

Neuron, Volume 54

Supplemental Data

A Neural Circuit Model of Flexible

Sensorimotor Mapping: Learning

and Forgetting on Multiple Timescales

Stefano Fusi, Wael F. Asaad, Earl K. Miller, and Xiao-Jing Wang

Supplemental Text

Generalizing to Novel Stimuli

We assumed that the strengths of synaptic inputs are modified in the learning process independently for each of the visual stimuli. If presynaptic neural pools that are selective for different stimuli are non-overlapping, each stimulus activates only one of the mutually exclusive neural pools. Since presynaptic activity is required for long-term modification, the synapses onto the decision circuit are updated independently for each stimulus. On the other hand, if the presynaptic neural pools do overlap, then for those neurons which are activated by both sensory stimuli (A and B), their g_{Left} and g_{Right} to Left and Right populations in the decision circuit will remain the same in spite of the learning process. Those cells do not contribute to the winner-take-all decision dynamics, given that the competition between Left and Right is driven by the differences in the inputs and not by the total input. Experimental evidence suggests that neural population responses do overlap for different stimuli in the late stages of the visual pathway (e.g. in IT (Miyashita, 1993; Tanaka, 1996)). Interestingly, the assumption of an overlapping population code gives rise to the following prediction. In the experiment of (Asaad et al., 1998), in each session two of the stimuli presented to the monkey are new. This means that, since neural recordings started only after months of behavioral training, by the time cortical recordings started, the monkey had already seen hundreds of stimuli. The probability that a typical neuron was never activated by the hundreds of stimuli seen by the monkey is small. For example if we assume that each stimulus activates a randomly chosen fraction f of neurons, then the probability that a neuron is never activated decreases exponentially with the number of stimuli which have been presented to the monkey (in particular it is proportional to $(1-f)^p$ where p is the total number of patterns seen since the beginning of training). If this neuron is activated by at least one stimulus used during behavioral training, this stimulus was associated with equal probability to Left and Right across blocks of trials. According to our model, this implies that the slow learning components to Left and Right are expected to be balanced. Consequently, the strong reset to chance level performance should be observed also for novel stimuli, even after the first errors in a session. Indeed the analysis of the performance for novel stimuli indicates that the reset is observed already in the first block of trials (Fig. 8F). The probability that the monkey responds correctly after one mistake preceded by an arbitrary number of correct trials is 0.54 ± 0.01 when all blocks are considered, and it is not significantly different from the probability 0.58 ± 0.03 of responding correctly when only the first blocks of trials are analyzed.

Mean Field Analysis of Learning

The mean field approach has been used to fit the model to the data (see below) and it can be summarized as follows: we consider one stimulus and we assume that the correct association is Left. The input configuration at any trial t is described by the probability distribution $p_t(c_L, c_R)$. Following each stimulus presentation, $p_t(c_L, c_R)$ is updated according to:

$$p_{t+1}(c_L, c_R) = P_L \left(\frac{c_L - q_+^R}{1 + q_+^R}, \frac{c_R}{1 - q_-^R} \right) p_t \left(\frac{c_L - q_+^R}{1 + q_+^R}, \frac{c_R}{1 - q_-^R} \right) + \left[1 - P_L \left(\frac{c_L}{1 - q_-^{NR}}, \frac{c_R}{1 - q_-^{NR}} \right) \right] p_t \left(\frac{c_L}{1 - q_-^{NR}}, \frac{c_R}{1 - q_-^{NR}} \right) \quad (1)$$

The equations are obtained by summing all realizations which arrive at c_L, c_R after one trial. It is possible to arrive there in two ways: 1) if Left is chosen (i.e. the correct response), then the synapses to Left start from c'_L and they are potentiated $c'_L \rightarrow c_L = c'_L + (1 - c'_L)q_+^R$ (the initial point c'_L is expressed as a function of the target c_L in the formula above), analogously the synapses to Right are depressed; 2) if Right is chosen (second term, $P_R = 1 - P_L$), then both synapses to Left and to Right are depressed. The equations are exact and they are solved numerically. At reversal, we have the initial condition:

$$p_0(c_L, c_R) = \delta(c_L)\delta(c_R - 1)$$

After t presentations, the probability that L is chosen is:

$$P_t = \iint_0^1 dc_L dc_R p_t(c_L, c_R) P_L(c_L, c_R)$$

This probability is plotted in Figure 4A. Figure 4D shows the performance after any error, and not after reversal, and it is computed in the same way the performance of Figure 4A, except for the initial condition that is now:

$$p_0(c_L, c_R) = \delta(c_L)\delta(c_R)$$

This choice is a consequence of the fact that after one error there is a strong reset and both c_L and c_R start from a value which close to zero, no matter where they were before the error (q_-^{NR} obtained from the fit is close to one).

Slow Components Are Balanced when the Motor Responses Are Equally Rewarded

The slow components c_L^s, c_R^s at equilibrium are denoted by $c_{L,eq}^s, c_{R,eq}^s$ respectively, where (Brunel et al., 1998):

$$c_{L,eq}^s = \frac{f_L(P_L r_+^R + (1-P_L)r_+^{NR})}{f_L(P_L r_+^R + (1-P_L)r_+^{NR}) + (1-f_L)(P_L r_-^{NR} + (1-P_L)r_-^R)}$$

$$c_{R,eq}^s = \frac{(1-f_L)(P_L r_+^{NR} + (1-P_L)r_+^R)}{(1-f_L)(P_L r_+^{NR} + (1-P_L)r_+^R) + f_L(P_L r_-^R + (1-P_L)r_-^{NR})}$$

where f_L is the fraction of trials in which Left is rewarded across the entire experiment, when a particular stimulus is presented (the index for the stimulus is again not indicated to simplify the notation).

This distribution is a stable steady state of the learning dynamics. We now consider a strongly biased case in which $c_L^s \gg c_R^s$ and hence $P_L = 1$. We also assume that $f_L = 1/2$, as for the stimuli whose associations are reversed in the experiment. In such a case, the expressions simplify and the c values converge toward the equilibrium values given by

$$c_{L,eq}^s = \frac{r_+^R}{r_+^R + r_-^{NR}}, \quad c_{R,eq}^s = \frac{r_+^{NR}}{r_+^{NR} + r_-^R}$$

If the slow learning rates r_{\pm}^R, r_{\pm}^{NR} are chosen such that $c_{R,eq}^s > c_{L,eq}^s$ when $P_L = 1$, then when $P_L = 0$, $c_{R,eq}^s < c_{L,eq}^s$ and the slow components will always converge to a perfect balance. Indeed, if Left wins more than Right, as in the case in which $P_L = 1$, the asymptotic configuration toward which the slow components converge, will favor Right because $c_{R,eq}^s > c_{L,eq}^s$. Conversely, when Right wins more than Left the situation will be reversed and the asymptotic value will be biased towards Left. Hence, when $P_L = 1$, c_L^s will decrease and tend towards a configuration with $P_R > P_L$. However, as soon as $P_R > P_L$ because c_R^s surpasses c_L^s , then $c_{R,eq}^s < c_{L,eq}^s$, so that Left will be favored. If the learning rates are small enough (to avoid oscillations of synaptic strengths), the only equilibrium solution will be $c_R^s = c_L^s$ and $P_L = 1/2 = f_L$. The learning rates used in the simulations for the slow components are: $r_+^R = r_-^R = 1.5 \times 10^{-4}$, $r_+^{NR} = r_-^{NR} = 0.002$. The exact values are not important provided that they satisfy the condition $c_{R,eq}^s > c_{L,eq}^s$ when Left wins ($P_L = 1$):

$$\frac{r_+^{NR}}{r_+^{NR} + r_-^R} > \frac{r_+^R}{r_+^R + r_-^{NR}} \Leftrightarrow \frac{r_+^{NR}}{r_+^R} > \frac{r_-^R}{r_-^{NR}}$$

Notice that in order to satisfy this condition, r_+^{NR} cannot be zero, indicating that for the slow components an anti-Hebbian term is necessary.

Expectancy of Reward

Similarly to (Suri and Schultz, 1999) we assume that the learning rates $q_{\pm}^{R,NR}$ depend on the expectancy of reward ξ which is updated every trial. In case of reward: $\xi \rightarrow \xi(1 - \varepsilon_+) + \varepsilon_+$, otherwise $\xi \rightarrow \xi(1 - \varepsilon_-)$. The quantity ξ is close to 1 if at least $1/\varepsilon_+$ trials were mostly rewarded, close to 0 if there were many mistakes; therefore it represents expected rewards. We used a Monte Carlo to determine the functional

dependence of $q_{\pm}^{R,NR}$ on ξ , which can best fit the model with monkey's behavioral data. It turned out that the dependence which minimizes the χ^2 is approximately linear and that for the best fit, both the learning rates for reward and no reward case should decrease by 20% as ξ goes from 0 to 1. $\varepsilon_+ = 0.03, \varepsilon_- = 0.2$.

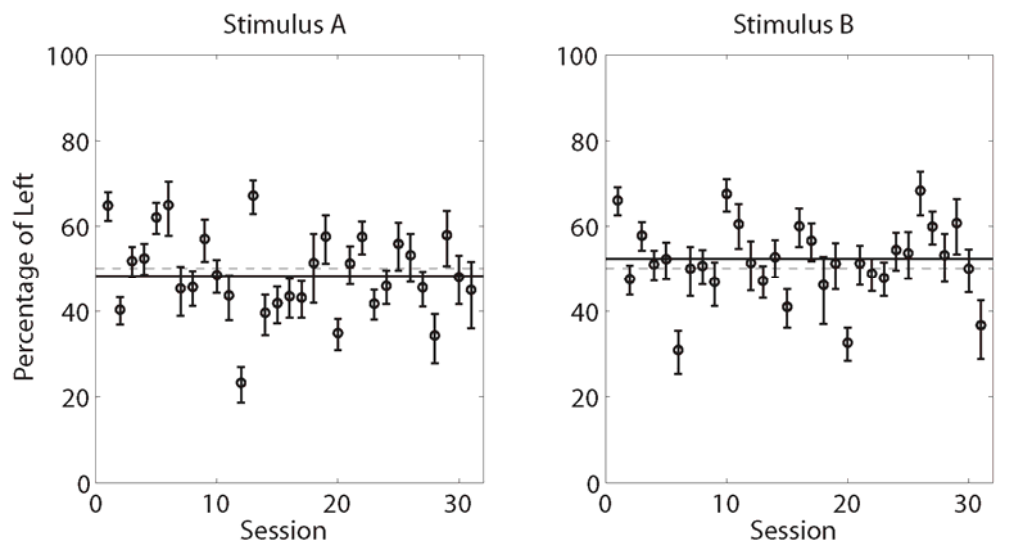


Figure S1.

Absence of bias for left or right saccade: percentage of cases in which the monkey makes a left saccadic movement after one mistake for all sessions, separately for stimulus A (left) and B (right). All points are close to 50%, indicating the absence of any bias in the choice. The average across all sessions (solid line) is not significantly different from 50% (dashed line) for both stimulus A (48.8 ± 0.8) and B (52.3 ± 0.8). The percentage has been plotted separately for each session because the visual cues are different in every session.

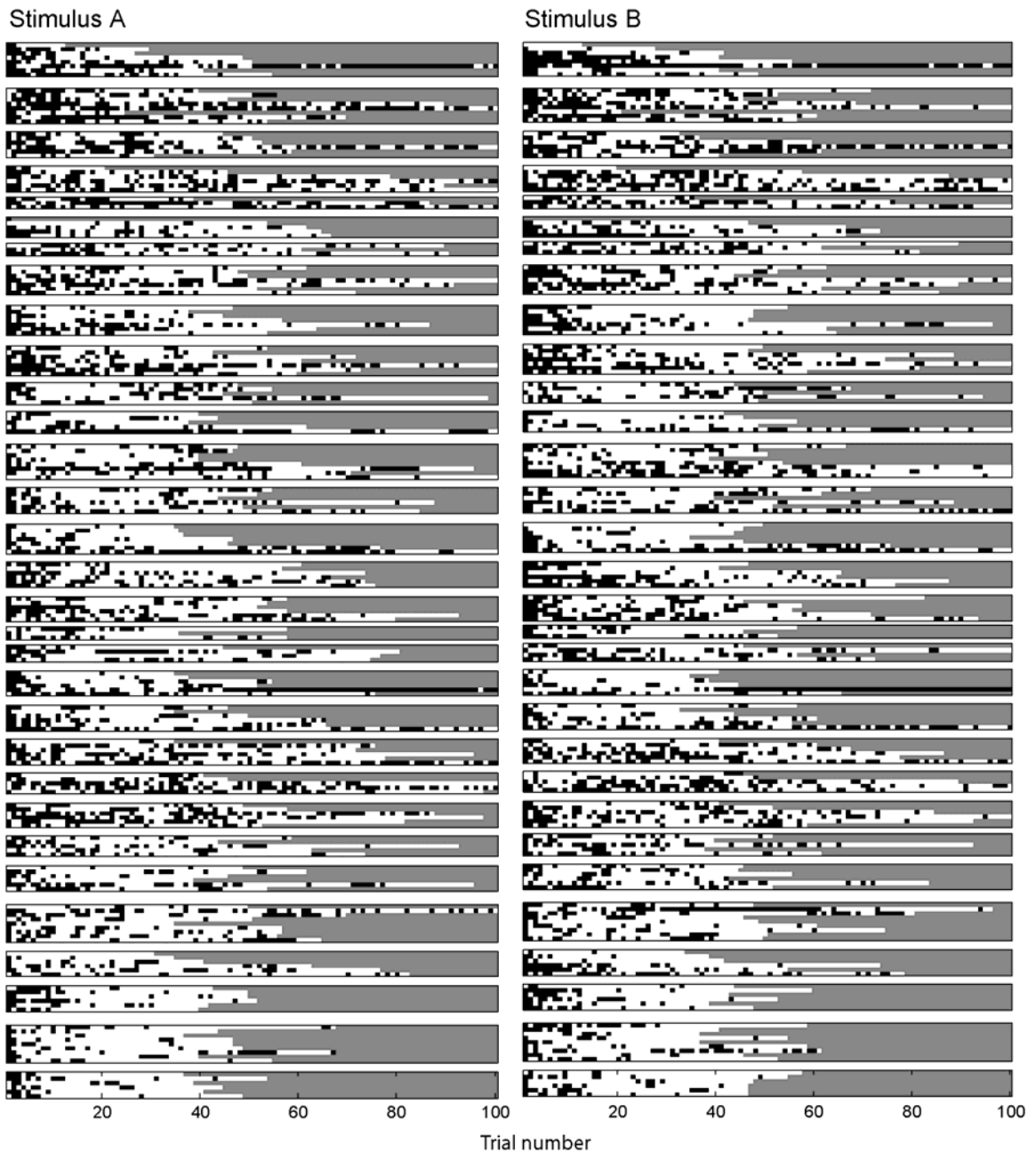


Figure S2.

Behavioral raster plots for all sessions, for stimulus A (left) and stimulus B (right). Each box contains one session and each line corresponds to one block in which the stimuli are consistently associated with one motor response. White dots are correct trials, black are mistakes. When the block terminates the dots are gray. There is no evidence in these rasters that the monkey abruptly switches from incorrect (old) to correct (new) associations, which would correspond to a mostly black bar followed by a mostly white bar.

Supplemental References

Brunel, N., Carusi, F., and Fusi, S. (1998). Slow stochastic Hebbian learning of classes in recurrent neural networks. *Network* 9, 123–152.

Miyashita, Y. (1993). Inferior temporal cortex: where visual perception meets memory. *Annu. Rev. Neurosci.* 16, 245–263.</jrn>

Suri, R.E., and Schultz, W. (1999). A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* 91, 871–890.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139.