Learning and Foraging in Robot-bees

Andrés Pérez-Uribe **Beat Hirsbrunner**
Parallelism and Artificial Intelligence Group
Computer Science Institute, University of Fribourg
Chemin du Musée 3, CH-1700 Fribourg, Switzerland
Andres.PerezUribe@unifr.ch, http://www-iiuf.unifr.ch/~aperezu

Abstract

Honey-bees have long served as a model organism for investigating insect navigation and collective behavior: they exhibit division of labor and are an example of insect societies where direct communication between workers enable cooperation in the task of collecting nectar and pollen for the colony. However, honey-bees seem to learn about their environment progressively before becoming foragers and displaying the very complex collective behaviors that have inspired researchers interested in collective intelligence. Motivated by recent researches by biologists and neuroscientists on the individual learning in honey-bees, we have implemented a hebbian-learning model and tested it in a foraging task with an autonomous mobile robot (a robot-bee). Then, we used a second learning model that merges unsupervised learning and reinforcement learning techniques. We present some experimental results, as well as the advantages and disadvantages of both models, and describe future directions of research.

1. Introduction

Engineers have recently considered nature as a source of inspiration for developing novel approaches and solutions to their problems. Basically, they have tried to understand and conceptualize the desired overall features or behaviors present in living organisms. Characteristics, such as evolution, fault tolerance, and adaptation have always been very interesting to engineers (Sipper et al., 1997). Social insect societies of ants, termites, wasps, and bees have interested researchers for their emergent complex behavior at the colony-level (Bonabeau et al., 1997). Such behavior is the result of the interaction of many individuals doted with simple behaviors and simple learning capabilities. In addition to being a decentralized system, insect societies exhibit flexibility and robustness, two desirable features for any engineering solution (Kube and Bonabeau, 2000).

Honey-bees have long served as a model organism for investigating insect navigation (McFarland, 1999). Researchers in artificial intelligence have also been interested in honey-bees, particularly for their collective behavior. Insect societies exhibit division of labor and cooperate through direct and indirect (stigmergy) communication schemes (Bonabeau et al., 1997). Honey-bees are an example of insect societies, where direct communication between workers enable cooperation in the task of collecting nectar and pollen for the colony (Gould and Gould, 1995). However, it has been recently proved that honey-bees seem to learn about their environment progressively before becoming foragers (Capaldi et al., 2000) and displaying the very complex collective behaviors that have inspired researchers interested in collective intelligence.

Motivated by recent researches on the individual learning process of "young" honey-bees, we have implemented and tested two learning models using a bee-like robot (i.e., a robot-bee) in a foraging task. Indeed, we found that learning in a second species of bees, the bumble-bees has been widely studied (Montague et al., 1995). Therefore, we first implemented a learning model previously developed by neuroscientists and tested it with our robot-bee in a foraging task, conducted in a workspace that contained two artificial flowers. Then, we used a second learning model that merges unsupervised learning (Dayan, 1999) and reinforcement learning techniques (Sutton and Barto, 1998).

This paper is organized as follows: Section 2 briefly describes some learning and foraging behaviors in honey-bees and recent studies that indicate that such species learn about their environment progressively before becoming foragers. Section 3 describes learning and foraging in bumble-bees. We focus on the fact that neuroscientists have identified and modeled a neuron in bumble-bees that appears to play an important role in classical conditioning experiments. Section 4 presents learning and foraging in robot-bees. We describe the model, the implementation, and test of two learning models: a hebbian learning, and, an unsupervised and reinforcement learning model. Finally, Section 5 presents some conclusions and future work.

2. Learning and foraging in honey-bees

Karl von Frisch (von Frisch, 1993), one of the pioneers in studying the complex behavior of honey-bees (*Apis mellifera*) found that bees are responsive to color when foraging for food. Indeed, it appears that bees learn and remember cues related to food in a very particular way: learning of one aspect (smell, color, shape, landmarks) of the flowers is tightly constrained in time. Color is learned in the 3 seconds that precede ingestion of nectar, odor is learned while the bee is on the flower, and landmarks are learned in the seconds after feeding. Moreover, foraging bees appear to learn those characteristics as a package. If for example, the smell is changed experimentally, then the whole package has to be re-learn (McFarland, 1999).

When a forager has learned and remembered information about food, it transmits this information to other bees in the hive via a dance. The angle between the axis of the dance and the vertical is the same as the angle between the source of food and the sun.

Bees, like many other nesting animals, primarily use learned visual features of the environment to guide their movement between the nest and foraging sites. Srinivasan et al. (Srinivasan et al., 2000) have recently found that a bee's odometer is driven visually and that honey-bees transmit environmental clues that appear along the way to food to other bees using the well-known honey-bee's dance.

Recent studies showed that honey-bees realize repeated "orientation" flights before becoming foragers (Capaldi et al., 2000). Researchers used harmonic radars to study orientation flights and found that they enable honey-bees to view the hive from different viewpoints. Moreover, researchers found that in those orientation flights, honey-bees hold the trip duration, but with increased experience they fly faster, so that they cover increasingly larger areas. Finally, Dukas suggested in 1994 that honey-bees, spend a significant portion of their life span learning and improving on their central task of collecting floral reward (Dukas and Visscher, 1994).

To summarize, honey-bees seem to learn about their environment progressively before becoming foragers and displaying the very complex collective behaviors that have inspired researchers working in the domain of collective intelligence.

3. Learning and foraging in bumble-bees

Bumble-bees (*Bombus Pennsylvanicus*) have many particular features. For example, individual worker bumble-bees are almost exclusively engaged in a single task: collecting nectar and pollen for the colony. Worker bumble-bees are sterile and thus not concerned with acquiring mates or reproductive decisions. They are largely free from predation, and unlike honey-bees, they do not communicate with each other about resources. This last feature renders bumble-bees ideal for studying learning behaviors like animal choice behavior and the evolution of decision-making processes (Real, 1991).

In the early 1990's Real and colleagues (Real, 1991) performed a series of experiments on bee foraging on artificial blue and yellow flowers. In one series of experiments, all the blue flowers contained 2 μl of nectar, 1/3 of the yellow flowers contained 6 μl of nectar, and the other 2/3 of yellow flowers contained no nectar. They observed that about 85% of the bees' visits were to blue flowers, thus avoiding risk.

M. Hammer (Hammer, 1993) identified a neuron (called VUMmx1) that delivers reward information during classical conditioning experiments with bees. More recently, Montague and colleagues (Montague et al., 1995) developed a simulator of a foraging bee to test a neural model and a form of predictive hebbian learning. The neural model they implemented is based on recent work on the VUMmx1 neuron which has widespread projections to odor processing regions of the honeybee brain and whose activity represents the reward of the gustatory stimuli. The predictive hebbian learning model made use of neuromodulatory influences to bias actions and control synaptic plasticity in a way beyond standard correlational mechanisms. Montague found that real bees were more variable than the modeled bees, but obtained results similar to those reported by Real. Moreover, he also switched the constant nectar flowers from blue to yellow and noticed the switching in the flower preference of his modeled bees, just as observed by Real in the experiments with real bees.

4. Learning and foraging in robot-bees

Motivated by the studies of behavior and learning in real and simulated bees, we have realized two robot-bee experiments based on the hebbian learning model of Montague (Montague et al., 1995) and a neurocontroller architecture that implements incremental unsupervised categorization of color visual information and trial-and-error learning of behaviors (Pérez-Uribe and Sanchez, 1997, Pérez-Uribe and Sanchez, 1999, Pérez-Uribe, 1999a).

We have used an autonomous mobile robot with a color CCD camera to implement a robot-bee (*Khepera Coloris*) and placed it in a workspace that models a field of flowers, where the robot-bee wanders in search of food and learns to choose between two species of flowers with different colors, blue and green, similar to Real and Montague's experiments (the experiments with real bees were realized with blue and yellow artificial flowers).

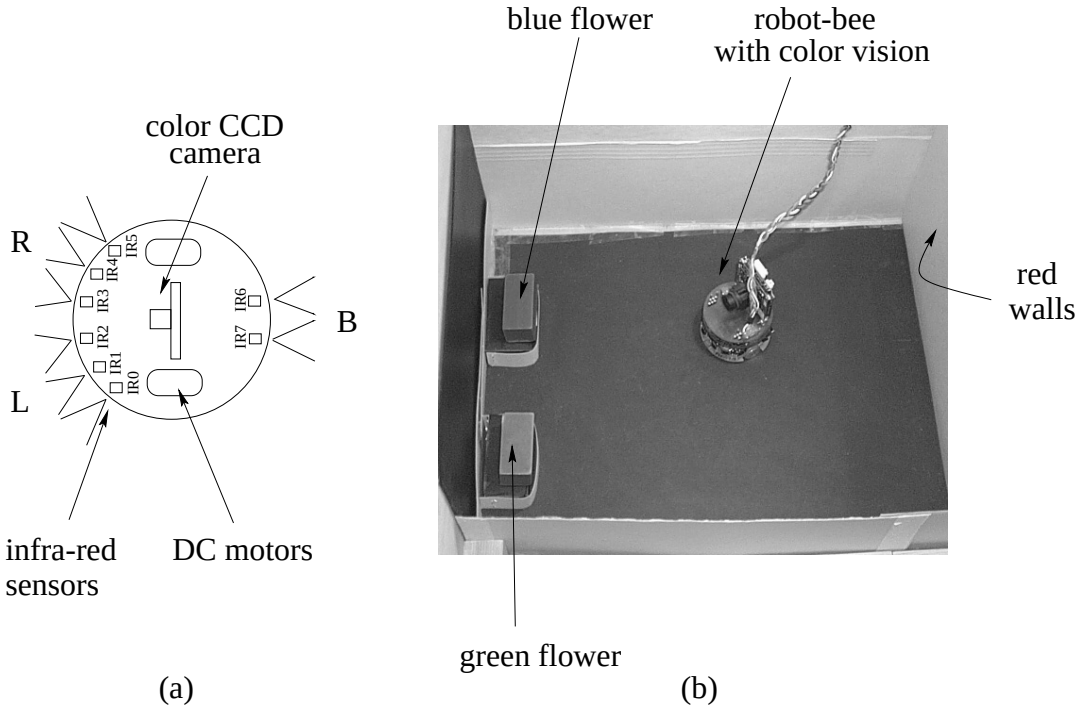


Figure 1: Foraging robot-bee setup. (a) A Khepera mobile robot with a color CCD camera is used to implement a robot-bee. (b) Workspace where the robot-bee wanders in search of food and learns to choose between two species of flowers.

4.1 Experimental setup

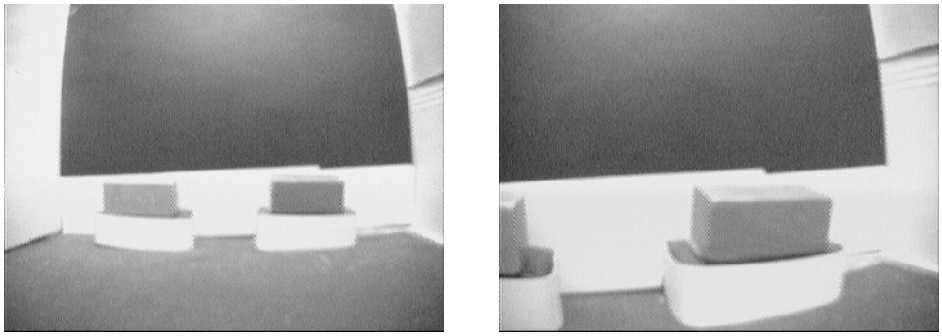
We have used the Khepera mobile robot developed at the Microcomputing Laboratory, Swiss Federal Institute of Technology in Lausanne (EPFL) (Mondada et al., 1993) to implement a robot-bee. It has a circular shape (Figure 1a), with a diameter of 55 mm, a height of 30 mm, and a weight of 70 g. It has two wheels controlled by two DC motors, and eight infrared sensors. The mobile robot contains a Motorola 68331 microcontroller, 256 Kbytes of RAM, 512 Kbytes of ROM, and a RS-232 serial link at 9600, 19200, or 38400 Baud. On-board rechargeable batteries provides around 30 minutes of autonomy. The Khepera robot is modular: several extension turrets have been developed including color CCD cameras, linear vision systems, grippers, artificial retinas, auditory systems, etc.. We have used a K2D video turret developed by K-Team (K-team, 1995a). It is a color CCD camera of 500(H)x582(V) pixels, wired PAL video output, and automatic light adaptation or auto iris.

The microcontroller provides infra-red readings with a 10-bit precision. In our experiments, we have used three 8-bit values as sensor readings: L (left front), R (right front) and B (back), corresponding to a preprocessing of the eight infra-red sensor signals. These 8-bit values encode a real number in the range [0 1). The speed of the two motors can be set in steps of 8 mm/s. The maximum speed is around 1 m/s (K-team, 1995b).

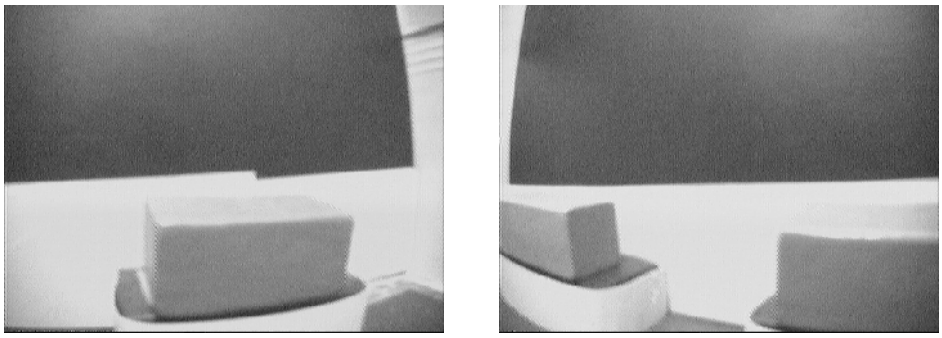
The arena we have used for the robot-bee foraging experiments is a box of 35cm $\times$ 30 cms with light red walls (Figure 1b). Two blocks made of wood, one blue and one green, represent two types of flowers where the robot-bee can collect nectar. Those blocks lie over two bases in order to enable the color CCD camera to sense them when the robot is close. The surface of the bases is black (a neutral color) like the floor and the front wall, but their sides are light red to enable the infra-red sensors to detect them (Figure 2 shows several snapshots of the artificial-flowers and the workspace as seen by the robot-bee from different positions in the arena).

When the color camera senses one of the two flowers (i.e., it senses a minimum percentage of blue or green) and the activation of the infra-red sensors exceeds a certain threshold value, it means that the robot-bee has encountered or "landed" on a flower, and receives its nectar in the form of a reinforcement signal.

Our robot-bee is insensitive to the red color. Indeed, it has been found that honey-bees have a well developed color vision that goes into the ultraviolet, but is insensitive to the red light, and that bees follow landmarks on the flowers called *honey guides* visible only in ultraviolet light (McFarland, 1999). Our robot-bee has three color sensing neurons: given a snapshot of the environment, one neuron gives the change in the percentage of blue, a second neuron gives the change in the percentage of green, and the third neuron gives the change in the percentage of other (neutral) colors. The R,G,B components of each pixel (coded in 8 bits) of a snapshot



(a) (b)



(c) (d)

Figure 2: Various snapshots of the artificial-flowers as seen by the robot-bee from different positions in the arena. (a) Artificial-flowers seen by the robot-bee from the starting position of the foraging experiments. (b) Snapshot of the vision of the robot-bee facing the blue-flower. (c) and (d) Closer snapshots of the artificial-flowers.

of the environment are obtained by a workstation that hosts an external grabber connected to the robot by a cable with unroller. To determine the blue and green colors we perform a thresholding preprocessing. Typically, we have defined the blueness of the blue flower as the color sensed by the robot-bee with RGB components $R < 220$, $G < 200$, and $B > 200$, and the greenness of the green flower as the color with RGB components $R < 220$, $B < 200$, and $G > 200$.

4.2 Hebbian learning experiment

In our first experiment we used the setup described above to implement a hebbian-learning foraging robot-bee. To control the robot-bee, we implemented Montague's hebbian learning model with neuromodulatory influences (Montague et al., 1995). In such model (See Figure 3), a single linear neuron P receives information representing the changes in the percentage of blue (x_b), green (x_g) and neutral (x_n) inputs from the visual field of the robot-bee (i.e., the output of the neurons B,G, and N), weighted by w_b, w_g, and w_n, respectively. The output of the P neuron $\delta(t)$ is defined as follows:

$$\delta(t) = r(t) + V(t) - V(t-1) \ , \qquad (1)$$

where,

$$V(t) = w_b(t)x_b(t) + w_g(t)x_g(t) + w_n(t)x_n(t) \ , \qquad (2)$$

$r(t)$ is a reward signal which takes a value that is a function of the activation of the S-neuron (See Figure 3) that "senses nectar" in the robot-bee. Such function has been obtained experimentally with real bees (Montague et al., 1995) and accounts for risk-aversion in bumble-bees (Real, 1991). It has the form of a positive decelerating function of the nectar volume. In our experiments, the robot-bee receives a reward of $r(6\mu l) = 1.0$ when encountering 1/3 of the green flowers, $r(0\mu l) = 0.0$ when encountering 2/3 of the green flowers, and $r(2\mu l) = 0.7$ when encountering any blue flower.

The $\delta(t)$ output of the P neuron represents an ongoing comparison between $V(t-1)$ and $r(t) + V(t)$. That difference is known as the *temporal difference error* in reinforcement learning techniques. If $\delta(t) > 0$ neuron P labels transition in the sensory input as *better than expected*, or *worse than expected* if $\delta(t) < 0$.

The weight values w_b and w_g are updated according to:

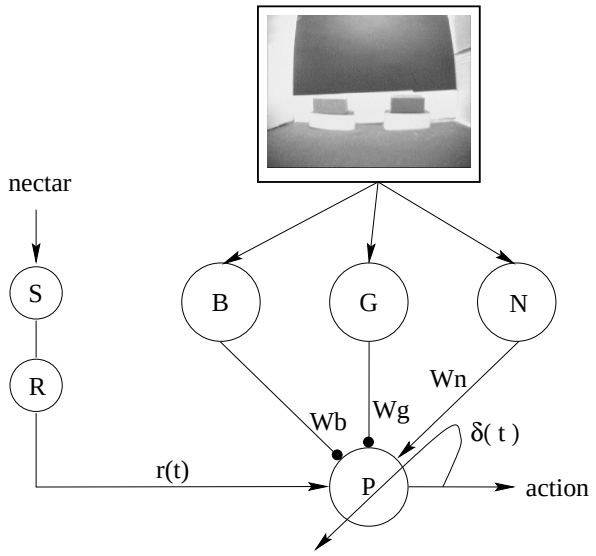


Figure 3: Montague'es hebbian learning model with neuromodulatory influences. Neurons B, G, and N sense blue, green, and neutral colors in the environment. The black dots mean modifiable interconnection weights. Weight w_n was fixed at -0.5.

$$\Delta w(t) = \lambda x(t-1)\delta(t) \ , \qquad (3)$$

where λ is a learning rate. The value of λ is modulated by neurons that become active or signals specifically associated with a flower encounter, such as touch-sensitive neurons and odourant detectors with appropriate thresholds (Montague et al., 1995). In our experiments λ was set to 0.9 when the robot-bee "landed" on an artificial-flower and to 0.0 otherwise. Therefore, weight changes only occurred during encounters with flowers (moreover, it is assumed that $V(t) = 0$ when the robot-bee gathers nectar at time t). This corresponds to a dopamine-neuron based neuromodulation of plasticity (Schultz et al., 1997). This makes part of our study of a new class of learning algorithms which are being developed by neuroscientists based on experimental work with monkeys. We are particularly interested to use such new algorithms and to test them with autonomous mobile robots given their potential for prediction and learning of sequential behavior (Suri and Schultz, 1998).

4.2.1 Robot-bee's resulting biased random walk

Each trial of the hebbian-learning robot-bee foraging task started by placing the robot-bee in the back of the arena facing the artificial flowers. The starting angle of the robot-bee was kept within 30 degrees approximately, but no special care was taken on the initialization angle.

At each time step t, the $\delta(t)$ output of the P neuron was used to choose between two possible actions: *go forward* and *reorient randomly*. The re-orienting action was taken with a probability $P_r(\delta(t)) = 1/(1 + e^{-m\delta(t)+b})$, where m was set to 3.0 and b was set to 0.1 (the slope m represents the amount of noise in the decision function). The re-orienting action consisted in a random change in heading from approximately -30^o to 30^o. The resulting path of the hebbian-learning robot-bee resembles a biased random walk similar to chemotaxing in bacteria (See Figure 4). Such decision making is called *klinokinesis* (Montague et al., 1995).

In these experiments, we have supplied the robot-bees with a pre-wired *basic reflex* to speed-up the trials: whenever the robot-bee senses *no green* and *no blue*, it turns right or left with 50% of probability, until it senses a minimum of green or blue. This *innate* behavior enabled the robot-bee to focus on moving through the artificial flowers, since our main interest was not navigation but learning in foraging robot-bees.

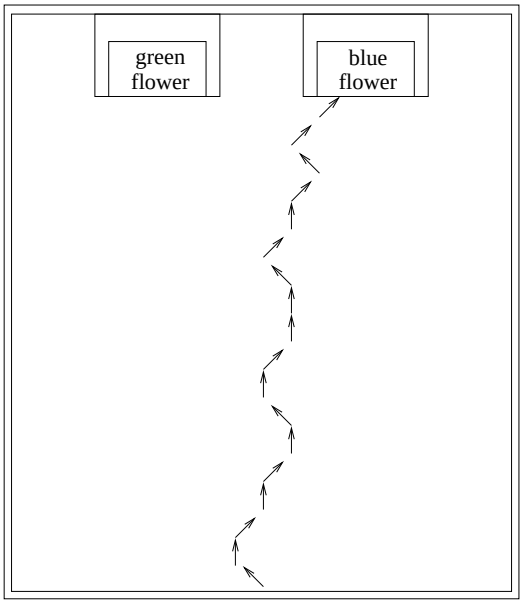


Figure 4: Typical "flight" of a robot-bee in the hebbian learning experiment.

4.2.2 Robot-bee's foraging results

We performed a number of runs of variable number of trials or robot-bee "flights" using Montague's hebbian learning model. The weight's w_b and w_g are set to zero at the beginning of every run as if a new bee were used for the experiment. In Table 1 we show the resulting percentage of blue-flower encounters in seven runs of variable number of trials and the percentage of visits to rewarding green-flowers during those runs. The last row of Table 1 shows that the overall percentage of blue-flower visits was only 45%. Indeed, we found that our robot-bee with hebbian learning did not achieved a clear

blue-flower preference in the reported experiments. This may be due to a difference with the real experiments reported in (Real, 1991) where each bee visits in average 40 flowers during a flight and not a single flower as in our robot-bee experiments. Our experiments resemble more those realized by Montague with a simulated bee in a simulated environment, since he considered a single flower encountering for each simulated-bee flight. However, in his model, Montague used many blue and yellow flowers randomly scattered, though, in our workspace there are only two "big" flowers, one of each species.

In reality, what may be a problem in our experimental setup is the fact that the robot-bee is "too big" compared to the size of the flowers. To see this, lets consider that we set the weight W_b such that $W_b \gg W_g$ as if it had learned it by experience, and then, consider that the robot-bee is approaching the blue flower (as in Figure 2c). The robot-bee does not change its orientation if it senses a continuous increase in the percentage of blue in the arena while approaching the blue flower (i.e. $x_b(t) > x_b(t-1)$). When it reaches the blue flower, the weight $W_b(t)$ is augmented, given that both $x_b(t-1)$ (the preceding change in the percentage of blue in the arena) and $\delta(t)$ (the output of P) are positive values. The weight $W_g(t)$ is decreased if $x_g(t-1)$ sensed a decrease in the percentage of green color.

A problem occurs when the robot-bee approaches the blue flower as in Figure 2d, because, in such case, the robot-bee senses an increase in the percentage of green and a decrease in the percentage of blue, when reaching the blue-flower. The result is that the learning algorithm does not work as expected.

Moreover, our experimental setup suffers from another problem concerning the proportion of variable-rewarding flowers. In the experiments with real bees (Real, 1991) 1/3 of variable-rewarding flowers provided $6\mu l$ of nectar, and the other 2/3 of such flowers provided no nectar. To simulate this ratio, we used a random number generator to determine to provide or not to provide nectar after an encounter with a variable-rewarding green flower (i.e., we delivered nectar only if rnd() $< 2/3$ when reaching a variable-reward flower). However, given that we were not able to realize hundreds of trials, the resulting ratio of rewarding green-flowers was quite different from 1/3 or 33% as shown in Table 1.

We have thus performed a run of 40 trials (20 extra trials were realized after those reported in run 7 of Table 1) to compute the ratio of variable-rewarding flowers, and found that it was approximately 30%, which was what we expected. However, the resulting ratio of blue-flower visits was of only 42.5%.

Given that our experimental setup was not appropriate to the hebbian learning model, we decided to use a less biologically plausible, but similar learning model that merges ideas from unsupervised learning and reinforcement learning techniques.

run	**blue-flower encounters**	**rewarding green-flowers**	**trials per run**
1	70 %	33 %	10
2	30 %	14 %	10
3	50 %	50 %	10
4	46.6 %	37.5 %	15
5	30 %	55 %	15
6	45 %	45 %	20
7	35 %	45 %	20
	44 %	40 %	100

Table 1: Percentage of blue-flower encounters in seven runs of variable number of trials and percentage of visits to rewarding green-flowers. The last row shows the overall percentage of blue-flower visits, the overall percentage of visits to rewarding green-flowers, and the total number of trials performed.

4.3 Unsupervised and reinforcement learning experiment

In a second experiment we have provided our robot-bees with a neurocontroller architecture based on the paradigms of unsupervised learning (Dayan, 1999) and reinforcement learning (Sutton and Barto, 1998). Such neurocontroller enables the robot-bee to autonomously categorize the input data it receives from its environment, to handle with the *stability-plasticity* trade-off (i.e., how can it preserve what it has previously learned, while continuing to incorporate new knowledge), and, finally, to generalize between similar situations and develop a proper policy for action selection, based on an evaluative reinforcement signal. The main difference between this second experiment and the previous one lies in that categorization will enable the robot-bee to have a less stochastic "flight" when foraging. Indeed, categorization of the visual information should enable the robot-bee to anticipate the consequences of its actions and to react differently and more deterministically when facing different situations, that is, when being far or near the flowers, when facing the blue or the green flower, etc. The unsupervised and reinforcement learning robot-bee model is similar to the one in Figure 3, except that the B and G neurons do not provide changes in the proportion of blue and green colors, but simply the percentage of blue and green color in a snapshot of the environment. Moreover, no N-neuron is used, and the P-neuron is implemented as a state-action value function instead of a simple linear neuron.

4.3.1 Learning of categories

Categorization refers to the process by which distinct entities are treated as equivalent. It is considered as one of

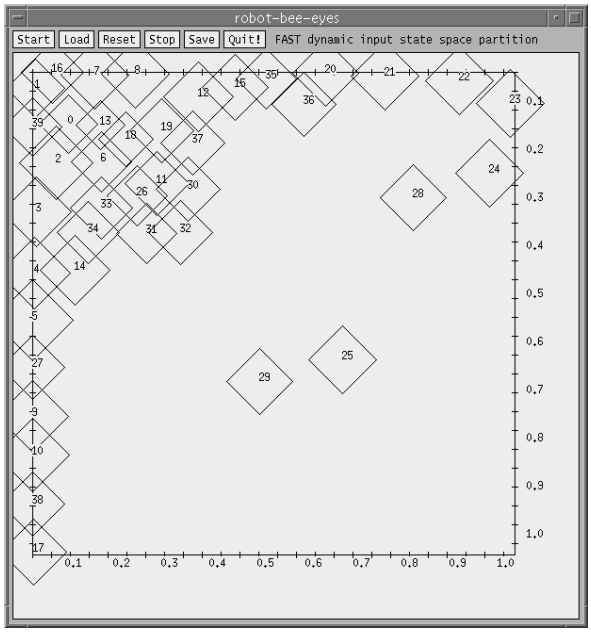


Figure 5: Typical robot-bee's (learned) categorization of its visual information in the foraging task. The x-axis corresponds to the activation of the sensor neurons responsive to the blue color, whereas the y-axis corresponds to the activation of the sensor neurons responsive to the green color.

the most fundamental cognitive activities because categorization allows us to understand and make predictions about objects and events in our world. This is essential in real bees, for instance, to be able to handle the constantly changing activation of their numerous photoreceptors in each eye.

To enable our robot-bee to autonomously categorize the input data it receives from its environment, we used an incremental unsupervised learning model called FAST (Flexible Adaptable-Size Topology) (Pérez-Uribe, 1999b) based on the premises of the Adaptive Resonance Theory (ART) (Carpenter and Grossberg, 1995). Categories or clusters are determined by the network itself, based on correlations of the inputs. At the outset, no neurons or categories are activated in the network. Input patterns are then presented and the network adapts through application of the FAST algorithm, which is driven by three processes: learning, incremental growth, and pruning. The learning mechanism adapts the neuronal reference vectors; as each input vector P is presented to the network, the distance $D(P, W_j)$, between P and every reference vector W_j, is computed. If $D(P, W_j) < T_j$ (the threshold of neuron j), W_j is updated as follows:

$$W_{ji}(t+1) = W_{ji}(t) + \alpha * T_j * (P_i - W_{ji}(t)) \ , \qquad (4)$$

where α is a learning parameter in the range $(0, 1]$. In our implementation, the Manhattan distance $D(P, W_j)$ is used as a measure of similarity between the reference vectors and the current n-dimensional input:

$$D(P, W_j) = \sum_{i=1}^{n} \mid P_i - W_{ji} \mid \qquad (5)$$

This distance gives rise to the diamond-shaped sensitivity region of neuron j shown in Figure 5. The activation of neuron j also entails an exponential decrease in its threshold value:

$$T_j(t+1) = T_j(t) - \gamma(T_j(t) - T_{min}) \ , \qquad (6)$$

where γ is a gain parameter in the range $(0, 1]$, and T_{min} is the minimal threshold. This threshold modification decreases the size of the sensitivity region for neurons in high-density regions of the vector space.

When an input vector P lies outside the sensitivity regions of all currently operational neurons, a new neuron is added (incremental growth mechanism), with its reference vector set to P and its threshold set to an initial value T_{ini}. The value of T_{ini} should also be decreased in order to avoid having sensitivity regions contained within other sensitivity regions after successive deactivation-activation phases, a situation which causes instability. The pruning mechanism was deactivated in these experiments (See (Pérez-Uribe, 1999b) for more details on the algorithm).

In our experiments with the robot-bee, the FAST incremental unsupervised learning system dynamically categorizes the information provided by the B and G neurons, which output the percentage of blue and green (as previously defined) given a particular snapshot of the environment. A maximum number of 40 categories was selected for the foraging robot-bee task. The values of the FAST learning parameters were initialized as follows:

T_{ini}	T_{min}	γ	L_r
0.1	0.025	0.01	0.2

In Figure 5, we show a typical example of the resulting adaptive categorization of FAST in our robot-bee learning and foraging task. The rhombi of Figure 5 represent the sensitivity regions of the 40 FAST neurons on the percentage-of-blue vs percentage-of-green plane.

Neuron 21, for example, categorizes the sensor readings that activate the x_b-input by 70% to 85%, and the x_g-input by less than 15%. In other words, neuron 21 is activated when the robot-bee lies very close to the blue flower as in in Figure 2b. Similarly, neuron 25 will be activated when the robot-bee lies near the two flowers, and neurons 0, 1, 16, or 39 when it is near its starting point facing the flowers (as in Figure 2a). The FAST neurons thus allow the robot-bee, for example, to start moving towards a preferred flower by changing its orientation in a particular direction. The distribution of the FAST neurons along the axis in the category-space

(Figure 5) is due to the fact that from the very first trials, the robot-bee prefers moving towards one of the flowers. Indeed, if we observe a typical robot-bee behavior (Figure 6), when the robot-bee starts a new trial, it rapidly chooses to move to one side (preferably towards the right) and then it proceeds moving facing one of the flowers. Thus, it rarely faces a point between the flowers (except, when it starts a new trial), which will activate neurons lying in the diagonal of the category-space (e.g., between neurons 30 to 32, and, 25 and 29).

4.3.2 Learning of behaviors

Reinforcement learning techniques were used to allow the system to selectively retain the actions that maximize the received reward over time. The reinforcement learning module associates a particular action (by means of a state-value function) to each environmental situation, determined by the robot-bee's neural vision system and the adaptive categorization module.

The robot-bee foraging task is treated in discrete time steps. At each time step t, the learning system receives some representation of the environment's *state*, it *tries* an action, and one step later it is *reinforced* by receiving a scalar evaluation (i.e., a reward or a punishment) and finds itself in a new state.

In our experiments, the robot-bee takes a snapshot of the environment, then neurons B and G outputs the percentage of blue and green in the environment, and the FAST module selects a category for such environmental situation. The reinforcement learning module associates a $Q(s,a)$ value to each FAST category s, which represents the expected cumulative reinforcement the robot-bee can receive when choosing action a from the environmental state s.

In particular, we used SARSA learning with *eligibility traces* or SARSA(λ) (Sutton and Barto, 1998), a kind of memory that serve as a temporary record of the occurrence of an event, such as visiting a state or taking an action (Sutton and Barto, 1998)). In this method, all the Q action values are modified after every interaction with the environment as follows:

$$Q(s,a) = Q(s,a) + \alpha[r + \gamma Q(s',a') - Q(s,a)]e(s,a) \ , \quad (7)$$

where s is the state, a is a possible action, s' and a' are the corresponding possible next state and action, r is the reward, α is a step-size parameter, and $e(s,a)$ is the *eligibility trace* of action a taken from state s. The eligibility trace is set to 1.0 when action a is taken from state s and updated as $e(s,a) = \gamma^{\lambda} e(s,a)$ after every interaction with the environment. γ is the discount factor and λ is a constant value in the range $(0,1]$. The values of the SARSA(λ) learning parameters were initialized as follows:

run	percentage of blue-flower encounters
1	85%
2a	45%
2b	53.3%
3	95%
4	85%

Table 2: Percentage of blue-flower encounters in four runs of 20 trials or robot-bee "flights". The percentage value for run 2b corresponds to 30 trials.

α	λ	γ
0.6	0.9	0.9

In our experiments, we initialized the Q-table to zeros everywhere, and we used an ϵ-greedy action selection mechanism, with ϵ equal to 0.05. In other words, our system exploits with a probability of 95%, and explores only 5% of the time.

In the robot-bee's experiments with the neurocontroller, we determined four possible actions: *go forward*, *turn right and go forward*, *turn left and go forward*, and *go backward*, with a fixed speed of 40mm/s. The first three actions are associated to particular situations during learning, while the last action is only used as part of a pre-wired *basic reflex* for wall avoidance, codified as the following reactive behavior: follow the direction of the least-activated sensors during 4 times 5 action-selection steps. This basic reflex is activated when one of the infra-red sensor exceeds a certain threshold value.

The robot-bee is punished (it receives a reinforcement of '-1') when it crashes or turns on itself. The robot holds two variables *dl* and *dr* that indicate if the system moves left or right "too much" in order to detect when the robot turns on itself (See (Pérez-Uribe, 1999a) for more details). When the robot-bee encounters a blue flower, it is rewarded as in the last experiment, but if the robot-bee encounters a green flower, it receives no reward (i.e., $r = 0.0$). Future experiments will consider the use of this learning model for risk-aversion studies as we did before.

In Figure 6, we show the typical action-selections taken by our robot-bee in different positions in the workspace after 20 trials or "flights". Such actions clearly show a blue-flower preference in our foraging robot-bee.

4.3.3 Robot-bee's foraging results

We performed four runs of 20 trials or robot-bee "flights". Each trial started by placing the robot-bee in the back of the arena facing the artificial flowers, as in the hebbian-learning experiments. The starting angle of the robot-bee was kept within 30 degrees approximately, but no special care was taken on the initialization angle.

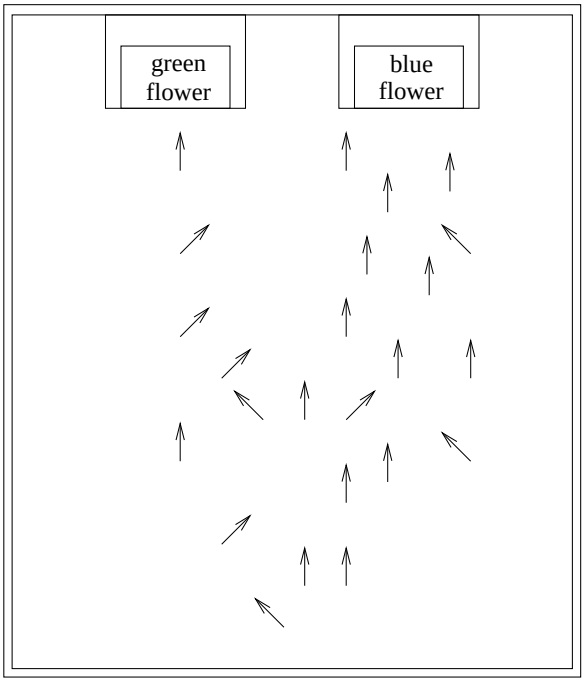


Figure 6: Typical robot-bee's (learned) foraging behavior in the unsupervised and reinforcement learning experiment. The arrows indicate the robot-bee's preferred action in different positions of the arena, facing the front.

The robot-bee successfully learned an appropriate state-space partition (adaptive categorization) and a perception-to-action mapping (action selection), such that the robot-bee was able encounter one of the two flowers while avoiding the walls. Indeed, the robot-bee preferred the blue flower, that is, the flower that provided some nectar (reward) instead of the green flower that did not provided any nectar (reward). In Table 2 we show the percentages of blue-flower encounters of the robot-bee during the four runs of 20 trials. In the second run, the robot-bee visited the zero-reward green flower more than average during the first trials; this shows that depending on the first flower encounters it can take longer to learn a blue flower preference, however, this behavior will be nevertheless achieved after some more trials as shown in run 2b (See Table 2) where we performed 10 extra trials that raised the overall percentage of blue-flower encounters from 45% to 53.3%. Indeed, 70% of such extra trials ended with an encounter with the (rewarding) blue flower.

5. Conclusions

We have presented two learning experiments in foraging robot-bees. In our experiments we used an autonomous mobile robot with a color CCD camera, which is placed in a workspace with two blocks of wood, one blue, and one green, simulating two different species of flowers that provide different rewarding nectar. In a first set of experiments we used a hebbian-learning model with neuromodulatory signals previously developed by neuroscientists. This model is biologically plausible and introduces the promising idea of using dopamine-like neural signals that modulate learning (Schultz et al., 1997). We are particularly interested to use such new algorithms and test them with autonomous mobile robots given their potential for prediction and learning of sequential behavior (Suri and Schultz, 1998).

The resulting foraging behavior is a biased random walk similar to that observed in bacteria. We performed a number of runs of variable number of trials or robot-bee "flights" using such model and obtained different results than those reported by neroscientists working with simulated bees, and biologists working with real bees. This may be due to some differences in the experimental setup: while in the real and simulated experiments there are many flowers scattered in the arena, in our workspace only two flowers can be chosen by the robot-bee in the foraging task, and the robot-bee appears to be "too big" compared to the size of the flowers, which causes the learning algorithm not to work as expected. Moreover, in our experiments, a robot-bee can only visit a single flower per trial, while in the real experiments, a honey-bee can visit up to 40 different flowers in a single flight.

Even, if we were very interested in using the hebbian learning model because of its very simple form, we moved to a second set of experiments, where we used a learning model based on unsupervised and reinforcement learning techniques. The main difference between this second experiment and the previous one lies in that unsupervised categorization enables the robot-bee to have a less stochastic "flight" when foraging. Indeed, categorization of the visual information enables the robot-bee to anticipate the consequences of its actions and to react differently and more deterministically when facing different situations. The learned foraging behavior is a clear blue-flower preference (i.e., the rewarding flower).

We are planning new experiments considering the use of the second learning model for risk-aversion studies as we did before, and to merge both learning models to provide powerful learning capabilities to our mobile robots.

Acknowledgements

We thank Fabrice Chantemargue for his help with the Khepera color vision system, the experimental setup, the image processing, and the graphical user interface. This work is supported by the Fonds national suisse de la recherche.

References

Bonabeau, E., Théraulaz, G., Deneubourg, J.-L., Aron, S., and Camazine, S. (1997). Self-Organization in Social Insects. *Trends in Ecol. Evol.*, 12:188–193.

Capaldi, E., Smith, A., Osborne, J., Fahrbach, S., Farris, S., Reynolds, D., Edwards, A., Martin, A., Robinson, G., Poppy, G., and Riley, J. (2000). Ontogeny of orientation flight in the honeybee revealed by harmonic radar. *Nature*, 403:537–540.

Carpenter, G. and Grossberg, S. (1995). Adaptive resonance theory (ART). In Arbib, M., (Ed.), *Handbook of Brain Theory and Neural Networks*, pages 79–82. MIT Press.

Dayan, P. (1999). Unsupervised learning. In *The MIT Encyclopedia of the Cognitive Sciences*. The MIT Press.

Dukas, R. and Visscher, P. (1994). Lifetime learning by foraging honey bees. *Animal Behaviour*, 48:1007–1012.

Gould, J. and Gould, C. (1995). *The Honey Bee*. Scientific American Library, New York.

Hammer, M. (1993). An identified neuron mediates the unconditioned stimulus in associative learning in honeybees. *Nature*, 366:59–63.

K-team (1995a). *Khepera K2D Video Turret User Manual*. K-Team SA, Lausanne, Switzerland. Version 1.1.

K-team (1995b). *Khepera User Manual*. K-Team SA, Lausanne, Switzerland. Version 4.06.

Kube, C. and Bonabeau, R. (2000). Cooperative transport by ants and robots. *Robotics and Autonomous Systems*, 30(1-2):85–101.

McFarland, D. (1999). *Animal Behaviour*, chapter 23. The complex behaviour of honey-bees, pages 421–433. Addison Wesley Longman, Harlow, England, third edition.

Mondada, F., Franzi, E., and Ienne, P. (1993). Mobile robot miniaturization: A tool for investigating in control algorithms. In *Proceedings of the Third International Symposium on Experimental Robotics*, Kyoto, Japan.

Montague, P., Dayan, P., Person, C., and Sejnowski, T. (1995). Bee Foraging in uncertain environments using predictive hebbian learning. *Nature*, 377:725–728.

Pérez-Uribe, A. (1999a). *Structure-Adaptable Digital Neural Networks*, chapter 6. A Neurocontroller Architecture for Autonomous Robots, pages 95–116. Swiss Federal Institute of Technology-Lausanne, Ph.D Thesis 2052.

Pérez-Uribe, A. (1999b). *Structure-Adaptable Digital Neural Networks*, chapter 4. Adaptive Categorization: from Biology to Hardware, pages 33–62. Swiss Federal Institute of Technology-Lausanne, Ph.D Thesis 2052.

Pérez-Uribe, A. and Sanchez, E. (1997). Structure-Adaptable Neurocontrollers: A Hardware-Friendly Approach. In José Mira, R. M.-D. and Cabestany, J., (Eds.), *Biological and Artificial Computation: From Neuroscience to technology*, pages 1251–1259, Lecture Notes in Computer Science 1240, Springer Verlag.

Pérez-Uribe, A. and Sanchez, E. (1999). A Digital Artificial Brain Architecture for Mobile Autonomous Robots. In Sugisaka, M. and Tanaka, H., (Eds.), *Proceedings of the Fourth International Symposium on Artificial Life and Robotics AROB'99*, pages 240–243, Oita, Japan.

Real, L. (1991). Animal Choice Behavior and the Evolution of Cognitive Architecture. *Science*, 253:980–986.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275:1593–1599.

Sipper, M., Sanchez, E., Mange, D., Tomassini, M., Pérez-Uribe, A., and Stauffer, A. (1997). A Phylogenetic, Ontogenetic, and Epigenetic View of Bio-Inspired Hardware Systems. *IEEE Transactions on Evolutionary Computation*, 1(1):83–97.

Srinivasan, M., Zhang, S., and Altwein, M. (2000). Honeybee navigation: Nature and Calibration of the "Odometer.". *Science*, 287(5455):851–853.

Suri, R. and Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Experimental Brain Research*, 121:350–354.

Sutton, R. and Barto, A. (1998). *Reinforcement Learning: An Introduction*. The MIT Press.

von Frisch, K. (1993). *The Dance Language and Orientation of Bees*. Harvard University Press, Cambridge, MA, reprint edition. (translated by L.E. Chadwick from "Tanzsprache und Orientierung der Bienen", Springer Verlag, 1965).