Learning pattern recognition and decision making in the insect brain

R. Huerta

BioCircuits Institute, University of California San Diego, La Jolla 92032, USA.

Abstract. We revise the current model of learning pattern recognition in the Mushroom Bodies of the insects using current experimental knowledge about the location of learning, olfactory coding and connectivity. We show that it is possible to have an efficient pattern recognition device based on the architecture of the Mushroom Bodies, sparse code, mutual inhibition and Hebbian leaning only in the connections from the Kenyon cells to the output neurons. We also show that despite the conventional wisdom that believes that artificial neural networks are the bioinspired model of the brain, the Mushroom Bodies actually resemble very closely Support Vector Machines (SVMs). The derived SVM learning rules are situated in the Mushroom Bodies, are nearly identical to standard Hebbian rules, and require inhibition in the output. A very particular prediction of the model is that random elimination of the Kenyon cells in the Mushroom Bodies do not impair the ability to recognize odorants previously learned.

Keywords: pattern recognition; decision making; learning; memory formation.
PACS: 87.85.dq, 87.85.Ng, 87.55.de, 87.55.kh, 87.57.nm, 87.64.Aa, 87.85.D-

INTRODUCTION

The process of deciding what action to take based on the current and future expected external/internal state is typically called decision making [1, 2, 3, 4, 5]. There are two key critical information processing components ubiquitous in the decision making process: i) the prediction of one's action on the environment, *i.e.*, regression, and ii) a pattern recognition problem to discriminate situations, *i.e.*, classification. Both tasks require models to substantiate the action of decision making, and the processes and mechanisms by which those models are learned reveal plausible mechanistic explanations of learning in the brain [6, 7, 8].

In this paper we want to elaborate on the decision making mechanisms that require learning using the most primitive form of all sensory modalities: chemical sensing. This is the sensory modality that coexisted with all forms of life on earth, from the living bacterias to the human brain and remains puzzling and enigmatic despite being so primordial. The insect brain is our choice to understand the underpinnings of learning because they rely on the olfactory modality and they are simpler than the mammalian counterparts. Moreover, the main brain areas dealing with olfactory processing are fairly well known due to the simplicity of the structural organization [9, 10, 11, 12, 13, 14, 15, 16], the nature of the neural coding [17, 18, 19, 20, 21, 22, 23, 24, 25, 26], the advent of the genetic manipulation techniques that isolate brain areas during the formation of memories [27, 28, 29, 30], and the extensive odor conditioning experiments that shed light into the dynamics of learning during discrimination tasks [31, 32, 33, 34, 35, 6].

The main areas where we will concentrate our efforts to understand learning are the

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 101-119 (2013); doi: 10.1063/1.4776507

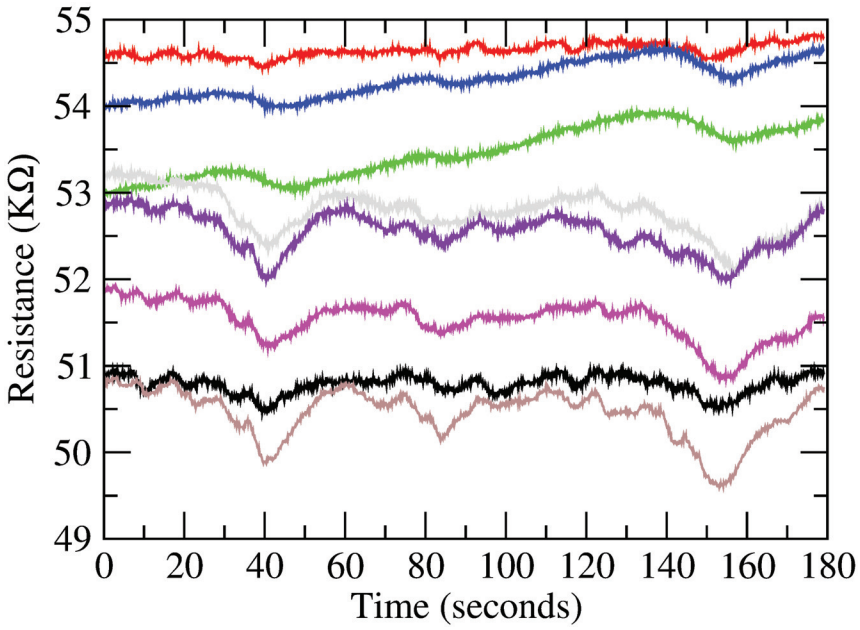

FIGURE 1. Recordings using an artificial sensor array [43, 48, 49, 50, 44] during carbon monoxide presence for 180 seconds in a wind tunnel under turbulent flow.

mushroom bodies [36, 37, 38, 5]. These are responsible for memory formation [36]. There are two additional layers of critical importance for odor processing just in front of the mushroom bodies which are the antennas and the antennal Lobes, but, although memory traces are present [39, 40, 41, 42], their primary function might be signal processing, feature extraction or information filtering [41]. Each of those processing layers are very different in their anatomical and physiological properties. Therefore, since our goal is to understand the mechanisms of learning, we first direct our efforts to the memory formation in the mushroom bodies (MBs) before understanding what specific aspects of the antennal lobe (AL) and the antenna are important.

THE STRUCTURAL ORGANIZATION OF THE INSECT BRAIN

The nature of the olfactory stimulus is stochastic due to the unreliable information carrier. The wind transports gases by turbulent flows that induces complex filaments of gas (see sensor responses in Fig. 1 in [43, 44, 45] and also recordings using a ionization detector in [46, 47]). The nature of the olfactory information differs very markedly from other sensory modalities like vision or audition. The information is intermittent and unreliable, yet evolution has provided to these primitive nervous systems the ability to extract all the necessary information for survival. The brain modules involved in pattern recognition in olfaction are the antennas, the antennal lobes (ALs) and the mushroom bodies (MBs).

Early code

The sensors in the antenna are called the olfactory receptor cells. They are also present in mammals [51] and we still do not have the sensor technology capable of reaching their reaction times, selectivity and stability [48, 49, 50]. Each type of olfactory receptor cell in the antenna connects to a specific glomerulus in the AL [52, 53, 54]. Thus, a chemosensory map of receptor activity in the antenna is represented in the AL. This genetically encoded architecture induces a stimulus-dependent spatial code in the glomeruli [55, 56, 57, 23, 58]. Moreover, the spatial code is maintained across individuals of the same species [59] as would be expected given the genetic structure. In principle this peripheral olfactory structure already seems to be able to discriminate among odors at this early stage. However, the ability to discriminate depends on the number of possible odors, their concentrations, and the complexity in the presence of mixtures [25].

Temporal dynamics in the Antennal Lobe

The antennal lobe receives the input from the olfactory receptor cells that deliver the information into particular sets of glomeruli. The neural network in the AL is made of projection neurons (PNs), which are excitatory, and lateral neurons (LNs), which are mostly inhibitory. The PNs and the LNs connect to each other via the glomeruli. The glomeruli structure induces a bipartite graph of connections that contrasts to the standard directed Bernoulli-induced graphs typically used in AL models [60, 61, 62, 63, 64, 65] with a few exceptions [66]. Moreover, the connections via the glomeruli may be complicated enough because they can be presynaptic [67, 68].

The odor stimuli processed by insects are not constant in time because insects move and the odor plumes flow through the air. The coding mechanism of the AL has to deal with this because the insect needs to detect the odor class, the source and the distance to the odor [16, 69] (see Fig. 2). Since the early works of [70, 71, 72] many experiments have demonstrated the presence of spatio-temporal patterns in the first relay station of the olfactory system of invertebrates and vertebrates [73, 74, 75, 76, 77, 22, 78, 79, 80]. This dynamics results from the interplay of excitatory and inhibitory neurons [74, 81, 82]. There is some debate about the function of temporal coding in behavior, because individuals react faster solving discrimination task than the structure of the temporal code indicates [83, 84, 85]. However there is evidence that by blocking inhibition in the AL insects lack the ability to discriminate between similar odors [33, 74]. And, second, the distance to the source of the odorant is encoded in the intermittentcy of the turbulent flow [43, 44, 45, 46, 47]. The further away from the source the sensors are, the slower the peak frequency of the recordings becomes. Thus, these two facts point out at the need for temporal code to solve pattern recognition (what gas) and regression estimation (how far and where). In fact, these two different functions may use separate pathways in the insect brain [69].

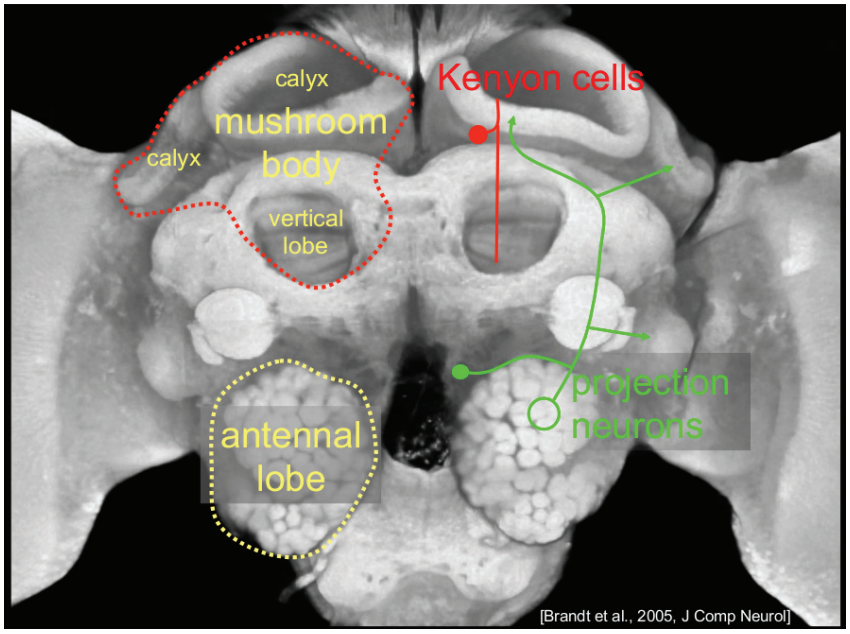

FIGURE 2. Anatomy of the honey-bee brain (courtesy of Robert Brandt, Paul Szyszka, and Giovanni Galizia). The antennal lobe is circled in dashed yellow and the MB is circled in red. The projection neurons (in green) send direct synapses to the Kenyon cells (KCs) in the calyx which is part of the MB.

The Formation of Memories in the Mushroom Bodies

Gain control. It is known that increasing odor concentrations recruit increasing numbers of glomeruli [86] and the activity of the projection neurons within the AL [87, 88, 89]. From the behavioral point of view, increasing the synapses of the local inhibitory neurons makes gases repulsive and to change the number of excitatory synapses makes odorants attractive [90]. This study shows how important the tight equilibrium of the excitatory and inhibitory network is in the AL. A simple transduction of the PN activity into the MB would increase the number of active KCs as well. However, mean firing rates of the PNs that send the output to the MBs have constant firing rates regardless of the gas concentration [91] and recordings of the PNs in the MBs show concentration independence[92][1]. Moreover, the *drosophila* shows that calcium activity is also independent of odor concentration in the KC neurons [30]. Therefore, the activity of the Kenyon cells KCs (see Fig. 2) appears to be heavily regulated and has been shown to generate sparse activity in Honeybees and Locust [94, 95, 96], which is consistent with the overwhelming predictions of associate memory and pattern recognition models [97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107].

[1] Despite there is evidence of gain control at least at the MB level, we still are missing good controls for the concentration of the gases delivered in the antenna [93]. For example, for 1-Hexanol dilution on mineral oil over 1% the concentration of the gas in the air saturates at 200-400 particle per millions.

Learning. Most importantly, the MBs undergo significant synaptic changes [108, 109] and it is known for quite a long time that play a key role in learning odor conditioning [110, 111, 112, 113, 27, 114, 115, 116]. This situates the MBs as the center of learning in the insects. A notion that has been promoted by Martin Heisenberg all along [36].

Inhibition. Another important factor to be able to organize learning and in particular pattern recognition in the Mushroom Bodies is inhibition or more specifically lateral inhibition in the output layer. The notion of lateral inhibition to improve information output has been around for a while [117]. Without lateral inhibition in the output it is not possible to organize competition to have neurons responding for a particular set of stimulus [118, 104]. In fact, it has been shown that there is strong lateral inhibition in the β-lobes in the MBs in [119] which is consistent with standing theoretical models [117, 120, 118, 104, 121].

Temporal code. We have argued that to estimate the distance to the source, temporal processing is required due to the turbulent transport of the gas (see Fig. 1). Further analysis of the data in [95] shows that at the early stages of the processing, right after the stimulus induces a reaction in the insect brain, the MBs can have better ability to discriminate. At later stages, the receptive fields or sensitivity of the MBs become broader [122]. Perhaps at this level slow lateral excitation between the KCs may better encode temporal information of the plume [123]. This implies that the discrimination and recognition of the gas may happen quickly, but the gas concentration estimation requires temporal integration over long time scales as shown in gas source localization using artificial sensor arrays [43, 48, 49, 50, 44].

THE COMPUTATIONAL ORGANIZATION OF THE INSECT BRAIN

In the previous section we tried to succinctly summarize some of the most relevant facts that are needed to build a pattern recognition device in olfaction. These are not by any means all of them and are not necessarily fully consistent with each other, but despite their differences there is more coherence than dissonance with the elements required to have an efficient pattern recognition device. In Fig. 3 we depict the basic model that we use to analyze the computational properties of the MBs.

The simplest model. If we want to understand first the role that the connectivity of the insect brain plays in pattern recognition problems, one has to chose the simplest possible model that complies with the integration properties of neurons. The basic concept is that whenever there is sufficient synaptic input arriving into a neuron, it is going to fire, respond, or transmit information to another group of neurons. A classic model of a neuron that is still successfully used today is the McCullough-Pitts neuron [124]. It is remarkable that it is still used despite being 70 years old and it is used to get estimates of the degree and strength of connections of network architectures to be implemented in

more realistic models[2]. The McCullough-Pitts (MP) neuron is expressed as

$$y_j = \Theta\Big(\sum_{i=1}^{N_{\rm AL}} c_{ji} x_i - \theta_{\rm KC}\Big) \quad j = 1,2,...,N_{\rm KC}. \tag{1}$$

$\mathbf{x}$ is the state vector of the AL neurons (see Fig. 3). It has dimension $N_{\rm AL}$, where $N_{\rm AL}$ is the number of AL neurons. The components of the vector $\mathbf{x} = [x_1, x_2, ..., x_{N_{\rm AL}}]$ are 0's and 1's. $\mathbf{y}$ is the state vector for the KC layer or Calyx; it is $N_{\rm KC}$ dimensional. The c_{ij} are the components of the connectivity matrix which is $N_{\rm AL} \times N_{\rm KC}$ in size; its elements are also 0's and 1's. $\theta_{\rm KC}$ is an integer number that gives the firing threshold in a KC. The Heaviside function $\Theta(\cdot)$ is unity when its argument is positive and zero when its argument is negative. This model can be generalized by replacing the Heaviside function by a nonlinear increasing function and can also be recast in the format of an ordinary differential equation to obtain the a Grossberg-type [126, 127] or Wilson-Cowan models [128].

Advantages and challenges. The MP model is adequate to answer limits in performance of pattern recognition devices for fast operation which is sufficient to account for the fast reliable code observed in the AL [129, 83, 84, 85]. It is also very useful to establish the equivalence with classical pattern recognition devices like the support vector machines (SVMs) [130, 131, 132, 133, 134]. However, it fails at comprehending the role of time in the brain [135, 136, 137, 138, 139] and thus by itself cannot easily solve the regression or distance-to-source estimation problem. Even if the system can recognize efficiently objects, it has to be controlled and regulated within the circuit itself and from other brain areas. This is a challenging problem that we do not address in this paper and requires models with a proper description of the time scales in the brain.

Information Conservation in the Mushroom Bodies

Hypothesis. The main hypothesis is that the Mushroom Bodies are a large screen where one can discriminate objects much more easily. The theoretical basis for discrimination on a large screen to discriminate more easily was already proposed by Thomas Cover [140] and later within the framework of support vector machines [141]. In addition, sparse code is a very useful component to achieving a powerful pattern recognition device as observed in the MBs and as already mentioned the theoretical support for sparse code is extensive [97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107].

Sparse code. The evidence of sparse code in the Calyx is found in the locust [95, 96] and the honeybee [94, 95]. The prevailing theoretical idea to make the code stable over time from the AL to the MB is using forward inhibition [142, 106, 96]. In what follows we will assume that neural circuits are placed in an stable sparse mode.

The AL-MB circuit as an injective function. As we are not addressing the temporal aspects of the system for now, the input for our classification system is an early sin-

[2] See for example the transition from a MP model in [104] to a realistic spiking model in [106] and the model of learning in [125] that resembles closely the model in [105].

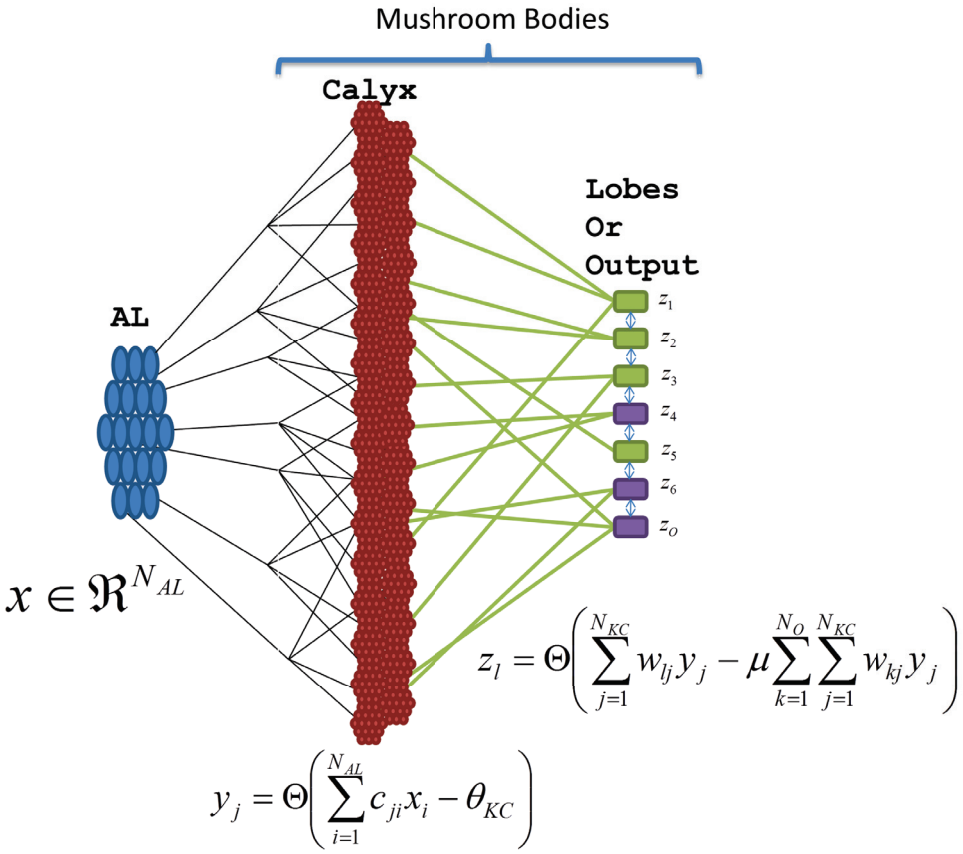

FIGURE 3. The equivalent model of the MBs. We denote by $\mathbf{x}$ as the AL code. $\mathbf{y}$ is the code on the Calyx or the KC neurons, that sometimes we will refer to $\mathbf{y} = \Phi(\mathbf{y})$ in the context of SVMs, and $\mathbf{z}$ is the output of the MBs. Note that all the output neurons inhibit each other with a factor μ.

gle snapshot of information when the antenna hits the plume. The hypothesis for the nonlinear transformation from the AL to the MB is then that every such snapshot or codeword in the AL has a unique corresponding codeword in the MB: The nonlinear transformation needs to be an injective function at least in a statistical sense. In [143] it was proposed to select the parameter values that allow constructing such an injective function from the AL to the KC layer with very high probability.

To determine the statistical degree of injectivity of the connectivity between the AL and KC, we first calculate the probability of having identical outputs given different inputs for a given connectivity matrix: $P(\mathbf{y} = \mathbf{y}' \,|\, \mathbf{x} \neq \mathbf{x}', C)$, where C is one of the possible connectivity matrices (see [143] for details) and the notation $\mathbf{x} \neq \mathbf{x}'$ is $\{(\mathbf{x}, \mathbf{x}') : \mathbf{x} \neq \mathbf{x}'\}$. We want this probability, which we call the probability of *confusion*, to be as small as possible, on the average over all inputs and over all connectivity matrices.

We write this average as $P(\text{confusion}) = \left\langle \langle P(\mathbf{y} = \mathbf{y}' \,|\, \mathbf{x} \neq \mathbf{x}', C) \rangle_{\mathbf{x} \neq \mathbf{x}'} \right\rangle_C$, where $\langle \cdot \rangle_{\mathbf{x} \neq \mathbf{x}'}$ is the average over all non-identical input pairs $(\mathbf{x}, \mathbf{x}')$, and $\langle \cdot \rangle_C$ is the average over all connectivity matrices C. This gives us a measure of *injectivity*, the opposite of confusion,

as

$$I = 1 - P(\text{confusion}), \tag{2}$$

The closer I is to 1, the better is our statistically injective transformation from the states $\mathbf{x}$ of the AL to the states $\mathbf{y}$ of the KCs.

There are two parameters of the model that can be adjusted using the measure of injectivity. One is the probability p_C of having a connection between a given neuron in the AL and a given KC. The second is the firing threshold θ_{KC} of the KCs. Fixed parameters in the model are the probability p_{AL} of having an active neuron in the AL layer, the number N_{AL} of input neurons, and the number N_{KC} of KCs. p_C and θ_{KC} can be estimated using the following inequality

$$I \leq 1 - \{{p_{\mathrm{KC}}}^2 + (1 - p_{\mathrm{KC}})^2 + 2\sigma^2\}^{N_{\mathrm{KC}}}, \tag{3}$$

where p_{KC} is the firing probability of a single neuron in the KC layer. It can be calculated for inputs and connection matrices generated by a Bernoulli process with probabilities p_{AL} and p_C as

$$p_{KC} = \sum_{i=\theta_{KC}}^{N_{AL}} \binom{N_{AL}}{i} (p_{AL} p_C)^i (1 - p_{AL} p_C)^{N_{AL} - i}. \tag{4}$$

where the summatory starts at the the threshold level at which the neurons can fire. This probability has variance (σ^2) for all the prior probabilities of the inputs $\mathbf{x}$ and connectivity matrices. This type of connectivity can be very unstable for perturbations of activity in the input[144]. As can be seen in Fig. 4 where small variations of the probability of activation of AL neurons can lead to a very sharp change in the MBs [145]. This unstability makes necessary to have gain control mechanisms to regulate the sparse activity as proposed in [142, 106, 96, 146] via forward inhibition or by synaptic plasticity [147]. The regulation of sparseness via plasticity from the AL to the MB is an unlikely mechanism to generate sparseness because it actually reduces the information content on the KCs [148].

The formula for the probability of confusion can be intuitively understood if we assume that the activity of every KC is statistically independent from the activity of the others. If so, the probability of confusion in one output neuron is the sum of the probability of having a one for two inputs plus the probability of having a zero for both: $p_{\mathrm{KC}}^2 + (1 - p_{\mathrm{KC}})^2$. Thus, the probability of confusions in all N_{KC} output neurons is $(p_{\mathrm{KC}}^2 + (1 - p_{\mathrm{KC}})^2)^{N_{\mathrm{KC}}}$ in the approximation of independent inputs. This bound on I should be close to unity for any set of parameter values we choose. The inequality for the measure of injectivity becomes an equality for sparse connectivity matrices.

Information preservation versus discrimination and stability. If one takes realistic physiological values [143] one can summarize the expression of confusion just in terms of $n_{KC} \ll N_{KC}$, which is the total number of simultaneously active neurons of the KC layer, as follows:

$$P(\mathbf{x} = \mathbf{x}') \propto e^{-2 \cdot n_{KC}}. \tag{5}$$

which means that ideally to improve injectivity, the system should be placed as far as it can from sparse code reaching the maximum at $n_{KC} = N_{KC}/2$. However, first,

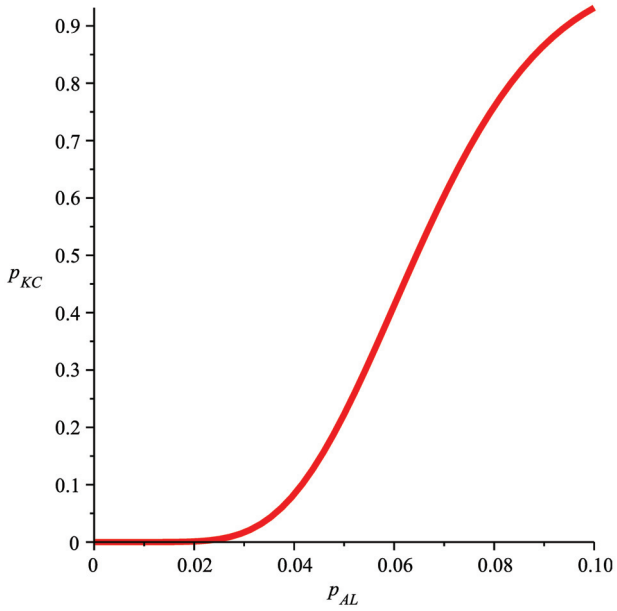

FIGURE 4. Probability of activation of the KC neurons as a function of the probability of the activation off the AL neurons. $N_A = 1000, p_C = 0.15$ and $\theta_{KC} = 10$.

the expression of injectivity in Eq. (3) saturates very rapidly, and, second, in terms of classification performance or memory storage one wants the opposite [97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107]. And in terms of stability as shown in Fig. 4 using realistic parameter values it is difficult to place the MB activity in moderate levels of activity between 10 and 50 percent for all possible inputs and for all the concentration levels (see [145] for details). In fact, a 5% variation in activity in the AL can switch the KCs from sparse activity to having almost all neurons responding to the input. This instability in the statistics of the input is not a desirable property of a pattern recognition device.

Learning Pattern Recognition in the Mushroom Bodies

Rationale. The evidence of learning odor conditioning in the MBs has mounted over the years. Although there has been plasticity shown in the AL network [41], the role played there is data tuning or preprocessing for the pattern recognition device, which is the MB. As we will show the MBs can be shown to be nearly equivalent to Support Vector Machines (SVM) [130] not only in terms of architecture but also in terms of learning. The data tuning or preprocessing in the dynamical system which is the AL can be shown to improve the performance of the pattern recognition device [44]. How this can be carried out in biologically plausible manner remains a mystery.

Beyond olfaction. Despite our main effort has been on olfaction, models of learning in the MBs have increased recently due to the multimodal nature of the MBs [149]. For

example, Wu and Gao's model of decision making of the visual information has the the center of the decision making in the MBs too [8]. The MBs not only have olfactory information but also contextual information, making the MB an integrative center that takes about 35% of the neurons in the insect brain [150].

The model. In [104, 105] we propose a basic model of learning in the MBs which is based on the MP neurons where the key component is to have the output neurons of the MB inhibiting each other (see Fig. 3) as

$$z_l = \Theta\Big(\sum_{j=1}^{N_{\rm KC}} w_{lj}\cdot y_j - \mu \sum_{k=1}^{N_O}\sum_{j=1}^{N_{\rm KC}} w_{kj}\cdot y_j\Big),\ \ l=1,\ldots,N_{\rm O}. \tag{6}$$

Here, the label O denotes the MB Lobes and μ denotes the level of inhibition between the output neurons. The output vector $\mathbf{z}$ of the MB lobes has dimension $N_{\rm O}$. The $N_{\rm KC}\times N_{\rm O}$ connectivity matrix are subjected to Hebbian learning [151] but implemented in a stochastic form. Synaptic changes do not occur in a deterministic manner [152, 153]. Axons are believed to make additional connections to dendrites of other neurons in a stochastic manner, suggesting that the formation or removal of synapses to strengthen or weaken a connection between two neurons is best described as a stochastic process [154, 153].

Output coding and classification decision. We do not know where the final decision or odor classification is taking place in the insect. It is even possible that we will never know because the neural layers from the periphery to the motor action are connected by feedback loops. This intricate connections make difficult to isolate areas of the brain during the realization of particular functions. What we can argue from the theoretical point of view is that the decision of what type of gas is presented outside in the antenna can take place in the output neurons of the MBs, $\mathbf{z}$, with a high odor recognition performance. This performance is much higher than any other location of the layers involved in olfactory processing.

Inhibition in the output. We hypothesized that mutual inhibition exists in the MB lobes and, in joint action with 'Hebbian' learning is able to organize a non-overlapping response of the decision neurons [104]. Recently this hypothesis was verified in [119] showing that the inhibition in the output neurons is fairly strong, and in [108] where plasticity was found from the KC neurons into the output ones.

Reinforcement Signal. One important aspect of learning in the insect brain is that it is not fully supervised. The reward signal is delivered to many areas of the brain as good (octopamine[3]), $r=+1$, or bad (dopamine), $r=-1$. The Mushroom Bodies are innervated by huge neurons that receive direct input from the gustatory areas [155]. They play a critical role in the formation of memories [5] and the can remain activated for long periods of time releasing octopamine into not only the MBs but also the ALs and other areas of the brain. In addition, the delay between the presence of the stimulus and the reward has an impact[156] in learning memories. The learning rules that one can use to have the system learn to discriminate are not unique [105]. For example one can

[3] Note that in the mammalian brain is just the opposite.

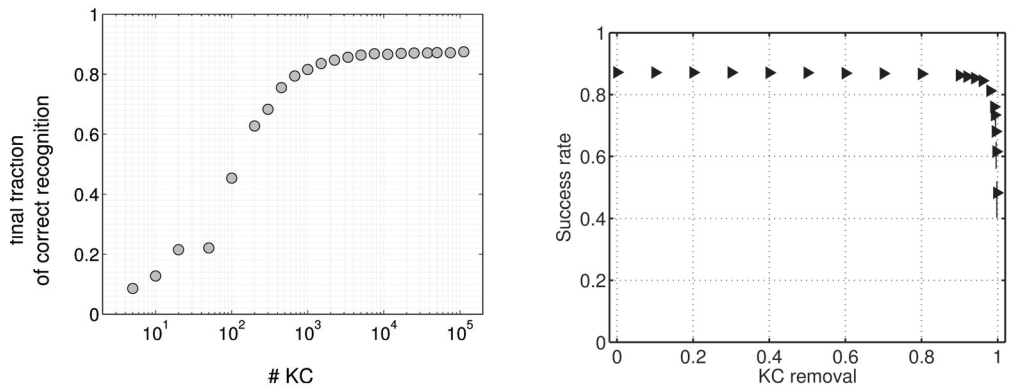

FIGURE 5. (LEFT) Accuracy in the classification of the MNIST handwritten digits for different sizes of the MB. (RIGHT) Success rate as function of random elimination of KCs.

use rules similar to [157] that are contained in the following expression:

$$\Delta w_{ij} \propto y_i\, r(US) z_j \,\text{with}\, P(US, y_i, z_j), \tag{7}$$

where the changes of the synaptic connections between the KCs and the output neurons depend on the activity of both layers and the reward signal $r(US)$ with some probability $P(US, y_i, z_j)$ that depends on the state of the MB and the nature of the unconditional stimulus, or reward signal. US denotes what behavioral experimentalists call unconditional stimulus that is our reward signal. Note that in [105] the values of $P(US, y_i, z_j)$ have a significant impact in the performance of the classifier.

Impact of MB size on accuracy and Robustness. The brain's ability to learn better is thought to be positively correlated with larger brains [158]. Larger brains consume more energy but memorize better and can survive in more complex environments. In [105] we investigated the models given by Eqs. (1,6) and apply them to the MB to solve a very well-known problem in pattern recognition: the MNIST dataset. The MNIST dataset is made of 60,000 handwritten digits for training the model and 10,000 for test [159]. Despite the digits are not obviously gases. The representation of the information in the MB is multimodal [149], so we can analyze the ability to recognize better by exploring larger brain sizes and provide a direct comparison with pattern recognition methods in machine learning. The main results are shown in Fig. 5 where we can show that the ability to have better accuracy in the recognition of digits with increasing brain sizes. The other interesting results of that investigation is the robustness of the MBs to damage or elimination of the KCs. On the right panel of Fig. 5 we can see that one has to eliminate above 99% of the KC neurons to observe a serious impairment of the performance in pattern recognition. This is another prediction. If the insect has learned and been trained previously, damage of the Calyx will not degrade its performance.

EQUIVALENCE BETWEEN THE MUSHROOM BODIES AND SUPPORT VECTOR MACHINES

Using the inhibitory SVM formalism proposed in [134], the synaptic input arriving into an output neuron, $\hat{z}_k$, can be expressed as

$$\hat{z}_k(\mathbf{x}) = \sum_j w_{kj}\Phi_j(\mathbf{x}) - \mu \sum_l \sum_j w_{lj}\Phi_j(\mathbf{x}) = \sum_j w_{kj} y_k - \mu \sum_l \sum_j w_{lj} y_j, \tag{8}$$

where $\Phi_j(\mathbf{x})$ is the nonlinear function that projects the AL code $\mathbf{x}$ into the KC neurons or $\mathbf{y}$. The response of this neuron is a threshold function on $\hat{z}_k$. For the purposes of the SVM what matters is the value of the synaptic input, $\hat{z}_k$, so we will concentrate on its analysis. To make the notation more compact let us write

$$\hat{z}_k(\mathbf{x}) = \hat{z}_k(\mathbf{y}) = \langle \mathbf{w}_k, \mathbf{y} \rangle - \mu \sum_l \langle \mathbf{w}_l, \mathbf{y} \rangle.$$

Note that to make $\hat{z}_k(\mathbf{x}) = \hat{z}_k(\mathbf{y})$ implies that during learning in the SVM there will not be learning from the projections of the AL to the MB.

For the sake of simplicity we consider that the SVM will classify a binary problem. A particular stimulus, $\mathbf{x}$, has a label $r = +1$ for positive label and $r = -1$ for negative label. Now since both the SVM and the honeybee learn by examples let us say that there are a total of N stimulus/examples, $\mathbf{y}_i$, with their corresponding labels, $r_i = +1, -1$ with $i = 1, \ldots, N$. The idea is that to have the classifier working properly then $r_i \hat{z}_k(\mathbf{y}_i) \geq 0$ for all the examples. However, a key concept in SVMs is that the SVM output needs to be above some margin such that $r_i \hat{z}_k(\mathbf{y}_i) \geq 1$. The margin value of 1 is standard although we can chose any value one likes. The most important thing to understand is that the examples belonging to different classes are sufficiently separated from each other. The next important aspect of SVMs is the loss function which is expressed as

$$\min_{\mathbf{w}_k} L = \min_{\mathbf{w}_k} \left(\frac{1}{2} ||\mathbf{w}_k||^2 + C \sum_{i=1}^{N} \max \left\{ 0, 1 - r_i \langle \mathbf{w}_k, \mathbf{y}_i \rangle \right\} \right). \tag{9}$$

The first term is called the regularization term, which is an upper bound to the generalization error [132], the second term corresponds to the measure of classification error using the hinge loss function. The hinge loss is not the most plausible error function because it is known that most of the population are risk averse and, second, the honeybees give more importance to strong odor concentrations than lower ones [160]. The implications of these two empirical observations should lead to interesting consequences that are left for further work, but the intensity of learning could be manipulated by making a a variable margin as $r_i \hat{z}_k(\mathbf{y}_i) \geq \rho(c)$ with c the gas concentration, and the hinge loss function could be replaced by another one with different weights for $r = +1$ and -1.

Following the concept of neural inhibition developed in [134], we can now write a multiclass setting for all the output neurons as

$$\min_{\mathbf{w}_1, \ldots, \mathbf{w}_O} \quad \left(\frac{1}{2} \sum_{i=1}^{O} \langle \mathbf{w}_i, \mathbf{w}_i \rangle + C \sum_{j=1}^{N} \max \left\{ 0, 1 - r_j \langle \mathbf{w}_i, \mathbf{y}_j \rangle - \mu \sum_{k=1}^{O} \langle \mathbf{w}_k, \mathbf{y}_j \rangle \right\} \right), \tag{10}$$

where O is the number of output neurons that compete with each other and the scalar value of the inhibition is optimal for $\mu = 1/O$ [134].

For the sake of simplicity and without loss of generality, we solve for $O = 2$. So we can write the loss function as

$$\min_{\mathbf{w}_1,\mathbf{w}_2} \left(\frac{1}{2}\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \frac{1}{2}\langle \mathbf{w}_2, \mathbf{w}_2 \rangle \right.$$
$$\left. + C\sum_{j=1}^{N} \max\left\{0, 1 - r_j \frac{1}{2}\langle (\mathbf{w}_1 - \mathbf{w}_2), \mathbf{y}_j \rangle\right\} + C\sum_{j=1}^{N} \max\left\{0, 1 - r_j \frac{1}{2}\langle (\mathbf{w}_2 - \mathbf{w}_1), \mathbf{y}_j \rangle\right\} \right). \tag{11}$$

As already mentioned synaptic changes do not occur in a deterministic manner [152, 153, 154]. To solve the problem (9) one may chose a gradient of the primal as in [161, 162, 163]. The gradient in this case is

$$\frac{\partial E}{\partial \mathbf{w}_k}\Big|_{\mathbf{y}_i} = \begin{cases} \mathbf{w}_k - \frac{C}{2} r_i \mathbf{y}_i & \text{if } r_i \hat{z}_k(\mathbf{y}_i) \leq 1, \\ \mathbf{w}_k & \text{otherwise,} \end{cases} \tag{12}$$

with k taking values 1 and 2. So the regularization term induces a continuous process of synaptic removal that it is well known to improve the generalization ability of the pattern recognition device. This is an important message in the sense that too much memory allows learning all of the data samples used for training but then fails on a new set of examples or stimulations. So a healthy level of memory removal boosts the ability to induce an abstract model. The second term of the gradient indicates that when the example is properly classified above some margin there is nothing to be done. On the other hand, if the stimulus is not properly classified then the synaptic changes have to be modified according to $\Delta w_{kj} \propto y_j r_i$. In other words, the activity levels of the KCs and the sign is determined by the reward signal. The main differences respect to the learning rule in Eq. (7) is that when the stimulus is properly classified above a margin no further changes are required in the connections.

CONCLUSION AND DISCUSSION

The honeybee has no more than a million neurons [150, 164]. 35% of those are in the MBs, which is the main location of learning in the insect brain. Another 20% of those are olfactory sensors, which gives a significant weight on the olfactory modality. Then, in between the olfactory receptor cells and the MBs, the AL just constitutes a 2% of the insect brain. This is the area where the information of the antenna is heavily compressed and then relayed into the MB with a significant reduction in the activity levels. Why is the AL is so important despite being son small compared to other brain areas? What is it doing with the signal: extracting dynamical features, normalizing the activity levels, decorrelating in time different stimulus? We do not know yet but our argument is that to provide an answer to these questions we first need to understand how the MBs work during learning and execution. Once we know, then we can determine aspects of the AL processing that improve the performance in pattern recognition and eventually in decision making.

Evolution and engineers. It is remarkable that when one asks engineers what problems need to be solved in pattern recognition of gases, they propose feature extraction methods to interpret the spatio-temporal signal form the sensors and a classifier and regressor to discriminate between gases and to estimate the concentrations [165, 166, 167]. The bioinspiration is not present in these arguments but yet the insect olfactory system appears to be doing just that. Preprocessing the olfactory receptor signals to extract a sparse code that will be passed to a classifier that resembles a support vector machine. In addition computational models even using this seemingly small number of neurons (a million) are extremely demanding in regular computers. Fortunately we also have alternative simulation methods based on graphics processing units (GPUs). GPUs now allow 10 to 100 fold speed-ups in simulations [168, 169] which makes the simulation of insect brains in full size and real time a possibility, removing the biases of scaled-down simplified models.

The MBs as SVMs. It is also remarkable that in contrast to the mainstream mindset that considers Artificial Neural Networks (ANN) as biologically inspired, the reality is that the paradigmatic back-propagation algorithm has yet to be found in the brain. Support Vector Machines, on the other hand, that have become the gold standard of pattern recognition due to its simplicity and nice properties during convex optimization, are actually biologically plausible, fit perfectly in the general scheme of the insect brain, and explains plasticity as a gradient of a loss function proposed by Vapnik [132]. An expert in statistical learning theory that probably thinks that insects are annoying living things rather than a fascinating puzzle of learning.

Role of Models in Neuroscience. Computational neuroscientists put incredible efforts in building computational realistic models to bridge the gap between theory and neural systems[4]. In the process of building these models they manage to reproduce a large variety of experimental observations that later are often rendered with diminished value due to the lack of predictive power, complexity of the systems and the models themselves. Our approach has been: first to understand the function, which is odor discrimination, pattern recognition and regression; second, to identify the neural architecture that solves the problem; and, third, understand the neural code if data is available. Then, taking that knowledge as constraints, we solve a pattern recognition problem and determine what minimal and simple additional key ingredients are needed to complete the task. We predicted for example strong inhibition in the output of the MB and Hebbian learning from the KCs to the output as it was later found. Another prediction derived from this type of model is robustness. As we can see in Fig. 5 the MB model can sustain heavy damage on the KCs without impairing the ability to classify incoming odors. Obviously, if the Calyx is heavily damaged the ability to learn deteriorates, but the recall power of previously stored memories is retained.

About time and Hebbian reversal. The question of how to use time effectively to better solve classification problems is still puzzling. Even though we know that training dynamical systems together with SVMs can improve performance of the classifiers, the plasticity rules are fairly unrealistic from the biological point of view. Moreover, we

[4] An inspection of ModelDB database illustrates this very clearly `http://senselab.med.yale.edu/modeldb/ListByModelName.asp?c=19&lin=-1`

still do not know whether Hebbian plasticity can actually be reversed in the presence of dopamine or octopamine [170, 157], but from the model and pattern recognition perspective the reversal of Hebbian learning needs to to be present to correct those synapses that are providing the wrong output. So a reversal of the spike timing dependent plasticity rule has to be somehow present when reinforcement signal like dopamine or octopamine is activated.

ACKNOWLEDGMENTS

The author acknowledges partial support from NIDCD R01DC011422-01, and to all the researchers that over the course of several decades have made extraordinary progress in understanding how insects solve the computational problem of olfaction.

REFERENCES

1. S. Grossberg, and W. E. Gutowski, *Psychol. Rev.* **94**, 300–318 (1987).
2. M. Hsu, I. Krajbich, C. Zhao, and C. F. Camerer, *J. Neurosci.* **29**, 2231–2237 (2009).
3. M. I. Rabinovich, R. Huerta, P. Varona, and V. S. Afraimovich, *PLoS Comput. Biol.* **4**, e1000072 (2008).
4. J. I. Gold, and M. N. Shadlen, *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
5. R. Menzel, *Nat. Rev. Neurosci.* **13**, 758–768 (2012).
6. C. R. Gallistel, S. Fairhurst, and P. Balsam, *Proc. Natl. Acad. Sci. USA* **101**, 13124–13131 (2004).
7. B. H. Smith, R. Huerta, M. Bazhenov, and I. Sinakevitch, *Distributed Plasticity for Olfactory Learning and Memory in the Honey Bee Brain*, Springer Netherlands, 2012, pp. 393–408.
8. Z. Wu, and A. Guo, *Neural Netw.* **24**, 333 – 344 (2011).
9. C. G. Galizia, S. L. McIlwrath, and R. Menzel, *Cell Tissue Res* **295**, 383–394 (1999).
10. M. Mizunami, J. M. Weibrecht, and N. J. Strausfeld, *J. Comp. Neurol.* **402**, 520–537 (1998).
11. K. Ito, K. Suzuki, P. Estes, M. Ramaswami, D. Yamamoto, and N. J. Strausfeld, *Learn Mem.* **5**, 52–77 (1998).
12. S. M. Farris, and N. J. Strausfeld, *J. Comp. Neurol.* **439**, 331–351 (2001).
13. J. D. Armstrong, K. Kaiser, A. Müller, K.-F. Fischbach, N. Merchant, and N. J. Strausfeld, *Neuron* **15**, 17–20 (1995).
14. N. J. Strausfeld, *J. Comp. Neurol.* **450**, 4–33 (2002).
15. C. G. Galizia, and B. Kimmerle, *J. Comp. Physiol. A* **190**, 21–38 (2004).
16. J. G. Hildebrand, and G. M. Shepherd, *Annu. Rev. Neurosci.* **20**, 595–631 (1997).
17. R. F. Galan, M. Weidert, R. Menzel, A. V. M. Herz, and C. G. Galizia, *Neural Comput.* **18**, 10–25 (2006).
18. G. Laurent, *Science* **286**, 723–728 (1999).
19. G. Laurent, *Nat. Rev. Neurosci.* **3**, 884–895 (2002).
20. G. Laurent, *Trends Neurosci.* **19**, 489–496 (1996).
21. T. Faber, J. Joerges, and R. Menzel, *Nature Neurosci.* **2**, 74–78 (1999).
22. R. Abel, J. Rybak, and R. Menzel, *J. Comp. Neurol.* **437**, 363–383 (2001).
23. C. G. Galizia, J. Joerges, A. Küttner, T. Faber, and R. Menzel, *J. Neurosci. Meth* **76**, 61–69 (1997).
24. C. G. Galizia, and R. Menzel, *Nature Neurosci.* **3**, 853–854 (2000).
25. N. J. Vickers, T. A. Christensen, T. C. Baker, and J. G. Hildebrand, *Nature* **410**, 466–470 (2001).
26. K. C. Daly, R. F. Galán, O. J. Peters, and E. M. Staudacher, *Front. Neuroeng.* **4** (2011).
27. T. Zars, M. Fischer, R. Schulz, and M. Heisenberg, *Science* **288**, 672–675 (2000).
28. T. Zars, R. Wolf, R. Davis, and M. Heisenberg, *Learn Mem.* **7**, 18–31 (2000).
29. T. Zars, *Curr. Opin. Neurobiol.* **10**, 790–795 (2000).
30. Y. Wang, N. J. D. Wright, H. F. Guo, Z. Xie, K. Svoboda, R. Malinow, D. P. Smith, and Y. Zhong, *Neuron* **29**, 267–276 (2001).

31. M. E. Bitterman, R. Menzel, A. Fietz, and S. Schäfer, *J. Comp. Psychol.* **97**, 107–119 (1983).
32. B. H. Smith, G. A. Wright, and K. C. Daly, "Learning-Based Recognition and Discrimination of Floral Odors," in *Biology of Floral Scent*, edited by N. Dudareva, and E. Pichersky, CRC Press, 2005, chap. 12, pp. 263–295.
33. J. S. Hosler, K. L. Buxton, and B. H. Smith, *Behav. Neurosci.* **114**, 514–525 (2000).
34. B. H. Smith, *Behav. Neurosci.* **111**, 1–13 (1997).
35. B. H. Smith, C. I. Abramson, and T. R. Tobin, *J. Comp. Psychol.* **105**, 345–356 (1991).
36. M. Heisenberg, *Nat. Rev. Neurosci.* **4**, 266–275 (2003).
37. M. Heisenberg, *Learn Mem.* **5**, 1–10 (1998).
38. R. L. Davis, *Neuron* **11**, 1–14 (1993).
39. R. F. Galán, S. Sachse, C. G. Galizia, and A. V. M. Herz, *Neural Comput.* **16**, 999–1012 (2004).
40. L. Rath, C. G. Galizia, and P. Szyszka, *Eur. J. Neurosci.* **34**, 352–360 (2011).
41. F. F. Locatelli, P. C. Fernandez, F. Villareal, K. Muezzinoglu, R. Huerta, C. G. Galizia, and B. H. Smith, *Eur. J. Neurosci.* ((in press)).
42. A. Acebes, J.-M. Devaud, M. Arnés, and A. Ferrús, *J. Neurosci.* **32**, 417–422 (2012).
43. M. Trincavelli, A. Vergara, N. Rulkov, J. Murguia, A. Lilienthal, and R. Huerta, *AIP Conference Proceedings* **1362**, 225 (2011).
44. R. Huerta, S. Vembu, M. Muezzinoglu, and A. Vergara, "Dynamical SVM for Time Series Classification," in *Pattern Recognition*, edited by A. Pinz, T. Pock, H. Bischof, and F. Leberl, Springer Berlin / Heidelberg, 2012, vol. 7476 of *Lecture Notes in Computer Science*, pp. 216–225.
45. M. Muezzinoglu, A. Vergara, N. Ghods, N. Rulkov, and R. Huerta, "Localization of remote odor sources by metal-oxide gas sensors in turbulent plumes," in *Sensors, 2010 IEEE*, IEEE, 2010, pp. 817–820.
46. J. A. Farrell, J. Murlis, X. Long, W. Li, and R. T. Cardé, *Env. Fluid Mech.* **2**, 143–169 (2002), 10.1023/A:1016283702837.
47. K. A. Justus, J. Murlis, C. Jones, and R. T. Cardé, *Env. Fluid Mech.* **2**, 115–142 (2002), 10.1023/A:1016227601019.
48. A. Vergara, S. Vembu, T. Ayhan, M. Ryan, M. Homer, and R. Huerta, *Sens. & Actuat. B: Chem.* **166**, 320–329 (2012).
49. A. Vergara, R. Calavia, R. M. Vázquez, A. Mozalev, A. Abdelghani, R. Huerta, E. H. Hines, and E. Llobet, *Analytical Chemistry* **84**, 7502–7510 (2012).
50. A. Vergara, M. Muezzinoglu, N. Rulkov, and R. Huerta, *Sens. & Actuat. B: Chem.* **148**, 298–306 (2010).
51. L. Buck, and R. Axel, *Cell* **65**, 175–187 (1991).
52. Q. Gao, B. Yuan, and A. Chess, *Nature Neurosci.* **3**, 780–785 (2000).
53. L. B. Vosshall, A. M. Wong, and R. Axel, *Cell* **102**, 147–159 (2000).
54. K. Scott, R. Brady, A. Cravchik, P. Morozov, A. Rzhetsky, C. Zuker, and R. Axel, *Cell* **104**, 661–673 (2001).
55. V. Rodrigues, *Brain Res.* **453**, 299–307 (1988).
56. P. G. Distler, B. Bausenwein, and J. Boeckh, *Brain Res.* **805**, 263–266 (1998).
57. J. Joerges, A. Küttner, C. G. Galizia, and R. Menzel, *Nature* **387**, 285–288 (1997).
58. C. G. Galizia, K. Nägler, B. Hölldobler, and R. Menzel, *Eur. J. Neurosci.* **10**, 2964–2974 (1998).
59. C. G. Galizia, S. Sachse, A. Rappert, and R. Menzel, *Nature Neurosci.* **2**, 473–478 (1999).
60. E. Av-Ron, and J. F. Vibert, *Biosystems* **39**, 241–250 (1996).
61. C. Linster, and B. H. Smith, *Behav. Brain Res.* **87**, 1–14 (1997).
62. M. Bazhenov, M. Stopfer, M. Rabinovich, R. Huerta, H. D. Abarbanel, T. J. Sejnowski, and G. Laurent, *Neuron* **30**, 553–567 (2001).
63. M. Bazhenov, M. Stopfer, M. I. Rabinovich, H. D. I. Abarbanel, T. Sejnowski, and G. Laurent, *Neuron* **30**, 569–581 (2001).
64. A. Capurro, F. Baroni, S. B. Olsson, L. S. Kuebler, S. Karout, B. S. Hansson, and T. C. Pearce, *Front. Neuroeng.* **5** (2012).
65. C. Assisi, and M. Bazhenov, *Front. Neuroeng.* **5** (2012).
66. H. Belmabrouk, T. Nowotny, J.-P. Rospars, and D. Martinez, *Proc. Natl. Acad. Sci. USA* **108**, 19790–19795 (2011).
67. S. Olsen, and R. Wilson, *Nature* **452**, 956–960 (2008).

68. C. M. Root, K. Masuyama, D. S. Green, L. E. Enell, D. R. Nässel, C.-H. Lee, and J. W. Wang, *Neuron* **59**, 311–321 (2008).
69. M. Schmuker, N. Yamagata, M. Nawrot, and R. Menzel, *Front. Neuroeng.* **4** (2011).
70. J. Kauer, *J. Physiol.* **243**, 695–715 (1974).
71. M. Meredith, *J. Neurophysiol.* **56**, 572–597 (1986).
72. J. S. Kauer, and J. White, *Annu. Rev. Neurosci.* **24**, 963–979 (2001).
73. R. W. Friedrich, and G. Laurent, *Science* **291**, 889–894 (2001).
74. M. Stopfer, and G. Laurent, *Nature* **402**, 664–668 (1999).
75. M. Stopfer, S. Bhagavan, B. H. Smith, and G. Laurent, *Nature* **390**, 70–74 (1997).
76. M. Wehr, and G. Laurent, *Nature* **384**, 162–166 (1996).
77. T. A. Christensen, B. R. Waldrop, and J. G. Hildebrand, *J. Neurosci.* **18**, 5999–6008 (1998).
78. D. P. Wellis, J. W. Scott, and T. A. Harrison, *J. Neurophysiol.* **61**, 1161–1177 (1989).
79. Y. W. Lam, L. B. Cohen, M. Wachowiak, and M. R. Zochowski, *J. Neurosci.* **20**, 749–762 (2000).
80. R. I. Wilson, G. C. Turner, and G. Laurent, *Science* **303**, 366–370 (2004).
81. S. Sachse, and C. G. Galizia, *J. Neurophysiol.* **87**, 1106–1117 (2002).
82. R. I. Wilson, and G. Laurent, *J. Neurosci.* **25**, 9069–79 (2005).
83. D. W. Wesson, R. M. Carey, J. V. Verhagen, and M. Wachowiak, *PLoS Biol.* **6**, e82 (2008).
84. G. A. Wright, M. Carlton, and B. H. Smith, *Behav. Neurosci.* **123**, 36–43 (2009).
85. M. F. Strube-Bloss, M. P. Nawrot, and R. Menzel, *J. Neurosci.* **31**, 3129–3140 (2011).
86. M. Ng, R. D. Roorda, S. Q. Lima, B. V. Zemelman, P. Morcillo, and G. Miesenbock, *Neuron* **36**, 463–474 (2002).
87. S. Sachse, and C. G. Galizia, *Eur. J. Neurosci.* **18**, 2119–2132 (2003).
88. C. E. Reisenman, T. A. Christensen, and J. G. Hildebrand, *J. Neurosci.* **25**, 8017–8026 (2005).
89. A. F. Silbering, R. Okada, K. Ito, and C. G. Galizia, *J. Neurosci.* **28**, 13075–13087 (2008).
90. A. Acebes, A. Martin-PeÃśa, V. Chevalier, and A. Ferrús, *J. Neurosci.* **31**, 2734–2745 (2011).
91. M. Stopfer, V. Jayaraman, and G. Laurent, *Neuron* **39**, 991–1004 (2003).
92. N. Yamagata, M. Schmuker, P. Szyszka, M. Mizunami, and R. Menzel, *Front. Syst. Neurosci.* **3**, 16 (2009).
93. J. E. Cometto-Muniz, W. S. Cain, and M. H. Abraham, *Chemical Senses* **28**, 467–477 (2003).
94. P. Szyszka, M. Ditzen, A. Galkin, C. G. Galizia, and R. Menzel, *J. Neurophysiol.* **94**, 3303–3313 (2005).
95. J. Perez-Orive, O. Mazor, G. C. Turner, S. Cassenaer, R. I. Wilson, and G. Laurent, *Science* **297**, 359–365 (2002).
96. M. Papadopoulou, S. Cassenaer, T. Nowotny, and G. Laurent, *Science* **332**, 721–725 (2011).
97. D. Marr, *J. Physiol.* **202**, 437–470 (1969).
98. D. Marr, *Proc. R. Soc. Lond. B Biol. Sci.* **176**, 161–234 (1970).
99. D. Marr, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **262**, 23–81 (1971).
100. D. Willshaw, and H. C. Longuet-Higgins, "Associative memory model," in *Machine Intelligence*, edited by B. Meltzer, and O. Michie, Edinburgh University Press, 1970, vol. 5.
101. S. Amari, *Neural Netw.* **2**, 451–457 (1989).
102. J.-P. Nadal, and G. Toulouse, *Network* **1**, 61–74 (1990).
103. G. Palm, and F. Sommer, *Network* **3**, 177–186 (1992).
104. R. Huerta, T. Nowotny, M. Garcia-Sanchez, H. D. I. Abarbanel, and M. I. Rabinovich, *Neural Comput.* **16**, 1601–1640 (2004).
105. R. Huerta, and T. Nowotny, *Neural Comput.* **21**, 2123–2151 (2009).
106. T. Nowotny, R. Huerta, H. D. I. Abarbanel, and M. I. Rabinovich, *Biol. Cybern.* **93**, 436–446 (2005).
107. V. Itskov, and L. F. Abbott, *Phys. Rev. Lett.* **101**, 018101 (2008).
108. S. Cassenaer, and G. Laurent, *Nature* **448**, 709–713 (2007).
109. R. Okada, J. Rybak, G. Manz, and R. Menzel, *J. Neurosci.* **27**, 11736–11747 (2007).
110. M. Heisenberg, A. Borst, S. Wagner, and D. Byers, *J. Neurogenet.* **2**, 1–30 (1985).
111. J. Mauelshagen, *J. Neurophysiol.* **69**, 609–625 (1993).
112. J. S. de Belle, and M. Heisenberg, *Science* **263**, 692–695 (1994).
113. J. B. Connolly, I. J. Roberts, J. D. Armstrong, K. Kaiser, M. Forte, T. Tully, and C. J. O'Kane, *Science* **274**, 2104–2107 (1996).
114. A. Pascual, and T. Préat, *Science* **294**, 1115–1117 (2001).
115. J. Dubnau, L. Grady, T. Kitamoto, and T. Tully, *Nature* **411**, 476–480 (2001).

116. R. Menzel, and G. Manz, *J. Exp. Biol.* **208**, 4317–4332 (2005).
117. S. Grossberg, *Math. Biosci.* **4**, 255 – 310 (1969).
118. R. O'reilly, *Neural Comput.* **13**, 1199–1241 (2001).
119. S. Cassenaer, and G. Laurent, *Nature* **482**, 47–52 (2012).
120. S. Grossberg, and N. Schmajuk, *Psychobiology* (1987).
121. B. Smith, R. Huerta, M. Bazhenov, and I. Sinakevitch, *Honeybee Neurobiology and Behavior* pp. 393–408 (2012).
122. F. B. Rodríguez, and R. Huerta, *Biol. Cybern.* **100**, 289–297 (2009).
123. T. Nowotny, M. I. Rabinovich, R. Huerta, and H. D. I. Abarbanel, *J. Comput. Neurosci.* **15**, 271–281 (2003).
124. W. S. McCulloch, and W. Pitts, *Bull. Math. Biophys.* **5**, 115–133 (1943).
125. J. Wessnitzer, J. Young, J. Armstrong, and B. Webb, *J. Comput. Neurosci.* **32**, 197–212 (2012).
126. S. Grossberg, *Proc. Natl. Acad. Sci. USA* **58**, 1329–1334 (1967).
127. Y. Cao, and S. Grossberg, *Neural Netw.* **26**, 75–98 (2012).
128. H. R. Wilson, and J. D. Cowan, *Kybernetik* **13**, 55–80 (1973).
129. S. Krofczik, R. Menzel, and M. P. Nawrot, *Front. Comput. Neurosci.* **2** (2009).
130. T. Nowotny, and R. Huerta, "On the equivalence of Hebbian learning and the SVM formalism," in *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*, IEEE, 2012, pp. 1–4.
131. K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, *IEEE Trans. Neural Netw.* **12**, 181–201 (2001).
132. V. N. Vapnik, *The nature of statistical learning theory*, Springer-Verlag New York, Inc., New York, NY, USA, 1995.
133. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2007, 1st ed. 2006. corr. 2nd printing edn.
134. R. Huerta, S. Vembu, J. M. Amigó, T. Nowotny, and C. Elkan, *Neural Comput.* **24**, 2473–2507 (2012).
135. Z. Li, and J. Hopfield, *Biol. Cybern.* **61**, 379–392 (1989).
136. R. Gütig, and H. Sompolinsky, *Nature Neurosci.* **9**, 420–428 (2006).
137. M. I. Rabinovich, R. Huerta, and V. Afraimovich, *Phys. Rev. Lett.* **97**, 188103 (2006).
138. M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H. D. Abarbanel, and G. Laurent, *Phys. Rev. Lett.* **87**, 068102 (2001).
139. P. Arena, L. Patano, and P. S. Termini, *Neural Netw.* **32**, 35 – 45 (2012).
140. T. Cover, *IEEE Trans. Electron Comput.* **14**, 326 (1965).
141. C. Cortes, and V. Vapnik, *Mach. Learn.* **20**, 273–297 (1995).
142. S. Luo, R. Axel, and L. Abbott, *Proc. Natl. Acad. Sci. USA* **107**, 10713–10718 (2010).
143. M. Garcia-Sanchez, and R. Huerta, *J. Comput. Neurosci.* **15**, 5–17 (2003).
144. T. Nowotny, and R. Huerta, *Biol. Cybern.* **89**, 237–241 (2003).
145. T. Nowotny, ""Sloppy engineering" and the olfactory system of insects," in *Biologically Inspired Signal Processing for Chemical Sensing*, edited by S. Marco, and A. Gutierrez, Springer, 2009, vol. 188 of *Studies in Computational Intelligence*.
146. C. Assisi, M. Stopfer, G. Laurent, and M. Bazhenov, *Nature Neurosci.* **10**, 1176–1184 (2007).
147. L. A. Finelli, S. Haney, M. Bazhenov, M. Stopfer, and T. J. Sejnowski, *PLoS Comput. Biol.* **4**, e1000062 (2008).
148. M. García-Sánchez, and R. Huerta, *Neurocomputing* **58**, 337–342 (2004).
149. J. Wessnitzer, and B. Webb, *Bioinspiration & Biomimetics* **1**, 63 (2006).
150. M. Giurfa, *Curr. Opin. Neurobiol.* **13**, 726 – 735 (2003).
151. D. Hebb, *The Organization of Behavior*, Wiley, 1949.
152. C. D. Harvey, and K. Svoboda, *Nature* **450**, 1195–1200 (2007).
153. L. F. Abbott, and W. G. Regehr, *Nature* **431**, 796–803 (2004).
154. H. S. Seung, *Neuron* **40**, 1063–1073 (2003).
155. M. Hammer, *Nature* **366**, 4 (1993).
156. P. Szyszka, C. Demmler, M. Oemisch, L. Sommer, S. Biergans, B. Birnbach, A. F. Silbering, and C. G. Galizia, *J. Neurosci.* **31**, 7229–7239 (2011).
157. S. Dehaene, and J. P. Changeux, *Prog. Brain Res.* **126**, 217–229 (2000).
158. E. C. Snell-Rood, D. R. Papaj, and W. Gronenberg, *Brain Behav. Evol.* **73**, 111–128 (2009).
159. Y. LeCun, and C. Cortes, The mnist database, Available online (1998).

160. C. Pelz, B. Gerber, and R. Menzel, *J. Exp. Biol.* **200**, 837–847 (1997).
161. J. Kivinen, A. Smola, and R. Williamson, *Signal Processing, IEEE Transactions on* **52**, 2165–2176 (2004).
162. T. Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Proceedings of the twenty-first international conference on Machine learning*, ACM, 2004, p. 116.
163. Y. Singer, and N. Srebro, "Pegasos: Primal estimated sub-gradient solver for SVM," in *In ICML*, 2007, pp. 807–814.
164. R. Menzel, and M. Giurfa, *Trends Cogn. Sci.* **5**, 62–71 (2001).
165. R. Gutierrez-Osuna, *Sensors J., IEEE* **2**, 189–202 (2002).
166. S. Marco, and A. Gutiérrez-Gálvez, *IEEE Sens. J.* **12(11)**, 3189–3213 (2012).
167. J. Fonollosa, A. Gutierrez-Galvez, A. Lansner, D. Martinez, J. Rospars, R. Beccherelli, A. Perera, T. Pearce, P. Vershure, K. Persaud, and S. Marco, *Procedia Computer Science* **7**, 226 – 227 (2011).
168. T. Nowotny, M. K. Muezzinoglu, and R. Huerta, *Int. J. Innov. Comput.* **7**, 3825–3838 (2011).
169. T. Nowotny, *BMC Neurosci.* **12**, P239 (2011).
170. J. Houk, J. Davis, and D. Beiser, *Models of information processing in the basal ganglia*, MIT press, 1994.