Supplementary information

Covariance based plasticity

The decision-making network is composed of two populations of Poisson neurons: each neuron is characterized by its firing rate and the spike count of a neuron in a trial (1 sec) is randomly drawn from a corresponding Poisson distribution. Action in a trial is determined by the total number of spikes in each population: the chosen action corresponds to the population that fires more spikes in a trial\(^1\).

At the end of the trial, the firing rate of each of the neurons (in the two population) is updated according to:

\[
f(t + 1) = f(t) + \eta \cdot R(t) \cdot (s(t) - f(t)) \quad (S1)
\]

where \(f(t)\) is the firing rate in trial \(t\), \(\eta = 0.07\) is the learning rate, \(R(t)\) is the reward delivered in trial \(t\) \((R(t) \in \{0,1\}\) in our simulations\) and \(s(t)\) is the measured (realized) firing rate in that trial, that is the spike count in the 1 sec trial. The initial firing rate of all simulated neurons is set to 2.5Hz. Because of the Poisson stochasticity in the spike count, the firing rate of each neuron follows a different trajectory.

Note that because on average, the empirical firing rate is equal to the true firing rate, \(f(t) = \langle s(t) \rangle\), changes in the firing rate are driven, on average, by the covariance of reward and the empirical firing rate: \(<\Delta f(t)> = \langle f(t + 1) - f(t) \rangle = \eta \cdot \text{cov}(R(t), s(t))^2\).

The network model was tested in the operant learning task of Fig. 1. A session was terminated (without further analysis) if the model was not able to choose the better option more than 14/20 times for at least 200 trials. This occurred on 20% of the sessions. Simulated neurons were
excluded due to low spike rate if the mean spike count was lower than 1 for all blocks. This occurred on 0.03% of the sessions.

The estimated Q-values in Fig. S1 were computed from the actions and rewards of the covariance model by assuming the Q-learning model (Eqs. 1 and 2), see Materials and Methods.

**Comparison with permuted spike counts**

In this section we considered the experiment and analysis described in\(^3\). That experiment consisted of four blocks, each associated with a different pair of reward probabilities, (0.72, 0.12), (0.12, 0.72), (0.21, 0.63) and (0.63, 0.21), appearing in a random order, with the better option changing location with each block change. The number of trials in a block was preset, ranging between 35 and 45 with a mean of 40 (this is unlike the experiment described in Fig. 1, in which termination of a block depended on performance).

First, we used Eqs. (1) and (2) to model learning behavior in this protocol. Then, we estimated the Q-values according to choice and reward sequences, as in Fig. 1. These estimated Q-values were used for regression of the spike counts of the random-walk, motor cortex, auditory cortex, and basal ganglia neurons in the following way: each spike count sequence was randomly assigned to a particular estimation of a pair of Q-values from one session. The spike count sequence was regressed on these estimated Q-values. The resultant t-values were compared with the t-values of 1,000 regressions of the spike-count, permuted within each block, on the same Q-values. The p-value of this analysis was computed as the percentage of t-values from the permuted spike-counts that were higher in absolute value than the t-value from the regression of the original spike count. The significance boundary was set at p<0.025\(^3\). Neurons with at least one significant regression
coefficient (rather than exactly one significant regression coefficient) were classified as action-value modulated neurons\textsuperscript{3}.

The results of this analysis are depicted in Fig. S3 for the random-walk (a), motor cortex (b) auditory cortex (c) and basal ganglia (d) neurons. Top left depicts the t-values from regressions of the original spike-count on the estimated Q-values, where green triangles denote those neurons that were significant in the permutation analysis. Bottom right histogram denotes the fraction of neurons that passed the permutation test. For all four groups, the number of action-value neurons was larger than expected by chance.
Supplementary figures

**Figure S1** Covariance based plasticity model

(a) An example of operant learning of the covariance model in the operant task of Fig. 1a. Legend is the same as in Figure 1a. (b) Two example covariance based plasticity neurons from the experimental session in (a). The estimated Q-values (red and blue) were computed from the choices and rewards in (a) (see Materials and Methods). Each estimated Q-value appears with the simulated neuron which had a significant regression coefficient on it. Legend is the same as in Fig 2a. (c) and (d) Population analysis. Same as in Figs. 1d and 1e. The two simulated neurons in (b) are denoted by squares in (c). Estimation in (d) is based on 500 sessions with 2,000 simulated neurons in a session. Legend is the same as in Fig. 1d and 1e, respectively. Error bars denote standard error of the mean.
Figure S2 Regression on reward probabilities. For each of the four data-sets (a) random-walk neurons, (b) motor cortex neurons (c) auditory cortex neurons and (d) basal ganglia neurons, spike counts in the last 20 trials in each block were regressed on reward probabilities in those blocks. Top-left of each panel denotes the t-values of the regressions of individual neurons and bottom right histograms denote the population statistics. Number of neurons used in (a), (b) and (c) and (d) is the same as in Figs. 2, 3, 4 and 5, respectively. Legend is the same as in Fig. 1d and 1E, respectively. Note that for this analysis we considered significance using the threshold of $p<0.05$. By contrast, in the same analysis was used with a significance threshold of $p<0.01$. For comparison, when considering the basal ganglia neurons with a significance threshold of $p<0.01$, the number of neurons that are erroneously classified as action-value neurons decreases from $37\%\pm3.3\%$ to $26\%\pm3\%$. 
Figure S3 Detrending analysis. Following, we conducted a multiple linear regression analysis using the following regression model:

\[ s(t) = \beta_0 + \beta_1 Q_1(t) + \beta_2 Q_2(t) + \beta_3 t + \beta_4 C(t) + \beta_5 C(t-1) + \beta_6 R(t) + \beta_7 R(t-1) + \epsilon(t) \]

Where \( s(t) \) is the spike count in trial \( t \), \( Q_1(t) \) and \( Q_2(t) \) are the estimated action-values in trial \( t \), \( C(t) \) and \( C(t-1) \) are the actions chosen in trial \( t \) and \( t-1 \), respectively, \( R(t) \) and \( R(t-1) \) are the rewards in trial \( t \) and \( t-1 \), respectively, \( \epsilon(t) \) is the residual error in trial \( t \) and \( \beta_0-\beta_7 \) are the regression parameters. (a), (b), (c) and (d) denote the random-walk neurons, motor cortex neurons, auditory cortex neurons and basal ganglia neurons, respectively. Top-left, t-values from regressions of the spike-count on the regression variables. As in, the significance boundaries for the t-values are 2.64, corresponding to \( p<0.01 \) (as opposed to \( p<0.05 \) elsewhere). Bottom right histograms denote the population statistics. Note, however, that the significance criterion is more stringent and the expected total number of identified action-value neurons by chance is only 2%. Number of neurons used in (a), (b) (c) and (d) is the same as in Figs. 2, 3, 4 and 5, respectively. Legend is the same as in Fig. 1d and 1e, respectively.
Figure S4 Unbiased identification of action-value neurons. Following\textsuperscript{6}, we considered an unbiased identification of action-value neurons. (a), (b) (c) and (d) denote the random-walk neurons, motor cortex neurons, auditory cortex neurons and basal ganglia neurons, respectively. The t-values for the different neurons are identical to Fig. S2. The f-value of each neuron was computed from the regression and a neuron was considered as non-significant (black dot) if $f > 0.01$, denoted by the circle in Top-left panels. For the significant neurons, the dashed lines define 8 equal-angle sectors, each corresponding to a different classification of the neuron. Bottom-right is the population analysis. Note that the expected total number of identified significant neurons by chance is only 1%. Number of neurons used in (a), (b) (c) and (d) is the same as in Figs. 2, 3, 4 and 5, respectively. Legend is the same as in Fig. 1d and 1e, respectively.
Figure S5 Comparison with permuted spike counts, following\(^3\) (a), (b) (c) and (d) denote the random-walk neurons, motor cortex neurons, auditory cortex neurons and basal ganglia neurons, respectively. Top-left, t-values from regressions of the original spike-count on the estimated Q-values. Green triangles, significant modulation by action-value according to the permuted spike-count analysis. Bottom-right, fraction of neurons significantly modulated by action-value across the population (0.05, expected by chance, is marked by a horizontal dashed line). Error bars are standard error of the mean. Number of neurons used in (a), (b) (c) and (d) is the same as in Figs. 2, 3, 4 and 5, respectively. Legend is the same as in Fig. 1d and 1e, respectively. Note that in all four cases the two t-values are correlated. This results from the correlation between \(Q_1(t)\) and \(Q_2(t)\) caused by the reward schedule in\(^3\).
Figure S6 Analysis with autoregressive coefficients. Following⁷, we conducted a multiple linear regression analysis using the following regression model:

\[ s(t) = \beta_0 + \beta_1 Q_1(t) + \beta_2 Q_2(t) + \beta_3 C(t) + \beta_4 R(t) + \beta_5 C(t) \cdot R(t) + \beta_6 CV(t) + \beta_7 s(t - 1) + \beta_8 s(t - 2) + \beta_9 s(t - 3) + \epsilon(t) \]

Where \( s(t) \) is the spike count in trial \( t \), \( Q_1(t) \) and \( Q_2(t) \) are the estimated action-values in trial \( t \), \( C(t) \) is the action chosen in trial \( t \), \( R(t) \) is the reward in trial \( t \), \( C(t) \cdot R(t) \) is the interaction between choice and reward in trial \( t \), where both are expressed as binary with values \{-1, 1\}, \( CV(t) \) is the value of the action that was chosen on trial \( t \), \( s(t - 1), s(t - 2), s(t - 3) \) are the spike counts one, two and three trials prior to current trial, respectively, \( \epsilon(t) \) is the residual error in trial \( t \) and \( \beta_{0-9} \) are the regression parameters. \((a), (b), (c)\) and \((d)\) denote the random-walk neurons, motor cortex neurons, auditory cortex neurons and basal ganglia neurons, respectively. Top-left, t-values from regressions of the spike-count on the regression variables. The significance boundaries for the t-values are 2.3, corresponding to \( p<0.025 \). Bottom-right, fraction of neurons significantly modulated by action-value across the population (0.05, expected by chance, is marked by a horizontal dashed line). Error bars are standard error of the mean. Number of neurons used in \((a), (b), (c)\) and \((d)\) is the same as in Figs. 2, 3, 4 and 5, respectively. Legend is the same as in Fig. 1d and 1e, respectively. Note that in all four cases the two t-values are correlated. This results from the correlation between \( Q_1(t) \) and \( Q_2(t) \) caused by the reward schedule in⁷.
Figure S7 Erroneous identification of policy neurons as action-value neurons. To study the potential erroneous identification of policy neurons as action-value neurons, we considered Poisson policy neurons, neurons whose firing rate is fully determined by the choice: in trials, in which the animal chose ‘left’ and ‘right’ their firing rate was 2 Hz and 1 Hz, respectively. When regressing the spike count of these choice neurons (1000 neurons×52 sessions) on the estimated Q-values we find that such neurons are more likely to be classified as action-value neurons than as policy neurons (light red). This is likely due to the bias in the analysis identified by6. The permutation analysis decreases the number of policy neurons identified as action-value neurons (dark red) but a substantial fraction of these policy neurons is still erroneously identified as action-value neurons. Error bars are SEM.
References


