Bee foraging in uncertain environments using predictive hebbian learning

P. Read Montague*, Peter Dayan†, Christophe Person* & Terrence J. Sejnowski‡

* Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA
† CBCL, Department of Brain and Cognitive Sciences, E25-201, MIT, Cambridge, Massachusetts 02139, USA
‡ Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037 and Department of Biology, University of California, San Diego, La Jolla, California 92037, USA

RECENT work has identified a neuron with widespread projections to odour processing regions of the honeybee brain whose activity represents the reward value of gustatory stimuli1-3. We have constructed a model of bee foraging in uncertain environments based on this type of neuron and a predictive form of hebbian synaptic plasticity. The model uses visual input from a simulated three-dimensional world and accounts for a wide range of experiments on bee learning during foraging, including risk aversion. The predictive model shows how neuromodulatory influences can be used to bias actions and control synaptic plasticity in a way that goes beyond standard correlational mechanisms. Although several behavioural models of conditioning in bees have been proposed4-7, this model is based on the neural substrate and was tested in a simulation of bee flight.

Real and colleagues8-11 performed a series of experiments on bumblebees foraging on artificial blue and yellow flowers whose colours were the only predictor of the nectar delivery. They examined how bees respond to the mean and variability of this delivery in a foraging version of a stochastic 'two-armed bandit' problem12,13. In one series of experiments, all the blue flowers contained 2 μl of nectar, ⅓ of the yellow flowers contained 6 μl, and the remaining ⅔ of the yellow flowers contained no nectar at all. In practice, 85% of the bees’ visits were to the constant yield blue flowers despite the equivalent mean return from the more variable yellow flowers. In a second series of experiments, Real showed that bumblebees will forage equally from each flower type if the mean reward from the variable type is made sufficiently large.

In the honeybee suboesophageal ganglion, an identified neuron, VUMmx1, delivers information about reward during classical conditioning experiments1. This neuron projects widely to brain regions involved in odour processing, becomes active in response to sucrose applied to the antennae and proboscis, and its firing can substitute for the unconditioned stimulus in a classical conditioning experiment. Specifically, presentation of an
odourant followed by artificial stimulation of VUMmx1 through an electrode caused subsequent presentation of the odourant alone to elicit a stereotyped behavioural response which indicated that the bee expects to receive nectar.

We have compared the behaviour of real bees to a neural model of bee learning based on VUMmx1 in a simulated three-dimensional arena containing a field of flowers possessing the same reward distributions as described above. In the model architecture, shown in Fig. 1. P was a simple linear unit receiving convergent sensory information representing the changes in the percentage of blue (xy), yellow (xy) and neutral (xy) inputs from the visual field, weighted by wB, wY and wN, respectively. In the presence of nectar (activity along r(t)), the output of the linear unit P was

\[ \delta(t) = r(t) + V(t) = r(t) + V(t) - V(t-1) \]  

(1)

where

\[ V(t) = w(t) \cdot x(t) = w_B x_B(t) + w_Y x_Y(t) + w_N x_N(t) \]  

(2)

The output of P, \( \delta(t) \), thus represents an ongoing comparison of \( V(t-1) \) and the sum \( r(t) + V(t) \), and, in the absence of reward, \( r(t) = 0 \), P's output begins transitions in sensory input as 'better than expected' (\( \delta(t) > 0 \)) or 'worse than expected' (\( \delta(t) < 0 \)). We interpret the sign of \( \delta(t) \) as changes in neuromodulator release about some basal level\(^{14,15} \).

After landing on a flower and receiving some volume of reward, the output of P, \( \delta(t) \), controlled learning according to

\[ \Delta w(t) = \lambda x(t-1) \delta(t) \]  

(3)

where \( \lambda \) is a learning rate. This learning rate is similar to one first proposed by Konorski\(^{16} \) and is a form of Hebb rule\(^{7} \) in which the postsynaptic factor \( \delta(t) \) carries a sign permitting the system to learn predictions rather than correlations\(^{18,22} \). VUMmx1 is more highly activated by sensory cues that predict reward than by cues that do not predict reward and could have the same function as \( \delta(t) \) in our model\(^{1} \). and 'N' represented changes in the percentage of blue, yellow and neutral colours in the visual field, hence their outputs reflect normalized responses. Such derivatives, typical of early sensory pathways, could be constructed anywhere along the path from the sensory units to P. Any portion of the field of view that did not contain blue or yellow was taken as neutral (white region in b). These neurons influenced the output of a diffusely projecting linear unit P through weights \( w_B, w_Y \) and \( w_N \), which were adaptively weighted at 0.5. As the model bee changed its heading and/or its height above the field, changes in the activity in these neurons would ensue. Changing the weights upon encounters with flowers permitted them to represent information about the predictive relationship between the sensory input and the amount of nectar\(^{19,23,30} \). Response of neuron R that delivers information about the reward to P and receives input from a sucrose (nectar) sensitive neuron S. The functional form of this relationship is derived from an empirically determined utility function for bumblebees\(^{8,10} \).

METHODS. The model bee had a single eye with a fixed field of view that was varied from 20° to 30°. This arrangement was not meant to represent the eye of the bee, but instead represents the use of visual information only from the central portion of the visual field along the direction of motion of the bee. At each iteration, the output of P influenced the decision of the bee to reorient randomly (Fig. 2). If no reorientation occurred, the bee took one step along its direction of flight, otherwise, it chose a random change in heading from -90° to +90° and then took a step (stepsize = 0.05). A landing was registered once the model bee's attitude was less than 0.05. At that point, the flower intersected by the direction of flight was selected. After landing on a flower, a reward is given according to the volume of nectar in the selected flower and activity in r(t) resulted (d). The weights (wB, wN) were adjusted only on encounters with flowers and otherwise influenced the decision of the bee to reorient through their influence on the fluctuating output of neuron P. After the model bee landed on a flower, received a nectar reward (or not), and had the sensory weights updated, it was randomly repositioned at the top of the simulated arena with a random initial heading. This explains the pattern of results shown in Fig. 4 at high altitudes.

Once the model bee lands on a flower (Fig. 1), we use an empirically derived 'utility' curve\(^{8,10} \) to determine the value of different volumes of nectar (Fig. 1c,d). We interpret this curve as an equilibrium measure of the value of different volumes of nectar that does not specify detailed dynamics of how an individual bee might estimate nectar volumes or their subjective utility (see ref. 6). We made the simplifying assumptions that the utility of the nectar in a given flower could be assessed in one iteration and that the sensory component is V(t) = 0 as the model gathers nectar at time t. At this time r(t) takes on the value prescribed by the utility curve and V(t-1) is driven exclusively by the current flower colour. In this way, \( \delta(t) \) takes on the form \( r(t) + V(t) - V(t-1) = r(t) - V(t-1) \) making equation (3) equivalent to the Rescorla—Wagner rule for classical conditioning\(^{3,23} \).

In the absence of reward (r(t) = 0), we used the output of P to bias actions. At each time t, \( \delta(t) \) determined whether the bee continued on its present heading for its next movement or whether it randomly reoriented (tumbled) before its next movement (Fig. 2a). This model is reminiscent of a biased random walk with \( \delta(t) \) choosing the probability of randomly changing direction. In chemotaxing bacteria, such decision-making is called klinokinesia\(^{32} \). In various invertebrate systems, neuromodulator delivery influences motor behaviours and thus choice\(^{27} \). However, we are making the novel proposal that VUMmx1 or its visually sensitive congeners also have such effects. The use of a signal such as \( \delta(t) \) to learn to choose actions appropriately leads to an asynchronous version of dynamic programming\(^{22} \), an engineering technique that can be used to find optimal sequences of actions to achieve a goal.

In the results reported here, we assumed that weight changes only occurred during encounters with flowers. Thus, the output from P was used continuously to guide actions but plasticity was gated. This could occur by modulation of the learning rate (\( \lambda \) in equation (3)) by neurons that become active in response.
FIG. 2. Simulations of bee foraging behaviour using predictive hebbian learning, a, Influence of $\delta(t)$ (labelled Delta) on the decision to reorient. The adaptable ($w_1$, $w_2$) and non-adaptable ($w_3$) sensory weights influence the decision to reorient through their influence on the size and sign of $\delta(t)$. $P[\delta(t)]$ is the probability of reorienting for a given value of $\delta(t)$ and has the form $1/(1 + \exp(m - b))$. The slope $m$ of the linear region of this curve determines the amount 'noise' in the decision function: dashed line $m = 3$, line with points $m = 30$. In the results presented in this paper, $m$ was varied from 5 to 45 and $b$ varied from 0.1 to 5.0. b, Fraction of visits to blue flowers for real and model bees: weights updated only on flower encounters. Each trial represents about 40 flower visits averaged over 5 real bees and exactly 40 flower visits for a single model bee. Trials 1-15 for the real and model bees had blue flowers as the constant type (2 μl in all flowers), the remaining trials had yellow flowers as constant. Data from real bees are shown as dashed line, data for simulated bees shown as connected points (learning rate $\lambda = 0.9$). The control trace (fluctuating trace around line at 0.5) shows the sampling behaviour for the model bee where $w_1 = w_2 = 0.5$ and no changes in the weights were made upon encounters with flowers. The control trace shows that the model for output is not biased toward either flower by other constraints associated with the chosen representation of the model bee and the arena. The real bees were more variable than the model bees and tended to sample the constant flowers at a slightly higher rate (85% compared to a range of 73%-85%). The real bees also had a slight preference for blue flowers which can be seen after the switch of reward distributions at trial 15 (ref. 7). For the model bee, the rewards were stochastically delivered so there was no effect of revisits on the effective variance of either the constant or variable flowers. In practice, this assumption was not controlled; however, moderate increases in variance for the constant flower would not influence dramatically the behaviour of the model.

to signals specifically associated with a flower encounter, such as touch-sensitive neurons and odourant detectors with appropriate thresholds. Similar results were obtained when the learning rate was varied continuously according to the height of the bee above the field of flowers (unpublished data).

FIG. 3. Tradeoff between the mean and variance of nectar delivery for bee foraging. a, b, Compare indifference points for the model bees against those for real bees for different learning rates. The coordinates of indifference points are taken as the mean and variance for which the bee samples equally from both the constant and variable flowers. The constant flower contained 0.5 μl of nectar. For the variable flower, the variance was fixed and the mean slowly increased until the percentage of visits to the constant flower was less than 50% over 80 flower visits; the mean and variance were then used as an indifference point. The stochastic nature of the model bee movement and the delivery of reward results in the observed spread in the indifference points. a, Indifference points for $\lambda = 0.05$. b, Indifference points for $\lambda = 0.9$. In both a and b, the data for real bees are shown as points connected by a solid line (units for mean and variance are μl and μl/s, respectively).
This tradeoff between the expected value of reward and its uncertainty is not, however, universal. In choice experiments in honeybees, Greggers and Menzel found that honeybees can be induced to maximize their return by choosing a single flower. This paper was one of the first to show how learning might affect foraging based on the Rescorla-Wagner model of conditioning. We extend this view, showing how the learned weights can be used to choose appropriate actions and how the resulting action choices influence the learning. In addition, the actions taken by the model are also influenced by the structure of the simulated environment in which it moved. Figure 4 shows how the output of the diffuse neuron $P$ was influenced by the structure of the foraging problem faced by the model bee. Hence, although learning (weight changes) was one major influence in the behaviour exhibited by the model, the structure of the environment played an important role in shaping its behavioural decisions.

The success of the model in accounting for the choice behaviour of bumblebees allows us to connect the performance of a simple neural system using known anatomical and physiological constraints and descriptions of a behaviour previously explained mainly in terms of a decision-theoretical framework for minimizing risk (ref. 8 but see ref. 6). Visual, gustatory and VUMmx1 inputs converge on the antennal lobes and the mushroom bodies, thus making these regions good candidate sites for the learning rule in equation (3). Delivery of octopamine, which is released by VUMmx1, directly to these regions can substitute for the firing of VUMmx1 in some conditioning experiments.

There is good evidence for similar predictive responses in primate mesencephalic dopaminergic systems. Hence, dopamine delivery in primates may be used by target neurons to guide action selection and learning, suggesting the conservation of an important functional principle, albeit differing in its detailed implementation.

Received 27 February; accepted 18 August 1995.


ACKNOWLEDGEMENTS. We thank P. S. Churchland, A. Dayan, A. Pouget, D. Rosien, S. Quarti and M. L. Bennett for their helpful comments and criticisms on earlier versions of this work. We also thank P. Yates and D. Egelman for help with computer simulations. This work was supported by the National Institute of Mental Health (P.R.M., T.J.S.), Center for Theoretical Neuroscience at Baylor College of Medicine (P.R.M.), Howard Hughes Medical Institute (T.J.S.), Natural Science and Engineering Research Council (Canada) (P.D.), Science and Engineering Research Council (UK) (P.D.).