

Learning and forgetting visuo-motor associations in changing environments

Stefano Fusi^{1,3}, Wael F. Asaad², Earl K. Miller², Xiao-Jing Wang³

¹ Institute of Physiology, University of Bern, Switzerland, ² Department of Brain and Cognitive Sciences and the Center for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA ³ Volen Center for Complex systems, Brandeis University, Waltham, Massachusetts, USA

What motor response should be associated with a given visual stimulus can be entirely different depending on the specific behavioral context. Primates have a remarkable ability to adapt to new environments and to find the responses which are rewarded in each context. Flexible sensori-motor mapping is at the core of cognitive control of behavior, and depends critically on the prefrontal cortex. The goal of this work is to propose a theoretical framework for learning sensori-motor associations, that combines a recurrent decision-making cortical network model with reward-gated Hebbian synaptic plasticity.

The process of learning visuo-motor associations has been studied experimentally in [1], where the authors trained two monkeys to respond with a left or a right saccade to a visual stimulus. The rewarded associations were changed in an unpredictable way, and the monkey had to learn them by trial and error. In particular two visual stimuli (A and B) were initially associated to Left and Right saccadic movements (L and R) respectively. Periodically the associations were reversed (from AL and BR to AR and BL and vice-versa) without any notice, and the monkey had to forget the old associations and to learn the new ones. Here we propose a model which captures quantitatively the process of learning and forgetting the visuo-motor associations and we confirm predictions based on the data of [1].

The relevant facts inferred from the analysis of the experimental data are: 1) when the associations are reversed, the monkey forgets quickly the

old associations and then slowly learns the new ones; 2) the monkey responds randomly after each mistake, regardless of the previous history; 3) becoming aware that the association has been reversed for one stimulus does not help the monkey to respond correctly to the other stimulus; 4) 29% of the recorded cells in pre-frontal cortex respond selectively to the direction of the planned saccadic movement and the directional preference of the cell does not change with learning.

The core of the model is a decision making network of integrate-and-fire neurons with realistic recurrent synaptic connections (mediated by AMPA, GABA and slow NMDA receptors) as in [2]. Two subpopulations of excitatory neurons represent the direction selective neurons of experimental point 4. When a visual stimulus is presented, the two populations compete and one of them eventually wins, expressing the monkey's decision. The parameters are tuned such that the network chooses randomly one of the two saccadic movements with equal probability, provided that the inputs activated by the visual stimuli to the two populations are perfectly balanced. Learning is modelled by the dynamics of the total inputs to the two competing populations for the two visual stimuli. From experimental fact 3, we know that the variables corresponding to different stimuli can be studied separately. For each pair of variables we introduce learning rates for potentiating and depressing the total inputs for the two possible responses in the presence and in the absence of reward. Each total input is restricted to a given interval, reflecting the fact that synapses are bounded and the neural activity varies in a limited range. This restriction makes the memory forgetful [3], i.e. the mnemonic trace of the past experiences decays exponentially with their age and old visuo-motor associations can be forgotten. The model reproduces the behavioral data when: the input to the selected/non-selected motor response population is potentiated/depressed when the reward is obtained; both inputs to the two populations are depressed in the absence of reward. The learning rate in the latter case must be much higher than in the presence of reward to reproduce experimental observations 1 and 2. Introduction of a dependence of the learning rates on the reward expectancy produces only marginal improvement. Saturation of the total inputs is probably enough to capture the behavior observed in [1].

Tuning the parameters to achieve a probabilistic decision making as required by experimental fact 2 is not easy. Any quenched heterogeneity (e.g. random connectivity) can disrupt the mechanism underlying the stochastic-

ity of the choice. These problems can be overcome if we assume that learning occurs on all possible temporal scales. In particular, if we add slow components of learning, the memory window can be extended to span several blocks of trials in which the same stimulus is remembered to be associated with both motor responses. As the experiment has been designed to avoid any bias in the response (left and right saccades are rewarded with the same probability when many blocks are considered), the slow components will tend to create an input configuration which makes the two responses equally probable. The ‘fast’ components will then allow to learn the correct associations within each block. Each mistake resets the fast components, bringing the system back to the symmetric configuration determined by the slow components. We show that: 1) the slow components reflect the reward history for each response; 2) the slow components can compensate any initial bias and large quenched heterogeneities. Moreover we predict that if the associations are never reversed, then a single mistake should not lead to random behavior because the slow components will consistently bias only one motor response. In the experiment [1], for two stimuli the associations were never reversed. Our prediction is confirmed by the behavioral data corresponding to these two stimuli.

Conclusions: we provided a mechanistic explanation of the rich phenomenology of experiment [1]. Both the behavior of the monkey and our network dynamics reflect the statistics of rewarded associations on multiple time scales, as required to create an internal representation of an hierarchy of contexts. The strong reset following a mistake leads to a random behavior because on long time scales the environment keeps changing and two specific motor responses are equally rewarded. Probabilistic decision making is not a spontaneously emerging behavior because it requires fine tuning, rather it is dictated by the statistics of the environment. Learning on multiple time scales can also produce a power law decay of the mnemonic trace, which is the observed behavior in psychophysics and provides the optimal way of storing memories [5].

Supplementary material

Description of the model

The core of the model is a decision making network of integrate-and-fire neurons with realistic recurrent synaptic connections (mediated by AMPA, GABA_A and slow NMDA receptors) as in [2]. Two subpopulations of excitatory neurons represent the direction selective neurons which encode the planned motor response (i.e. the decision of the monkey). When a visual stimulus is presented, the two populations compete through a population of inhibitory neurons. One of the two excitatory populations eventually wins, expressing monkey's decision. The parameters are tuned in such a way that when the inputs activated by the visual stimuli to the two population are perfectly balanced, the monkey chooses randomly one of the two saccadic movement with equal probability. In general the probability of choosing one of the two responses can be approximated by a sigmoidal function of the difference between the inputs to the two competing populations. For example the probability of choosing Left when stimulus A is presented is:

$$p_{AL} = \frac{1}{1 + e^{-(c_{AL}-c_{AR})/\sigma}}$$

where c_L and c_R are the total synaptic conductances to the populations representing Left and Right saccadic responses respectively. σ is determined by the parameters of the neural dynamics.

Learning dynamics

Learning is modelled by the dynamics the total stimulus selective synaptic conductances (inputs) to the two competing populations. Following each trial, these inputs are updated according to the neural activity and to whether the monkey is rewarded or not. From the data analysis, we know that the inputs corresponding to the different stimuli can be studied separately (we know that the feedback the monkey gets when it responds to one visual stimulus does not affect the response to the other visual stimulus). For each pair of variables we introduce the learning rates for potentiating and depressing the total inputs for the two possible responses in the presence and in the absence of reward. Each total input is restricted to a given interval,

reflecting the fact that synapses are bounded and the neural activity varies in a limited range. In particular we assume that we have the following dynamics:

$$c_{vm} \rightarrow c_{vm} + q_+^r(1 - c_{vm}) - q_-^r c_{vm}$$

where c_{vm} is a variable representing the total synaptic conductance to the population expressing motor response $m(= L, R)$ when stimulus $v(= A, B)$ is presented. q_+^r and q_-^r are the learning rates for potentiating and depressing c_{vm} , and they depend on whether the monkey got reward or not ($r = R, NR$).

This restriction makes the memory forgetful [3], i.e. the mnemonic trace of the past experiences decays exponentially with their age and old visuo-motor associations can be forgotten (palimpsest property).

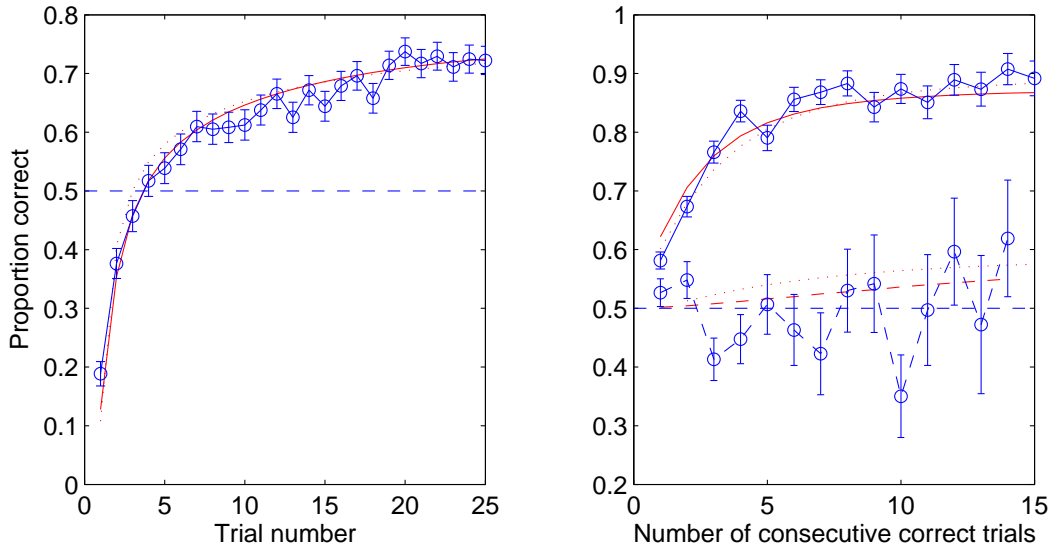


Figure 1 *Fitting the model to the behavioral data (blue): performance vs number of trials after a reversal of the associations (left) and performance following the two particular sequences of 15 trials described in the text. The red lines are the results of the simulations: the dotted/solid line corresponds to the best fit in the absence/presence of the expectancy of reward.*

The model could reproduce quantitatively the behavioral data when: the input to the population of the selected/non-selected motor response is potentiated/depressed when the reward is obtained; both inputs to the two populations are depressed in the absence of reward. The learning rate in

the latter case must be ten times higher than in the presence of reward to reproduce the fast reset observed in the experiment.

The effect of expectancy of reward

We introduced a dependence of the learning rates on the expectancy of reward. The expectancy of reward ξ^μ for stimulus μ is updated every trial. In particular, if the monkey is rewarded then: $\xi \rightarrow \xi(1 - \epsilon_+) + \epsilon_+$, otherwise: $\xi \rightarrow \xi(1 - \epsilon_-)$. ξ is close to 1 if the least $1/\epsilon_+$ trials were mostly rewarded, close to 0 if there were many mistakes. We then assumed that the learning rates can be modulated by ξ and we determined the functional form of this dependence by fitting the model to the data. We found that all the learning rates decrease linearly with the expectancy of reward. However the dependence is weak and the match between data and model was only marginally improved. Saturation of the total inputs is probably enough to capture the behavioral data when such a simple task is considered.

Figures illustrating the results

Reproducing the experimental behavior

We fitted the model to the behavioral data. In particular we wanted to reproduce 1) the average behavior following the reversal of the associations (Figure 1, left); 2) the behavior following a single mistake (i.e. when the monkey completes correctly the trial but it gives the wrong response) occurring after $n(=1, \dots, 15)$ consecutive correct trials (Figure 1, right, lower curve); 3) the behavior following $n(=1, \dots, 15)$ consecutive correct trials (Figure 1, right, upper curve). The average behavior of the simulated network was estimated with a mean-field approach, and the parameters have been fitted to the data by using a Montecarlo. The results are plotted in Figure 1.

Learning to respond randomly

The configuration of the slow learning components is determined by the statistics of reward across many different blocks. For this reasons the inputs to the populations corresponding to the left and right saccadic movements tend to balance if left and right are rewarded with the same probability. This is true for any initial state, also when it is strongly biased towards one of

the two responses. This is illustrated in Figure 2 where a network with fast and slow components is simulated. The learning rule of the fast and the slow components is the same except for the learning rates (100 times lower for the slow components) and the modification of the input to the non-selected population in case of no reward. In such a situation the input is potentiated instead of being depressed.

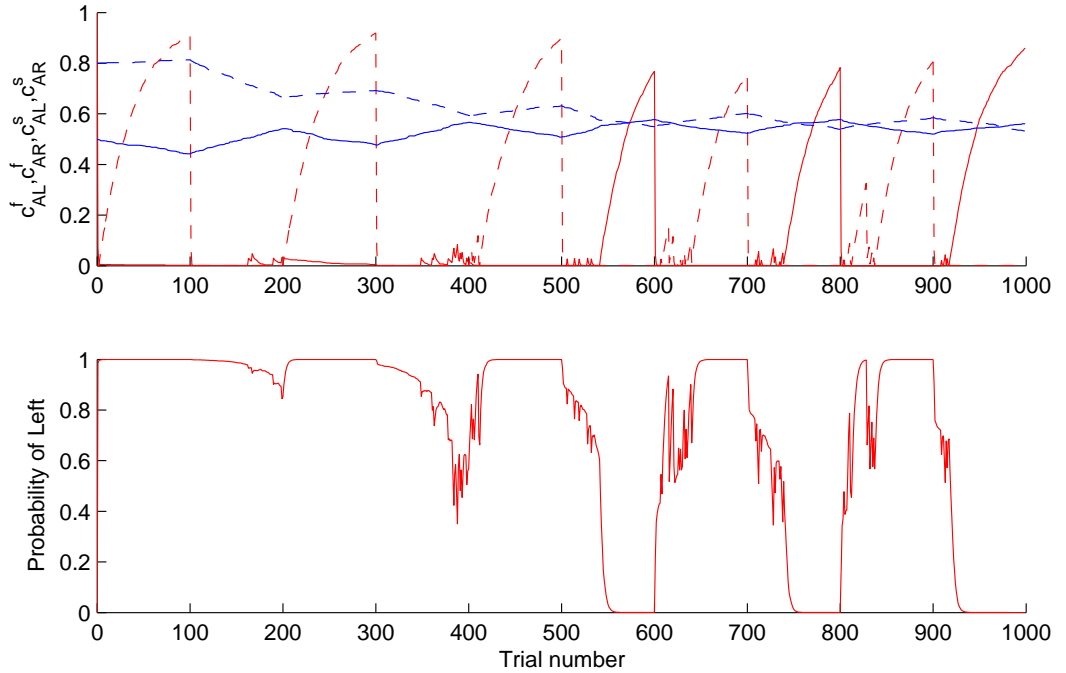


Figure 2 *Learning probabilistic decision making: the blue curves are the slow $c_{AL,AR}^s$ (solid, dashed), the red curves are the fast $c_{AL,AR}^f$. Both are plotted as a function of the number of trials for stimulus A (top). The corresponding probability of choosing L is reported in the bottom plot. The correct associations are reversed every 100 trials. The system starts from a situation which is strongly biased towards L. At the beginning L is always deterministically chosen. After a few reversals, the balance between the slow components is achieved (the blue curves tend to the same value).*

In general, if one motor response is more rewarded than the other, the network will choose that response with a higher probability that will depend on reward history on long time scales. This is illustrated in Figure 3, left

panel, where the probability of choosing left, is plotted against the probability that left is rewarded when many blocks are considered. Monkeys are known to be able to encode the probability that a motor response is rewarded (“matching behavior”, see e.g. [4]).

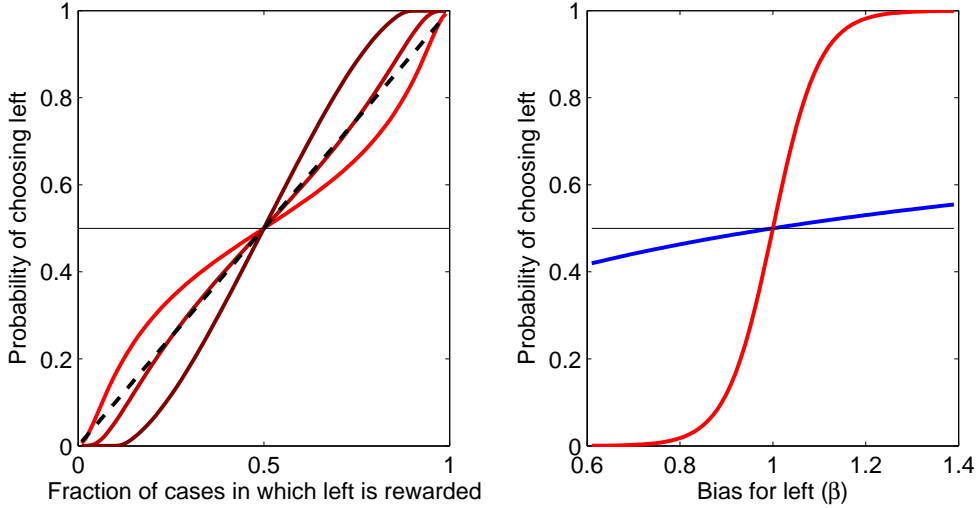


Figure 3 *Compensating biases due to heterogeneities. Left: the slow components encode the statistics of reward on long time scales: the probability of choosing Left is plotted against f_L , the probability that Left is rewarded (the statistics is estimated across many blocks). The three red curves correspond to three different values of σ ($\sigma = 0.15, 0.07, 0.015$, from lighter to darker). The dashed line corresponds to $p_L = f_L$. Right: compensating a quenched bias with learning. The probability of choosing left vs the bias towards Left ($\beta = 1$ corresponds to no bias) is plotted before (red) and after (blue) learning ($\sigma = 0.015$). In this case $f_L = 1/2$.*

The ability to encode the probability of reward does not depend much on the specific choice of the neural and the learning parameters. When a quenched bias β (> 1 or < 1) is always favoring/disfavoring one of the two motor responses (e.g. Left):

$$p_{AL} = \frac{1}{1 + e^{-(\beta c_{AL} - c_{AR})/\sigma}}$$

then learning will compensate this bias and produce a final balanced situation when Left and Right are equally rewarded. The final c_{AL} will be equal to c_{AR}/β , to produce motor responses with equal probabilities. This is illustrated in Figure 3, Right.

Predictions

Slow components become symmetric to reflect the fact that the probability of rewarding the two responses is the same for the stimuli whose associations are continuously reversed. For some other stimuli, the associations were never changed. We predicted that for these stimuli the slow components biasing the two motor responses should not be symmetric because each stimulus is consistently associated to one motor response only, throughout all the blocks (i.e. on long time scales). Hence a fast reset leading to random behavior, should not occur following a single mistake. This prediction has been confirmed by the data of [1] and it is illustrated in Figure 4.

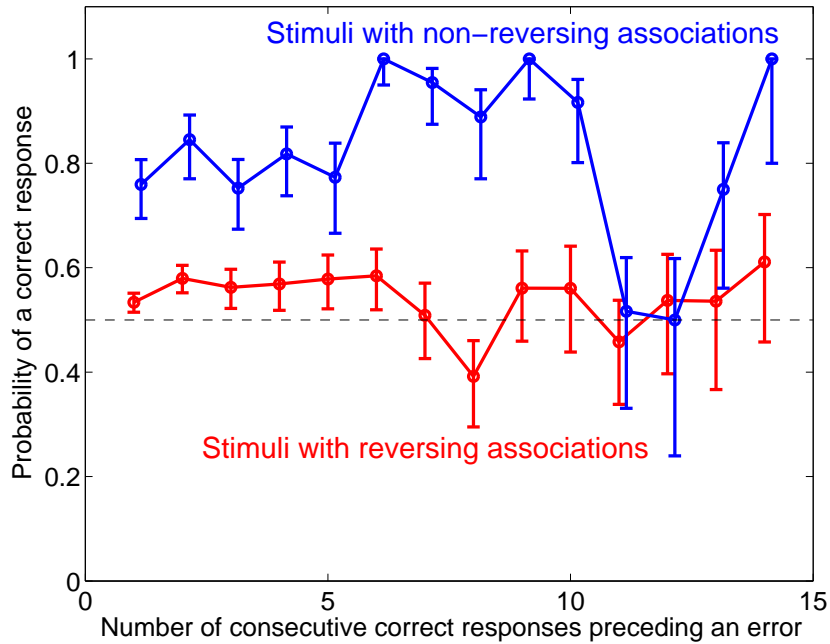


Figure 4 Behavior following an error: for the stimuli whose associations are continuously reversed (red), the monkey responds randomly, at chance

level, after each mistake, regardless the number of preceding correct trials (on the x -axis). For these stimuli the slow components of the inputs to the two competing populations (corresponding to Left and Right saccades) are symmetric. The stimuli which are consistently associated to only one motor response, the fast and slow components are biased towards a single motor response. A single mistake cannot reset this bias (blue line). Notice that now the mistakes include also the trials which are not completed properly to increase the statistics.

References

- [1] W.F. Asaad, G. Rainer, E.K. Miller, Neural activity in the primate prefrontal cortex during associative learning, *Neuron*, **21**, 1399-1407 (1998)
- [2] X.-J. Wang, Probabilistic decision making by slow reverberation in cortical circuits, *Neuron*, **36**, 955-968 (2002)
- [3] S. Fusi, Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates, *Biol. Cyb.*, **87**, 459-470 (2002)
- [4] L.P. Sugrue, G. S. Corrado, W.T. Newsome, Matching behavior and the representation of value in parietal cortex, *Science*, **304**, 1782-1787 (2004)
- [5] S. Fusi, P. Drew, L.F. Abbott, Cascade Models of synaptically stored memories, *Neuron*, **45**, 599-611, (2005)