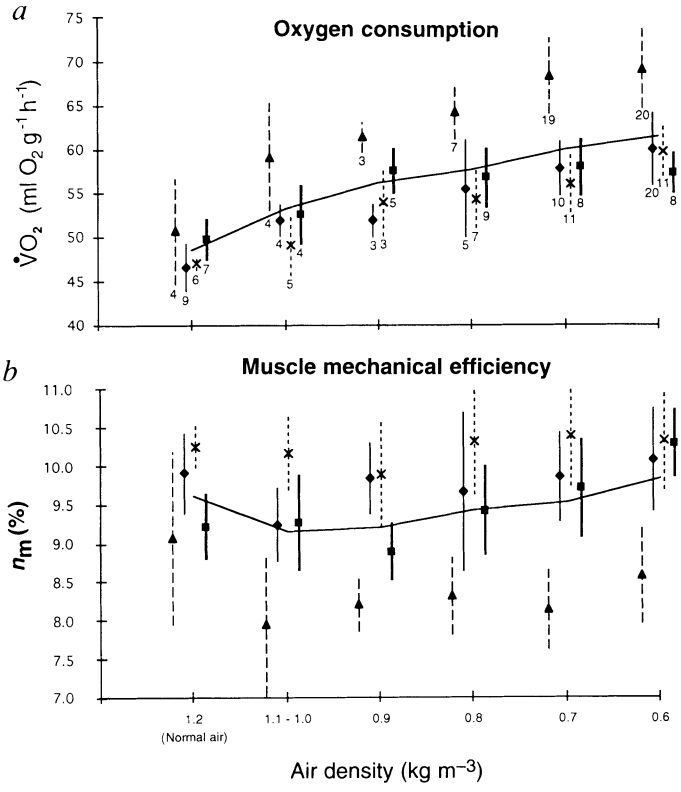

FIG. 4 *a*, Oxygen consumption and *b*, muscle mechanical efficiency in relation to air density reduction (symbols as in Fig. 1). Sample size is indicated and is reduced because only continuous traces of oxygen consumption longer than 5 s were used (hover-feeding events that were short or composed of intermittent feeding were not used). Oxygen consumption was determined using open flow mask respirometry, as described previously[2]. Birds were trained to feed through a cylindrical mask. During hover-feeding, air was drawn through the feeder mask and was sampled and analysed by an Applied Electrochemistry S-3A/I Oxygen Analyzer. Metabolic power input (P_{input}) was determined from the rate of oxygen consumption by assuming a conversion factor of 21.1 J ml^{-1} O_2 for carbohydrate utilization[15] and a respiratory quotient of 1 (ref. 16). Mechanical flight muscle efficiency (n_m) was estimated as $P_{per}/(0.9 \times P_{input})$, assuming that 90% of total metabolism is attributable to the two pairs of major flight muscles[17]. P_{per} was used because hummingbirds can probably store kinetic energy elastically during the deceleration phase of the wing stroke[2]. Calculated values of P_{zero} and associated muscle mechanical efficiency were implausibly high, for example, 349 W kg^{-1} of muscle with 25% mechanical efficiency at failure[18]. Repeated-measures ANOVA results indicate a highly significant density effect ($P<0.001$) for oxygen consumption but not for muscle mechanical efficiency ($P=0.078$). No trial effect was found for either variable.

Hummingbirds hovering in normal air already exhibit stroke amplitudes of 140–150°, and failure occurred universally near stroke amplitudes of 180°. By contrast, stroke amplitude for orchid bees hovering in heliox only reached 142°. Opposite wing interference and even impact is likely when stroke amplitude exceeds 180°. Thus limits for power production may be indicated by maximum stroke amplitude, a primary determinant of induced and total mechanical power during hovering flight. Further non-invasive studies of animal flight performance are necessary to evaluate such interactions between morphological design and the physiological limits inherent to skeletal muscle. □

Received 11 May; accepted 23 August 1995.

1. Suarez, R. K. *Experientia* **48,** 565–570 (1992).
2. Wells, D. J. *J. exp. Biol.* **178,** 39–57 (1993).
3. Hochachka, P. W. *Muscles as Molecular and Metabolic Machines* (CRC, Boca Raton, FL, 1994).
4. Norberg, U. M. *Vertebrate Flight* (Springer, New York, 1990).
5. Weis-Fogh, T. *J. exp. Biol.* **56,** 79–104 (1972).
6. Dudley, R. *J. exp. Biol.* **198,** 1065–1070 (1995).
7. Ellington, C. P. *Phil. Trans. R. Soc. Lond.* B **305,** 145–181 (1984).
8. Berger, M. *J. Ornithol.* **115,** 273–288 (1974).
9. Wells, D. J. *J. exp. Biol.* **178,** 59–70 (1993).
10. Dial, K. P. & Biewener, A. A. *J. exp. Biol.* **176,** 31–54 (1993).
11. Stevenson, R. D. & Josephson, R. K. *J. exp. Biol.* **149,** 61–78 (1990).
12. Josephson, R. K. *A. Rev. Physiol.* **55,** 527–546 (1993).
13. Suarez, R. K., Lighton, J. R. B., Brown, G. S. & Mathieu-Costello, O. *Proc. natn. Acad. Sci. U.S.A.* **88,** 4870–4873 (1991).
14. Greenewalt, C. H. *Smithson. misc. Collns* **144,** 1–46 (1962).
15. Brobeck, J. R. & DuBois, A. B. in *Medical Physiology* Vol. 2 (ed. Mountcastle, V. B.) 1351–1365 (Mosby, St Louis, MO, 1980).
16. Suarez, R. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **87,** 9207–9210 (1990).
17. Lasiewski, R. C. *Physiol. Zool.* **36,** 122–140 (1963).
18. Ellington, C. P. *J. exp. Biol.* **115,** 293–304 (1985).

ACKNOWLEDGEMENTS. We thank J. J. Bull, L. E. Gilbert, J. L. Larimer and M. J. Ryan for comments on the manuscript. This work was supported by an NIH NRSA and a University of Texas Reeder Fellowship.

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 92093, 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 stimuli[1,2]. 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 proposed[3–7], this model is based on the neural substrate and was tested in a simulation of bee flight.

Real and colleagues[8–11] 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' problem[12,13]. In one series of experiments, all the blue flowers contained 2 μl of nectar, $\frac{1}{3}$ of the yellow flowers contained 6 μl, and the remaining $\frac{2}{3}$ 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 suboesophogeal ganglion, an identified neuron, VUMmx1, delivers information about reward during classical conditioning experiments[1]. 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

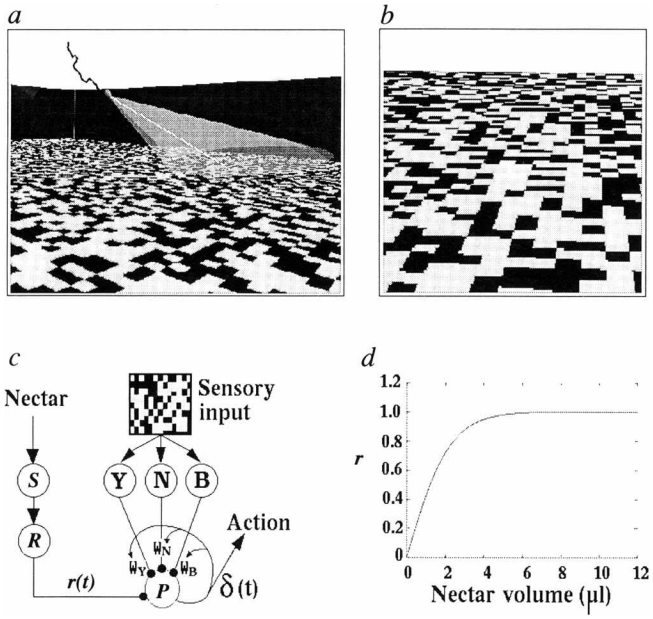

FIG. 1. Model of a foraging bee. *a*, Model bee moving about the three-dimensional arena along with a highlighted cone-shaped region indicating the field of view available to the simulated cyclopean eye. The arena was composed of blue and yellow squares (160 by 160) representing flowers of two different colours (shown here as black and grey). The model bee could move throughout the arena, though it reflected off the ceiling and walls. The trajectory of the bee is shown as a black trail. *b*, Image formed on the 200 by 200 pixel retina of the model bee illustrated in *a*. *c*, Architecture of the model. The output of neurons 'B', 'Y', 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. w_B and w_Y were adaptable and w_N was fixed 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]. *d*, 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 altitude 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 (w_B, w_Y) 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.

odourant followed by artificial stimulation of VUMmx1 through an electrode caused caused subsequent presentation of the odourant alone to elicit a stereotypic 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 (x_B), yellow (x_Y) and neutral (x_N) inputs from the visual field, weighted by w_B, w_Y and w_N, respectively. In the presence of nectar (activity along r(t)), the output of the linear unit *P* was

$$\delta(t) = r(t) + \dot{V}(t) \equiv r(t) + V(t) - V(t-1) \quad (1)$$

where

$$V(t) = w(t) \cdot x(t) = w_B(t)x_B(t) + w_Y(t)x_Y(t) + w_N(t)x_N(t) \quad (2)$$

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 labels 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:

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

where λ is a learning rate. This learning rate is similar to one first proposed by Konorski[16] and is a form of Hebb rule[17] 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].

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. 1*c*,*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[3–5,23].

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. 2*a*). 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 klinokinesis[24]. In various invertebrate systems, neuromodulator delivery influences motor behaviours and thus action choice[25–27], 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 (λ in equation (3)) by neurons that become active in response

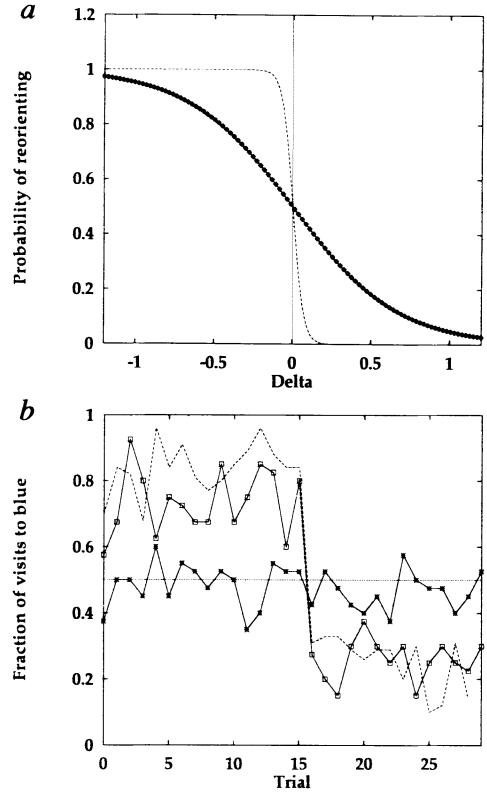

FIG. 2. Simulations of bee foraging behaviour using predictive hebbian learning. *a*, Influence of $\delta(t)$ (labelled Delta) on the decision to reorient. The adaptable (w_B, w_Y) and non-adaptable (w_N) sensory weights influence the decision to reorient through their influence on the size and sign of $\delta(t)$. $P(\delta(t))$ is the probability of reorienting for a given value of $\delta(t)$ and has the form $1/(1+\exp(mx+b))$. 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 (2μl 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 $w_B=w_Y=0.5$ 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[8], 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).

Figure 2*b* shows the behaviour of model bees compared with that of real bees[8] in the experiment testing the extent to which they prefer a constant reward to a variable reward of the same long-term mean. The constant and variable flower types were switched after trial 15. The behaviour of the model matches best

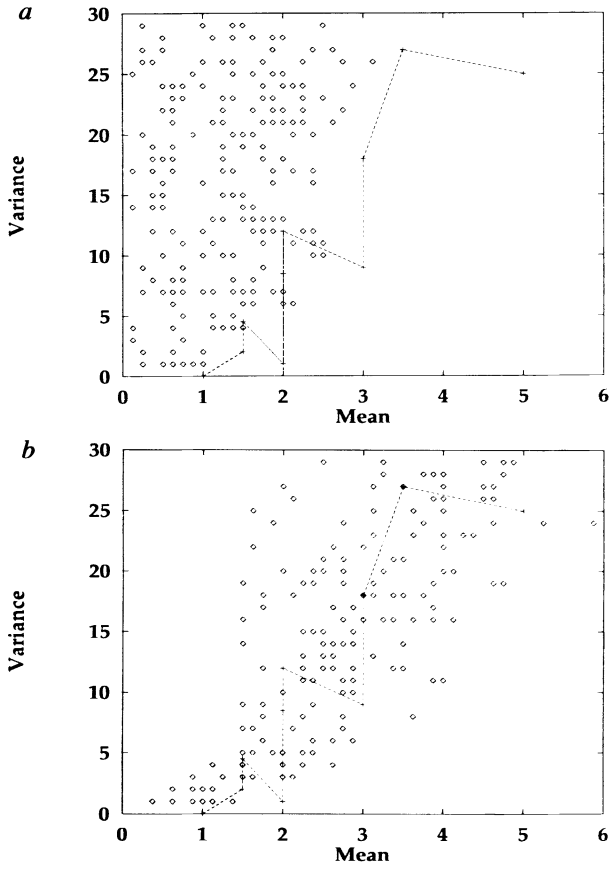

FIG. 3 Tradeoff between the mean and variance of nectar delivery for bee foraging. *a*, *b*, Compare indifference points for the model bees against those for real bees for different learning rates. The coordinates of indifference points are taken as the mean and variance for which the bee samples equally from both the constant and variable flowers. The constant flower contained 0.5μl 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[9] (units for mean and variance are μl and μl^2, respectively).

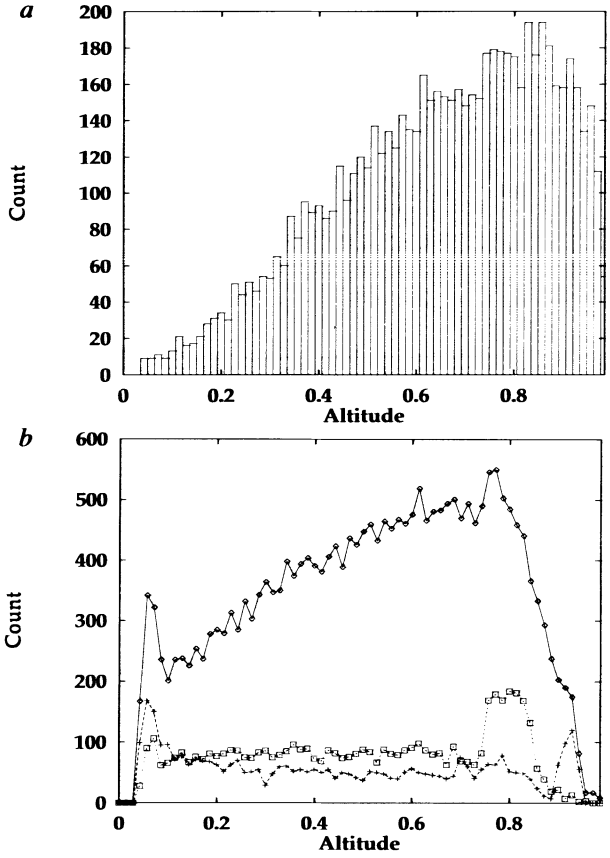

FIG. 4 Structure of foraging problem in an enclosed arena. The choice of a particular representation for the arena and the sensory apparatus of the model bee imposes a number of constraints that influence foraging in the simulated arena. The data presented in *a* and *b* are collected from a total of 1,200 flower landings and 41,735 steps of the model bee (stepsize = 0.05). The learning rate λ was 0.9. *a*, Cumulative histogram of the number of times the model bee viewed primarily neutral colour (>90% of visual field) as a function of altitude in the arena. As the model bee gets closer to the field of flowers, this happens far less frequently. Since w_N was fixed at −0.5, this histogram shows how often reorientations result from viewing regions outside the simulated field of flowers. *b*, Histograms of positive and negative fluctuations in $\delta(t)$ as a function of altitude. Top trace: distribution of all reorientations. Middle trace: distribution of negative fluctuations in $\delta(t)$ excluding those due to viewing primarily neutral colour ($\delta(t) < -0.1$). Bottom trace: distribution of positive fluctuations in ($\delta(t) > 0.1$). As the model bee approached the simulated field, a given change in orientation resulted in larger fluctuations in $\delta(t)$. In both panels, binsize = 0.014 (70 bins).

the observed data for a learning rate of $\lambda = 0.9$, suggesting that the real bee uses information over small time windows for controlling its foraging decisions[8]. These data show that for an equal sized reward between two behavioural choices, bees are biased to choose the more certain predictor of nectar. The uncertainty of a predictor can, however, be balanced by the expected size of the reward.

For example, as shown in Fig. 3, real bumblebees will forage equally on the constant and variable flower types if the mean reward from the variable type is made sufficiently large. This tradeoff behaviour was also exhibited by the model bee. For a given variance in reward from the variable flower, the mean reward was increased until the model bees foraged equally from each flower type: at that point the mean and variance was recorded as the point at which the model bee sampled indifferently from each type. Figure 3*a*,*b* show such indifference plots for two learning rates, 0.05 (*a*) and 0.9 (*b*). As before, the behaviour of real bees was matched best by the higher learning rate (0.9). The reason for the shape of the observed plots can be traced to the nonlinear response function in Fig. 1*d*, which provides diminishing returns for larger volumes of nectar.

This tradeoff between the expected value of reward and its uncertainty is not, however, universal. In choice experiments in honeybees, Greggers and Menzel[6] showed 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[28].

There is good evidence for similar predictive responses in primate mesencephalic dopaminergic systems[14,15,29]. 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.

1. Hammer, M. *Nature* **366,** 59–63 (1993).
2. Hammer, M. & Menzel, R. *J. Neuroscience* **15,** 1617–1630 (1995).
3. Couvillon, P. A. & Bitterman, M. E. *Anim. Learn. Behav.* **13,** 246–252 (1985).
4. Couvillon, P. A. & Bitterman, M. E. *Anim. Learn. Behav.* **14,** 225–231 (1986).
5. Couvillon, P. A., Lee, Y. & Bitterman, M. E. *Anim. Learn. Behav.* **19,** 381–387 (1991).
6. Greggers, U. & Menzel, M. *Behav. Ecol. Sociobiol.* **32,** 17–29 (1993).
7. Menzel, R., Greggers, U. & Hammer, M. in *Insect Learning: Ecology and Evolutionary Perspectives* (eds Papaj, D. R. & Lewis, A. C.) 79–125 (Chapman and Hall, London, 1993).
8. Real, L. A. *Science* **253,** 980–986 (1991).
9. Real, L. A. *Ecology* **62,** 20–26 (1981).
10. Harder, L. D. & Real, L. A. *Ecology* **68,** 1104–1108 (1987).
11. Real, L. A., Ellner, S. & Harder, L. D. *Ecology* **71** 1625–1628 (1990).
12. Berry, D. A. & Fristedt, B. *Bandit Problems: Sequential Allocation of Experiments* (Chapman and Hall, London, 1985).
13. Krebs, J. R., Kacelnik, A. & Taylor, P. *Nature* **275,** 27–31 (1978).
14. Montague, P. R., Dayan, P., Nowlan, S. J., Pouget, A. & Sejnowski, T. J. in *Advances in Neural Information Processing Systems 5* (eds Hanson, S. J., Cowan, J. D. & Giles, C. L.) 969–976 (Morgan Kaufmann, San Mateo, CA, 1993).
15. Montague, P. R. & Sejnowski, T. J. *Learn. Memory* **1,** 1–33 (1994).
16. Konorski, J. *Conditioned Reflexes and Neuron Organization* (Cambridge University Press, Cambridge, 1948).
17. Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
18. Samuel, A. L. *IBM. J. Res. Dev.* **3,** 211–229 (1959).
19. Sutton, R. S. & Barto, A. G. *Psych. Rev.* **88,** 135–170 (1981).
20. Sutton, R. S. & Barto, A. G. in *Learning and Computational Neuroscience* (eds Gabriel, M. & Moore J. W.) 497–538 (MIT Press, Cambridge, MA, 1987).
21. Sutton, R. S. *Machine Learn.* **3,** 9–44 (1988).
22. Barto, A. G., Sutton, R. S. & Watkins, C. J. C. H. *Technical Report 89–95,* (Computer and Information Science, University of Massachusetts, Amherst, MA, 1989).
23. Rescorla, R. A. & Wagner, A. R. in *Classical Conditioning II: Current Research and Theory* (eds Black, A. H. & Prokasy, W. H.) 64–69 (Appleton-Century-Crofts, New York, 1972).
24. Spudich, J. L. & Koshland, D. E. *Proc. natn. Acad. Sci. U.S.A.* **72,** 710–713 (1975).
25. Lockery, S. R. & Kristan, W. B. *J. comp. Physiol.* **168,** 165–177 (1991).
26. Goldstein, R. S. & Camhi, J. M. *J. comp. Physiol.* **168,** 103–112 (1991).
27. Harris-Warrick, R. M., Coniglio, L. M., Barazangi, N., Guckenheimer, J. & Gueron, S. *J. Neurosci.* **15,** 342–358 (1995).
28. Hammer, M. & Menzel, R. *Soc. Neurosci. Abstr.* **20,** 582 (1994).
29. Schultz, W., Apicella, P. & Ljungberg, T. *J. Neurosci.* **13,** 900–913 (1993).
30. Widrow, B. & Stearns, S. D. *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, 1985).

ACKNOWLEDGEMENTS. We thank P. S. Churchland, A. Dayan, A. Pouget, D. Raizen, S. Quartz and R. Zemel 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 (Can) (P.D.), Science and Engineering Research Council (UK) (P.D.).