A computational approach towards the ontogeny of mirror neurons via Hebbian learning

Lotte Weerts
Supervisors: dr. Sander Bohté and dr. Rajat Mani Thomas
University of Amsterdam
lotteweerts@gmail.com

ABSTRACT
It has been proposed that Hebbian learning could be responsible for the ontogeny of predictive mirror neurons in the premotor cortex (Keysers and Gazzola, 2014). Here, we show that an artificial neural network (ANN) that evolves via variation of Oja's rule (an implementation of Hebbian learning) is sufficient to account for the emergence of predictive mirror neurons. By extension, this work provides positive evidence for the association hypothesis, which states that mirror-like behavior in the motor cortex is a byproduct of associative learning.

Keywords
Mirror neurons, Hebbian learning, Associative learning, Computational model

INTRODUCTION
In the early 1990s, mirror neurons were discovered in the ventral premotor cortex of the macaque monkey (Di Pellegrino et al., 1992). These neurons fired both when the monkeys grabbed an object and when they watched another primate grab that same object. More recently, evidence has started to emerge that suggests the existence of mirror neurons with predictive properties (Keysers and Gazzola, 2014). These neurons fire when the action they encode is the action most likely to happen next, rather than the action that is concurrently being observed. Understanding how the brain is capable to predict future actions of others by means of a computational model could be of use in practical applications, such as self-driving cars that can predict actions of other cars. A critical question concerns how (predictive) mirror neurons have developed to behave the way they do. In other words: what is the ontogeny of mirror neurons?

Keysers and Gazzola (2014) proposed a mechanistic perspective on how mirror neurons in the premotor cortex could emerge due to associative learning, or more precisely, Hebbian learning. Hebbian learning explains learning as a change in synaptic strength of neurons that are concurrently active. To investigate whether or not Hebbian learning is sufficient to lead to the emergence of mirror neurons, we present a computational approach that implements the mechanics described by Keysers and Gazzola (2014). This involves the use of an artificial neural network (ANN) to simulate activity in the premotor cortex (PM) and the superior temporal sulcus (STS). The PM coordinates self-performed actions, whereas the STS is a region known to respond to the sight of body movements and the sound of actions. By exciting neurons in these two areas action execution and observation are simulated. The problem addressed in this work can be defined as follows: *can artificial neural networks that evolve via a local Hebbian-like learning rule, when exposed to action execution and observation, be sufficient to lead to the emergence of predictive mirror neurons?*

Here, we show that Oja's rule, an implementation of Hebbian learning, is sufficient to impose predictive mirror-like behavior. First, the proposed research question is considered in relation to other studies on the ontogeny of mirror neurons. Subsequently, three implementations of Hebbian learning will be discussed: the covariance learning rule, the BCM rule, and Oja's rule. Additionally, a variation of Oja's rule is presented. The computational model that describes the artificial neural network is introduced. A newly designed procedure for the quantitative analysis of mirror neuron-like behavior has been applied to the recorded activity of ANNs that evolve via several different learning rules. Finally, it is argued that an artificial neural network that evolves via a variation of Oja's rule is sufficient to explain the emergence of mirror neurons.

HEBBIAN LEARNING
The association hypothesis states that mirror-like behavior of neurons in the motor cortex arises due to the establishment of associations between sensory- and motor stimuli. Keysers and Gazzola (2014) proposed that Hebbian learning, a type of associative learning, could account for the emergence of mirror neurons. Their idea is illustrated by a model that describes connections between the premotor cortex (PM), the inferior posterior parietal cortex (area PF/PFG) and superior temporal sulcus (STS). Both the PM and PF/PFG play a role in action coordination, whereas the STS is a region known to respond to the sight of body movements and the sound of actions. The proposed model explains the emergence of predictive mirror neurons as a result of sensorimotor input of self-performed actions. Action execution is coordinated by activity in the PM. Subsequently, reafferrence occurs, which refers to the observation of self-performed actions. Once this observation has reached

the STS, neurons in the PM that encode the next action have already become active. If Hebbian learning is assumed, the simultaneous activity in the STS of the observed action and activity in the PM for the next action causes an association between neurons in both regions. Keysers and Gazzola (2014) suggest that such an increase in synaptic strength could account for the emergence of predictive mirror neurons.

Cooper et al. (2013) argued that Rescorla-Wagner (RW), a supervised learning rule, and not Hebbian learning, can cause the emergence of mirror neurons. Their computational model simulates an experimental study (Cook et al., 2010) that examined automatic imitation, which is thought to be an index for mirror neuron activity. Only RW was capable of simulating the data of Cook et al. However, Hebbian learning is not limited to the learning rule that was used by Cooper et al. (2013). Other implementations of Hebbian learning, for example Oja's rule (Oja, 1982), were not considered. Therefore, Hebbian learning should not be dismissed fully as a possible explanation for the emergence of mirror neurons.

COMPUTATIONAL MODEL

The artificial neural network created in this work simulates the PM and STS. The general structure of connectivity in the network (figure 1) is a result of three assumptions. First, the intermediate step of the PM/PFG has been omitted. Second, each neuron only represents one action phase. Note that it is not suggested that one action is simulated by one neuron in the brain. In fact, each node in the ANN should be viewed as a cluster of neurons rather than individual neurons. Third, PM → STS connections are modeled as inhibitory connections between PM and STS neurons that encode the same action. They are modeled as inhibitory because the net flux of activity from PM to STS is known to be net inhibitory. However, the brain contains less inhibitory neurons (20% of all neurons) than excitatory neurons. Therefore, the total number of inhibitory connections is reduced by only allowing PM → STS connections for the same action.

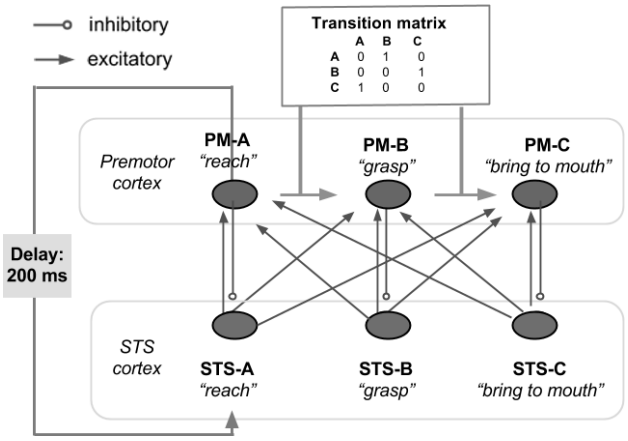


Figure 1: The structure of the two-layered ANN that simulates the PM and STS cortices.

The activity of one neuron is simulated as a function of the weighted activity of all presynaptic neurons, similar as in Cooper et al. (2013):

$$a_i(t+1) = \rho a_i(t) + (1-\rho)\sigma\big(I_i(t)\big)$$

$$I_j(t) = \sum_i w_{j,i} a_i(t-1)\, E_j + B_j + \mathcal{N}(0, \mu^2)$$

Here, ρ is a value between 0 and 1 that determines to which extent previous activity persists over time. $I_j(t)$ refers to the input to the neuron, which is transformed by a value between 0 and 1 by the sigmoid function σ. The weight matrix w contains the weights of the connections. When the artificial network is trained, w is altered. Additionally, the value of $I_j(t)$ depends on three factors: E_j, B_j and μ. E_j refers to the amount of stimulation caused by other sources than neurons within the network and can be either on or off. The input signal further consists of the activation bias B_j and Gaussian noise determined by μ. Each time step t represents 1 ms of activity. The parameters ρ, E_j, B_j and μ are constant in this model. The values used in our simulations are shown in table 1.

Parameter	ρ	E_j	B_j	μ
Value	0.990	4.0	2.0	2.0

Table 1: Parameters of activity calculation, from Cooper et al. (2013).

Keysers and Gazzola (2014) propose that mirror neuron activity emerges due to synaptic changes that result from simultaneous action execution and reafference. This implies that the simulation of these mechanics should consist of two separate phases. In the first phase, the training phase, action execution and reafference are simulated. During this phase, the weights are being updated. To simulate action execution, the PM neurons are activated by turning on the E_j variable, which persists for 300 ms (which is analogous to 300 time steps t in the model). After 300 ms, the execution transitions to the next action phase. A stochastic transition matrix determines the sequence of actions. To simulate action observation, an STS neuron that encodes the same action as a PM neuron will be activated 200 ms after activation of the PM neuron. The second phase, the testing phase, consists of the mere observation of actions and therefore only consists of STS activity (whilst still adhering to the probabilities in the transition matrix). Mirror neuron behavior is said to occur if PM neuron activity predicts future activities of STS neurons during the testing phase.

IMPLEMENTATIONS OF HEBBIAN LEARNING

Donald Hebb (1949) proposed that learning occurs on neuron level as a result of changes in synaptic strength due to concurrent activity of the post- and presynaptic neuron. Over time many computational implementations of Hebb's postulate have been proposed. Here, three will be discussed: the covariance rule, BCM and Oja's rule. Additionally, a variation of Oja's rule is presented.

The covariance rule is an implementation of Hebbian learning (Dayan and Abott, 2001) and is defined as follows:

$$\Delta w_{j,i} = \alpha \cdot \frac{1}{n-1} \sum_{i=0}^{n} (a_i - \langle a_i \rangle)(a_j - \langle a_j \rangle)$$

Here, $\langle x \rangle$ denotes the average over a particular time, whereas α refers to the learning rate. A disadvantage of the covariance rule is that the weights are unbounded. This can make the learning rule unstable (Dayan and Abott, 2001). BCM is a modification of the covariance rule that imposes a boundary on the synaptic strengths (Dayan and Abott, 2001):

$$\Delta w_{j,i} = \langle a_i \cdot a_j (a_j - \theta_v) \rangle$$

$$\Delta \theta_v = \langle a_j^2 - \theta_v \rangle$$

Here, θ_v is implemented as a sliding threshold, which causes the weights to be constrained from growing without bounds.

Oja's rule, introduced by Oja (1982), is an alternative learning rule that imposes a boundary on the weights:

$$\Delta w_{j,i} = \langle a_i \cdot a_j - \alpha(a_j^2) w_{i,j} \rangle$$

Oja's rule induces an online renormalization by constraining the sum of squares of the synaptic weights. More precisely, the boundary of the weights is inversely proportional to a, that is, $|w|^2$ will relax to a^{-1} (Oja, 1982). In contrast to BCM, Oja's rule allows for an explicit choice of the size of the boundary by choosing α.

One disadvantage of both BCM and Oja's rule is that baseline activity of neurons is not taken into account. This causes the weights to grow even if no significant spike in activity is measured. Therefore, we propose an alternative to Oja's rule that imposes a threshold θ on all activities of the neurons:

$$a_{x\prime} = \begin{cases} a_x & if\ a_x \geq \theta \\ 0 & otherwise \end{cases}$$

This threshold prevents the weights to change from unsubstantial activities. Note that the addition of the threshold only introduces zero correlation for certain periods of time. Therefore, it does not affect the size of bounds imposed by α. Choosing the proper threshold value can be viewed as intrinsic homeostatic plasticity, which refers to the capacity of neurons to regulate their own excitability relative to network activity (Turrigiano and Nelson, 2004).

ANALYSIS PROCEDURE

To quantitatively determine whether or not mirror neuron behavior has occurred in a simulation, the PM and STS activity in the testing phase have been transformed into transition matrices. By comparing the transition matrices, it can be determined to what extent the PM and STS activity is similar, and thus, to what extent mirror neuron behavior has emerged. The following procedure was used to calculate a transition matrix from recorded activity. First, a low-pass filter is applied to the activity signals of all neurons. Significant peaks are subtracted by setting a threshold equal to the mean plus one standard deviation. Subsequently, the delay between each peak at time t and the peaks of other neurons within $t + 3000$ is determined via a cross correlation. The neuron with the lowest positive delay d_i is selected as the closest successor. In the transition matrix, $t_{j,i}$ is increased by one. As a final step, the rows in the transition matrix are normalized. If this procedure is applied to both PM and STS activity, the similarity between activity can be quantified via the Frobenius norm, which gives error ϵ:

$$\epsilon = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (t_{j,i}^{sts} - t_{j,i}^{pm})^2}$$

A low ϵ indicates a high level of mirror neuron-like behavior. To determine whether or not a low retrieved error is significant, it is compared to the distribution of errors of randomly generated PM matrices compared to the calculated STS matrix.

RESULTS

Figure 2 shows the mean and standard deviation of the error ε of the transition matrices for 10 simulations of the covariance rule, BCM, Oja's rule and the thresholded Oja's rule after a training phase of 250.000 ms for one action sequence consisting of four action phases. All but Oja's rule with a threshold do not impose mirror neuron behavior, as the errors are above significance level. The instability of the covariance rule can be observed by the large variance in its errors. The BCM rule also does not impose mirror neuron behavior. A closer look at the activity of BCM (figure 2) shows that activity in PM neurons completely suppresses STS activity. This problem can be surpassed by Oja's rule by choosing a higher threshold for excitatory neurons ($\alpha_e = 10^{-1}$) than for inhibitory neurons ($\alpha_i = 1^{-1}$). The addition of the threshold to Oja's rule to account for baseline activity successfully produces the emergence of mirror-like behavior.

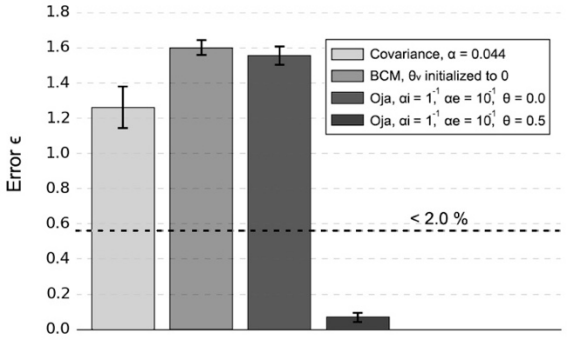
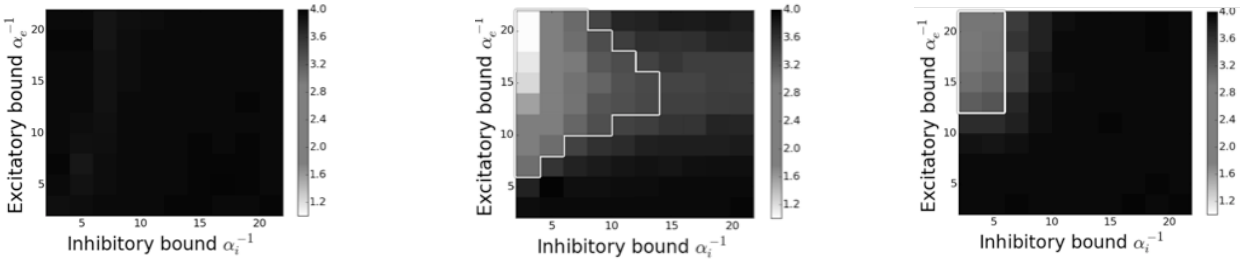


Figure 2: Average and standard deviation of error ϵ of 10 simulations of four learning rules. Each simulation consisted of 250.000 ms in both the training and testing phase, for a single action sequence of 4 action phases. The striped line indicates the value under which the norm is deemed to be significant (< 2% of 10 million random matrices).

A parameter space analysis was performed to see how different settings of the three parameters, namely α_e, α_i and threshold θ, change the outcome of applying Oja's rule. The results are shown in figure 3. If θ is relatively low (0.3, figure 3a), Oja's rule does not impose mirror neuron behavior at all. Figure 3b shows that $\theta = 0.5$ does impose mirror behavior if, in general, the bound of excitatory connections is higher than the bound of inhibitory connections. If both parameters are higher than approximately 15, no mirror behavior emerges. If θ is high (0.7, figure 3c), neuron mirror behavior is imposed, but to lesser extent than when the threshold is 0.5. Here, the excitatory bounds much exceed the inhibitory bounds to a larger extent for mirror neuron behavior to emerge.



3a: $\theta = 0.3$ 3b: $\theta = 0.5$ 3c: $\theta = 0.7$

Figure 3: Results of parameter space analysis. Each figure depicts results for different thresholds θ. The x-axis and y-axis show the bounds for inhibitory neurons (α_i^{-1}) and excitatory neurons (α_e^{-1}) respectively. The lower the error (denoted by a color range from black to white) the better the performance was for a parameter setting. The areas delineated by the white line are all retrieved errors that are lower than the errors of 98% of 10 million randomly generated matrices. If the model is set to these parameter settings, mirror neuron behavior emerges.

CONCLUSION

In previous work it has been proposed that Hebbian learning could be responsible for the ontogeny of predictive mirror neurons (Keysers and Gazzola, 2014). Here, we have shown that a variation of Oja's rule (an implementation of Hebbian learning) is sufficient to explain the emergence of mirror neurons. An artificial neural network that simulates the interactions between the premotor cortex (PM) and the superior temporal sulcus (STS) has been created. Different implementations of Hebbian learning have been compared in performance on a simple action sequence. Additionally, a parameter space analysis has been performed on the proposed thresholded Oja's rule to determine the sensitivity of the parameters on its performance.

We identified that from the learning rules considered, only the thresholded Oja's rule is sufficient to impose mirror neuron behavior. The other learning rules (covariance, BCM and the original Oja's rule) are subject to at least one of the following three limitations. First, the covariance rule is unbounded, which makes it unstable and biologically implausible. Second, BCM and Oja's rule are sensitive to baseline activity, which prevents mirror neurons from emerging. Third, if the inhibitory neurons are not bounded strong enough, as in BCM, PM inhibition causes a complete suppression of STS activations. This prevents associations between the PM and STS to occur. A variation of Oja's rule that uses a threshold and different bounds for inhibitory and excitatory synapses surpasses each of these limitations.

A parameter space analysis indicates that the parameters of the proposed thresholded Oja's rule must adhere to two constraints for mirror neuron behavior to emerge. First, the threshold value must be higher than the baseline activity, but lower than the highest peaks measured. Currently, this threshold is imposed as a fixed constant. An extension of this work would be to model homeostatic plasticity by dynamically determining the threshold value based on the overall network activity. Second, mirror neuron behavior can only be imposed if the bounds for the excitatory neurons are higher than those of the inhibitory neurons, otherwise inhibitory PM neurons completely suppress STS activity.

In conclusion, we have shown that a thresholded Oja's rule is sufficient to account for the emergence of mirror neurons in an artificial neural network that simulates interactions between the PM and STS cortices. Therefore, this work provides positive evidence for the proposal that Hebbian learning is sufficient to account for the emergence of mirror neurons. In the broader sense, this work promotes the idea that statistical sensory motor contingencies suffice in the explanation for the ontogeny of mirror neurons. Therefore, it can be regarded as positive evidence for the association hypothesis.

ROLE OF THE STUDENT

This study was performed by Lotte Weerts under supervision of Sander Bohté and Rajat Mani Thomas. The idea to create a computational model of the mechanics described by Keysers et al. (2014) was suggested by the supervisors. The student proposed the variant of Oja's rule. The design and implementation of the structure of the ANN and the analysis procedure, the processing of the results, formulation of the conclusions and writing were performed by the student.

REFERENCES

Cook, R., Press, C., Dickinson, A., and Heyes, C. (2010). Acquisition of automatic imitation is sensitive to sensorimotor contingency. *Journal of Experimental Psychology: Human Perception and Performance,* 36(4):840.

Cooper, R. P., Cook, R., Dickinson, A., and Heyes, C. M. (2013). Associative (not Hebbian) learning and the mirror neuron system. *Neuroscience letters,* 540:28–36.

Dayan, P. and Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, 91(1):176–180.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. John Wiley & Sons.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience Biobehavioral Reviews*, 34(4):575 – 583.

Keysers, C. and Gazzola, V. (2014). Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1644):20130175.

Oja, E. (1982). Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273.

Turrigiano, G. G. and Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience,* 5(2):97–107