
Article:
MaBouDi, H., Marshall, J.A.R. and Barron, A.B. (2020) Honeybees solve a multi-comparison ranking task by probability matching. Proceedings of the Royal Society B: Biological Sciences, 287 (1934). 20201525. ISSN 0962-8452

https://doi.org/10.1098/rspb.2020.1525

Honey bees solve a multi-comparison ranking task by probability matching

HaDi MaBouDi[1], James A.R. Marshall[1], Andrew B. Barron[1,2*]

1. Department of Computer Science, University of Sheffield, Sheffield, England

2. Department of Biological Sciences, Macquarie University, North Ryde, Sydney, Australia

* Corresponding author: Andrew B. Barron: andrew.barron@mq.edu.au

Abstract

Honey bees forage on diverse flowers which vary in the amount and type of rewards they offer, and bees are challenged with maximising the resources they gather for their colony. That bees are effective foragers is clear, but how bees solve this type of complex multi-choice task is unknown. Here we set bees a five-comparison choice task in which five colours differed in their probability of offering reward and punishment. The colours were ranked such that high ranked colours were more likely to offer reward, and the ranking was unambiguous. Bees' choices in unrewarded tests matched their individual experiences of reward and punishment of each colour, indicating bees solved this test not by comparing or ranking colours but by basing their colour choices on their history of reinforcement for each colour. Computational modelling suggests a structure like the honey bee mushroom body with reinforcement-related plasticity at both input and output can be sufficient for this cognitive strategy. We discuss how probability matching enables effective choices to be made without a need to compare any stimuli directly, and the utility and limitations of this simple cognitive strategy for foraging animals.

Key words. colour learning, ecological rationality, Multi-armed bandit task, mushroom body, probability matching, reinforcement learning.

Introduction

Foraging honey bees must gather concealed pollen and nectar from many cryptic and variable flower species, all of which vary in the quality and amount of reward. None of these probabilities are known to the foraging bee. This kind of problem has been described as a multi-armed bandit task [1, 2] since there are multiple options available all with unknown potential for payouts. Here we challenged honey bees with a controlled learning task that offered five options differing in the probability of offering reward and punishment to examine how bees solved this type of multiple-comparison task.

One solution would be to identify the option that offered the highest probability of reward and always pick that option (probability maximising) [3]. For example, if a bee learned that blue is rewarded 60% of the time and yellow 40% of the time, and if the bee could compare and rank these alternatives then she would maximise her return by picking blue all the time. An alternative solution involves no comparisons at all. Bees could simply learn the reinforcement properties of each option and match their choices accordingly (probability matching) [4-6]. In the above example, trained bees would choose blue 60% of the time and yellow 40% of the time. This would still result in bees preferring one colour over another, but this could be achieved without comparing or ranking them.

Probability maximising is economically rational since it offers the maximum possible reward where probabilities of different options are known and fixed. In scenarios where the reward probabilities are not known a-priori and have to be estimated from sampling and/or change during sampling probability matching offers a better solution to the trade-off of exploiting current resources while exploring other alternatives [2, 7]. In situations that demand simultaneous sampling of different alternatives while harvesting from the best estimated option (a problem faced by many generalist forager animals), probability matching may be the best foraging strategy [1, 2, 7, 8].

Probability maximising requires comparing and ranking different options [7, 9]. Bees do have the capacity to learn abstract relationships between stimuli [10-12]. A recent study showed Polistes wasps could pass a behavioural test considered indicative of the capacity for transitive inference [13]. This requires arranging stimuli in a ranked order [13] suggesting that *Polistes dominula* and *Polistes metricus* at least have a capacity for ranking stimuli. Probability matching involves no comparisons: it involves learning the reinforcement properties of each

stimulus only. In foundational comparative cognition Bitterman considered probability matching to be "less intelligent" than probability maximising [4, 14]. While that interpretation is open to debate, it is true that probability matching is computationally simpler than probability maximising [2, 7].

Whether honey bees choose by ranking stimuli by probability of reinforcement (probability maximising) or whether they learn the probability of reinforcement for different stimuli (probability matching), is presently unclear. Greggers and Menzel [15] trained honey bees to four different artificial feeders that varied in rate of sucrose delivery, and showed evidence of both probability matching and probability maximising strategies [15]. Fischer et al [16] and Keasar et al [2] tested honey bees and bumblebees (*Bombus terrestris*) and found that bees imperfectly matched their preferences to feeder probability of reward. Here we examined honey bees' choice strategy in a complex multi-option choice test in which bees were punished for wrong choices.

We set honey bees a five-armed bandit task. We challenged honey bees with a colour learning task in which five coloured stimuli were organised in a rank order such that when trained with pairwise presentations of colours only the higher ranked colour was rewarded, with the lower ranked colour being punished. The ranking was unambiguous, and it influenced the probability of each colour being rewarded or punished during training. Tests then explored whether honey bees demonstrated learning of the ranking of the colours (probability maximising), or whether they instead learned the reinforcement history of each separate colour (probability matching). Our data supported a probability matching strategy. To explore how bees might achieve a probability matching strategy we developed a neural network model capable of probability matching inspired by the bee brain. We discuss the consequences of these findings for our understanding of insect choice behaviour.

Material and methods

Animals and testing arena. Experiments were conducted at the Sheffield University Research Apiary containing four standard commercial hives of honey bees (*Apis mellifera*). To attract bees for our experiments we placed a gravity feeder of 20% sucrose solution (w/w) approximately 15m from the hives. Bees visiting this feeder were given individually distinctive marks on their abdomen and/or thorax using coloured Posca marking pens (Uni-Ball, Japan).

5 m from the gravity feeder (and further from the hives) we established the testing arena: a box (100 x 80 x 80 cm) of white expanded PVC foam boards with a UV-transparent Plexiglas roof. Bees could enter the arena through a transparent Perspex corridor (20 x 4 x 4 cm). Interior walls and floor were covered with a pink random dot pattern to provide a contrast between the colour of the bees and the background to assist video analysis (figure 1*a*).

Stimuli. Bees were trained to visit coloured stimuli inside the arena. These were disks (2.5 cm in diameter) of coloured paper covered with transparent laminate (figures 1*b* and S1). The colours of stimuli were selected to cover the range of visible and discriminable colours for bees. (Figures S1*d* and S1*e*). Stimuli were placed on small inverted transparent plastic cups to raise them from the arena floor.

Pre-training phase. Marked bees were attracted from the gravity feeder by offering them a cotton bud soaked with 50% sucrose solution (w/w). Bees feeding from the cotton bud were gently moved to the entrance of the arena and given more 50% sucrose to drink to satiation. This was repeated until the bee flew independently to the entrance of the arena. Bees were trained to fly into the arena via the entrance tube to find drops of 50% sucrose placed on transparent disks of laminate on top of the plastic cups. Bees were released from the arena by lifting the roof. Once a bee flew by herself into the arena to feed, she was selected for the training phase.

Task and training. Bees were trained with five different coloured stimuli in a colour discrimination task. The five different colours were assigned an arbitrary rank (C1-5). Multiple stimuli were presented to them at each trial so that bees could visit a number of different colours in a trial before returning to the hive (capturing the ecology of bees foraging on patches on flowers). In each trial a bee was presented with eight stimuli: four of one colour and four of a different colour (figure 1*a*). For any given pair of colours in a trial the colour with the lower rank was rewarded and the colour with the higher rank was punished, (table S1). To prevent bees from learning any configural cues, feeders were placed randomly within the arena. Feeders were rewarded with 10 µl sucrose solution 50% (w/w) and punished with 10 µl of saturated quinine hemisulphate solution.

In each trial bees were able to freely land on stimuli and to feed. 10 µl drops of 50% sucrose solution were replaced on depleted rewarded stimuli until the bee had fed to satiation and left the arena via the roof. Bees returned to the arena by their own volition. Typically, the inter-

trial interval was 5-10 minutes. After each trial, all stimuli were cleaned with 70% ethanol and water to remove any possible pheromonal cues left by the bee., then air-dried before reuse [17, 18].

Bees were assigned at random to one of two groups: A and B. For group A colours were assigned as: blue = C1, yellow = C2, pink = C3, orange = C4 and green = C5 with white used as a novel colour for testing. For group B the colour assignment was: green = C1, orange = C2, white = C3, yellow = C4 and blue = C5 with pink used as a novel colour in testing. Colours were pseudo-randomly selected to cover the range of colours from green to orange (figure S1). Note that similar colours in hexagon coordinates (figure S1) were not close to each other in the order of the protocol. For example, green and orange in group B were the first and second highest-ranked (C1 and C2) while they are far apart in the hexagon colour space (figure S1*c*). Moreover, the stimuli C2 and C4 were chosen such that they were not similar to C1 and C5, respectively. This does not let the bees use a similarity rule in responding to stimuli. Comparing behaviour of bees from groups A and B allowed us to explore for possible innate colour biases influencing choice.

Bees were randomly assigned to one of two protocols P1 and P2 (tables S2, S3). Although the pairs of stimuli were randomly ordered in both protocols, the number of presentations of each pair was consistent in both protocols (table S1).. Comparing performance between protocols allowed us to examine whether training sequence influenced performance. In training each bee was therefore assigned to one colour group (A or B) and one training protocol (P1 or P2). Over 18 training trials bees experienced all combinations of the five colours twice, with the exception that bees in training never experienced C2 paired with C4 (table S1). This pairing was excluded so that in the post training *transfer test* we could examine how bees responded to a colour pair they had never previously encountered. The 18 bouts of training presented bees with reward from C1 in eight out of eight trials; from C2 in 4 out of 6 trials, from C3 in 4 out of eight trials, from C4 for two out of six trials, and never from C5 (table S1).

Testing. Immediately following the training phase each bee was given three learning tests. The *learning test* presented bees with C1 and C5 – a colour combination they had previously experienced and that had the greatest difference in reinforcement history. The *transfer test* presented bees with C2 and C4 – a combination they never experienced in training. The *novel test* presented bees with white and pink, which was a choice between a colour they had

148 experienced to be rewarded and punished equally often in training trials and a colour they had
149 not experienced in training.

150 During all tests all stimuli offered 10 ul water. During tests bees were video recorded for 120
151 s and their landings on stimuli recorded, after which bees were released from the roof of the
152 arena. The sequence of the tests was randomised for each bee. Between tests bees were allowed
153 to feed in the arena on 10 ul sucrose drops placed on disks of transparent laminate so that we
154 maintained their motivation to visit the arena. As in training, stimuli were cleaned between
155 tests.

156 **Colour selection.** In this study we used colours that are distinguishable for bees [17, 19-22]
157 (figure S1*a*). We measured the spectral reflectance of the all colours used in the experiment
158 (figure S1*b*) following methods of the honeybees' receptors model demonstrated by Chittka
159 [23] and using the spectral sensitivity functions of the honeybees [24] (figure S1*c*). We
160 calculated the Euclidean distance between all pairs of colour in the bee colour space (figure
161 S1*d*). It has been established that colours differing by a Euclidian distance of 0.1 or more in
162 the hexagon coordinates are discriminable for bees [19, 22, 25]. Thus, bees were able to
163 distinguish all colours used in this study (figure S1*d*.) All used colours evoke different levels
164 of the activity from the visual receptors of honeybees (figure S1*e*). Bees also can use colour
165 contrast for target detection [26] and the colour contrasts for all stimuli are high enough to be
166 distinguishable by bees (figure S1*e*). In addition to these models, The bees' performance in the
167 training phase confirms that the bees could distinguish all pairs of colours (see tables S2&S3
168 for training bout order and figure S2 for results).
169

170 **Observation and video recording.** To record bee behaviour inside the arena iPhone 6 cameras
171 were positioned above the entrance viewing into the arena, and on the top of the arena viewing
172 down. The cameras were configured to record at 30 fps at a resolution of 720p (1,280x 720
173 pixels) in the training phase, and 240 fps in the testing phase. During training we recorded from
174 when the bee entered the arena until she was released from the roof. During tests we recorded
175 for 120 s from when she entered the arena. We developed an algorithm to analyse the flight
176 paths of bees from the videos. Bees' flight path was determined by extracting the x-y
177 coordinates of the bee's body and their direction of flight frame by frame (Video S1). The
178 algorithm was used to calculate the number of rewarding and punishing experiences each bee
179 had with each colour during the training and testing.

Only bees that completed the entire training and test sequences were included in the results. Of our 20 bees, only one bee did not complete the entire paradigm due to rain stopping the experiment on that day. We noted each time a bee landed on a stimulus in training, from which we determined how many times each bee encountered each colour as rewarded or punished. Given the spatial information of the stimuli in the view of the camera, we used a threshold flight speed classification to evaluate if bees had landed on a stimulus. This was considered a choice (correct or incorrect) if the bee' s body was located on the stimulus' border and her speed is less than the threshold obtained from all flight information (ref).

Data analysis and statistics. To evaluate bees' performance in the three different unrewarded tests, we analysed the proportion of correct choices estimated as landings on the rewarded stimuli divided by the total number of landings during the 120 s test. We examined the effect of colour group (A or B), protocol (P1 or P2), test type and colour ranking on performance using Wilcoxon signed rank and Wilcoxon rank-sum tests. In addition, Generalised linear mixed models (GLMM) were applied to the performance of the bees in the learning, transfer and novel tests to examine the effects of the protocol, the colour sets and the interaction between these factors on the bees' responses (table S4). Further, to evaluate the homogeneity of the behaviour of different groups of bees, the Brown-Forsythe test was used**.**

To further examine whether individual experience of the history of reward and punishment associated with each colour influenced degree of colour preference we calculated the number of visits to each colour that was rewarded and punished for each bee (figure S3). In our protocol bees freely visited and chose multiple feeders within each training bout and therefore each individual bee experienced a unique history of rewards and punishments for each colour.

For each bee we calculated a reinforcement index (R index) for each colour (C_i) as:

$$R_{c_i} = (\#rewarded\ choices\ of\ C_i - \#punished\ choices\ of\ C_i)/\#\ total\ choices$$

This index has a scale from -1 to +1, and colours that were experienced as punished more often than rewarded had negative R indices. Since the total number of choices both within and between trials varied between bees, the reinforcement index was normalised to the total number of choices each bee made across all of its trials. Hence, the reinforcement index as defined is independent of the motivation of bees to respond to stimuli, but also shows the relative preference of bees for all the colours. We evaluated the relationship between bees' R indices

and bees' performance in the transfer test using a Spearman's correlation test. All statistical tests were performed in MATLAB 2018 (MathWorks, Natick, MA, USA).

Results

In the learning test bees preferred C1 (always rewarded in training) to C5 (always punished) (figure 1*c*; Wilcoxon rank-sum test, z = 5.77, n = 20, p = 7.47e-10, chance level = 0.5). The transfer test presented bees with a colour combination not used in training (C2 vs C4). Here bees preferred C2 to C4 (figure 1*c*: Wilcoxon rank-sum test, z = 5.19, n = 20, p = 2.09e-07) demonstrating that bees preferred a stimulus with both a higher likelihood of reward and a higher ranking in a novel colour comparison. The total numbers of choices during the first 120 s of tests differed between bees and between tests (figure S3*a*; Kruskal-Wallis test, df=62, chi-chq=6.78, p=0.03) but we do not think this impacted the results. We obtained the same result even if the same number of choices was standardised across bees' and tests when calculating performance (figures *S*3*b and S*3*c*). This indicates the number of choices did not have any effect on bees' performance.

The degree of preference differed between the learning test and transfer tests (Wilcoxon signed rank test, z = 3.92, n = 20, p = 8.79e-05). The preference for C1 in the learning test was greater than the preference for C2 in the transfer test (figure 1*c*). This indicates that the degree of colour preference was not absolute (as would be expected if bees were probability maximising). The degree of preference was influenced by the probability of reward and punishment for different colours during training, which is consistent with probability matching.

The training protocols gave bees different numbers of rewarded and punished trials for each stimulus, and within a trial bees differed in how many visits they made to coloured stimuli before departing the arena. The R index was calculated from visit numbers, and R indices differed significantly for different colours (figure 2*a,b*, Wilcoxon rank-sum test, z > 3.74, n = 10, p < 1.82e-4). The R index for C3 was greater than zero (Wilcoxon rank-sum test, z = 4.16, n = 10, p = 3.07e-5 for group A; z = 5.21, n =10, p = 1.82e-7 for group B). Therefore even though C3 was paired with reward and punishment in 50% of the training trials, on average bees experienced C3 as rewarded more often than punished, indicating that bees visited C3 more in rewarded trials than punished trials. In the novel test bees preferred C3 more than a novel colour (figure 1*d*: Wilcoxon rank-sum test, z = 5.76, n = 20, p = 7.93e-09), perhaps as a consequence of this experience in training.

There was no difference in performance in the learning, transfer and novel tests of bees that had been trained with different colour contingencies (Groups A and B, figure S3*d*; Wilcoxon rank-sum test, z = -1.32, n = 10, p > 0.18; Generalized linear mixed model, p > 0.08; see table S1 for more details). There was also no difference in performance between bees from different training protocols in the learning, transfer and novel tests (P1 and P2, figure S3*e*; Wilcoxon rank-sum test, z = -1.32, n = 10, p > 0.18; Generalized linear mixed model, p > 0.16, table *S4*). The Dimulated data generated from a Monte Carlo algorithm (See Supplementary methods) confirmed that the result of the GLMM would be consistent for a larger sample size with 60 observations (table S5). Also, to make sure each group of 5 bees that were trained to a specific colour set and a protocol showed statistically homogenous behaviour in terms of performance in the tests, the Brown-Forsythe test was applied. All four groups of bees were statistically homogenous (Brown-Forsythe test, df1=3 , df2=16; in the learning test: F=1.25, p=0.32; in the transfer test: F=0.08, p=0.96; in the novel test: F=1.82, p=0.18). Further, for all three tests there was no significant effect of the interaction between the protocol and the colour set used in the experiment on bee performance (p > 0.14, tables S4 and S5). In addition to the homogeneity of the data, power analyses of the statistical tests were calculated for the bees' performances in the tests (figure S3*f*). The power analyses support that an appropriate number of bees were chosen for the experiment.

Bees' performance in the transfer test was positively correlated with the difference in their R indices for C2 and C4 (figure 2*c*; Spearman's correlation test, rho=0.4, n=20, p=0.046). This finding is consistent with each individuals' degree of preference in the transfer test being influenced by their specific experience of reward and punishment for the two colours in training. It is not consistent with bees comparing and ranking colours by reward likelihood to make a choice.

A simple neural network model sufficient for learning history of reinforcement, consistent with the neuroanatomy of the bee brain.

To explore how learning of the history of reinforcement of colours in this 5-colour learning task might be achieved by honeybees we developed a neural network model inspired by the

neurobiology of colour coding and colour learning in bees [29-32] (figure 3*a*). Full details of the model are given in Supplementary Methods.

In this model, three different types of receptors, Short (S)- ,medium (M)- and long (L)-wavelength-sensitive photoreceptors are stimulated by the light reflected by each colour stimulus, which we quantified as the spectral reflectance function of each stimulus (figures S1*b* and S1*e*). Axons from colour receptors project to the medulla and make inhibitory connections with transmedullary (TM) neurons [33-36]. In bees the L-receptors project to the lamina also, [37] but in this model we consider medulla processing only.

We considered a simple circuit such that one transmedullary neuron is activated by one type of receptor only. Transmedullary neurons exhibit high spontaneous activity and receive inhibitory signals from receptors, therefore the three different types of transmedullary neurons respond to colours by decreasing their firing rate from the spontaneous rate [33]. The transmedullary neurons send either excitatory or inhibitory signals to the third order neurons (amacrine or large field neurons) in the next layer of the medulla [33, 38, 39] (figure 3*a*). Following the algorithm presented in Vasas et al. [30], in the model their synaptic weights (L) were estimated from empirical neurophysiological data to reproduce the diverse activity of colour sensitive neurons reported by Kien and Menzel [31, 32] (figure S5).

W in figure 3*a* describes the matrix connectivity between the third-order neurons and Kenyon cells (KCs) of the mushroom body in the collar region. Kenyon cells output to a single mushroom body output neuron (MBON) in the alpha lobe of the mushroom body $D(c)$ through a vector of synaptic weights, V. The protocerebral-calycal tract feedback pathway (figure 3*a*, red) takes inputs from all Kenyon cells and sends a feedback inhibitory signal to the collar region [40-44]. In the model this pathway is represented as a single neuron that contributes feedback gain control to the system and maintains sparse coding across the Kenyon cell population.

In the model, a single reinforcement neuron (figure 3*a*, yellow) modulates strengths of synaptic connectivity at both the input and output of the Kenyon cells in response to both reward and

punishment. This is responsible for the changes in activity of the network during training with rewarded and punished coloured stimuli (supplementary methods, Equations 6-8).

The output of the MBON $D(c)$ was used to evaluate the performance of the model. $D(c)$ has a tonic firing rate, which is decreased by punishment, and increased by reward. Following training, maximal performance of the model was judged as an increase in firing rate of $D(c)$ to maximum to a colour that had been rewarded in training, or a decrease in firing rate of $D(c)$ to minimum to a colour that had been punished in training.

We assessed the performance of the model in a simple colour discrimination task. The model was presented with any pair of monochromatic colours between 300 to 700nm, one was rewarded and one punished. Prior to training the model did not differentiate between any two colours (no difference in output from $D(c)$). Following 10 training trials with the rewarded and punished colours the model was able to discriminate different colours (figure 3*b*) [26, 45]. We then trained the model using protocol P1 or P2 (table S2). Performance of the model closely matched responses of bees in the learning test (figure 3*a*; compare with figure 1*c*). Following training the model also successfully discriminated between a novel pair of colours and preferred the colour that had been more often associated with reward in training, similar to that of honey bees (figure 3*c*; compare with 1*c*).

The model did not differentiate between a novel colour and a colour that had been equally reinforced and punished during training (figure 3*c*). Honey bees, by contrast, preferred a colour that had been paired with reward in 50% of trials (C3) over a novel colour (figure 1*d*). But, as we noted above, bees made more visits to C3 when rewarded than when punished and hence the mean R index for C3 was positive (figures. 2*a,b*). In this respect training of bees differed from training in the model. When training the model the R index of a colour paired equally with punishment and reward was zero. We therefore added an additional test, and found that the model was able to successfully discriminate a colour that had been paired with reward on 66% of trials during training from a novel colour (figure S6, Wilcoxon rank-sum test, z = 5.31, n = 20, p = 5.62e-8).

The model allowed us to examine which elements of our network were necessary for this form of colour discrimination learning. Sparseness of colour coding in the KCs strongly affected the

performance of the model: Dense coding of colours in the KC population reduced the ability to learn to discriminate colours by reward and punishment (figure S7).

Plasticity at both the input (W) and output (V, figure 3*a*) of the Kenyon cells was essential for the model to correctly discriminate all colours by their history of reinforcement. In training if the weights in connection matrix W (between the third-order neurons and Kenyon cells) were fixed bees could learn to prefer a colour that was always rewarded over a colour that was always punished in training (figures S7*b,c*), but their performance in the transfer test (a comparison between one stimulus reinforced at 66% and one reinforced at 33%) was reduced when compared to the full model (figure 3*c* and figures S7*b,c*), and depended on both the perceptual similarity and the reinforcement history of the colours. Plasticity in connection matrix W decorrelates the activity of KC for any two presented colours that differ in reward history, even if the population activity of third-order neurons are highly correlated (figure S5). Hence, post training distinctive groups of KCs separately encode the colour information of each colour. This increases the ability of KCs activated by different colours to drive different levels of activity in the MBON from $D(c)$ due to changes in connectivity resulting from different reward/penalty ratios associated with each colour during training.

If connection weights in W were fixed the degree of difference between the pattern of KC activation to two different colours correlated with the perceptual difference between the two colours (figure S1). With this limitation the response of the model to any specific colour post training was sensitive to both the reinforcement history of that colour during training, and the reinforcement history of similar colours during training (figures S7*b,c*).

Discussion

Our data show that in a five-armed bandit task bees matched their colour choices in tests to the probability of each colour being rewarded in training (figure 1*c*, 2*c*). This behaviour can be explained simply by them learning the properties of each stimulus, and does not require comparison between options.

Probability matching may be the ecologically optimal solution to this kind of task if reinforcement probabilities for each option are unknown or could change [1, 2, 7]. Matching choice probability to the history of reinforcement offers a simple and effective solution to the

360 explore/exploit trade off in which an animal must optimise across exploiting a current resource
361 type or patch, or moving to and sampling alternatives [1, 7]. Under some circumstances, a
362 probability matching strategy could appear as floral constancy if one flower type were
363 significantly more rewarded than others in a patch, or as floral majoring and minoring if two
364 alterative flower types differed strongly in reward likelihood [2, 7].

365 Probability matching as a solution to a five-armed bandit task is computationally parsimonious.
366 Niv et al [7] have argued that reinforcement learning is sufficient for probability matching.
367 The mushroom body is an important brain region for learning [46-48], and octopaminergic and
368 dopaminergic neuromodulatory neurons are essential for plastic adjustment of connection
369 strengths in the mushroom body circuit in response to reinforcement [49-54]. Distinct but
370 interacting dopaminergic and octopaminergic neurons encode reward and punishment in the
371 fly brain [50], and it is likely that something similar occurs in honey bee brains [51, 52].

372 In our model we considered the insect mushroom body and its visual inputs (figure 3*a*) as a
373 very simple reinforcement learning system to explore the feasibility of learning to probability
374 match in insects. In the model plasticity at both the input (calyx) and output (lobes) of the
375 mushroom body was necessary for the model to effectively learn the history of reinforcement
376 for different colours. Plasticity at the mushroom body input was needed for the system to be
377 able to decorrelate the neural representations of perceptually similar colours in order to learn
378 independent reinforcement histories for them (figure S7). Reinforcement-related plasticity at
379 the mushroom body calyx and lobes is feasible for insects. Hammer [55, 56] argued the
380 modulatory neuron VUMmx1 mediates learning of sugar reward in honey bees. This neuron
381 innovates both the calyx and lobes of the mushroom body [55]. There are other modulatory
382 inputs (both inhibitory and excitatory) to the calyx of both bees and flies [52-54]. Strube-Bloss
383 [47] has also argued plasticity in the calyx may be important for learning to distinguish different
384 stimuli by reward and punishment.

385 Our model also emphasised the need for sparse coding of colour across the Kenyon cell
386 population for efficient learning (figure S7). Our mushroom body model is equivalent to a
387 three-layer associative network. Sparse coding in the middle layer of a three-layer network is
388 recognised to be an important feature for effective classification in such types of network [57,
389 58]. In honey bees sparse coding is supported by the GABAergic inhibitory feedback PCT
390 pathway from the mushroom body outputs to the calyx [29, 52, 59-61], and such feedback is
391 essential for complex discrimination in bees [41, 62].

We can explain bee's behaviour in this complex task without requiring the bees to compare the properties of any of the stimuli offered. This is perhaps a counter-intuitive way of thinking about behaviour in a choice test, but probability matching will give the appearance of choice and preference without the animal effecting any choice. This is, of course, not evidence bees or other insects *cannot* rank. As we noted above, Polistes wasps have solved a transitive inference task that controlled for reinforcement history [13] suggesting a capacity to compare and rank. Honey bees, however, failed at a similar task [63]. We note that the speed at which bees learn reinforcement history, and the effectiveness of this strategy in optimising performance in most foraging tasks may obviate the need for more complex choice strategies in most circumstances.

We also propose that given its simplicity, plausibility and ubiquity probability matching should be considered a most parsimonious explanation of behaviour in animal 'choice' assays. Under many circumstances in assays offering a binary-choice probability matching can give the appearance of overt comparison or even ranking of options to form a preference, when no comparison may be involved. Considering and controlling for the reinforcement history of different options would help to distinguish probability matching strategies from true ranking or comparison strategies.

Acknowledgements

We thank Neville Dearden and Michael Port for technical assistance. We thank Lars Chittka and Vera Vasas for their assistance measuring the spectra of coloured stimuli.

Fundings

HM and JARM are supported by the Engineering and Physical Sciences Research Council (grant no EP/P006094/1). ABB is supported by funding by an ARC Future Fellowship (FT140100452), a Leverhulme Visiting Fellowship, and the Templeton World Charity Foundation (grant no. TWCF0266).

References

[1] Morimoto, J. 2019 Foraging decisions as multi-armed bandit problems: Applying reinforcement learning algorithms to foraging data. *J Theor Biol* **467**, 48-56.
[2] Keasar, T., Rashkovich, E., Cohen, D. & Shmida, A. 2002 Bees in two-armed bandit situations: foraging choices and possible decision mechanisms. *Behavioral Ecology* **13**, 757-765.
[3] Hull, C.L. 1930 Simple trial-and-error learning: A study in psychological theory. . *Psychological Review* **37**.
[4] Bitterman, M.E. 1975 The comparative analysis of learning. *Science* **188**, 699-709.
[5] Estes, W.K. 1959 The statistical approach to learning theory. In *Psychology: A study of a science. Study I: Conceptual and systematic. Volume 2: General systematic formulations, learning, and special processes* (ed. S. Koch). New York, McGraw Hill.
[6] Herrnstein, R. 1970 On the law of effect. *Journal of Experimental Animal Behavior* **13**, 243-266.
[7] Niv, Y., Joel, D., Meilijson, I. & Ruppin, E. 2002 Evolution of reinforcement learning in uncertain environments: a simple explanation for complex foraging behaviors. *Adaptive Behavior* **10**, 5-24.
[8] The Modelling Animal Decisions Group, Fawcett, T.W., Fallenstein, B., Higginson, A.D., Houston, A.I., Mallpress, D.E.W., Trimmer, P.C. & McNamara, J.M. 2014 The evolution of decision rules in complex environments. *Trends Cogn Sci* **18**, 153-161.
[9] Greggers, U. & Mauelshagen, J. 1997 Matching behavior of honeybees in a multiple-choice situation: The differential effect of environmental stimuli on the choice process. *Animal Learning & Behavior* **25**, 458-472. (doi:10.3758/bf03209852).
[10] Avarguès-Weber, A. & Giurfa, M. 2013 Conceptual learning by miniature brains *Proc. Roy. Soc. Lond. B* **280**. (doi:10.1098/rspb.2013.1907).
[11] Giurfa, M. 2013 Cognition with few neurons: higher-order learning in insects. *Trends in Neurosciences* **in press**.
[12] Giurfa, M., Zhang, S., Jenett, A., Menzel, R. & Srinivasan, M.V. 2001 The concepts of 'sameness' and 'difference' in an insect. *Nature* **410**, 930-933.
[13] Tibbetts, E.A., Agudelo, J., Pandit, S. & Riojas, J. 2019 Transitive inference in Polistes paper wasps. *Biol. Lett.* **15**, 20190015.
[14] Bitterman, M.E. 1965 Phyletic differences in learning. *American Psychologist* **20**, 396-410.
[15] Greggers, U. & Menzel, R. 1993 Memory dynamics and foraging strategies of honeybees. *Behav Ecol Sociobiol* **32**, 17-29.
[16] Fischer, M.E., P. A. Couvillon & Bitterman, M.E. 1993 Choice in honeybees as a function of the probability of reward. *Animal Learning and Behavior* **21**, 187-195.
[17] Li, L., MaBouDi, H., Egertová, M., Elphick, M.R., Chittka, L. & Perry, C.J. 2017 A possible structural correlate of learning performance on a colour discrimination task in the brain of the bumblebee. *Proc. Roy. Soc. Lond. B* **284**, 20171323.
[18] Solvi, C., Al-Khudhairy, S.G. & Chittka, L. 2020 Bumble bees display cross-modal object recognition between visual and tactile senses. *Science* **367**, 910-912.
[19] Reser, D.H., Wijesekara Witharanage, R., Rosa, M.G.P. & Dyer, A.G. 2012 Honeybees (Apis mellifera) learn color discriminations via differential conditioning independent of long wavelength (Green) photoreceptor modulation. *PLoS ONE* **7**, e48577.
[20] Spaethe, J., Tautz, J. & Chittka, L. 2001 Visual constraints in foraging bumblebees: flower size and color affect search time and flight behaviour. *PNAS* **98**, 3898-3903.
[21] Giurfa, M. 2004 Conditioning procedure and color discrimination in the honeybee *Apis mellifera*. *Naturwissen.* **91**, 228-231.

[22] Avarguès-Weber, A., de Brito Sanchez, M.G., Giurfa, M. & Dyer, A.G. 2010 Aversive Reinforcement Improves Visual Discrimination Learning in Free-Flying Honeybees. *PLoS ONE* **5**. (doi:10.1371/journal.pone.0015370).
[23] Chittka, L. 1992 The color hexagon: a chromaticity diagram based on photoreceptor excitations as a generalized representation of color opponency. *J Comp Physiol A* **170**, 533-543.
[24] Peitsch, D., Fietz, A., Hertel, H., de Souza, J., Ventura, D.F. & Menzel, R. 1992 The spectral input systems of hymenopteran insects and their receptor-based colour vision. *J Comp Physiol A* **170**, 23-40.
[25] Dyer, A. & Chittka, L. 2004 Biological significance of discriminating between similar colours in spectrally variable illumination: bumblebees as a study case. *J Comp Physiol A* **190**, 105-114.
[26] Dyer, A.G., Spaethe, J. & Prack, S. 2008 Comparative psychophysics of bumblebee and honeybee colour discrimination and object detection. *J Comp Physiol* **194**, 617-627.
[27] Green, P. & MacLeod, C.J. 2016 SIMR:anR package for power analysis of generalizedlinear mixed models by simulation. *Methods in Ecology and Evolution* **7**, 493-498.
[28] Bates, D., Mächler, M., Bolker, B. & Walker, S. 2014 Fitting Linear Mixed-Effects Models using lme4. (doi:https://arxiv.org/abs/1406.5823).
[29] Cope, A.J., Vasilaki, E., Minors, D., Sabo, C., Marshall, J.A.R. & Barron, A.B. 2018 Abstract concept learning in a simple neural network inspired by the insect brain. *PLoS Comput Biol* **14**, e1006435.
[30] Vasas, V., Peng, F., MaBouDi, H. & Chittka, L. 2019 Randomly weighted receptor inputs can explain the large diversity of colour-coding neurons in the bee visual system. *Scientific Reports* **9**, 8330.
[31] Kien, J. & Menzel, R. 1977 Chromatic properties of interneurons in the optic lobes of the bee. I. Broad band neurons. *Journal of Comoparative Physiology* **113**, 35-53.
[32] Kien, J. & Menzel, R. 1977 Chromatic properties of interneurons in the optic lobes of the bee. II Narrow band and colour opponent neurons. *J Comp Physiol* **113**, 35-53.
[33] Paulk, A.C., Dacks, A.M. & Gronenberg, W. 2009 Color processing in the medulla of the bumblebee (Apidae: *Bombus impatiens*). *J. Comp. Neurol.* **513**, 441-456.
[34] Suart, A.E., Borycz, J. & Meinertzhagen, I.A. 2007 The dynamics of signaling at the histaminergic photoreceptor synapse of arthropods. . *Progress in Neurobiology* **82**, 202-227.
[35] Bicker, G. 1999 Biogenic amines in the brain of the honeybee: cellular distribution, development and behavioural functions. *Microscopy Research and Technique* **44**, 166-178.
[36] Nässel, D.R. 1999 Histamine in the brain of insects: a review. *Microscopy Research and Technique* **44**, 121-136.
[37] Ribi, W.A. & Scheel, M. 1981 The second and third optic ganglia of the worker bee. *Cell Tissue Res.* **221**, 17-43.
[38] Meyer, E., Matute, C., Streit, P. & Nässel, D.R. 1986 Insect optic lobe neurons identifiable with monoclonal antibodies to GABA. . *Histochemistry* **84**, 207-216.
[39] Ehmer, B. & Gronenberg, W. 2002 Segregation of visual input to the mushroom bodies in the honeybee (Apis mellifera). *J. Comp. Neurol.* **451**, 362-373.
[40] Okada, R., Rybak, J., Manz, G. & Menzel, R. 2007 Learning-related plasticity in PE1 and other mushroom body-extrinsic neurons in the honeybee brain. *Journal of Neuroscience* **27**, 11736-11747. (doi:10.1523/jneurosci.2216-07.2007).
[41] Devaud, J.M., Papouin, T., Carcaud, J., Sandoz, J.C., Grünewald B & Giurfa, M. 2015 Neural substrate for higher order learning in an insect: mushroom bodies are necessary for configural discriminations. *Proceedings of the National Academy of Sciences of the United States of America.* **112**, E5854.

517 [42] Ganeshina, O. & Menzel, R. 2001 GABA-immunoreactive neurons in the mushroom
518 bodies of the honeybee: an electron microscopic study. *J. Comp. Neurol.* **437**, 335-349.
519 [43] Haehnel, M. & Menzel, R. 2010 Sensory representation and learning-related plasticity in
520 mushroom body extrinsic feedback neurons of the protocerebral tract. *Frontiers in Systems*
521 *Neuroscience* **4**, 161. (doi:10.3389/fnsys.2010.00161).
522 [44] Rybak, J. & Menzel, R. 1993 Anatomy of the mushroom bodies in the honey bee brain -
523 the neuronal connections of the alpha-lobe. *J. Comp. Neurol.* **334**, 444-465.
524 (doi:10.1002/cne.903340309).
525 [45] Dyer, A.G. & Neumeyer, C. 2005 Simultaneous and successive colour discrimination in
526 the honeybee (*Apis mellifera*). . *J Comp Physiol* **191**, 547-557.
527 [46] Menzel, R. 2001 Searching for the memory trace in a mini-brain, the honeybee.
528 *Learning & Memory* **8**, 53-62.
529 [47] Strube-Bloss, M.F., Nawrot, M.P. & Menzel, R. 2011 Mushroom body output neurons
530 encode odor-reward associations. *J Neurosci* **31**, 3129-3140. (doi:10.1523/jneurosci.2583-
531 10.2011).
532 [48] Heisenberg, M. 2003 Mushroom body memoir: from maps to models. *Nature Reviews*
533 *Neuroscience* **4**, 266-275.
534 [49] Schwaerzel, M., Monastirioti, M., Scholz, H., Friggi-Grelin, F., Birman, S. &
535 Heisenberg, M. 2003 Dopamine and octopamine differentiate between aversive and
536 appetitive olfactory memories in Drosophila. *Journal of Neuroscience* **23**, 10495-10502.
537 [50] Cognigni, P., Felsenberg, J. & Waddell, S. 2018 Do the right thing: neural network
538 mechanisms of memory formation, expression and update in Drosophila. *Curr Opin*
539 *Neurobiol* **49**, 51-58.
540 [51] Perry, C.J. & Barron, A.B. 2013 *Neural mechanisms of reward in insects*543-562 p.
541 [52] Søvik, E., Perry, C.J. & Barron, A.B. 2015 Insect reward systems: Comparing flies and
542 bees. In *Advances in Insect Physiology* (pp. 189-226.
543 [53] Burke, C.J., Huetteroth, W., Owald, D., Perisse, E., Krashes, M.J., Das, G., Gohl, D.,
544 Silies, M., Certel, S. & Waddell, S. 2012 Layered reward signalling through octopamine and
545 dopamine in Drosophila. *Nature* **492**, 433-437. (doi:10.1038/nature11614).
546 [54] Waddell, S. 2010 Dopamine reveals neural circuit mechanisms of fly memory. *Trends in*
547 *Neurosciences* **33**, 457-464.
548 [55] Hammer, M. 1993 An identified neuron mediates the unconditioned stimulus in
549 associative olfactory learning in honeybees. *Nature* **366**, 59-63.
550 [56] Hammer, M. 1997 The neural basis of associative reward learning in honeybees. *Trends*
551 *in Neuroscience* **20**, 245-252.
552 [57] Bazhenov, M., Huerta, R. & Smith, B. 2013 A computational framework for
553 understanding decision making through integration of basic learning rules. *Journal of*
554 *Neuroscience* **33**, 5686-5697.
555 [58] Huerta, R., Amigo, J.M., Nowotny, T. & Elkan, C. 2012 Inhibition in multiclass
556 classification. *Neural Computation* **24**, 2473-2507.
557 [59] Bicker, G., Schäfer, S. & Kingan, T. 1985 Mushroom body feedback interneurones in
558 the honeybee show GABA-like immunoreactivity. *Brain research* **360**, 394-397.
559 [60] Grünewald, B. 1999 Morphology of feedback neurons in the mushroom body of the
560 honeybee *Apis mellifera*. *J. Comp. Neurol.* **404**, 114-126.
561 [61] Grünewald, B. 1999 Physiological properties and response modulations of mushroom
562 body feedback neurons during olfactory learning in the honeybee *Apis mellifera*. . *J Comp*
563 *Physiol A* **185**, 565-576.
564 [62] Boitard, C., Devaud J-M, G, I. & M, G. 2015 GABAergic feedback signaling into the
565 calyces of the mushroom bodies enables olfactory reversal learning in honey bees. *Frontiers*
566 *in Behavioral Neuroscience* **9**, 198.

567 [63] Benard, J. & Giurfa, M. 2004 A test of transitive inferences in free-flying honeybees:
568 unsuccessful performance due to memory constraints. *Learning and Memory* **11**, 328-336.

569

Figures

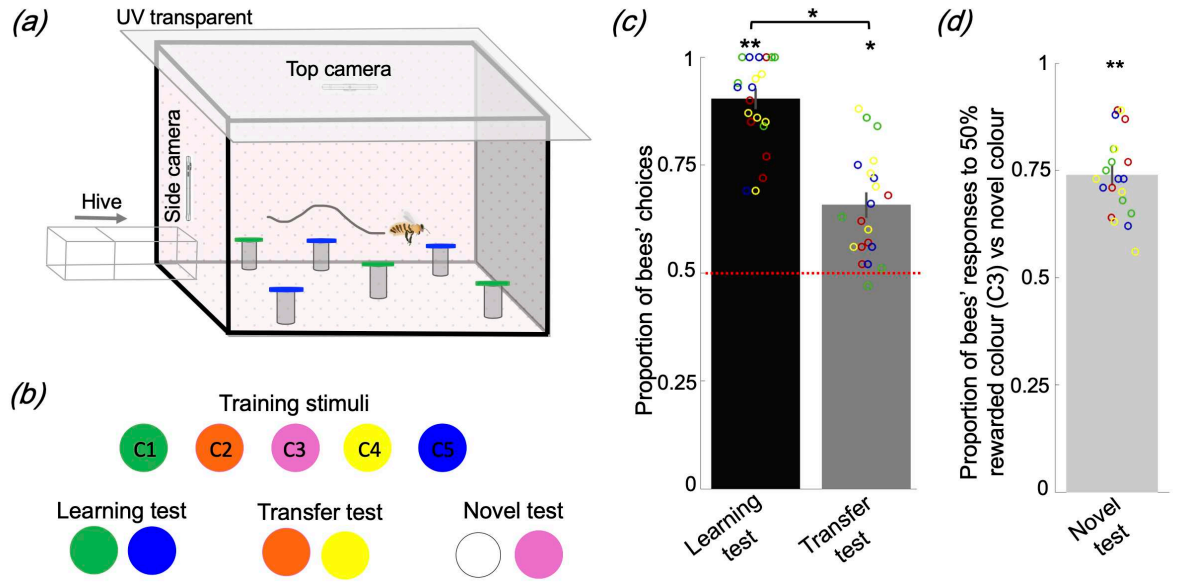
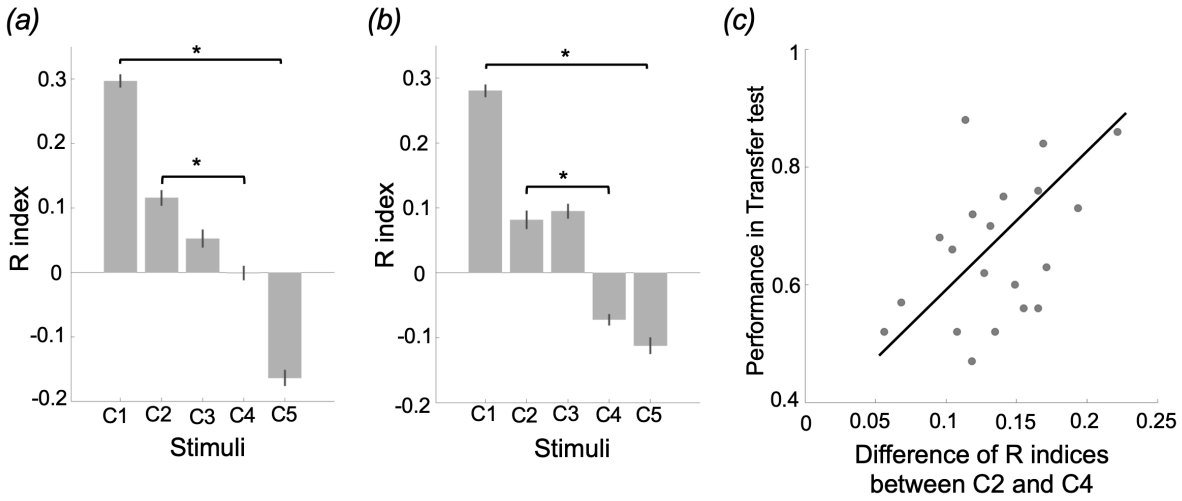


Figure 1. ***a)*** The experimental paradigm. Each bee received 18 training trials. In each training trial, stimuli of one colour offered 10 ul sucrose solution and the other offered 10 ul of quinine hemisulphate. ***b)*** Five different colours were used in this study. The colours differed in the proportion of training trials in which they offered reward and punishment (rewarded at 100%, 66%, 50%, 33%, 0% of trials, table 1). Following training bees received three unreinforced tests. In the *learning test* bees were presented with the colours that had always been rewarded and always punished in training. In the *transfer test* bees were tested with a colour combination that had not been used in training. In the *novel test* bees were presented with a novel colour and a colour that had been rewarded in 50% of training trials. ***c, d)*** Performance on the tests. Bars show mean frequency with which bees landed on stimuli they had experienced as more rewarded during training. *d)* In the novel test, bees chose the trained colours (C3) more than the novel colour. Values for each individual bee are shown by small empty circles in different colours: Red: group A, Protocol P1; Green: group A, Protocol P2; Blue: group B, Protocol P1; Yellow: group B, Protocol P2 (table S6). N = 20. Vertical lines: s.e.m. Dashed line indicates performance expected at random. * indicates $p<0.05$ and ** indicates $p<0.005$.



587

588 **Figure 2.** ***a,b)*** Mean R index for bees for each colour (*a* = group A, *b* = group B). Bars show
589 mean and s.e.m. of the 10 bees experience of each colour in the training phase. * indicates
590 $p<0.05$ for comparing the R index of a pair of colours used in the learning test (C1 vs C5) and
591 the transfer test (C2 vs C4). ***c)*** Relationship between performance in the transfer test and
592 difference in R index between the two colours in the transfer test. Line indicates the linear fit
593 to the data (Spearman's correlation test, rho = 0.4, n = 20, p = 0.046).

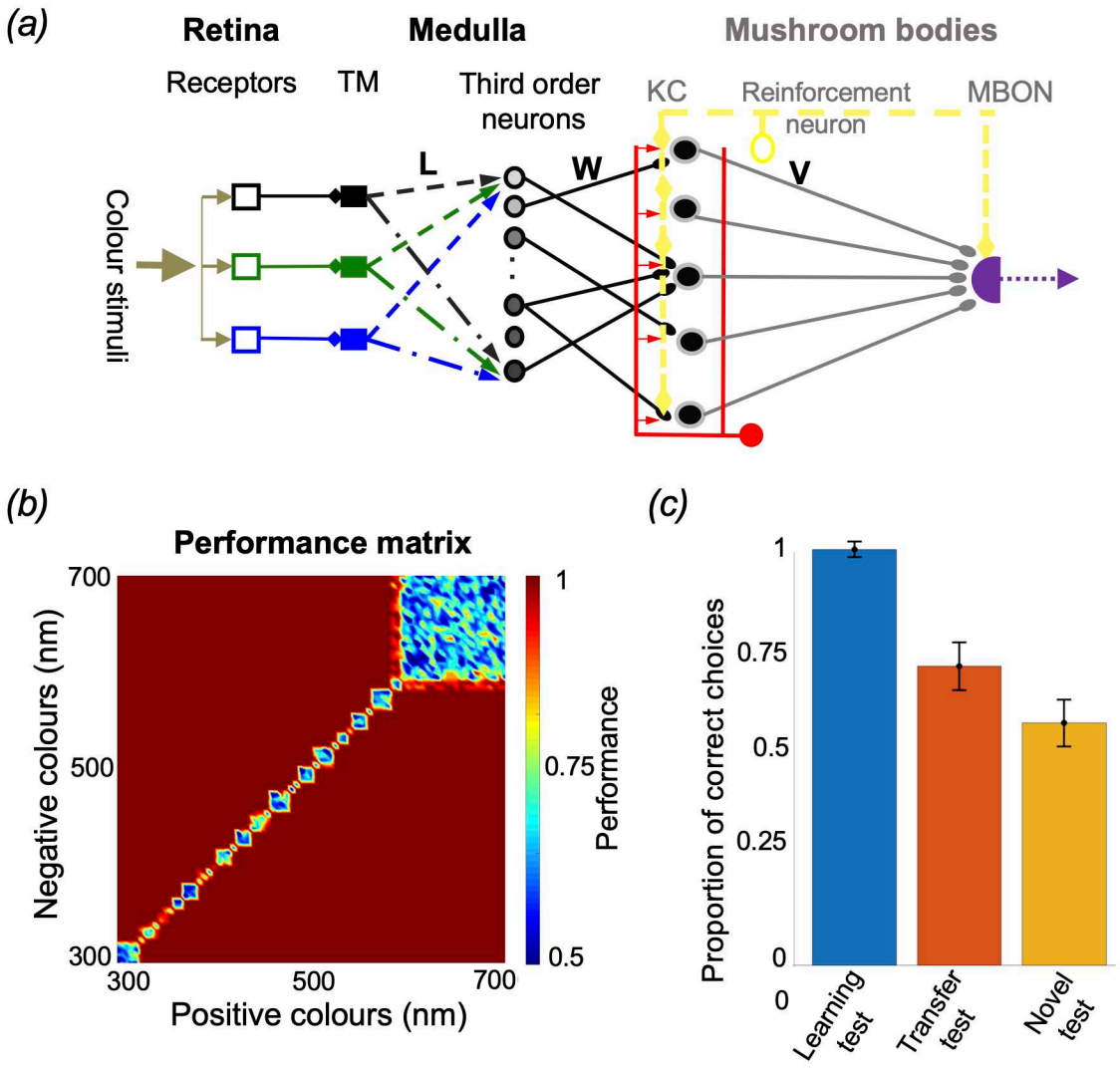


Figure 3. ***a)*** Network topology of the model of colour learning in the bee brain. Colour information is transferred from the three different types of photoreceptors to the medulla. Each receptor sends an inhibitory signal to only one type of transmedullary neuron (TM). Next, TM neurons randomly connect with the third order neurons via a connection matrix L. The vector connectivity of L was estimated from empirical data to produce a wide diversity of colour sensitive neurons. Colour information was transferred from the medulla to the Kenyon cells of the mushroom bodies (MB) via connection matrix W. Kenyon cells output to a single neuron (MBON) in the alpha lobe of the mushroom bodies via connection matrix V. The inhibitory feedback PCT pathway (red) maintains low excitability and sparse coding across the Kenyon cell population. A neuromodulatory neuron firing in response to both reward and punishment projects to both the input and output of the MB. There it acts to alter synaptic connectivity in both regions (W and V) in response to reward and punishment according to the proposed learning rule (see Supplementary Material and Method). ***b)*** The performance matrix of the

608 model in a simple colour discrimination task. The colour at each matrix's element displays the
609 performance of the model to two monochromatic colours whose wavelengths are presented on
610 the x and y axes. Following 10 trials with one colour associated with reward and one with
611 punishment the model is able to discriminate all combination of colours if their wavelength
612 distances are larger than 5 nm. ***c)*** Model's performance following training with protocol P1 or
613 P2 (tables S2&S3). Bars were calculated from the performance of the model for 50 different
614 initial parameters that simulated 50 different model bees that were trained to the training
615 stimuli.

616

Supplementary Material for:

Honey bees solve a multi-comparison ranking task by probability matching

HaDi MaBouDi[1], James A.R. Marshall[1], Andrew B. Barron[1,2]

1. Department of Computer Science, University of Sheffield, Sheffield, England

2. Department of Biological Sciences, Macquarie University, North Ryde, Sydney, Australia

Supplementary Method:

A neural network model for learning the history of reinforcements in bees

a) Model of colour coding in the visual sensory system of bees

Three different types of photoreceptors (i.e. short, medium and long wavelength sensitive) are simulated by the light reflected by the stimulus. The amount of light absorbed by these photoreceptors was estimated as

$$P_m = R\int_{300}^{700} \mathrm{I}_c(\lambda)S_m(\lambda)D(\lambda)d\lambda \qquad \text{Eq. 1}$$

where R describes the overall sensitivity of the receptor, $I_c(\lambda)$ is the spectral reflectance function of the stimulus, c, $S_m(\lambda)$ is the spectral sensitivity function of each receptor (short, medium and long), and $D(\lambda)$ is the illumination. Based on prior research on bees' photoreceptors, in this study we set perfectly uniform illumination ($D(\lambda) = 1$ at all λ) and the sensitivity factor $R = 6$.

Because of the reported non-linearity in the response of photoreceptors, the receptors' response to a colour c , $E_m(c)$ is calculated from the quantum catch, $E_m(c) = P_m/(P_m + 1)$.

Next, each photoreceptor sends an inhibitory signal to only one type of transmedullary cell in the medulla (TM). Hence the firing rate response of $m - th$ TM neurons is calculated with

$$r_m^{TM}(c) = r_0 + v_m E_m(c) \qquad \text{Eq. 2}$$

where r_0 is the high spontaneous activity of TM cells. v_m is the inhibitory synaptic weights of pre-synapse TM cells which take values between [-1, 0].

The TM cells send either excitatory or inhibitory signals to the third-order cells in the next layer (figure 3*a*) of the medulla through the matrix of synaptic weights $L = \left(L_{m,k}\right)$. The response of the $k - th$ third-order cells is modelled as

$$r_k^A(c) = F\left(\sum_{m=1}^{3} L_{m,k} r_m^{TM}(c) + b^A;\ \alpha, \beta\right) \qquad \text{Eq. 3}$$

where $F(x; \alpha, \beta) = A_0/\left(1 + e^{-\alpha(x-\beta)}\right)$ is the activation function and b^A represents the spontaneous activity of third-order neurons. The parameters α, β control the sensitivity of the third-order neurons to the total presynaptic input. Following Vasas et al [1], the synaptic weights of the these third-order cells, $\left(L_{m,k}\right)$, were fitted to the empirical data [2, 3] (figure S5) using a gradient descent algorithm as described in Vasas et al [1]. Previously, Vasas et al. [1] showed that in a similar model random connectivity between transmedullary neurons and third-

order neurons was sufficient to describe all the documented types of colour sensitive neurons in medulla: narrow or broad-band neurons that respond to a small or large range wavelength of the light and colour opponent neurons that respond to multiple ranges of light wavelengths with either excitation or inhibition [1].

b) Model of learning in the mushroom bodies

After fitting the third-order cells' responses to the empirical data, the firing rate of Kenyon cells (KC) of the mushroom bodies (figure 3*a*) is modelled as:

$$r_p^{KC}(c) = F\left(\sum_{k=1}^{22} W_{p,k,} r_k^A(c) - I_f\,;\ \alpha, \beta\right) \qquad \text{Eq. 4}$$

where $W_{p,k}$ is the synaptic connectivity between $k-th$ third order neurons and $p-th$ KC in the mushroom bodies. The outputs of all KC stimulate a single mushroom body output neuron, $D(c)$ through a vector of synaptic weights, $V = (v_1, v_2, \ldots v_i);\ i = 1{:}\,N_{kc}$. The activity of this neuron in response to the colour c is expressed by

$$r^D(c) = F\left(\sum_{m=1}^{N_{kc}} V_m . r_m^{KC}(c)\,;\ \alpha, \beta\right) \qquad \text{Eq. 5}$$

For simplicity, we set the same activation function, $F(x; a, b)$ for all firing rate models of third-order cells, KC and the decision neuron (Eqs. 3, 4 and 5) with fixed parameters: $A_0 = 100$ spike/ sec and $\alpha = 0.05$ and $\beta = 50$.

$I_{f=} \sum_{m=1}^{N_{kc}} Q_m . r_m^{KC}(c)$ is the input of the inhibitory feedback pathway to the KC that is obtained from the average activity of Kenyon cells. The synaptic weights Q_m control the sparseness of activity of the KC. Higher values of Q_m increase inhibitory input to the KCs (Eq. 4). This results in reduced or no activity in some KC. Hence, the population activity of KC becomes more sparse. Setting Q_m to 0 effectively models the system without any inhibitory feedback to the KC, which produces dense activity in the KC population.

We propose that the difference in responses of $D(c)$ to the rewarding and punishing stimuli must be increased during the training phase. Hence, the optimal synaptic weights, $V^{opt}$and W^{opt}, could be obtained from maximizing the cost function that represents the difference between the activity of $D(c)$ to all pairs of the positive (c_p) and negative stimuli (c_n)

$$\langle W, V\rangle^{opt} = \underset{\langle C_p, C_n\rangle}{argmax}\ [r^D(c_p) - r^D(c_n)]. \qquad \text{Eq. 6}$$

If we assume $E = r^D(c_p) - r^D(c_n)$ in the equation 6, as the cost function, we have

$$\frac{\partial E}{\partial V_i} = r_i^{KC}(c_p) f\left(\sum_{m=1}^{N_{kc}} V_m r_m^{KC}(c_p)\,; a, b\right) - r_i^{KC}(c_n) f\left(\sum_{m=1}^{N_{kc}} V_m r_m^{KC}(c_n)\,; a, b\right) \quad \text{Eq. 7}$$

and

$$\frac{\partial E}{\partial W_{i,j}} = V_j\, r_i^A(c_p) f\left(\sum_{k=1}^{22} W_{p,k,} r_k^A(c_p) - I_f; a, b\right) f\left(\sum_{m=1}^{N_{kc}} V_m r_m^{KC}(c_p); a, b\right) - V_j\, r_i^A(c_n) f\left(\sum_{k=1}^{22} W_{p,k,} r_k^A(c_n) - I_f\,; a, b\right) f\left(\sum_{m=1}^{N_{kc}} V_m r_m^{KC}(c_n)\,; a, b\right) \quad \text{Eq. 8}$$

where $f(.)$ is the derivation of the activation function F.

Finally, the reinforcement neuron (figure 3a) makes a reward- or punishment- modulated connection with $D(c)$ and KC. $\delta(t)$ presents the reinforcement signal that depends on whether a stimulus is paired with reward or punishment ($\delta(t) = 1$), or $\delta(t) = 0$ for when no reinforcement is presented with the stimulus. Hence, the synaptic weights of the last two layers, V and W might increase or decrease from stimuli presented at step t to $t + 1$ according to the following learning rules:

$$V_i^{t+1} = V_i^t + \eta\, \frac{\partial E}{\partial V_j}\, \delta(t) \quad \text{Eq. 9}$$

and

$$W_{i,j}^{t+1} = W_{i,j}^t + \eta\, \frac{\partial E}{\partial W_{i,j}}\, \delta(t) \quad \text{Eq.10}$$

where η is the time constant that controls that learning rate at which the weights change.

Power Analysis

To assess the effect of sample size on the outcomes of GLMM (table S4), a simulated data based on the bees' performance in the experimental tests was generated using a Monte Carlo algorithm. The number of observations within groups was artificially increased to 15 bees (60

bees in total) through the simulation (there were 5 observations at each group in the current data). The R Packages SIMR and LME4 were used to simulate and evaluate the extended data [4, 5].

References

[1] Vasas, V., Peng, F., MaBouDi, H. & Chittka, L. 2019 Randomly weighted receptor inputs can explain the large diversity of colour-coding neurons in the bee visual system. *Scientific Reports* **9**, 8330.
[2] Kien, J. & Menzel, R. 1977 Chromatic properties of interneurons in the optic lobes of the bee. I. Broad band neurons. *Journal of Comoparative Physiology* **113**, 35-53.
[3] Kien, J. & Menzel, R. 1977 Chromatic properties of interneurons in the optic lobes of the bee. II. Narrow band and colour opponent neurons. . *J Comp Physiol* **113**, 17-34.
[4] Green, P. & MacLeod, C.J. 2016 SIMR:anR package for power analysis of generalizedlinear mixed models by simulation. *Methods in Ecology and Evolution* **7**, 493-498.
[5] Bates, D., Mächler, M., Bolker, B. & Walker, S. 2014 Fitting Linear Mixed-Effects Models using lme4. (doi:https://arxiv.org/abs/1406.5823).

Supplementary Figures:

(a)

Stimuli	C1	C2	C3	C4	C5	test
Human colour						
RGB values (R, G, B)	(5, 255, 5)	(250, 120, 10)	(255, 190, 220)	(250, 250, 0)	(5, 5, 255)	(250, 250, 250)
Hexagon loci (x,y)	(0.47, 0.05)	(0.31, -0.09)	(0.30, 0.24)	(0.53, -0.09)	(0.22, 0.29)	(0.38, 0.22)

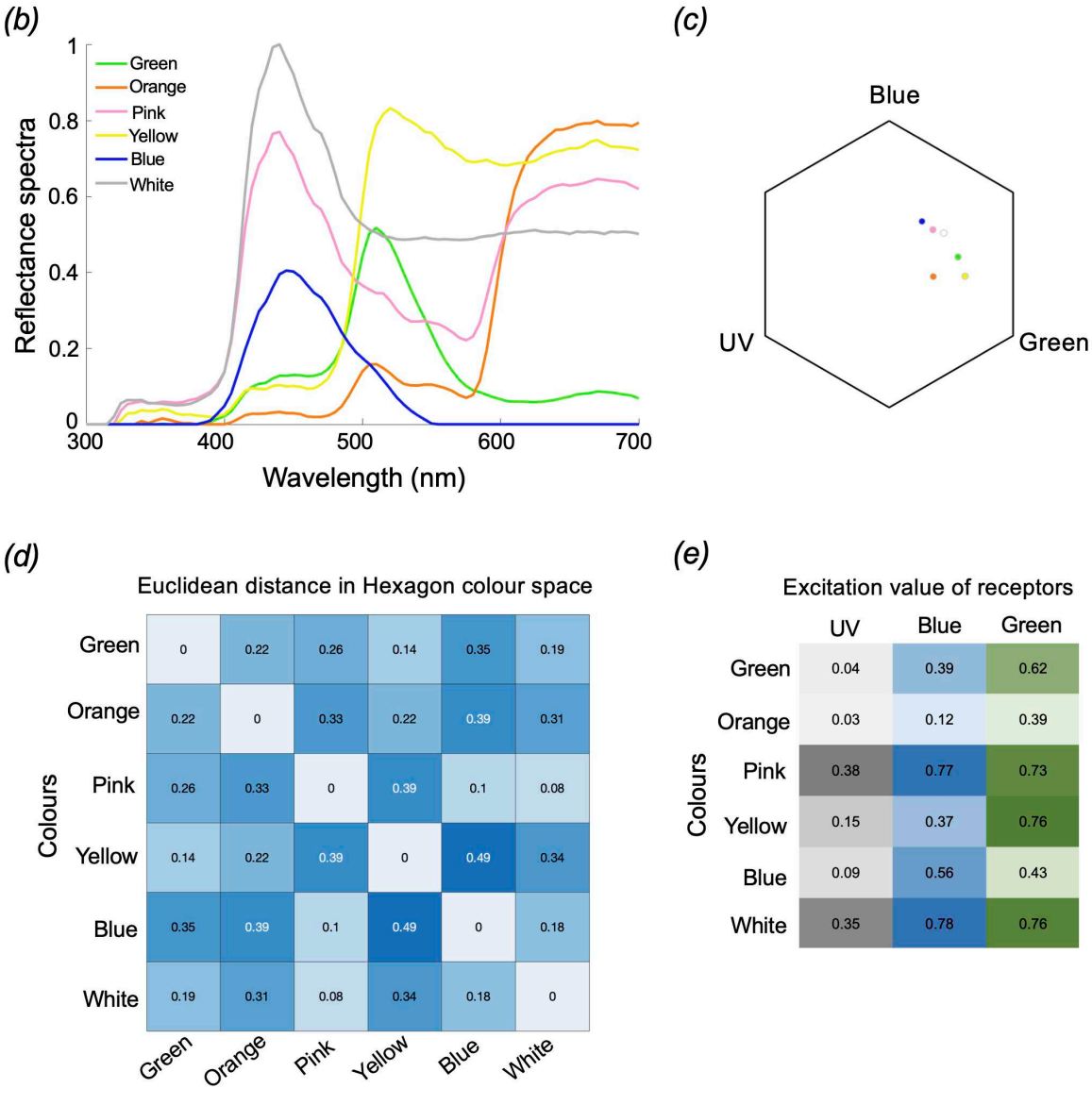


Figure S1. Colour information for stimuli used in this study. ***a)*** The stimuli used in this experiment with RGB values, and the hexagon loci in the honey bee visual space. ***b)*** Relative spectral reflectance plot of the stimuli used in the training phase and the tests. ***c)*** Loci of colours in a model of bee colour space, describing the range of colours a bee can see given their three photoreceptors maximally sensitive to UV, blue, and green wavelengths. Dots indicate each of

the stimuli coded by colour as perceived by humans. ***d)*** The Euclidean distance between each pair of coloured stimuli was calculated in the hexagon colour space. Smaller distance indicates more similarity between colours for a bee. ***e***) The activity of each honeybee receptor (UV, blue and green) evoked by each colour. Darker cell shadings indicate higher activation of the receptor.

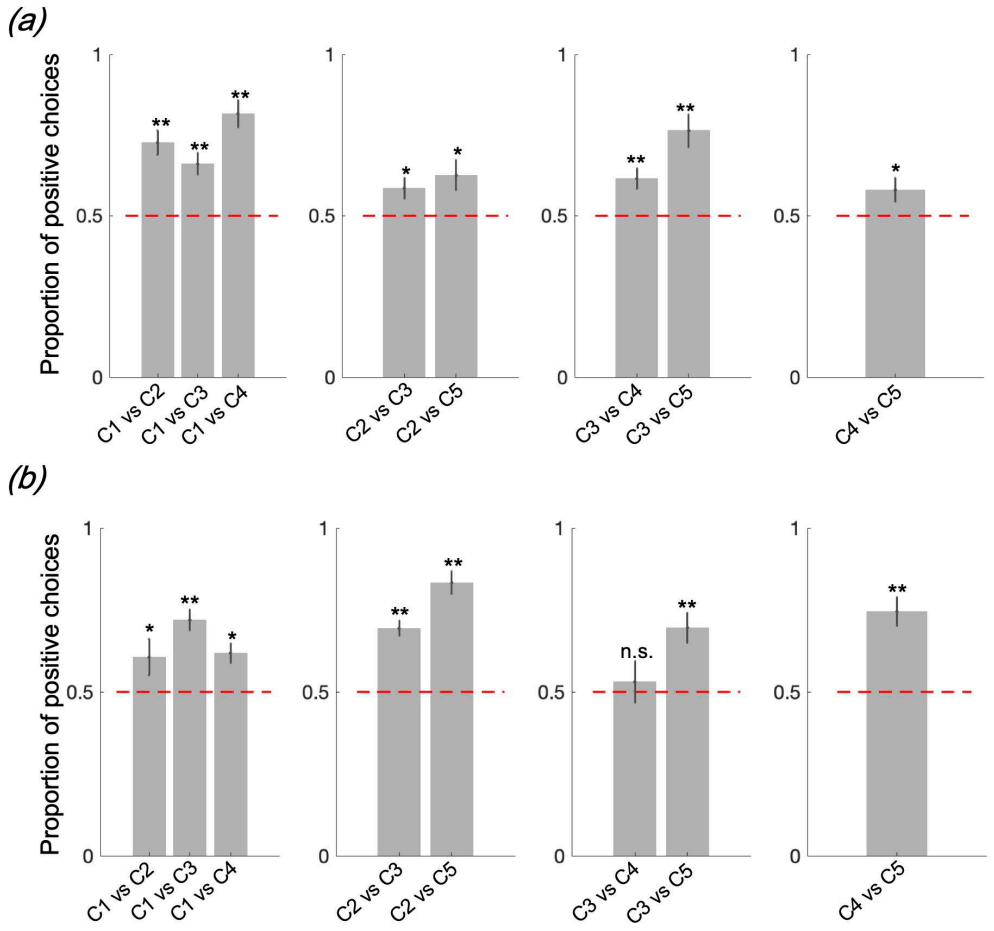


Figure S2. Bees' responses to different pairs of colours at the end of the training phase. Bees' responses to each pair of stimuli on the last presentation of each pair of colours during training. The responses of bees trained to protocol A and B are presented in panels ***a*** and ***b***, respectively. For group A (panel ***a***) the colours were assigned as: C1=blue, C2=yellow, C3=pink, C4= orange and C= green. For group B (panel **b**) the colour assignment was: C1=green, C2=orange, C3=white, C4=yellow and C5=blue. The left bar-graphs shows the proportion choosing C1 over colours C2 (at bout 10 in protocol P1 and at 17 in protocol P2), C3 (at bout 15 in protocol P1 and at 14 in protocol P2) and C4 (at bout 12 in protocol P1 and at 16 in protocol P2). The second left graphs shows the proportion selecting C2 over C3 (at bout 17 in protocol P1 and at 12 in protocol P2) and C5 (at bout 13 in protocol P1 and at 15 in protocol P2). The third graph shows the proportion choosing C3 over C4 (at bout 14 in protocol P1 and at 11 in protocol P2) and C5 (at bout 11 in protocol P1 and at 18 in protocol P2). Finally, the last bar-graph shows the performance of bees in choosing C4 over C5 at bout 16 in protocol P1 and at bout 18 in protocol P2. This figure confirms that the bees were able to discriminate between all pairs of colours used in the experiment. Bars show mean and standard error. Asterisks show the rewarding colours was chosen more than the chance level. * indicates $p < 0.05$ and ** indicates $p < 0.005$.

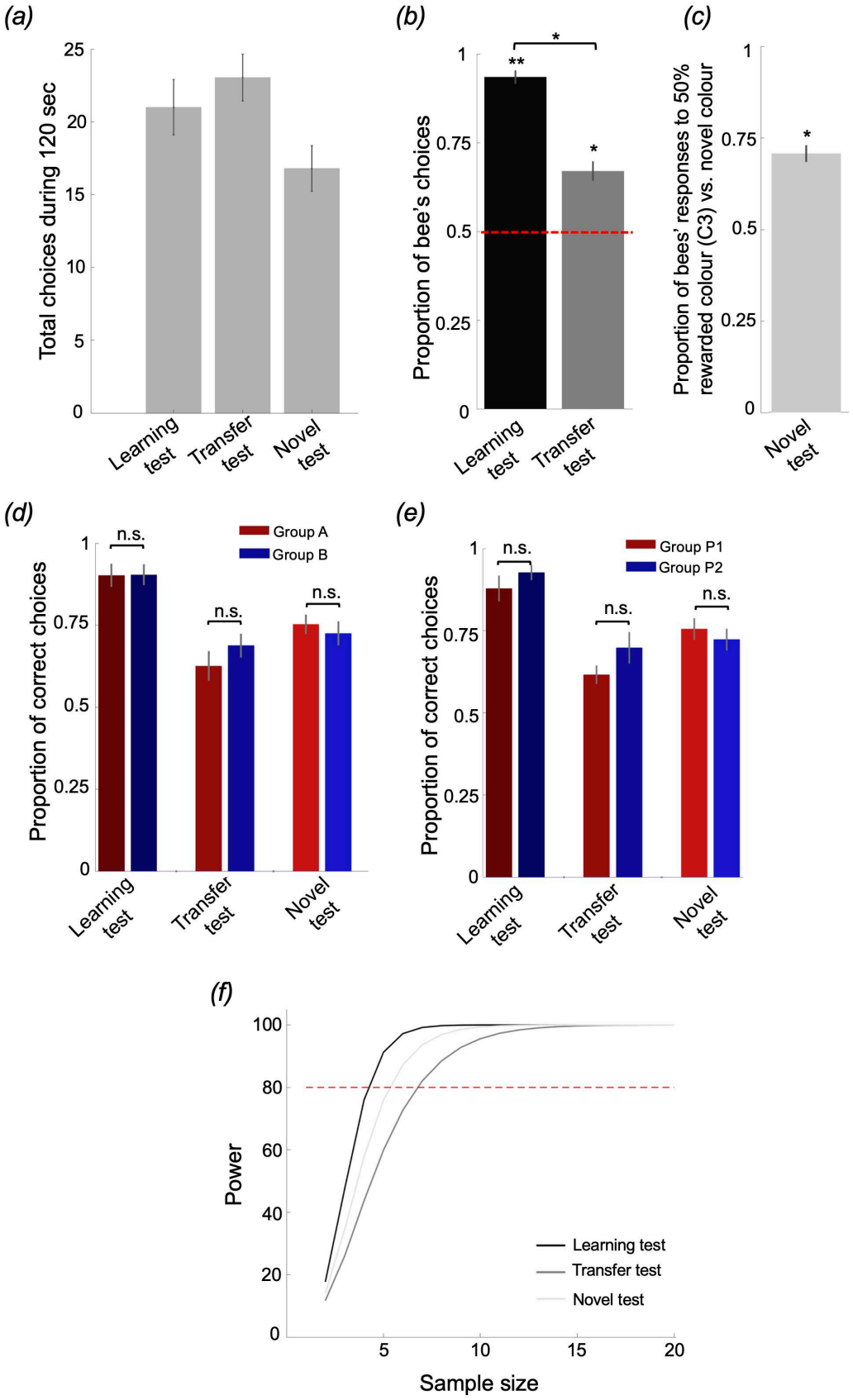


Figure S3. **Bees' performance in the tests.** ***a***) the bar-graph shows the total choices of the bees during the first 120 sec of the learning, transfer and novel tests. ***b, c)*** Bees' performance

calculated from the first 10 choices only of each bee in the learning test, transfer and novel test. Bars show mean frequency with which bees landed on stimuli they had experienced as more rewarded during training. Dashed line indicates performance expected at random. ***c)*** In the novel test, bees chose the trained colours (C3) more than the novel colour. Values for each individual bee are shown by small empty circles in different colours: Red: group A, Protocol P1; Green: group A, Protocol P2; Blue: group B, Protocol P1; Yellow: group B, Protocol P2. There was no significant difference between ***(d)*** groups of bees trained with different colours (A vs B) and ***(e)*** groups of bees trained with different order of colour presentation (P1 vs P2). Further GLM analyses confirm that bees' responses in the three tests did not differ between protocols or the colour sets used in the experiment (Tables *S1*). Vertical lines: s.e.m. * indicates $p<0.05$ and ** indicates $p<0.005$.***(f)*** Results of power analyses for each behavioural test assuming mean and variance in performance seen in this study. Our sample size was 20 and these power analyses suggest we had more than acceptable power (80% is a threshold power that is commonly accepted).

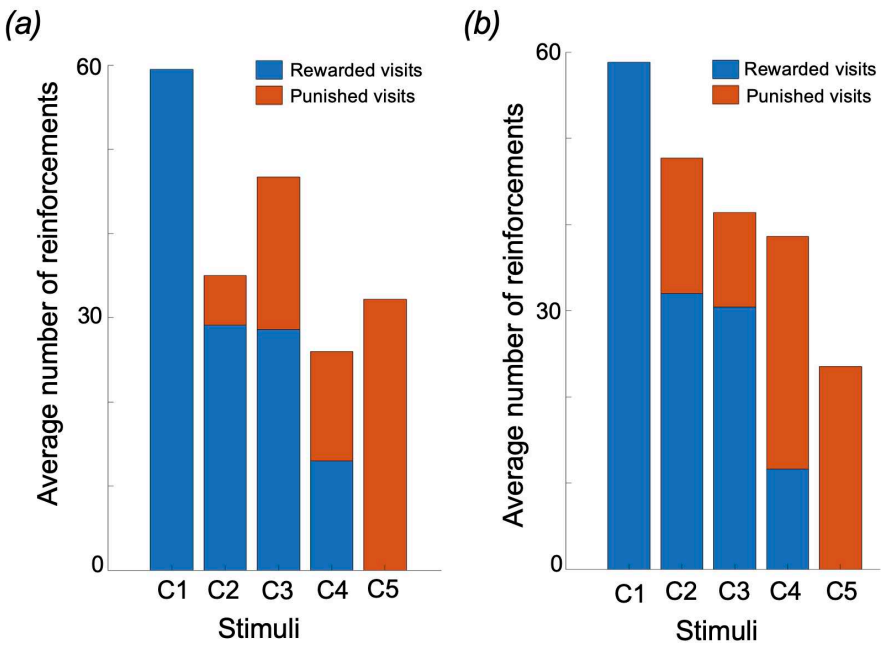


Figure S4. Number of visits to stimuli that were reinforced and punished in the training trials**.** The stacked bars show the average number of punished and rewarded visits that bees experienced from each colour in the training phase ***a***) Group A; ***b***) Group B).

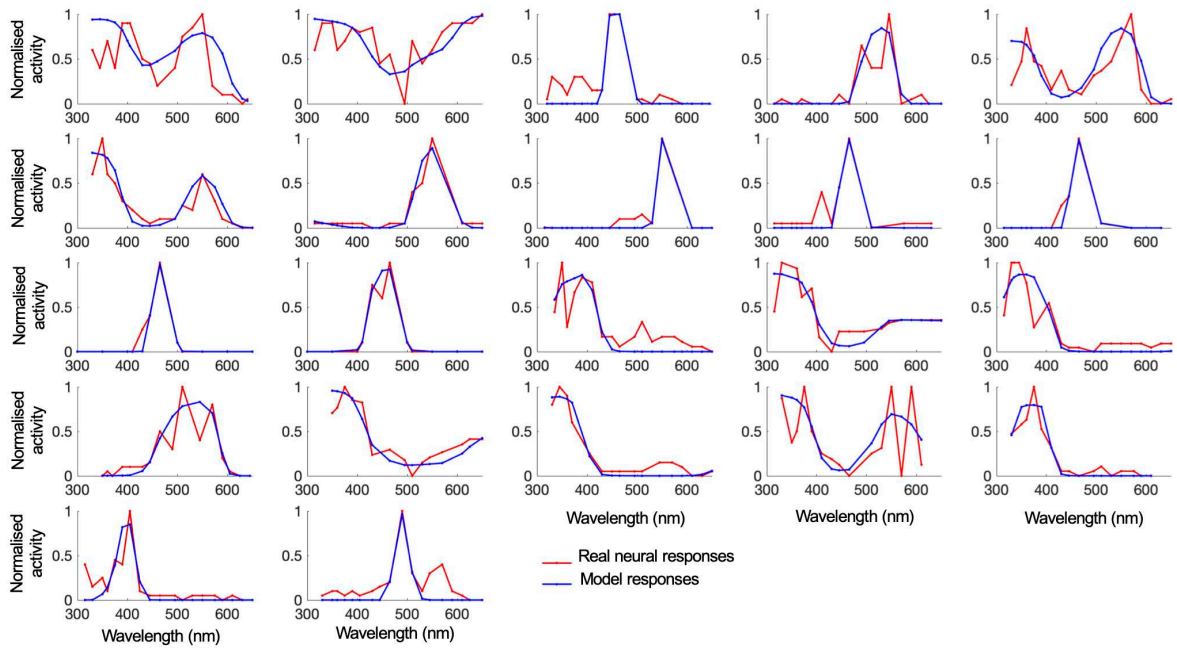


Figure S5. Diverse activities of medulla neurons generated by the model. Plots show the spectral tuning curves of third-order neurons. The red curves show the empirically measured spectral tuning curves reported in [2, 4]. The blue curves show the model responses of the third-order neurons in the medulla to different colours.

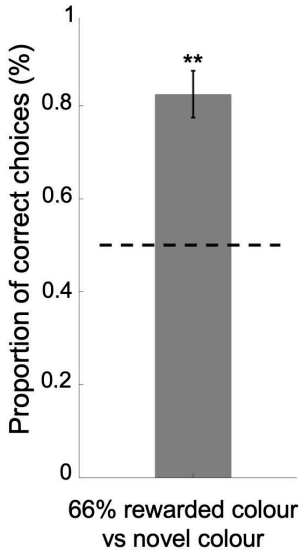


Figure S6. Model's performance in discriminating 66% rewarded colour from the novel colour. Mean ± SE of the model's performance in discriminating a colour rewarded 66% of time in training from a novel colour. Data generated from 50 different initial parameters that simulated 50 different model bees following training with protocol P1 or P2 (Table *2*).

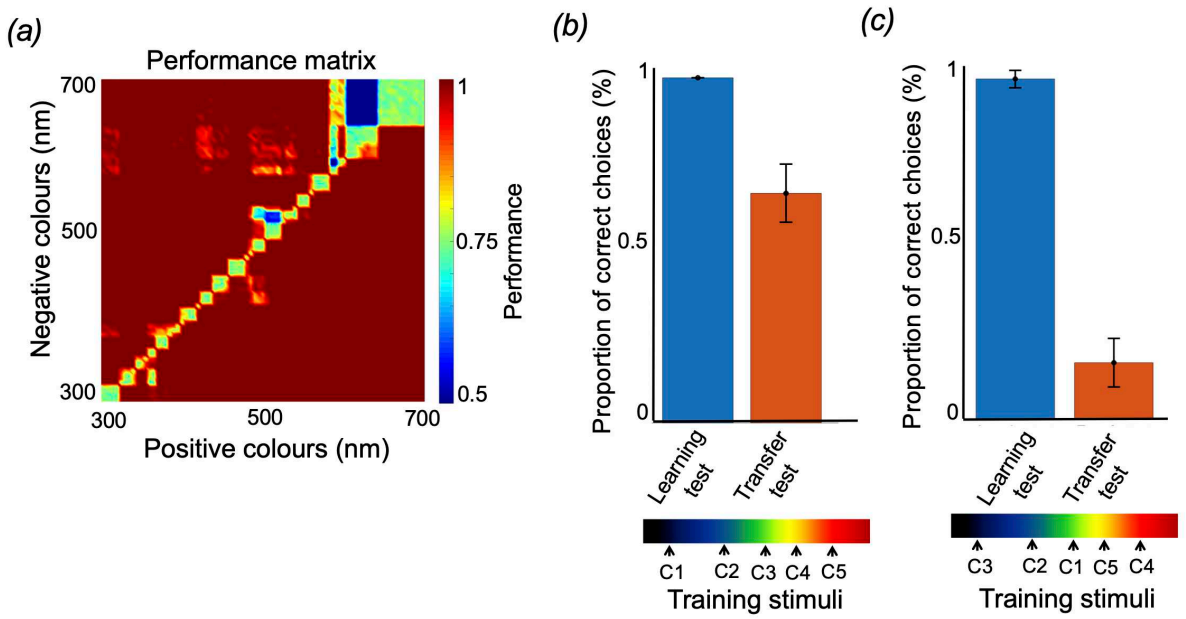


Figure S7. ***a*)** The performance of the model with low sparseness in the Kenyon cell population. The matrix shows the performance of the model in the colour discrimination task. The colour of each matrix element displays the performance of the model to two monochromatic colours whose wavelengths are presented on the x and y axes. The performance of the model in colour discrimination is reduced when the number of KCs activated by any colour is increased (compare with figure *3b*). ***b* & *c*)**. Performance of the model with connection plasticity in layer V only (figure 3). The model was trained in the history of reinforcement paradigm with two different colour sets, shown beneath panels b and c. This model with no plasticity in connection matrix W (figure 3) is not able to reproduce the performance of bees in the transfer test (compare with figures 1*c* and 3*c*). The performance of the model in the transfer test is sensitive to the set of training colors. The model chose a 66% rewarded colour more than a 66% punished colour if the five training colours were ordered from low wavelength to high wavelength by reinforcement history during training (***b***). However the model did not learn to prefer the 66% rewarded colour if training colours were randomly assigned a reinforcement history (***c***). Bars were calculated from the performance of the model for 50 different initial parameters each initial parameter provides the responses of a model bee.

Supplementary Tables

Stimulus	% trials rewarded	% trials punished	Colour Pairs
C1	100	0	C1 vs C2 C1 vs C3 C1 vs C4 C1 vs C5
C2	66	33	C2 vs C1 C2 vs C3 C2 vs C5
C3	50	50	C3 vs C1 C3 vs C2 C3 vs C4 C3 vs C5
C4	33	66	C4 vs C1 C4 vs C3 C4 vs C5
C5	0	100	C5 vs C1 C5 vs C2 C5 vs C3 C5 vs C4
(Novel colour)	0	0	…

Table S1. Summary of training trials. Each colour was paired with all others, except for pairing C4 with C2. This pairing was excluded from the training procedure to use as a novel pair in the transfer test. In each pairing the higher ranked colour (red) was rewarded and the lower ranked colour punished (blue).

Protocol P1

# Bouts	Stimuli at each bout
1	C1 vs C2
2	C3 vs C5
3	C1 vs C4
4	C2 vs C5
5	C3 vs C4
6	C1 vs C3
7	C4 vs C5
8	C2 vs C3
9	C1 vs C5
10	C1 vs C2
11	C3 vs C5
12	C1 vs C4
13	C2 vs C5
14	C3 vs C4
15	C1 vs C3
16	C4 vs C5
17	C2 vs C3
18	C1 vs C5

Protocol P2

# Bouts	Stimuli at each bout
1	C3 vs C5
2	C1 vs C2
3	C1 vs C4
4	C2 vs C5
5	C1 vs C3
6	C3 vs C4
7	C1 vs C5
8	C4 vs C5
9	C2 vs C3
10	C1 vs C5
11	C3 vs C4
12	C2 vs C3
13	C4 vs C5
14	C1 vs C3
15	C2 vs C5
16	C1 vs C4
17	C1 vs C2
18	C3 vs C5

Tables S2 & S3. Training sequences. Two different protocols for training were used. Each protocol was created by ensuring bees experienced each pair of colours in each half of the 18 training bouts. Otherwise colour pairings were ordered randomly. 10 bees were trained with P1, and 10 with P2.

Dependent variable	Fixed factors	Estimate	SE	tStat	DF	pValue
Bees's performance in the learning test	Intercept	0.0154	0.0025	6.1524	16	1.3909e-05
	Protocol	-0.0022	0.0015	-1.461	16	0.16
	Colour set	-0.0028	0.0015	-1.824	16	0.08
	Protocol:Colour set	0.0014	0.0009	1.5208	16	0.14
Bees's performance in the transfer test	Intercept	0.0183	0.0056	3.215	16	0.005
	Protocol	-0.00017	0.0036	-0.049	16	0.96
	Colour set	-1.28e-05	0.0036	-0.003	16	0.99
	Protocol:Colour set	-0.0010	0.0022	-0.449	16	0.65
Bees's performance in the novel test	Intercept	0.0108	0.0036	2.990	16	0.008
	Protocol	0.0012	0.0023	0.517	16	0.61
	Colour set	0.0012	0.0023	0.548	16	0.59
	Protocol:Colour set	-0.0004	0.0014	-0.314	16	0.75

Table S4. Summary of generalised liner mixed models (GLMM) of original data. GLMM exams the effect of the training protocol, the colour set and their interaction on bees' responses in the learning, transfer and novel tests. The dependent variables for each GLMM were the performance of the bees in the learning test, transfer test and novel tests (proportion of correct choices, figure S3). Bee index was included in the models as random factors.

Dependent variable	Fixed factors	Estimate	SE	tStat	DF	pValue
Bees's performance in the learning test	Intercept	0.0158	0.0027	5.777	16	7.62-09
	Protocol	-0.0017	0.0017	-1.036	16	0.30
	Colour set	-0.0029	0.0017	-1.674	16	0.09
	Protocol:Colour set	0.0011	0.0010	1.091	16	0.27
Bees's performance in the transfer test	Intercept	0.0162	0.0057	2.815	16	0.004
	Protocol	-0.0014	0.0036	-0.391	16	0.69
	Colour set	-0.0018	0.0036	-0.4949	16	0.62
	Protocol:Colour set	-0.0022	0.0023	-0.984	16	0.32
Bees's performance in the novel test	Intercept	0.0132	0.0039	3.338	16	0.008
	Protocol	-0.0002	0.0025	-0.082	16	0.93
	Colour set	-0.0001	0.0025	-0.053	16	0.95
	Protocol:Colour set	0.0006	0.0015	0.405	16	0.68

Table S5. Summary of generalised liner mixed models (GLMM) of simulated data. GLMM evaluates the effect of the training protocol, the colour set and their interaction on *simulated data* that was generated from the original bees' responses in the learning, transfer and novel tests assuming the sample size was 15 bees per groups per protocol (60 bees in total). The simulated data was obtained from Monte Carlo simulation based on the bees' performance in the test for more sample size (see supplementary methods). The dependent variables for each GLMM were the performance of the bees in the learning test, transfer test and novel tests (proportion of correct choices). Bee index was included in the models as random factors.

Bee index	Protocol	Group	learning test (%)	Transfer test (%)	Novel test (%)
1	P1	A	90	57	87
2	P1	A	77	56	89
3	P1	A	85	68	77
4	P1	A	100	62	71
5	P1	A	72	52	64
6	P1	B	84	51	65
7	P1	B	94	84	75
8	P1	B	100	86	77
9	P1	B	100	47	68
10	P1	B	100	63	80
11	P2	A	93	72	62
12	P2	A	100	66	73
13	P2	A	100	52	88
14	P2	A	93	75	71
15	P2	A	69	56	73
16	P2	B	95	88	80
17	P2	B	87	70	89
18	P2	B	86	73	63
19	P2	B	85	60	70
20	P2	B	96	76	56

Table S6. Bees' performance in the experiment tests. The table shows the responses of each individual bee to the learning, transfer and novel test (see figure 1).

Supplementary Video

Video S1: Sample video of honey bee in transfer test.