Published in final edited form as:

Nat Neurosci. 2009 May ; 12(5): 655–663. doi:10.1038/nn.2304.

Reinforcement learning can account for associative and perceptual learning on a visual decision task

Chi-Tat Law and Joshua I. Gold

Department of Neuroscience, 116 Johnson Pavilion, 3610 Hamilton Walk, University of Pennsylvania 19104-6074

Abstract

We recently showed that improved perceptual performance on a visual motion direction-discrimination task corresponds to changes not in how sensory information is represented in the brain but rather how that information is interpreted to form a decision that guides behaviour. Here we show that these changes can be accounted for using a reinforcement learning rule to shape functional connectivity between the sensory and decision neurons. We modelled performance based on the readout of simulated responses of direction-selective sensory neurons in the middle temporal area (MT) of monkey cortex. A reward prediction error guided changes in connections between these sensory neurons and the decision process, first establishing the association between motion direction and response direction and then gradually improving perceptual sensitivity by selectively strengthening the connections from the most sensitive neurons in the sensory population. The results suggest a common, feedback-driven mechanism for some forms of associative and perceptual learning.

Introduction

Perceptual sensitivity to simple sensory stimuli can improve with training ¹. Despite the prevalence of this phenomenon, known as perceptual learning, its neural basis remains incompletely understood, especially for vision. Improvements in sensitivity to simple visual stimuli can be accompanied by changes at various stages of visual processing in the brain, from primary visual cortex to higher-order sensory-motor areas ²⁻⁴. However, little is known about the mechanisms that drive these changes. Here we present and explore the hypothesis that some of these changes can be driven by a reinforcement signal that generates a selective readout of the most informative sensory neurons to form a perceptual judgment that guides behaviour.

Reinforcement learning describes how to learn by trial and error to act in a manner that maximizes reward and minimizes punishment ⁵. One popular form of reinforcement learning uses a reward-prediction error to drive learning. This error signal, which compares predicted and actual rewards, can help to form associations between sensory input and

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Author Contributions

CL and JIG planned the study and wrote the manuscript together. CL implemented the model.

rewarded actions and is thought to be reflected in the phasic activity of midbrain dopamine neurons 6, 7 (but see refs. 8, 9). A similar dopamine-based signal can help to drive changes in auditory cortex, and some forms of visual perceptual learning seem to require a reinforcement signal 10, 11. However, how such signals relate to the neural changes responsible for enhanced perceptual sensitivity are not known.

We examined whether a reinforcement-learning rule based on a reward-prediction error could account for both associative and perceptual learning on a direction-discrimination task. Previously we trained monkeys to decide the direction of random-dot motion and respond with a saccadic eye movement 4. The monkeys first learned the association between motion direction and saccadic response and then gradually learned to make accurate direction decisions using weaker motion stimuli. These behavioural improvements corresponded to changes in the motion-driven responses of neurons in the lateral intraparietal area (LIP), which reflects aspects of sensory, motor, cognitive, and reward processing including formation of the direction decision 12-17. In contrast, there was no apparent change in the motion-driven responses of neurons in the middle temporal area (MT), which encodes motion information used to solve the task 18, 19.

We propose that reinforcement signals first help to establish functional connections from a population of MT-like sensory neurons to a population of LIP-like decision neurons that interpret the sensory information to determine the saccadic response. The same mechanisms then further refine these connections to weigh more strongly inputs from the most informative sensory neurons, thereby improving perceptual sensitivity. We show that this model can explain the time course and asymptotic behaviour of both associative and perceptual learning, changes related to the readout of the sensory representation, and the establishment and progression of motion-sensitive responses in neurons that form the decision. The results suggest that reinforcement learning might play a general role in both establishing and shaping patterns of connectivity critical for forming perceptual decisions.

Results

A decision-making, reinforcement learning model

The model is based in part on a pooling scheme that relates the activity of a population of MT-like neurons to a perceptual decision about motion direction, and in part on a reinforcement learning rule that evaluates and adjusts this pooling process (which for weak stimuli represents a “partially observable” or “hidden” state) at the end of each trial according to the reward outcome (Fig. 1) 20-22.

The pooling model is composed of three stages that correspond to the sensory representation in MT, the pooling of the MT responses, and the formation of the direction decision based on the pooled response in LIP (Fig. 1, orange arrows; see Supplemental Fig. 1 for alternative implementations). We modelled MT as a population of 7200 neurons with 36 different direction tuning functions distributed uniformly around the full 360°, with trial-by-trial responses to the motion stimulus and interneuronal correlations based on previously measured values 4, 23, 24. The cumulative, motion-driven responses from each MT neuron (x_i) were then pooled by LIP as a weighted sum:

$$y = \sum w_i x_i, \quad [1]$$

where w_i is the pooling weight assigned to the i^{th} MT neuron. Like previous studies, we assumed that this pooling process is not perfect and corrupted the pooled response, y , with Gaussian noise 21. The direction decision was made based on the arithmetic sign of the pooled response: rightward (i.e., between -90 and 90° , as in Fig. 3) for >0 , leftward otherwise.

The pooling weights from MT to LIP were random initially and then adjusted according to a reinforcement-learning rule after each trial (Fig. 1, green arrows). We used a simple delta rule that adjusts the pooling weights on trial $n+1$ (w_{n+1}) based on the weights on trial n (w_n) and an update term (Δw) that depends on a reward prediction error 20, 25, 26,

$$\mathbf{w}_{n+1} \leftarrow \mathbf{w}_n + \Delta \mathbf{w} \quad [2]$$

$$\Delta \mathbf{w} = \alpha \mathbf{C} (r - mE[r]) (\mathbf{x} - nE[\mathbf{x}]) \quad [3]$$

where α is the learning rate (which, in principle, could change with time or differ for correct and error trials, although we do not do so here), C describes the choice (-1 for leftward, 1 for rightward), r is the reward outcome for the current trial (1 if a reward was given, 0 otherwise), $E[r]$ is the predicted probability of making a correct (rewarded) choice given the pooled response (y) for that trial, $\mathbf{x}$ is the vector of MT responses, $E[\mathbf{x}]$ is the vector of baseline MT responses (which we modelled as the responses to 0% coherence motion; the population mean could also be used), and m and n are binary variables that determine the exact form of the rule used (in the main text we use $m=1$ and $n=0$; for alternatives see Table 1 and Supplemental Fig. 2). The difference between actual and predicted reward is called the reward prediction error and is equal to $r - E[r]$. In general, weight adjustments are based on the correlation between the reward prediction error and $\mathbf{x}$. C determines the sign of the adjustments. Neurons that respond more strongly to rightward motion will tend to have more positive weights because the reward prediction error tends to be positive on correct, rightward-choice trials ($C(r - E[r])=1 - E[r]$) and incorrect, leftward-choice trials ($C(r - E[r])=-E[r]$). Conversely, neurons that respond more strongly to leftward motion will tend to have more negative weights.

This formulation assumes that a single LIP neuron (or pool of neurons) forms the decision variable, the sign of which determines choice. Alternatively, one can consider two pools of decision neurons corresponding to each of the two choices 27. In this case, the decision variable is computed as the difference between the two pooled responses (with independent noise sources), and Eq. 3 is used to update two separate sets of weights (see **Supplemental Methods**). This formulation gives comparable results to the single decision pool (Supplemental Fig. 3), and therefore we use the simpler scheme here, which allows us to present and analyze a single set of weights for each simulation.

The term $E[r]$ can be thought of as an estimate of the subject's confidence at the time of the decision that the decision is correct and a reward will be obtained 28. We computed this

estimate using the pooled response y : the sign of y determined the decision, whereas the absolute magnitude of y reflected confidence in that decision. Specifically, y was modelled after LIP and therefore assumed to be proportional to the log-odds ratio that rightward is the correct choice, given the sensory evidence x and equal priors 17, 29, 30. Thus, the estimated probability of a correct decision was computed as $1/(1+e^{-\beta|y|})$, where β is a proportionality constant that can be estimated using a sequential estimation technique (or directly from the recent history of choices; see Supplemental Fig. 2) 31.

After each update, we normalized $\mathbf{w}$ by $(\sum w_i^2/w_{amp})^{1/2}$ to keep the vector length of $\mathbf{w}$ constant throughout training (because $\sum w_i^2 = w_{amp}$). This weight normalization, which is thought to be a common feature of cortical plasticity 32, enhanced the stability of the model by preventing any of the pooling weights from growing without bound. Moreover, it prevented the model from learning by indiscriminately increasing the overall magnitude of the weights and therefore the magnitude of the pooled response (y in Eq. 1) to overcome the effects of the decision noise. The value of w_{amp} was chosen to give a pooled response similar to responses of LIP neurons in trained monkeys (Supplemental Fig. 9e).

We simulated trial-by-trial performance of the model using the exact sequences of stimulus conditions (i.e., motion directions, coherences, and viewing durations for each monkey) used for training in our previous study 4. We then compared the model to the real behavioural, MT, and LIP data.

Comparison of the model and behavioural data

The monkeys learned from experience to associate a given direction of motion with a particular eye-movement response. We quantified this form of associative learning using the lapse rate (errors for high-coherence stimuli; Fig. 2a,b, gray symbols). We assume that these errors reflect, in part, incomplete knowledge of the sensory-motor association but not limitations in perceptual processing, because the high-coherence stimuli are easily discriminable even for naïve subjects, or failure to perform the general task requirements, because this analysis included only trials in which the monkey maintained fixation during motion viewing and then responded with an appropriate eye movement (87.4% of all trials for monkey C, 88.7% for Z). For both monkeys, the lapse rate declined rapidly over the first week or so of training, quickly reaching an asymptotic value of near zero (Fig. 2a).

The monkeys also became increasingly sensitive to weak motion signals over many months of training. We quantified these improvements in perceptual sensitivity using the discrimination threshold (the motion coherence corresponding to ~81% correct responses at a viewing duration of 1 sec; Fig. 2a,b, black symbols). Threshold was measured concurrently with lapse rate for all but the earliest training sessions, when performance was poor and only high-coherence stimuli were used. Thresholds decreased gradually, starting when low-coherence stimuli were introduced and continuing well after the monkeys had acquired the visuomotor association (i.e., near-zero lapse rates; Fig. 2b) 4.

The model can account for both the associative and perceptual changes. Using the same sequences of trials experienced by the monkeys, we used the model to generate a simulated sequence of choices during learning. We then computed the lapse rate (Fig. 2b, gray

symbols, using 250-trial blocks) and discrimination threshold (Fig. 2b, black symbols, using 1,000-trial blocks) from the simulated behavioural data. Like for the monkey data, lapse rates declined rapidly to near zero (time constants, τ_{la} , of exponential fits to the behavioural lapse rate, mean [68% confidence intervals, or CI] in number of trials for monkey C: 1,082 [1,080 1,175] for real data, 2,532 [1,839 3,658] for simulated data; monkey Z: 9,205 [9,139 9,564] for real data, 6,132 [6,000 6,539] for simulated data). Discrimination threshold improved more gradually, eventually reaching lower asymptotes comparable to those reached by the monkeys (time constants, τ_{th} , of exponential fits to behavioural threshold, mean [68% CI] in number of trials for monkey C: 28,317 [26,945 29,839] for real data and 26,841 [26,170 27,326] for simulated data; monkey Z: 18,422 [17,339 20,500] for real data, 22,467 [22,016 22,620] for simulated data). The lower asymptotes were also similar (mean [68% CI] in percent coherence for monkey C: 13.2 [12.7 13.7] for real data and 11.8 [11.7 12.0] for simulated data; monkey Z: 21.5 [19.7 22.5] for real data and 21.0 [20.3 21.5] for simulated data).

The time course of learning for the model depended critically on the learning rate (α in Eq. 2). We simulated performance of the model for both monkeys for a range of learning rates (α ,) and estimated τ_{la} and τ_{th} from exponential fits. Both parameters decreased as the learning rate increased (Fig. 2c, black symbols), indicating that simulated performance improved more rapidly with training using larger learning rates. The values of τ_{la} and τ_{th} estimated from the monkeys' behavioural data (solid lines in Fig. 2a; asterisks in Fig. 2c) were consistent with the relationship between τ_{la} and τ_{th} of the model for a particular learning rate for monkey C (linear regression, $H_0: [\tau_{la}, \tau_{th}]_{\text{monkey}} = [\tau_{la}, \tau_{th}]_{\text{model}}, p=0.56$). The match was not as good for monkey Z ($p<0.05$), whose lapse rate declined more slowly than for any of the other monkeys we have trained on this task 4, 33, possibly reflecting factors other than knowledge of the sensory-motor association like distractibility (e.g., monkey Z took longer to attain fixation and broke fixation more often than monkey C during the first ~50 training sessions; see Supplementary Table 1 in ref. 4). Below we use the learning rate that provides the best-matching τ_{la} and τ_{th} between the model and the monkeys (Fig. 2b; $\alpha = 7 \times 10^{-7}$ for monkey C, and $\alpha = 1 \times 10^{-7}$ for monkey Z).

These results were robust to a variety of pooling schemes and reinforcement rules used in the model (Table 1 and **Supplemental Figs. 1 and 2**). In brief, similar results were found using one or two pools of decision neurons; additive, multiplicative, or both kinds of pooling noise; multiplicative, subtractive, or no normalization of the linear weights; and linear or non-linear pooling. Likewise, similar results were found using different learning rules, as long as they were based on a correlation between sensory input and a reward prediction error. A qualitatively different scheme, in which pooling weights remained constant but a decision bound on the pooled signal varied with training, was unable to reproduce the pattern of behavioural results.

Changes in pooling weights optimize behaviour

The improvements in simulated discrimination performance resulted from changes in the pooling weights (Fig. 3a). The association between motion direction and decision direction was established early in training, as lapse rates fell to near zero but performance to weaker

motion signals was still near chance (Fig. 2b, 1st arrow). At this early stage, pooling weights tended to be strongest but of opposite signs near the two directions of motion used in the discrimination task (Fig. 3a, second panel). As training progressed and simulated performance improved (Fig. 2b, 2nd and 3rd arrows), the pooling weights continued to evolve such that weights to more sensitive neurons tuned to $\sim 0^\circ$ became more positive, and weights to more sensitive neurons tuned to $\sim 180^\circ$ became more negative (Fig. 3a, third and fourth panels). Thus, the improvements in sensitivity to weak motion appeared to result from an increasingly selective readout of the more sensitive sensory neurons with training.

The learning rule (Eq. 2) appeared to guide the pooling weights to a form of optimal linear readout at the end of training. We computed the optimal pooling weights for the population of sensory neurons in our simulations using Fisher's linear discriminant analysis, which maximizes percent correct for the discrimination task for any linear decoder without knowledge about the correlation structure in MT (Fig. 3b; see also **Supplemental Methods** and Supplemental Fig. 4). We compared the pooling weights of the model (Fig. 3c; gray symbols) to the optimal pooling weights (red lines) for neurons with two different levels of sensitivity (top row, mean neurometric threshold = 8.7% coherence; bottom row, 16.5% coherence). In both cases, the pooling weights approached the forms predicted by the optimal scheme, with the largest absolute weights assigned to neurons tuned to the two discriminated directions of motion and then falling off gradually as a function of direction tuning and motion sensitivity. These optimal weights define a decision boundary that most effectively separates the high-dimensional, population MT responses for leftward versus rightward motion (Supplemental Fig. 5)

Relationship with choice probability in MT

Choice probability is a measure of the relationship between trial-by-trial fluctuations in the activity of individual neurons and choice behaviour 34, 35. For neurons in area MT, choice probability tends to be near chance early in training and then progresses steadily to values slightly but reliably above chance after training, consistent with a noisy contribution to the direction decision 4, 21, 35. The model shows a similar, steady increase of choice probability for neurons tuned to the two directions of motion with training, as the pooling weights of these sensory neurons are adjusted to drive the decision process more effectively (Fig. 4).

However, the changes in pooling weights alone were not sufficient to account for another key feature of the real MT choice probability data: a selective increase with training for the most sensitive neurons 4. As noted previously, the strength of readout (here implemented as the pooling weights) only partially determines choice probability. Another key factor is interneuronal correlations, which must be present for the activity of any one neuron in the population to be predictive of behaviour 21. We examined how choice probability was affected by the structure of interneuronal correlations, which we assumed remained static throughout training but could depend on the similarity of response properties of pairs of sensory neurons 4, 36, 37. These assumptions had little effect on the pooling weights learned by the model (Supplemental Fig. 4) but substantial effects on choice probability. When correlation strength depended on the similarity of direction tuning but not sensitivity

between pairs of neurons, simulated choice probability was insensitive to neurometric sensitivity throughout training (Fig. 4a,b). In contrast, when the strength of correlations between pairs of neurons in our model decreased as their direction tuning and sensitivity became less similar, simulated choice probability matched the MT measurements and increased selectively for the most sensitive neurons (Fig. 4c,d). Thus, dynamic pooling weights and static interneuronal correlations that both depended on neuronal sensitivity could together account for changes in the relationship between MT activity and choice behaviour throughout training.

Relationship with decision-related activity in LIP

The selective changes in the pooling weights caused improvements in the pooled response (y in Eq. 1) that were somewhat similar to changes in motion-driven responses of individual LIP neurons measured during training (Fig. 5a) 4. Like in LIP, the simulated pooled response reflected the direction decision throughout training, increasing from zero to more positive values with increasing viewing time on trials in which a rightward decision was made, and decreasing from zero to more negative values with increasing viewing time on trials in which a leftward decision was made. With training, the pooled response also became increasingly dependent on motion strength, increasing more steeply as a function of viewing time for rightward decisions and decreasing more steeply as a function of viewing time for leftward decisions.

However, there were two important differences between the pooled response in our model and LIP activity. First, the pooled response in our model grew roughly linearly as a function of viewing duration. In contrast, LIP activity tended to increase early in motion viewing but then saturate at a threshold level 4. This saturation is thought to represent the termination of a decision process, particularly for response-time versions of the task 12, 30. However, for our task, in which we controlled the viewing duration, a decision model with no such termination is sufficient to account for behavioural performance for the range of viewing durations used 4, 38. Moreover, an alternative learning model based on a dynamic termination process could not account for the behavioural data (Supplemental Fig. 1).

The second difference was that our model tended to overestimate the signal-to-noise ratio (SNR, the difference in mean responses to the two directions of motion divided by their common standard deviation) of individual LIP neurons throughout training. We computed the SNR of the pooled response of the model as a function of training for two coherence levels for each monkey (Fig. 5b, black symbols). We also computed the SNR of individual LIP responses using a receiver-operator characteristic (ROC) analysis that quantified how well the responses distinguished between the two direction choices, assuming Gaussian noise (Fig. 5b, red symbols). The SNR of both the simulated and real LIP responses grew consistently in a similar manner with training (correlation coefficient between SNRs of the simulated and real LIP responses for monkey C: $r_{51.2\%}=0.65$, $H_0: r=0$ using Fisher's Z transformation, $p<10^{-15}$; $r_{12.8\%}=0.32$, $p<10^{-6}$; monkey Z: $r_{99.9\%}=0.41$, $p<10^{-3}$; $r_{25.6\%}=0.17$, $p=0.15$), approaching the upper bound achieved by the optimal pooling weights (Fig. 3b and Fig. 5b, dotted line). However, the real LIP responses tended to have a slightly lower SNR than the simulated responses (paired t-test, $H_0: \text{SNR}_{\text{data}}-\text{SNR}_{\text{model}}=0$; $p_{51.2\%}=0.05$ and

$p_{12.8\%}=0.10$ for monkey C; $p_{99.9\%}<10^{-22}$ and $p_{25.6\%}<10^{-10}$ for monkey Z). One likely cause of this discrepancy is that our model represents a pooled LIP signal that is less noisy than individual neurons. Also, the simulated responses represent the difference in activity between two populations of LIP neurons supporting leftward and rightward decisions, respectively, and thus is less noisy than either component measured separately. Accordingly, an alternative model with separate LIP populations corresponding to the two direction decisions had SNRs that more closely matched the data (Supplemental Fig. 3f).

Perceptual learning on a fine discrimination task

The coarse discrimination task from our previous study, modelled above, used two alternative motion directions separated by 180° . In this case, the most informative neurons are those that respond most strongly to those directions (Fig. 3a; red and blue triangles) 39, 40. We tested whether the model can also account for improved perceptual sensitivity on a fine discrimination task, in which the two alternative motion directions are separated by a much smaller amount (Fig. 6a; red and blue triangles). In this case, the most informative neurons are not those that respond most strongly to the presented stimuli but rather are tuned $\sim 40^\circ$ away from those directions 34, 39, 40.

We performed similar simulations as before using the same learning rate as the coarse discrimination task ($\alpha=7\times 10^{-7}$), but trained the model using $\pm 10^\circ$ motion stimuli instead of 0° and 180° . Unlike the weight profiles for the coarse task, changes in the pooling weights for this task were not centred on neurons tuned to the direction of motion of the stimulus ($\pm 10^\circ$). Instead, the strongest weights developed around neurons tuned to values offset from the stimuli by $\pm \sim 40^\circ$ (Fig. 6a), consistent with previous reports and reflecting the direction tuning width of individual MT neurons (Supplemental Fig. 7) 34, 39, 40. By the end of training, which for this difficult task took approximately twice as many trials as for the coarse task, the weights were similar to the optimal readout (Fig. 6b,c). Thus, the learning model is not specific to the coarse discrimination task but instead can find the most informative neurons - which are not necessarily the most responsive neurons - to solve at least two different types of discrimination task.

Specificity of learning

A key feature of many forms of perceptual learning is that the improvements tend to be specific to the stimulus configuration used during training, including for coarse and fine direction discrimination tasks 4, 41, 42. This specificity helps to distinguish improvements in perceptual sensitivity from associative and other forms of learning. The stimulus configuration we used to train the monkeys for the coarse task tended to use a roughly horizontal axis of motion but varied somewhat from session-to-session, depending on the direction and spatial tuning properties of the MT and/or LIP neuron being recorded. Under those conditions, improvements in performance tended to be largest for sessions in which the axis of motion was similar to values used in previous sessions 4. Our model, trained on the same sequence of motion axes, showed similar specificity (Fig. 7a,b).

We examined the specificity of learning in more detail for both the coarse and fine discrimination tasks (Fig. 7c-f). After training the model using a single pair of simulated

motion directions, we measured both lapse rate and discrimination threshold using different directions. For the coarse task, lapse rates were mostly absent except 90° from the trained direction, consistent with the weight profiles that distinguish the two alternatives across a broad range of directions (e.g., Fig. 3a). In contrast, perceptual sensitivity degraded steadily as a function of distance from the trained directions, comparable to the decline in weights over that range. For the fine task, both lapse rates and discrimination threshold showed a higher degree of stimulus specificity, reflecting narrower weight profiles (e.g., Fig. 7c,e) and the overall greater difficulty of that task.

Predictions

Our model makes several testable predictions. First, these forms of perceptual learning are driven by a reward prediction error. According to the model, this error signal depends on both motion coherence (which generates the prediction) and reward feedback and should be present even after acquisition of the sensory-motor association. We predict that this error signal is encoded by brainstem dopaminergic neurons that encode a similar error signal during conditioning tasks 6, 7. Second, interneuronal correlations in MT should depend on neurometric sensitivity (Fig. 4c). Third, the specificity of learning on the coarse and fine discrimination tasks should depend on the direction axis in the manner shown in Fig. 7. Fourth, learning should be fastest if strong motion stimuli are used early in training and weaker stimuli later (consistent with ref. 43), because the error signal depends on the value of the pooled MT response and therefore is noisier for weaker motion, particularly early in training (Supplemental Fig. 8).

Discussion

We previously showed that in monkeys trained on a visual motion discrimination task, improvements in perceptual sensitivity corresponded to changes in motion-driven responses of neurons in area LIP, which represents the readout of motion information to form a direction decision, but not area MT, a likely source of that motion information 4, 12, 13, 18, 35, 44. Here we showed that a computational model that uses a reinforcement-learning rule to adjust pooling weights between MT-like sensory neurons and LIP-like decision neurons can account for both behavioural and neural changes observed during training.

Our model suggests that the changes we measured in LIP during training reflect an increasingly selective readout of the most informative MT neurons. In reality, the sensory evidence is likely provided by not just MT but also other motion-sensitive neurons in the brain, like those found in the middle superior temporal area (MST) 45. Likewise, LIP is just one of an interconnected network of brain regions, including the superior colliculus and frontal eye field, that represent and likely contribute to the formation of the direction decision 17. Therefore, our model is not informative about where in the brain the actual changes in connectivity occur. Rather, the model establishes principles governing how functional (i.e., direct or indirect) connectivity between areas like MT that provide the sensory evidence for the task and areas like LIP that form the decision is modified by experience.

Our simulations also provide a deeper understanding of the relationship between the noisy activity of MT-like neurons and behavioural choices. This relationship, called choice probability, appears to arise from both an appropriate readout scheme and a particular form of interneuronal correlations. Choice probability in our simulations matched real MT data and increased selectively for the most sensitive neurons only when pairwise correlations depended on the similarity of both the direction tuning and sensitivity of each pair of neurons 4. Several modelling studies have made similar assumptions about stronger correlations between neurons with similar response properties, possibly arising from common inputs to similarly tuned neurons 36, 46, 47. Consistent with this idea, response correlations in V1 have been shown to depend on the similarity of tuning between pairs of neurons 37. Pairwise correlations in MT have been shown to be stronger between neurons with similar direction tuning curves 24, but their relationship to neuronal sensitivity has not yet been examined systematically.

Our reinforcement-learning model used a simple delta rule to adjust the pooling weights based on a reward prediction error from the current decision. This kind of reward prediction error has been used extensively to account for learning behaviours that involves the establishment of sensory-response associations and is reflected in the phasic activity of midbrain dopamine neurons 6. These signals are likely driven at least in part by reward-related activity encoded in numerous cortical and subcortical regions including the orbitofrontal and anterior cingulate cortex, striatum, and ventral tegmental area 7. However, in our model reward was essentially a surrogate for whether the response was correct or not, suggesting that other kinds of feedback signals related more closely to errors than rewards might also play a role in driving learning 48. Further work is needed to determine which of these feedback signals are present during perceptual learning.

This simple model with a single learning rule can account for the time course, magnitude, and specificity of both associative and perceptual improvements. Early in training, the feedback reinforcement signal first establishes the functional connectivity in stimulus-response association, from neurons that represent the sensory stimulus to neurons that control the motor responses. This sensory-motor connectivity is further refined by the same learning mechanism to provide a more selective read-out of the most sensitivity sensory signals associated with that response, a form of channel reweighting thought to underlie several forms of perceptual learning 4, 49.

Other feedback signals, such as attention, have also been implicated in gating and/or guiding neural changes during perceptual learning 2, 43. We did not model effects of attention on learning explicitly in our model, mainly because the attention state of the animal was not manipulated or measured in our experiments. It has been suggested that the co-occurrence of attention and reward feedback is an important factor deciding which stimulus features are learned during perceptual learning 11. Our model provides a framework for addressing this important issue.

Methods

All monkey data are from ref. 4. The procedures in that study were carried out in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the University of Pennsylvania Institutional Animal Care and Use Committee (IACUC). Simulations in this study were performed using MATLAB (The Mathworks, Inc., Natick, MA).

MT-like responses

Our model contained MT-like neurons with responses to the random-dot stimulus that were based on previous measurements 4. We characterized the responses of each MT neuron measured in that study (141 for monkey C, and 106 for monkey Z) using four parameters (Supplemental Fig. 9) 50: 1) the linear coherence dependence (gain) of the neuron's mean response to motion presented in the neuron's preferred direction (k_p , in spikes/sec²/100% coh), 2) the gain to motion presented in the neuron's null direction (k_n , in spikes/sec²/100% coh), 3) the response to 0% coherence motion (k_0 , in spikes/sec), and 4) the variance-to-mean ratio of the response (ϕ , in spikes/sec). We assumed that the response of each neuron to a given stimulus was a normally distributed random variable with mean m and variance v given by:

$$m = T \{ k_0 + COH [k_n + (k_p - k_n) f(\theta|\Theta)] \}, \text{ where } f(\theta|\Theta) = e^{\frac{-(\theta - \Theta)^2}{2\sigma_\Theta^2}} \quad [4]$$

$$v = \phi m \quad [5]$$

where θ is motion direction (0° and 180° for the coarse task, $\pm 10^\circ$ for the fine task), COH is fraction coherence, T is stimulus duration (in sec), and $f(\theta|\Theta)$ is the Gaussian-shaped direction tuning curve with mean Θ (the preferred direction of the neuron) and tuning width σ_Θ (40° , the average tuning width measured in ref. 4). Using Eqs. 4 and 5 we could simulate the response of each neuron to a stimulus of any direction, coherence, and viewing duration.

For each simulation, we randomly selected 200 neurons, with replacement, from this library to form a group of MT neurons with the same direction tuning Θ . This group was then repeated 36 times, each group with a different direction tuning distributed uniformly around 360° (Θ from -170° to 180° in 10° increments), for a total of 7200 neurons in each simulation.

Responses of the MT-like neurons in our model were weakly correlated 23, 24. The response correlation between any pair of neurons i and j ($i, j = 1, 2 \dots 7200$) depended on the similarity of their direction tunings and sensitivity to motion and were defined by the correlation matrix R :

$$R = [\rho_{ij}] \quad [6]$$

$$\rho_{ij} = \begin{cases} g_{sen} \times g_{dir} & \text{if } i \neq j \\ 1 & \text{otherwise} \end{cases}, \quad [7]$$

where ρ_{ij} is the correlation coefficient between neuron i and neuron j that depends on two functions g_{sen} and g_{dir} , which describe the similarity of the sensitivity to motion and direction tuning of neurons i and j , respectively. For most of the simulations, we assumed that correlation magnitude decreases linearly as a function of the difference in ranked sensitivity and exponentially with the difference in direction tuning 36 (Fig. 4b). Thus, g_{sen} and g_{dir} are given by:

$$g_{sen} = \begin{cases} \rho_{max} - \frac{|sen_i - sen_j|}{b_{sen}} & \text{if } \rho_{max} > \frac{|sen_i - sen_j|}{b_{sen}} \\ 0 & \text{otherwise} \end{cases} \quad [8]$$

$$g_{dir} = e^{-\frac{|\Theta_i - \Theta_j|}{b_{dir}}}, \quad [9]$$

where ρ_{max} is the maximum correlation between neurons, sen_i is the ranked sensitivity (i.e., inverse of the neurometric threshold coherence, rank ordered for all 7200 neurons) of neuron i , and b_{sen} and b_{dir} determine the rate at which ρ_{ij} decreases. We set $\rho_{max}=0.5$, $b_s = 20$ and $b_\Theta = 30$ in order to obtain an average correlation of ~ 0.18 between pairs of similarly tuned neurons, a value similar to that reported previously 4, 23, 24. The values of ρ_{max} , b_s and b_Θ affect mainly the choice probability of MT neurons, and have very little effect on the time courses and asymptotic performance of learning (Supplemental Fig. 6). For one set of simulations (Fig. 4a), we assumed that the correlation between neurons does not depend on the difference in motion sensitivities. For this case, g_{sen} was a constant equal to 0.15, the average response correlation of MT neurons recorded in our and other previous studies 4, 23, 24.

Pooling scheme and decision formation

To simulate the stochastic, trial-by-trial responses for each MT-like neuron in the population, we first created a 7200-dimensional vector of independent normal deviates (i.e., normal random variables with 0 mean and unit variance), $\mathbf{z}$, one for each MT-like neuron. We transformed these independent normal deviates to a vector of correlated normal deviates, $\mathbf{r}$, by multiplying $\mathbf{z}$ by the square root of the correlation matrix, $\mathbf{U}$, computed with the Cholesky factorization method:

$$\mathbf{r} = \mathbf{U}\mathbf{z} \quad [10]$$

The responses of the MT-like neurons (x in Eq. 1) were then computed by scaling and shifting each element in $\mathbf{r}$ according to:

$$x_i = m_i + \sqrt{v_i} r_i, \quad [11]$$

where m_i and v_i are the mean and variance of the i^{th} neuron.

These MT responses were pooled by a decision process as a weighted sum. We made two assumptions about this pooling process. First, we assumed that the pooling process is not perfect and therefore corrupted the pooled signal (y in Eq. 1) with both additive ($\sim N(0, 5^2$ spikes/sec)) and multiplicative ($\sim N(0, 2y)$) noise (the effects of each noise source considered separately is described in Supplemental Fig. 1). This pooling noise is critical for modelling the pattern of changes in choice probabilities of MT neurons, because without it, the choice probability will be greater than chance (0.5) even in the earliest stage of training. The standard deviation of the additive noise was chosen such that the value of choice probability is close to 0.5 during the early stage of training (Supplemental Fig. 9d). The values and justifications of other model parameters are listed in Table S1. Second, the pooling process represents integrated MT activity (because of T in Eq. 4, x is a cumulative response). We assumed that this integration process did not change with training, as suggested by preliminary analyses of behaviour 4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank L. Ding, M. Nassar, B. Heasley, R. Kalwani, C-L. Teng, S. Bennur, and M. Todd for helpful comments on this manuscript and J. Zweigle for expert technical assistance. Supported by the Sloan Foundation, the McKnight Foundation, the Burroughs-Wellcome Fund, and NIH R01-EY015260 and T32-EY007035.

References

1. Gibson EJ. Perceptual learning. *Annu Rev Psychol.* 1963; 14:29–56. [PubMed: 13947716]
2. Gilbert CD, Sigman M, Crist RE. The neural basis of perceptual learning. *Neuron.* 2001; 31:681–697. [PubMed: 11567610]
3. Mukai I, et al. Activations in Visual and Attention-Related Areas Predict and Correlate with the Degree of Perceptual Learning. *J Neurosci.* 2007; 27:11401–11411. [PubMed: 17942734]
4. Law CT, Gold JJ. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat Neurosci.* 2008; 11:505–513. [PubMed: 18327253]
5. Sutton, RS.; Barto, AG. Reinforcement learning: an introduction. MIT Press; Cambridge, Mass.; 1998.
6. Schultz W. Behavioral theories and the neurophysiology of reward. *Annu Rev Psychol.* 2006; 57:87–115. [PubMed: 16318590]
7. Schultz W. Multiple dopamine functions at different time courses. *Annu Rev Neurosci.* 2007; 30:259–288. [PubMed: 17600522]
8. Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology (Berl).* 2007; 191:391–431. [PubMed: 17072591]
9. Redgrave P, Gurney K, Reynolds J. What is reinforced by phasic dopamine signals? *Brain Res Rev.* 2008; 58:322–339. [PubMed: 18055018]
10. Bao S, Chan VT, Merzenich MM. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature.* 2001; 412:79–83. [PubMed: 11452310]
11. Seitz A, Watanabe T. A unified model for perceptual learning. *Trends Cogn Sci.* 2005; 9:329–334. [PubMed: 15955722]
12. Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci.* 2002; 22:9475–9489. [PubMed: 12417672]
13. Hanks TD, Ditterich J, Shadlen MN. Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nat Neurosci.* 2006; 9:682–689. [PubMed: 16604069]

14. Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. *Nature*. 1999; 400:233–238. [PubMed: 10421364]
15. Andersen RA, Buneo CA. Intentional maps in posterior parietal cortex. *Annu Rev Neurosci*. 2002; 25:189–220. [PubMed: 12052908]
16. Colby CL, Goldberg ME. Space and attention in parietal cortex. *Annu Rev Neurosci*. 1999; 22:319–349. [PubMed: 10202542]
17. Gold JJ, Shadlen MN. The neural basis of decision making. *Annu Rev Neurosci*. 2007; 30:535–574. [PubMed: 17600525]
18. Salzman CD, Britten KH, Newsome WT. Cortical microstimulation influences perceptual judgements of motion direction. *Nature*. 1990; 346:174–177. [PubMed: 2366872]
19. Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci*. 1992; 12:4745–4765. [PubMed: 1464765]
20. Williams RJ. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*. 1992; 8:229–256.
21. Shadlen MN, Britten KH, Newsome WT, Movshon JA. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci*. 1996; 16:1486–1510. [PubMed: 8778300]
22. Dayan P, Daw ND. Decision theory, reinforcement learning, and the brain. *Cogn Affect Behav Neurosci*. 2008; 8:429–453. [PubMed: 19033240]
23. Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*. 1994; 370:140–143. [PubMed: 8022482]
24. Bair W, Zohary E, Newsome WT. Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J Neurosci*. 2001; 21:1676–1697. [PubMed: 11222658]
25. Barto AG, Anadan P. Pattern recognizing stochastic learning automata. *IEEE Transactions on Systems, Man, and Cybernetics*. 1985:360–374.
26. Loewenstein Y, Seung HS. Operant matching is a generic outcome of synaptic plasticity based on the covariance between reward and neural activity. *Proc Natl Acad Sci U S A*. 2006; 103:15224–15229. [PubMed: 17008410]
27. Mazurek ME, Roitman JD, Ditterich J, Shadlen MN. A role for neural integrators in perceptual decision making. *Cereb Cortex*. 2003; 13:1257–1269. [PubMed: 14576217]
28. Kepecs A, Uchida N, Zariwala HA, Mainen ZF. Neural correlates, computation and behavioural impact of decision confidence. *Nature*. 2008; 455:227–231. [PubMed: 18690210]
29. Gold JJ, Shadlen MN. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn Sci*. 2001; 5:10–16. [PubMed: 11164731]
30. Kiani R, Hanks TD, Shadlen MN. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci*. 2008; 28:3017–3029. [PubMed: 18354005]
31. Spiegelhalter DJ, Lauritzen SL. Sequential updating of conditional probabilities on directed graphical structures. *Networks*. 1990; 20:579–605.
32. Royer S, Pare D. Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature*. 2003; 422:518–522. [PubMed: 12673250]
33. Connolly PM, Bannur S, Gold JJ. Correlates of Perceptual Learning in an Oculomotor Decision Variable. *J Neurosci*. 2009 in press.
34. Purushothaman G, Bradley DC. Neural population code for fine perceptual decisions in area MT. *Nat Neurosci*. 2005; 8:99–106. [PubMed: 15608633]
35. Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci*. 1996; 13:87–100. [PubMed: 8730992]
36. Sompolinsky H, Yoon H, Kang K, Shamir M. Population coding in neuronal systems with correlated noise. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2001; 64:051904. [PubMed: 11735965]

37. Smith MA, Kohn A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci.* 2008; 28:12591–12603. [PubMed: 19036953]
38. Gold JJ, Shadlen MN. The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J Neurosci.* 2003; 23:632–651. [PubMed: 12533623]
39. Jazayeri M, Movshon JA. A new perceptual illusion reveals mechanisms of sensory decoding. *Nature.* 2007; 446:912–915. [PubMed: 17410125]
40. Hol K, Treue S. Different populations of neurons contribute to the detection and discrimination of visual motion. *Vision Res.* 2001; 41:685–689. [PubMed: 11248258]
41. Ball K, Sekuler R. Direction-specific improvement in motion discrimination. *Vision Res.* 1987; 27:953–965. [PubMed: 3660656]
42. Fahle M. Perceptual learning: specificity versus generalization. *Curr Opin Neurobiol.* 2005; 15:154–160. [PubMed: 15831396]
43. Ahissar M, Hochstein S. Task difficulty and the specificity of perceptual learning. *Nature.* 1997; 387:401–406. [PubMed: 9163425]
44. Newsome WT, Pare EB. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J Neurosci.* 1988; 8:2201–2211. [PubMed: 3385495]
45. Celebrini S, Newsome WT. Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *J Neurosci.* 1994; 14:4109–4124. [PubMed: 8027765]
46. Shamir, M.; Sompolinsky, H. Correlation codes in neuronal networks. In: Dietterich, TG.; Becker, S.; Ghahramani, Z., editors. *Advances in neural information processing systems*. MIT Press; 2002.
47. Yoon, H.; Sompolinsky, H. The effect of correlations on the fisher information of population codes. In: Kearns, MS.; Solla, SA.; Cohn, DA., editors. *Advances in Neural Information Processing Systems*. MIT Press; 1998. p. 167-173.
48. van Veen V, Carter CS. Error detection, correction, and prevention in the brain: a brief review of data and theories. *Clin EEG Neurosci.* 2006; 37:330–335. [PubMed: 17073172]
49. Doshier BA, Lu ZL. Mechanisms of perceptual learning. *Vision Res.* 1999; 39:3197–3221. [PubMed: 10615491]
50. Britten KH, Shadlen MN, Newsome WT, Movshon JA. Responses of neurons in macaque MT to stochastic motion signals. *Vis Neurosci.* 1993; 10:1157–1169. [PubMed: 8257671]

Figure 1.

Schematic of the decision-making, reinforcement-learning model. Decision-making is accomplished via a feedforward, linear pooling scheme (orange arrows): the motion stimulus causes noisy activation of a population of direction-selective neurons with varied response properties, including different direction tunings and sensitivities; next, the output of each neuron is weighed and then summed to produce a scalar decision variable y , which is further corrupted with decision noise; finally, the sign of the decision variable determines the choice (rightward for $y > 0$, leftward otherwise). Learning occurs via a feedback loop (green arrows): the difference between the reward outcome and the predicted probability of reward based on the value of the decision variable y forms the reward prediction error, which is used to update the pooling weights. Associative learning involves establishing appropriate weights as a function of direction tuning; e.g., positive weights for rightward-tuned neurons and negative weights for leftward-tuned neurons. Perceptual learning involves shaping the weights so that the neurons most informative about the given stimulus are weighed most strongly.

Figure 2.

Discrimination performance of the monkeys and model with training. **a**, Discrimination threshold (●; logarithmic scale on the left ordinate) and lapse rate (τ ; error rate at 99.9% coherence, linear scale on the right ordinate) with 68% CIs plotted as a function of trial number (computed using 1000-trial bins for discrimination threshold, and 250-trial bins for lapse rate) computed from psychophysical data of the two monkeys during training. Solid lines are best fits of a decaying single-exponential function **b**, Discrimination threshold and lapse rate computed from choice responses of the model, plotted as in panel **a**. Choice responses of the model were averaged over 10 simulations with learning rate (α in Eq. 3) 7×10^{-7} for monkey C and 1×10^{-7} for monkey Z. **c**, Time constants of exponential fits to lapse (τ_{la} , abscissa) and threshold (τ_{th} , ordinate) computed from psychophysical data (asterisks) and model simulations using learning rates 2.5, 3, 5, 7, and 9×10^{-7} for monkey C and 0.5, 0.75, 1, 1.5, and 2×10^{-7} for monkey Z (black symbols increasing in size for increasing learning rates). Error bars are s.e.m. over five simulations. Solid lines are weighted linear fits. Dashed lines are 95% confidence bounds.

Figure 3.

Changes in pooling weights with training on a coarse discrimination task. **a**, Pseudocolor plots of the pooling weights binned by the neurons' threshold sensitivity to the motion stimulus (based on the ability of an ideal observer to discriminate motion direction based only on the neural responses; 6%-80% coh in logarithmic increments of 1.06) and direction tuning (-170 - 180° in linear increments of 10°) for different training epochs. Color scale of weight magnitude is shown at the right. Gray horizontal bands correspond to neuronal sensitivities not found in the MT data set used to model the sensory responses 4. **b**, Optimal pooling weights for maximizing the percentage of correct responses (see Methods). **c**, Pooling weights (points) as a function of direction tuning for two different threshold sensitivities (top: 8.7% coherence; bottom: 16.5% coherence; arrows on the right of panel **a**) at different training epochs. Red lines indicate the weight profiles for optimal linear pooling from **b**. Error bars are s.e.m. over ten simulations. Blue and red triangles at the bottom of each panel indicate the directions of motion to discriminate.

Figure 4.

Changes in choice probability of neurons in the sensory representation with training **a**, Pseudocolor plot depicting the structure of noise correlations between one neuron with median neurometric sensitivity and tuned to 0° motion and other neurons in the population with different sensitivities and direction tunings. The correlations depend only on the difference in direction tuning between neurons. **b**, Choice probability of sensory neurons tuned to rightward motion (0°), computed for neurons binned by sensitivity (6%-80% coh in logarithmic increments of 1.06) for three different training epochs. Error bars are s.e.m. over ten simulations. Solid lines are weighted linear fits. Choice probability tended to increase with training but did not depend on sensitivity. **c**, **d**, Same conventions as in **a**, **b**, showing the results of simulations that used correlations that depended on the difference in both direction tuning and sensitivity between neurons. Choice probability tended to increase with training, particularly for the most sensitive neurons.

Figure 5.

Changes in the pooled responses with training **a**, Mean pooled responses as a function of viewing time (0.1-sec-wide bins) for different motion strengths (see legend) for rightward (solid lines) and leftward (dashed lines) correct direction decisions. **b**, SNRs of the pooled responses (black points; 1,000-trial bins) as a function of training at two motion strengths (top: above discrimination threshold; bottom: near threshold) for monkeys C (left) and Z (right). Red symbols are the SNR of LIP responses at the same motion coherences and viewing times for motion into versus out of the neuron's response field. Dotted lines are the SNR achieved by the optimal linear pooling weights (see Fig. 3b).

Figure 6.

Changes in pooling weights with training on a fine discrimination task. **a**, Pseudocolor plots of the pooling weights binned by the neurons' threshold sensitivity to the motion stimulus (6%-80% coh in logarithmic increments of 1.06) and direction tuning (-170 - 180° in linear increments of 10°) for different training epochs. Color scale of weight magnitude is shown at the right. Gray horizontal bands correspond to neuronal sensitivities not found in the MT data set used to model the sensory responses 4. **b**, Optimal pooling weights for maximizing the percentage of correct responses (see Methods). **c**, Pooling weights (points) as a function of direction tuning for two different threshold sensitivities (top: 8.7% coherence; bottom: 16.5%; arrows on the right of panel **a**) at different training epochs. Red lines indicate the weight profiles for optimal linear pooling from **b**. Error bars are s.e.m. over ten simulations. Blue and red triangles at the bottom of each panel indicate the directions of motion to discriminate.

Figure 7.

Specificity of learning. **a,b**, The difference between discrimination threshold measured for a given session and its 21-session running average for the model (filled circles) and real behavioural data (open circles) plotted against the familiarity of the axis of motion used in that session for data from monkeys C (**a**) and Z (**b**). Familiarity was computed as the z-score of the current direction with respect to the distribution of motion directions used before that session. Thus, less frequently used motion directions have larger z-scores 4. Lines are linear fits; positive slopes imply that performance was worse for less familiar axes of motion (H_0 : slope=0, $p < 0.05$ for all four regressions). **c-f**, The model was trained on a coarse (**c,d**) and fine (**e,f**) discrimination task using fixed axes of motion (dashed lines) and then tested (without further learning) using different axes of motion. The coarse task required

discrimination between the given direction axis and its 180° opposite. The fine task required discrimination between $\pm 10^\circ$ relative to the given direction axis. **c,e**, percent correct for high-coherence stimuli, a measure of knowledge of the sensory-motor association, plotted as a function of the axis of motion relative to the trained axis. **d,f**, discrimination threshold, a measure of perceptual sensitivity, plotted as a function of the axis of motion relative to the trained axis.

Table 1

Model variants. Bold indicates the version used in the main text. Statistical test is whether the model-generated lapse rates and discrimination thresholds have single-exponential time constants that match the data (t-test based on linear regression parameters for $H_0: [\tau_{th}, \tau_{la}]_{\text{monkey}} = [\tau_{th}, \tau_{la}]_{\text{model}}$)

		Models	$H_0: p > 0.05$	Figure
Pooling	Decision neurons	1) One pool of decision neurons representing right-left	Y	2, 3, S3
		2) Two pools of decision neurons representing right and left separately	Y	S3
	Decision noise	1) Additive and multiplicative	Y	2, 3
		2) Additive only	Y	S1a.1
		3) Multiplicative only	Y	S1a.2
	Weight normalization	1) Multiplicative	Y	2, 3
		2) Subtractive	Y	S1b.2
		3) None	Y	S1b.1
	Nonlinear pooling	1) 1	Y	2, 3, S1c
	(exponent of pooled	2) 1.19	Y	S1c
	signal)	3) 1.41	Y	S1c
		4) 2	N	S1c
	Decision bound	1) No decision bound	Y	2, 3
		2) Decision bound increases with training, static weights	N	S1d
Reinforcement learning	Learning rules	1) Correlation between reward prediction error (PE) and MT activity	Y	2, 3
		2) Correlation between reward PE and mean-subtracted MT activity	Y	S2a.1
		3) Correlation between reward value and mean-subtracted MT activity	N	S2a.2
		4) Correlation between reward value and MT activity	N	S2a.3
		5) Oja-like learning rule	Y	S2b
	Reward prediction	1) Sequentially updated estimate of logistic function	Y	2, 3
		2) Direct estimate of logistic function from previous 300 trials	Y	S2c