rely on the previous states of the neural system. Although the details concerning these states remains a mystery, as the observable outputs of this system, animal behaviors are believed to be an appropriate reflection of the internal activities in the neural system.

Chapter 4

Prediction and Experimental Design

4.1 Introduction

Our analysis results find evidence supporting the theory that the behavioral effects of activating individual descending neurons are context-dependent. To proceed from our current findings and explore the command hierarchy within nervous systems, we here discuss potential experimental designs and corresponding predictions. Despite the fact that part of the experiments or techniques may not be currently available, we aim at demonstrating approaches to searching for the potential biological factors contributing to this context-dependency, including influences from other neighboring descending neurons, the interactions between segments in the ventral nerve cord, and the firing rate of descending neurons.

4.2 Inhibiting Neighboring Descending Neurons

In each segment of the ventral nerve cord, multiple descending neuron are connected. The activities of these neighboring descending neurons could affect the functioning of the activated individual descending neurons. Thus, one possible hypothesis is that the influences from neighboring descending neurons contribute to the context-dependent encoding of individual descending neurons. The effects induced by activating individual descending neurons are modified by the activities of neighboring descending neurons. All neural activities are combined together to determine the outputs of the segment in the ventral nerve cord. Therefore, depending on the activities of neighboring descending neurons, the behavioral effects of activating individual descending neurons may vary.

To evaluate the validity of the proposed hypothesis that influences from neighboring descending neurons contribute this context-dependency, corresponding experiments are designed to inhibit the normal functioning of neighboring descending neurons, and remeasure animal behaviors when the descending neurons are activated. In the designed experiment, for each activated individual descending neurons and its connected segments, the neighboring descending neurons would be first identified. During the experiment, all these neighboring neurons would be inhibited and become incapable of functioning by genetic tools. Then, researchers can activate the individual descending neurons by optogenetic techniques, and record animal behaviors. By comparing these experimental results with the ones in which neighboring neurons are not inhibited, we can determine whether neighboring descending neurons participate in shaping the context-dependent encoding. If the proposed hypothesis were solid, the induced behaviors by the individual descending neurons would no longer depend on the previous behavioral states, even though the induced behavioral effects were context-dependent in the experiment without inhibiting neighboring descending neurons.

4.3 Isolating Segments in Ventral Nerve Cord

Apart from the influences from neighboring descending neurons, since the segments in the ventral nerve cord are interconnected, the interactions between these segments can also play a critical role in determining the context-dependent encoding of individual descending neurons. Thus, although each segment regulates the functioning of its corresponding body part, its functioning can be influenced by the activities of nearby segments. The interactions between these segments determine the final behavior outputs. Therefore, depending on the states of neighboring segments in the ventral nerve cord, the behavioral effects of activating individual descending neurons may vary.

To test the proposed hypothesis that interactions between segments in the ventral nerve cord contribute to the context-dependent encoding, in the designed experiment, all connections between segments would be removed, by surgeries or genetic tools, to isolate each segment from other influences. Researchers can then activate the individual descending neurons and record behavior outputs. The validity of the proposed hypothesis can be determined by comparing these experimental results with the ones in which segments are not isolated. If the proposed hypothesis were accurate, after the isolation of segments in the ventral nerve cord, the induced behavior outputs would no longer depend on previous behavioral states, given the fact that the encoding of descending neurons is context-dependent when segments in the ventral nerve cord are not isolated from each other.

4.4 Descending Neurons at Multiple Firing Rates

In our current experimental data-sets, although individual descending neurons were activated by optogentic techniques during the stimulation periods, the exact firing rates of descending neurons were not under accurate measurement or control. However, firing rate represents a significant part of neural activities, and might be responsible for the discovered context-dependent encoding of descending neurons. One potential hypothesis is that during each stimulation period, the optogenetic techniques may induce different firing rates in descending neurons, and lead to varied behavioral effects.

To evaluate the role of firing rates of descending neurons in determining behavior outputs, one of the designed experiments is to the record or control the exact spiking activities of descending neurons. By repeatedly activating individual descending neurons for a substantial number of times, the correspondence relations between firing rates and induced behavioral outputs can be built. If the proposed hypothesis were solid and firing rates of descending neurons did participate in affecting the final behavior outputs, different behaviors would correspond exactly to varied ranges of firing rates. On the other hand, in our current experiment, firing rates are bounded by the kinetics of channelrhodopsins, and may not cover the normal firing rates of descending neruons. The channelrhodopsin used in our current experiment, *CsChrimson*, has relatively slow kinetics and can only reliably drive spikes up to 20 Hz [18]. Thus, it would be interesting to explore the effects of descending neurons at higher spiking frequencies and investigate whether new phenotypes can be generated, by the application of other channelrhodopsins or techniques.

Chapter 5

Further Studies

5.1 Alternative quantifications of context-dependency

To investigate the context-dependent encoding of descending neurons, a significant first step is to quantitatively measure and characterize the context-dependency. Thus, one important part of the future studies would be to continue searching for appropriate parameters for the quantification of context-dependency.

In our current study, mutual information is applied to quantify the context-dependency between the pre-stimulation and intra-stimulation behavior probability distributions. In the next phase, it would be beneficial to continue improving this application of mutual information, in terms of both the selection of time intervals and the assignment of the upregulated regions. Despite the successful application of the mutual information, future studies should also focus on exploring the other potential parameters for the quantification of context-dependency. For example, we previously studied Jensen-

Shannon divergence between the behavior probability distributions in the adjacent intervals, and found that the calculated values always rise during the optogenetic stimulation. Thus, the Jensen-Shannon divergence can be a potential parameter for quantifying both the behavioral effects of the optogenetic stimulation and the extend of context-dependency.

5.2 Context-dependency in Multiple Scales

Currently, the context-dependency is discussed in terms of behavior probability distributions. Serving as a general quantification parameter of animal behaviors, probability distributions provide a relatively straightforward approach to exploring the changes in behaviors. However, the behavior probability distribution only constitutes one relatively small segment of animal behaviors. Other factors, like behavioral sequential execution, also play a significant role in shaping animal behaviors. Thus, exploring context-dependency within these factors would be a potential direction for future studies.

Previous studies have invested a substantial number of efforts in investigating the neural modulation of sequential behavior executions, like the suppression hierarchy among behavior executions within the grooming sequence in *Drosophila* [1]. Thus, one of the future directions is to expand the study of the context-dependency to multiple scales, including both the behavior probability distribution and transitions between behaviors. In our previous study, we found that depending on the previous behavioral sequences, the following behavioral transition probability would vary. Due to the limitation of the size of data sets, we are currently unable to proceed along this direction and explore the context-dependency in the behavioral sequences. Thus, quantifying the changes in the behavioral sequential execution should be a promising study direction for the future studies.

5.3 Properties of the Communication Channels

Serving as the information bottleneck, descending neurons provide useful resources for studying the properties of the communication channels. Thus, for future studies, a prospective field is how to generalize the properties of the communication channel based on the context-dependency, including both the competition between individual descending neurons and the integration of multiple neurons.

For the study of competitions between individual descending neurons, an important step is to understand the roles of individual descending neurons. In the neural system, the majority of synapses are chemical synapses, which are capable of various signaling, including mediating either excitatory or inhibitory effects on postsynaptic cells [32]. As illustrated in the Results section, we explore the behavioral effects upon the activation of the descending neurons connected to the same single segment in the ventral nerve cord. Thus, the next step can be to further explore the impacts of individual descending neurons, and investigate the roles of brain regions and segments in the ventral nerve cord, in determining behavior outputs.

Another interesting aspect is the integration of the activities of multiple descending neurons. Currently, a substantial number of GAL4 lines have been found to be capable of simultaneously expressing in multiple descending neurons. Thus, one of future directions is to obtain data sets where multiple descending neurons are simultaneously activated, and investigate whether new phenotypes are generated by the simultaneously stimulation to study the signal integration process of the communication channels. By repeating this process using larger assemblies of descending neurons, in the long term, the final goal is to investigate the general principals underlying the whole neural system and make corresponding predictions that could be tested in the experiments.

5.4 Conclusion

Within the frame of information theory, the data compression and the noisy-channel coding constitute two main topics, where there exists a substantial amount of evidence directly related to context-dependency. For example, arithmetic coding [33], a once widely used compression method for text files and images, relies heavily on the sequential order of codes and usually generates varied results under different contexts. For the future study, given such abundant materials and resources, it become relatively important to choose an appropriate level and scale. For example, in our current study, the animal behaviors instead of direct activities of neurons are studied. In the future, depending on both the technical developments and theoretical understanding, different choices might be made and more levels can be explored. In current phase, since the details of neural network are still largely unknown, scientists usually choose to start from particular cases and then try to generalize fundamental principles. This is also part of what we did and tried to do. However, in these cases, the biases of the samples should be carefully examined. For the future study, the combination of both solid basis of details and fundamental principles would be extremely important and useful.

Chapter 6

Summary

6.1 Results

Based on an unsupervised measurement of animal behaviors and optogenetic techniques, our study aims at exploring the context-dependency within the behavioral effects of activating individual descending neurons in *Drosophila*. To quantify this context-dependency, we first calculate the mutual information between probability distributions between pre-stimulation and intra-stimulation behaviors, and use the shuffling method to find 16 out of 98 GAL4 lines have significantly large mutual information. To better evaluate the context-dependency within each GAL4 line, we then use statistical method to estimate the true mutual information for GAL4 lines with 2 significant regions. Out of these 63 GAL4 lines, the largest estimated mutual information turns out to be 0.12 ± 0.01 bit. Given one of the upper bounds of the mutual information for two-region lines is $\ln n_{region} = 1$ bit, this value is relatively large. By comparing the experimental groups with control groups, 49 out of 63 GAL4 lines have larger mutual information in experimental groups than control groups. Based on information theory, we also study the general principles of the signal transmission within the neural network, and put forward possible theoretical hypotheses for the context-dependency. In the end, we discuss several potential experimental designs and directions for future studies.

6.2 Discussion

Scientists have been exploring how the brain controls behaviors for a long time. However, the lack of both quantification tools and the ability to independently control individual neurons have seriously restricted this exploration. Based on the optogenetic techniques, our study applied an unsupervised behavior analysis method to precisely quantify the behavioral effects upon neural activations. The combination of optogenetic techniques and an unsupervised measurement of animal behavior provides a unique opportunity to investigate how signals are encoded to generate behavior outputs.

The encoding process in the neural network has become a heated topic in recent years. Several articles have discussed context-dependency in the encoding process at multiple levels. For example, researchers have found the evidence of context-dependency in the encoding of olfactory behaviors in *Drosophila* [34], and the context-dependent control of vocal acoustics by individual muscles [35]. However, these researches are usually confined to particular systems, like olfactory system and the targeted behaviors are relatively simple, like attraction or aversion. Based on a newly developed unsupervised measurement of animal behaviors, our study focus on systematic behavioral impacts of neural stimulations, and concentrate on descending neurons to avoid the involvement of other unknown factors, like the brain.

In our current study, the corresponding segments of the ventral nerve cord are only considered to be the ones that descending neurons are directly linked to, based on the anatomical structures. However, these segments are interconnected, and can influence each other. Thus, it is relatively difficult to identify which segment plays the major role in generating the corresponding behaviors. Investigating the interactions between segments in terms of behavior generation can be a potential direction for future studies.

To explore the general principles of signal transmission within the neural system, a critical step is to obtain the distribution of the corresponding parameters for a substantial amount of data set. Although the current data set seems relatively smaller for generating meaningful principles. we focus on the properties in a small population of descending neurons. To generalize fundamental principles, more examination and comparisons are needed.

6.3 Conclusion

This thesis studies the context-dependency of the behavioral effects upon the optogenetic activations of descending neurons in *Drosophila*. Based on the analysis results, information theory is then applied to investigate the signal transmission mechanisms within neural networks, and possible theoretical hypotheses for the context-dependency. Combining both experimental results and theoretical analysis, this thesis aim to provide insights into the functioning of neural systems, and assist the search for the major principles governing the functions of the neural system.

References

[1] Andrew M Seeds *et al.*, “A suppression hierarchy among competing motor programs drives sequential grooming in Drosophila,” *eLife*, 2014.

[2] Anne C. von Philipsborn, Tianxiao Liu, Jai Y. Yu, Christopher Masser, Salil S. Bidaye, and Barry J. Dickson, “Neuronal Control of Drosophila Courtship Song,” *Neuron*, vol. 69, no. 3, pp. 509–522, 2011.

[3] Adam J Calhoun, Sreekanth H Chalasani, and Tatyana O Sharpee, “Maximally informative foraging by Caenorhabditis elegans,” *eLife*, 2014.

[4] David H. Gire, Vikrant Kapoor, Annie Arrighi-Allisan, Agnese Seminara, and Venkatesh N. Murthy, “Mice Develop Efficient Strategies for Foraging and Navigation Using Complex Natural Stimuli,” *Curr. Biol.*, vol. 26, no. 10, pp. 1261–1273, 2016.

[5] Simon Pick and Roland Strauss, “Goal-driven behavioral adaption in gap-climbing Drosophila,” *Curr. Biol.*, vol. 15, no. 16, pp. 1473–1478, 2005.

[6] Anke Borgmann and Ansgar Buschges, “Insect motor control: methodological advances, descending control and inter-leg coordination on the move,” *Curr. Opin. Neurobiol.*, vol. 33, pp. 8–15, 2015.

[7] Gordon J. Berman, William Bialek, and Joshua W. Shaevitz, “Predictability and hierarchy in Drosophila behavior,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 113, no. 42, pp. 11943–11948, 2016.

[8] Andrea Maesani, Pavan Ramdya, Steeve Cruchet, Kyle Gustafson, Richard Benton, and Dario Floreano, “Fluctuation-Driven Neural Dynamics Reproduce Drosophila Locomotor Patterns,” *PLoS Comput. Biol.*, vol. 11, no. 11, 2015.

[9] Roy E Ritzmann and Ansgar Buschges, “Adaptive motor behavior in insects,” *Curr. Opin. Neurobiol.*, vol. 17, pp. 629–636, 2007.

[10] CynthiaT. Hsu and Vikas Bhandawat, “Organization of descending neurons in Drosophila melanogaster,” *Sci. Rep.*, vol. 6, 2015.

[11] Salil S. Bidaye, Christian Machacek, Yang Wu, and Barry J. Dickson, “Neuronal Control of Drosophila Walking Direction,” *Science*, vol. 344, no. 6179, pp. 97–101, 2014.

[12] Tihana Jovanic *et al.*, “Competitive Disinhibition Mediates Behavioral Choice and Sequences in Drosophila,” *Cell*, vol. 167, pp. 858–870, 2016.

[13] Jonathan P. Britt, Ross A. McDevitt, and Antonello Bonci, “Use of channelrhodopsin for activation of CNS neurons,” *Curr. Protoc. Neurosci.*, 2012.

[14] Jacob G Bernstein, Paul A Garrity, and Edward S Boyden, “Optogenetics and thermogenetics: technologies for controlling the activity of targeted cells within intact neural circuits,” *Curr. Opin. Neurobiol.*, vol. 22, pp. 61–71, 2012.

[15] Wenjing Wang, Zahra Nossoni, Tetyana Berbasova, and Camille T. Watson, “Tuning the Electronic Absorption of Protein-Embedded All-trans-Retinal,” *Science*, vol. 338, no. 6112, pp. 1340–1343, 2012.

[16] Hidehiko K. Inagaki, Yonil Jung, Eric D. Hoopfer, Allan M. Wong, and Neeli Mishra, “Optogenetic control of freely behaving adult Drosophila using a red-shifted channelrhodopsin,” *Nat. Methods*, vol. 11, no. 3, pp. 325–332, 2014.

[17] John Y Lin, Per Magne Knutsen, Arnaud Muller, David Kleinfeld, and Roger Y Tsien, “ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation,” *Nat. Neurosci.*, vol. 16, pp. 1499–1508, 2013.

[18] Nathan C Klapoetke, Yasunobu Murata, Sung Soo Kim, Stefan R Pulver, and Amanda Birdsey-Benson, “Independent optical excitation of distinct neural populations,” *Nat. Methods*, vol. 11, no. 3, pp. 338–346, 2014.

[19] A.H. Brand and N. Perrimon, “Targeted gene expression as a means of altering cell fates and generating dominant phenotypes,” *Development (Cambridge, England)*, vol. 118, no. 2, pp. 401–415, 1993.

[20] Haojiang Luan, Nathan C. Peabody, Charles R. Vinson, and Benjamin H. White, “Refined Spatial Manipulation of Neuronal Function by Combinatorial Restriction of Transgene Expression,” *Neuron*, vol. 52, no. 3, pp. 425–436, 2006.

[21] Mayank Kabra, Alice A Robie, Marta Rivera-Alba, Marta Rivera-Alba, Steven Branson, and Kristin Branson, “JAABA: interactive machine learning for automatic annotation of animal behavior,” *Nat. Methods*, vol. 10, no. 1, pp. 64–67, 2013.

[22] JOHN C. FENTRESS and FRANCES P. STILWELL, “Grammar of a Movement Sequence in Inbred Mice,” *Nature*, vol. 244, pp. 52–53, 1973.

[23] Greg J. Stephens, Bethany Johnson-Kerner, William Bialek, and William S. Ryu, “Dimensionality and Dynamics in the Behavior of C. elegans,” *PLoS Comput. Biol.*, vol. 4, no. 4, p. e1000028, 2008.

[24] Alexander B. Wiltschko *et al.*, “Mapping Sub-Second Structure in Mouse Behavior,” *Neuron*, vol. 88, pp. 1121–1135, 2015.

[25] Gordon J. Berman, Daniel M. Choi, William Bialek, and Joshua W. Shaevitz, "Mapping the stereotyped behaviour of freely moving fruit flies," *J. R. Soc. Interface*, vol. 11, no. 99, 2014.

[26] John Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 694–714, 1986.

[27] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[28] William Bialek, *Biophysics: Searching for Principles*. Princeton University Press, 2012.

[29] Kyle H. Srivastava *et al.*, "Motor control by precisely timed spike patterns," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 114, no. 5, pp. 1171–1176, 2017.

[30] Kaufman, L. and Rousseeuw, P.J., "Clustering by means of Medoids," *Statistical Data Analysis Based on the L1–Norm and Related Methods*, pp. 405–416, 1987.

[31] Hagan, M. T., H.B. Demuth, and M.H. Beale, *Neural Network Design*. Boston: PWS Publishing, 1996.

[32] Eric R. Kandel, James H. Schwartz, Thomas M. Jessell, and Steven A. Siegelbaum, *Principles of Neural Science*, 5th ed. McGraw-Hill Education, 2012.

[33] David J.C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[34] Laurent Badel, Kazumi Ohta, Yoshiko Tsuchimoto, and Hokto Kazama, "Decoding of Context-Dependent Olfactory Behavior in Drosophila," *Neuron*, vol. 91, pp. 155–167, 2016.

[35] Kyle H. Srivastava, Coen P.H. Elemans, and Samuel J. Sober, “Multifunctional and Context-Dependent Control of Vocal Acoustics by Individual Muscles,” *J. Neuroscience*, vol. 35, no. 42, pp. 14183–94, 2015.