Role of Top-Down Feedback in Unsupervised Learning in Biologically Realistic Networks

Leo Breston
Neuroscience Graduate Program
University of California, San Diego
La Jolla, CA 92093-0634
lbreston@ucsd.edu

Jacob Garret
Neuroscience Graduate Program
University of California, San Diego
La Jolla, CA 92093-0634
jcgarret@ucsd.edu

Oscar Gonzalez
Neuroscience Graduate Program
University of California, San Diego
La Jolla, CA 92093-0634
o9gonzal@ucsd.edu

Abstract

Top-down feedback has been proposed to play an important role in extracting salient features in our environment. This process involves selectively increasing the activity of specific sensory neurons. Though top-down feedback has been proposed as a mechanism involved in attention, its involvement in unsupervised learning is still relatively unknown. Here, we explore the role of top-down feedback in unsupervised learning in biologically realistic neurons. To do this, we used the Brian neuron simulation environment to construct a small network of leaky integrate and fire neurons and tested the effect of top-down feedback on the changes synaptic strengths of simulated sensory neurons. In our simplified network with synaptic weights regulated by spike-timing dependent plasticity, our network could learn to discriminate between two different input types. The addition of stable top-down feedback, however, did not affect the learning rate. Additionally, we found a prominent role for inhibitory activity in the ability of our network to learn to discriminate between inputs types. Overall, our study explores the roles of top-down feedback and inhibition in unsupervised learning.

1 Introduction

Feedback is ubiquitous in the brain. At every structural scale there are as many, if not more, feedback connections than feedforward ones. Top-down connections represent a common type of cortico-cortical feedback in which higher level areas project back to the lower ones upstream of them. As an example, V1 drives V2, which in turn relays back to V1 **[1].** These types of connections have been observed throughout the cortex and appear to be a fundamental circuit motif. It is believed that they an important role in both selective attention, by amplifying salient features at lower levels, and in working memory, by creating an oscillator which can maintain persistent activity, however their role in learning is poorly understood **[2].**

Most models of neural networks consider learning to be purely a feedforward process. In traditional, biologically plausible, unsupervised learning each neuron uses a rule, such as Hebbian plasticity or STDP, to modify its synapses until they correspond to a principle component its input. Intra-layer connections create competition between the neurons causing each one to learn a different component and successive layers learn increasingly abstract features by using the activity of the preceding layer as their input **[3].** While this process works well without any top-down connections, previous work has suggested that feedback may increase the selectivity of learned features **[4].** Still unclear, is how it affects the network's speed of acquisition.

Understanding the effect of top-down feedback on network learning rate is important because current architectures require a long time and many examples for their synaptic weights to converge. We hypothesize that feedback may boost this rate in two ways. First, it might act as an excitatory recurrent input which increases the activity of the neurons relevant to a given pattern, thereby increasing their individual learning rate. Second, if the higher level neurons have a larger receptive than the lower ones they may be able to transfer information about global correlations back down to the lower level. In this paper we use a simple three-layer network to investigate these hypotheses. Unfortunately, we didn't observe any change in learning rate, however future models with different types of plasticity and network parameters might yield better results. For now, it remains an open question.

2 Methods

We used the Brain 2.0 Python package to construct our neurons and networks. Our network was comprised of leaky integrate-and-firing (LIF) neurons similar to those from **[3]**. Additionally, our network contained an input layer, hidden layer, and output layer, each with eight, four, and two LIF neurons respectively. LIF neurons within the input layer were modeled by equation 1, where V_0 is the input stimulus and τ_{input} is the input neuron time constant. LIF neurons within the hidden and output layers were modeled by equations 2 and 3, where g_e is the synaptic conductance (strength), E_e is the synaptic reversal potential, V_r is the resting membrane potential, E_l is the leak reversal potential, and τ_m and τ_e are the membrane and synaptic time constants respectively. The values for these constants are provided in table 1.

$$\frac{dV}{dt} = \frac{(V_0 - V)}{\tau_{input}} \qquad (1)$$

$$\frac{dV}{dt} = \frac{(g_e * (E_e - V_r) + E_l - V)}{\tau_m} \qquad (2)$$

$$\frac{dg_e}{dt} = -\frac{g_e}{\tau_e} \qquad (3)$$

Table 1: LIF Neuron Constants

Variable	**Value**
V_0	0 mV to 2 mV
τ_{input}	5 ms
E_e	0 mV
V_r	-60 mV
E_l	-74 mV
τ_m	10 ms
τ_e	5 ms

The architecture varied between the layers of our network. There were no lateral excitatory synapses formed within a given layer, however, within the hidden and output layers, there were inhibitory synapses. For initial simulations, network connectivity contained all-to-all connections between layers, and within layers with inhibition. There were no inter-layer inhibitory synapses formed. The entire network architecture is shown in figure 1A (red are inhibitory, and blue are excitatory connections). Briefly, the input layer formed excitatory synapses onto the hidden layer neurons in an all-to-all fashion. Hidden layer neurons then formed excitatory synapses onto the output layer neurons, in addition to inhibitory synapses onto other neurons within the hidden layer. Output neurons formed inhibitory synapses on each other, and, in the case of top-down feedback, the output neurons formed excitatory synapses on hidden layer neurons. Stimulus patterns into the input layer varied between stimulus presentations. For a given trial, one set of input layer neurons received excitatory stimulation (filled circles in figure 1) while the other input neurons received none. In the following trial, the previously unstimulated neurons would receive the stimulus and the others would not. This pattern was repeated throughout the simulation.

Finally, excitatory synapses between the hidden layer neurons and output layer neurons were regulated by spike-timing dependent plasticity (STDP). All synaptic weights were initialized to random values. STDP synapses were modeled in the same way (eq. 4 & 5). The values of the constants for the equations below are provided in table 2. We did not implement STDP in inhibitory synapses.

$$\frac{dA_{pre}}{dt} = -\frac{A_{pre}}{\tau_{pre}} \qquad (4)$$

$$\frac{dA_{post}}{dt} = -\frac{A_{post}}{\tau_{post}} \qquad (5)$$

Table 2: STDP Constants

Variable	**Value**
A_{pre} (initial)	0.005
τ_{pre}	100 ms
A_{post} (initial)	-0.00476
τ_{post}	100 ms

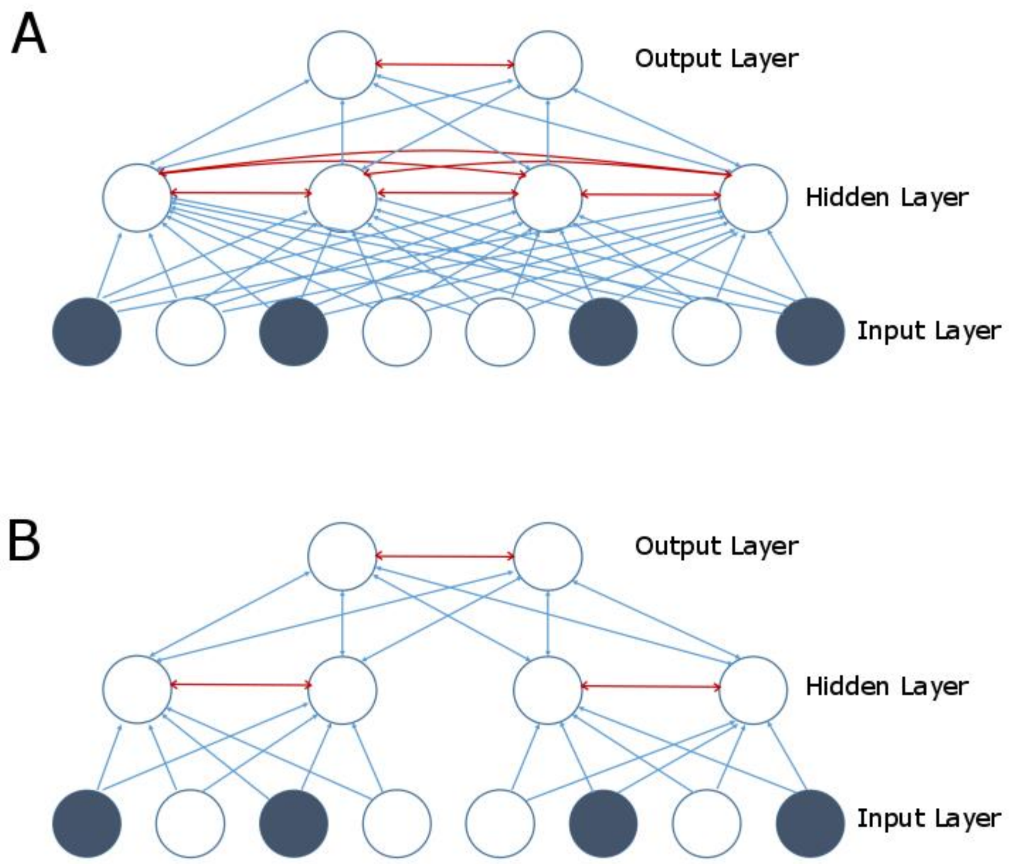

Figure 1 Network configurations for testing input discrimination in top-down feedback condition.

3 Results

3.1 All-to-All connectivity and not feedback

Beginning with our all-to-all network configuration, as shown in figure 1A, we tested the ability of our network to properly distinguish between two different sets of stimuli. We found that out all-to-all network with STDP regulated synaptic weights was able to learn to distinguish between two different patterns of stimulation (figure 2). As shown in figure 1A, the stimulation pattern to the input neurons began with stimulation to the first, third, sixth, and eight neuron. This pattern was followed by a brief period of silence and then the second, fourth, fifth, and seventh neurons were stimulated. Initially, not all neurons within the hidden layer responded to the input layer stimulation (first trial in figure 2B). However, by the third set of stimulations, the network firing pattern in the hidden layer converged to a specific sequence of firing. The output layer neurons, however, did not show alternating patterns of activity as seen in the hidden layer neurons (figure 2C). Rather than forming distinct response patterns, the output layer behaved in a more winner take all condition, in that only one of the two neurons remained active during the periods of stimulation. This may arise from strong lateral inhibition between the two output layer neurons. Taken together, the response in the hidden layer suggests that the network was able to discriminate between different stimulation patterns in the input layer.

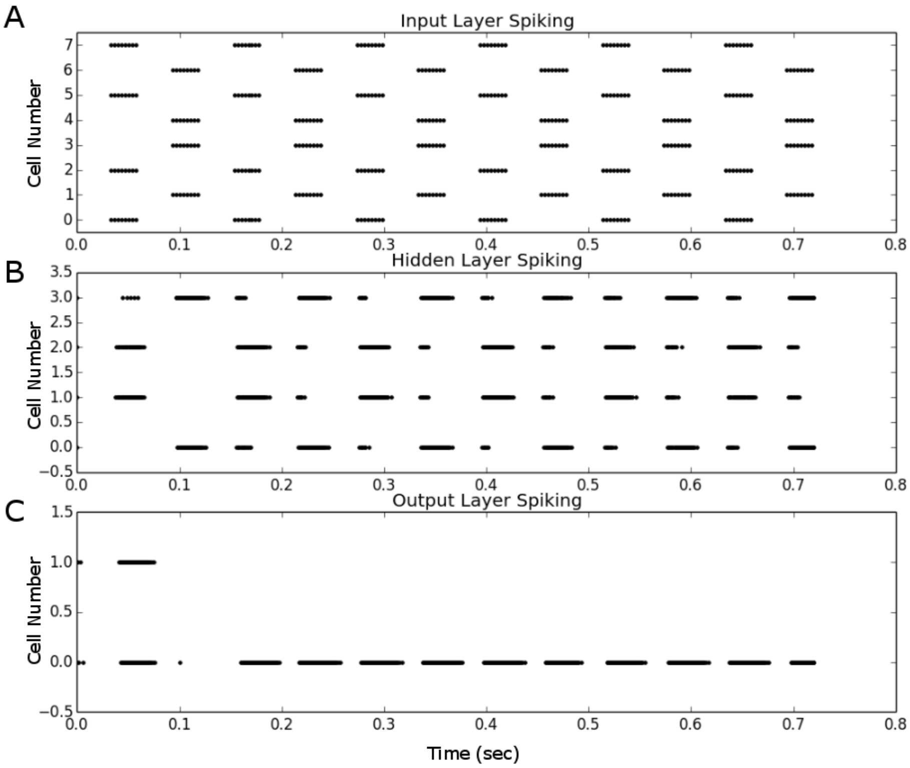

Figure 2 Network spiking patterns in response to varied input layer stimulation patterns.

3.2 All-to-All connectivity with feedback

We next tested the effect of adding top-down feedback on the learning rate of our network. To do this, we added excitatory connections from the output layer to the hidden layer. Having added the feedback, we tested the network with the same input patterns as described above. With strong top-down feedback, our network was no longer able to discriminate between the input patterns (figure 3). As shown in figure 3A, the input layer responded as expected to the input stimulus, however, the hidden layer no longer learned how distinguish between stimulus types. Rather, specific neurons within the hidden layer became overly active, and suppressed activity in the remainder of neurons within that layer (figure 3B). Interestingly, the activity within the over active neurons of the hidden layer continued even during the brief periods of silence in the input layer, suggesting that the network may have entered an unstable, hyperexcitable state. Additionally, the output layer also exhibited hyper excitability similar to that seen in the hidden layer (figure 3C). Unlike the activity output layer in the network without feedback, in this version of the network, both output neurons had continuous spiking pattern.

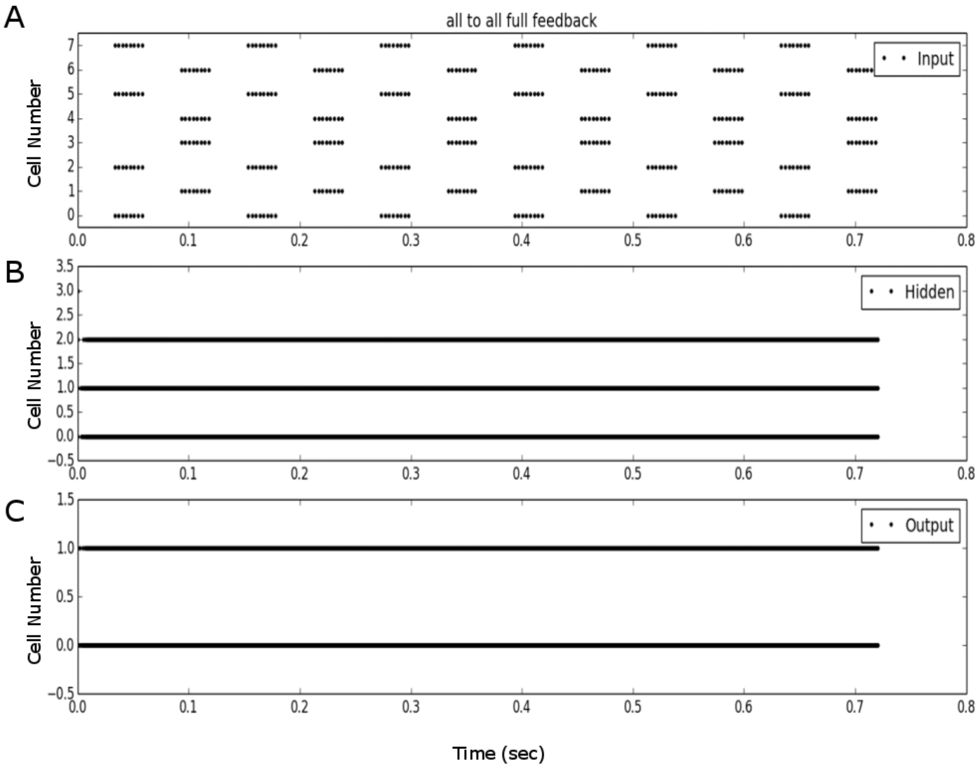

Figure 3 Neuron spiking pattern for the network with strong top-down feedback.

Since strong top-down feedback prevented the network from accurately distinguishing between stimulus patterns, we reduced the strength of the feedback to see if the strength of the feedback could affect the learning of the network. As shown in figure 4, weaker top-down feedback allowed the network to once again learn to discriminate between the two stimulus patterns. Once again, the learning in our network took place in the hidden layer. As shown in figure 4B, the hidden layer slowly converged to specific firing patterns in response to the different stimulus patterns. The rate at which the hidden layer converged to a specific firing pattern did not, however, differ from the rate of convergence in the network without top-down feedback (compare figure 2 and figure 4).

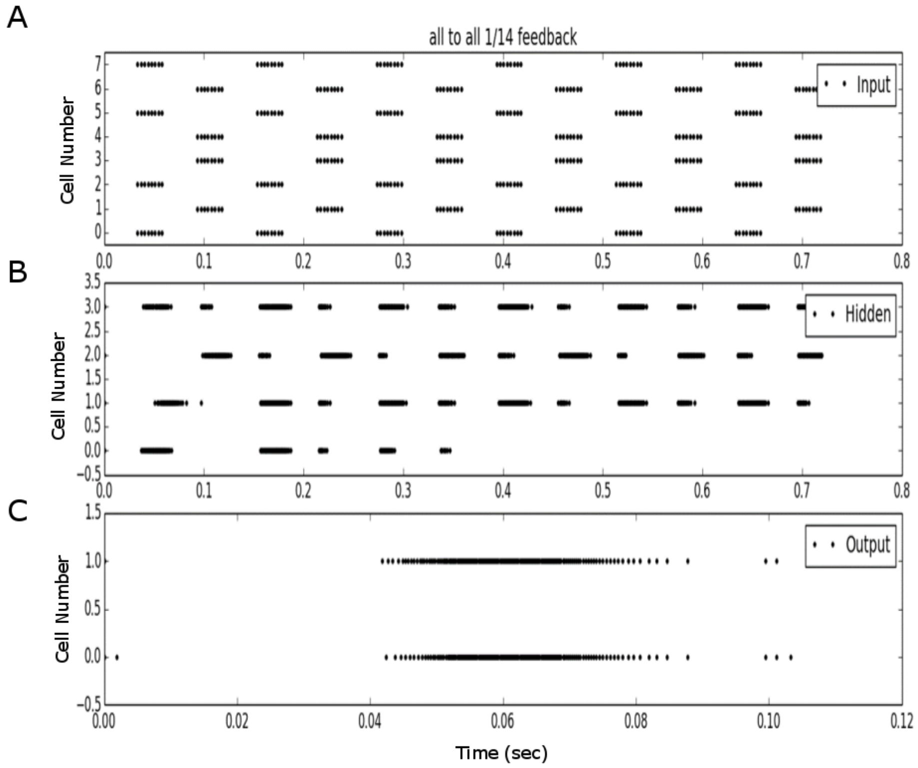

Figure 4 Neuron spiking patterns in a network with weaker top-down feedback.

3.2 Role of intra-layer inhibition

Because the amount of excitability seems to have an effect on the ability of the network to correctly distinguish between the different stimuli patterns, we next tested the effect of intra-layer inhibition of the networks ability to learn. We started by first removing all inhibition within the hidden and output layers. Complete removal of the intra-layer inhibition prevented the hidden layer from learning to distinguish between the two stimuli patterns (figure 5). As shown in figure 5B, similar to previous simulations, the first two stimuli trails resulted in different firing patterns in the hidden layer. The spiking patterns then quickly converged to a network state where all hidden layer neurons fired simultaneously to the input layer stimuli. This is most likely due to the fact that there was no lateral inhibition between hidden layer neurons so there was no competition between the neurons. Similarly, the output layer neurons also showed simultaneous firing in response to the firing in both input and hidden layer neurons (figure 5C).

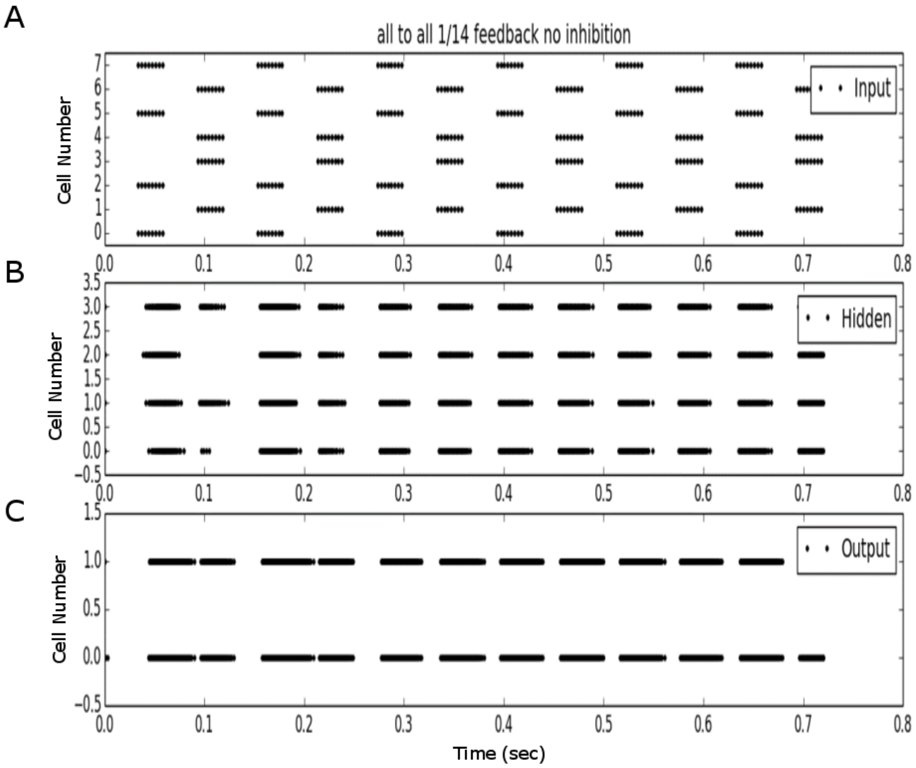

Figure 5 Neuron spiking patterns in a network with top-down feedback, but with no intra-layer inhibition.

Since the lack of inhibition prevented learning in the network, we then tested the effect of strong inhibition (figure 6). We found that the increased intra-layer inhibition also prevented the network from learning. Similar to the network with no inhibition, figure 6B shows that the hidden layer neurons were unable to distinguish between the stimuli patterns. However, unlike the network with no inhibition, in this network, only one hidden layer neurons would respond to the input layer stimuli. This is due to the high amount of competition between the hidden layer neurons that arises from the strong intra-layer inhibition. The output layer neurons only firing occasionally, in stark contrast to the previous network with no inhibition (figure 6C).

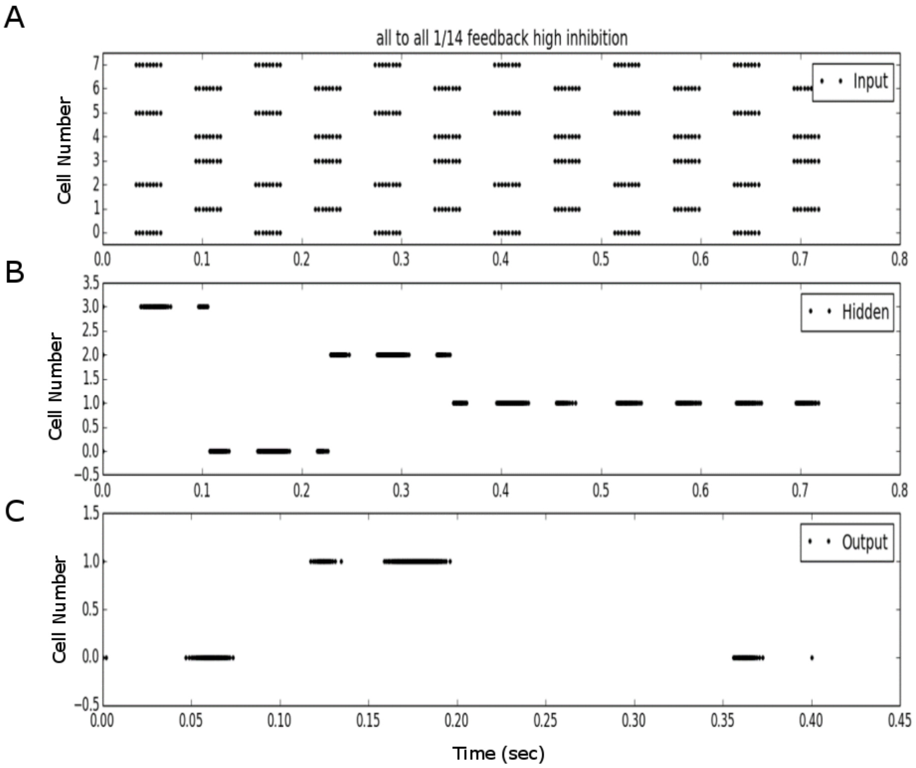

Figure 6 Neuron spiking in a network with top-down feedback and strong intra-layer inhibition.

3.2 Half-to-Half connectivity

Finally, we tested how learning in our network might be affected by reducing the connectivity of the network from all-to-all to half-to-half, meaning that the first four neurons of the input layer only formed synapses on the first two neurons of the hidden layer. Likewise, the last four neurons in the input layer would synapse on only the last two neurons of the hidden layer. A diagram of this network connectivity is presented in figure 1B. Similar to the all-to-all network with no top-down feedback, the half-to-half connectivity with no feedback was able to distinguish between the two different stimuli patterns (figure 7). Figure 7B shows that the hidden layer neurons gradually learned to respond differently to the input stimuli. The rate of learning in this network connectivity was similar to that seen in the all-to-all network (compare figure 2 and figure 7).

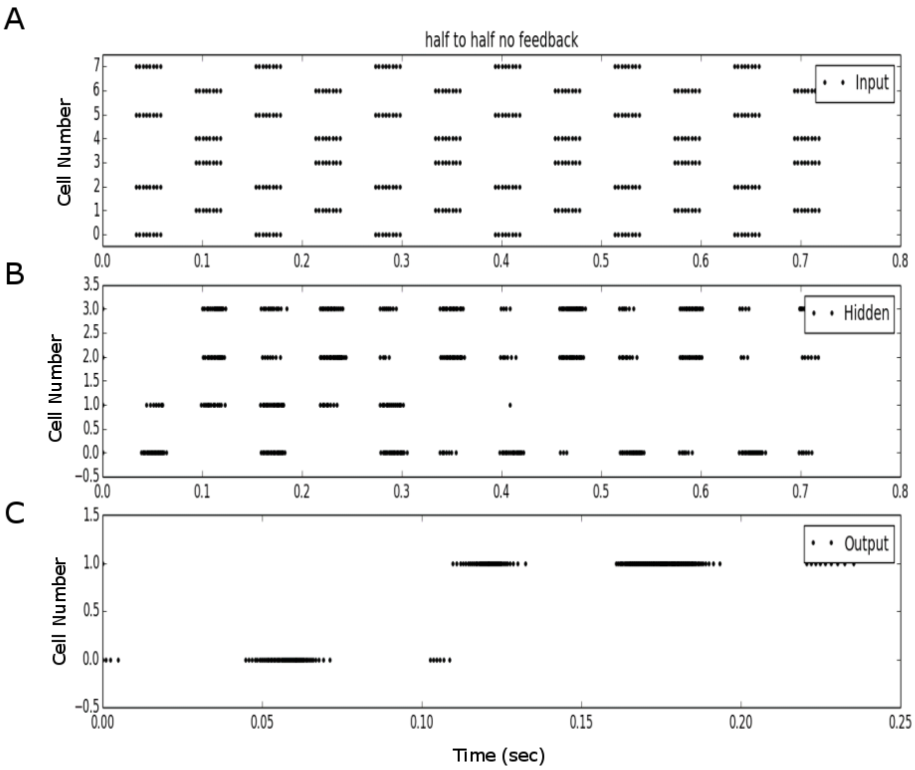

Figure 7 Neuron spiking patterns in a network with half-to-half connectivity without top-down feedback

The addition of the top-down feedback in the half-to-half network did not result in learning rate differences. This network was also able to learn to distinguish between the stimuli patterns (figure 8). However, in all these networks, the learning occurred at the level of the hidden layer. The output layer was unable to distinguish between the different stimuli patterns (figures 2-8 panel C).

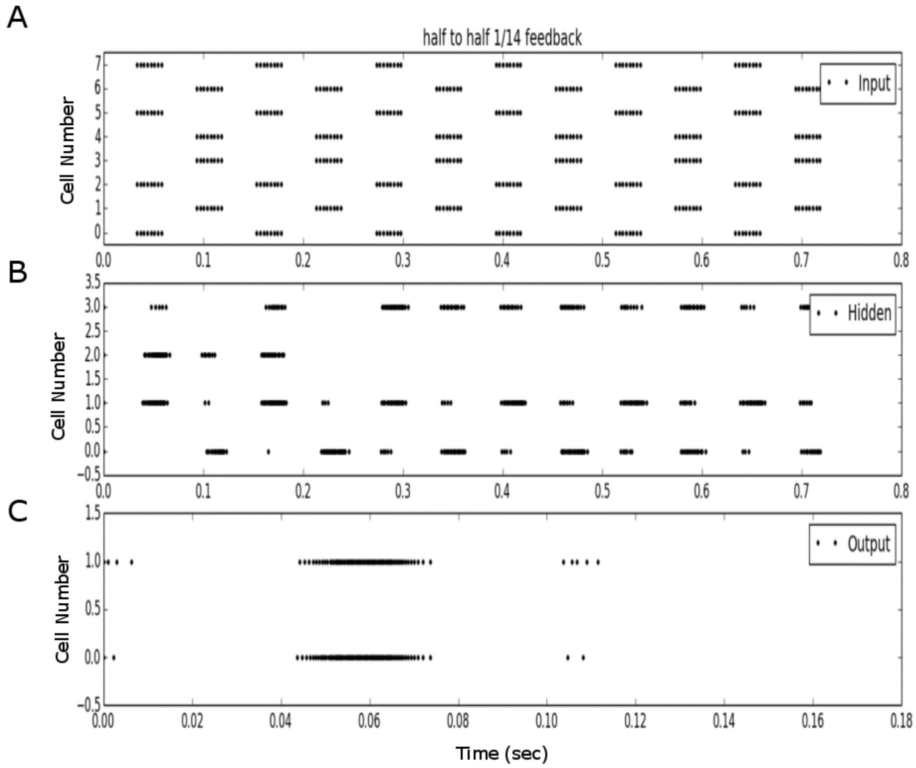

Figure 8 Neuron spiking patterns in a half-to-half network with top-down feedback

To quantify the differences in the learning rates due to the top-down feedback in both network configurations, we measured the number of trails before the hidden layer neurons converged to specific firing patterns as a function of the ratio of feedback to feedforward weight. As seen in figure 9, in either network configuration, the average number of steps to hidden layer convergence did not differ. This suggests that, in our network, top-down feedback from the output layer to the hidden layer did not significantly affect the learning rates.

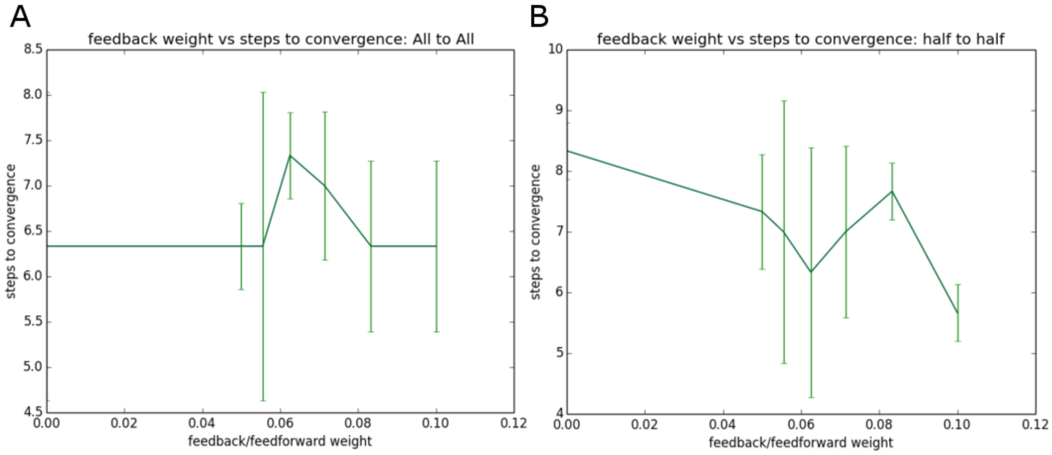

Figure 9 Mean number of step to hidden layer convergence as a function of feedback to feedforward weight ratio.

4 Discussion

In our model, the hidden layer was able to converge to specific, distinct patterns of spiking in response to either input pattern, when inhibition was kept within an optimal range and excitatory feedback was minimal. When mutual intra-layer inhibition was removed, uninhibited spiking of postsynaptic neurons resulted in all feedforward synaptic weights increasing. This led to a firing pattern in which all hidden layer and output layer neurons fired in response to either input pattern. Excessive inhibition resulted in a winner-take-all situation in which a single hidden layer neuron spiked in response to either input. These results demonstrated that an optimal amount of intra-layer mutual inhibition is necessary for learning with input discrimination to occur.

In addition, strong excitatory feedback from output neurons was able to disrupt input-specific convergence by causing persistent spiking in the hidden and output layers, akin to a seizure state. This did not occur with weaker feedback synapses, as these alone were not able to support spiking in hidden layer neurons. While in our model this persistent activity disrupted encoding, persistent spiking supported by recurrent excitatory input has been proposed to underlie working memory **[2]**.

Our model was not able to generate input-specific spiking in the output layer. This may be due to our use of spike-timing dependent plasticity, given that our input consisted of spike trains that would have more effectively increased synaptic weights with a spike rate-based model of plasticity. Likely because of the lack of input-specific spiking in output neurons, excitatory feedback to the hidden layer from these neurons did not improve the learning rate in hidden layer neurons.

References

1. P. Girard, J.M. Hupé, J. Bullier. “Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities.” J. Neurophysiol., 85 (2001), pp. 1328–1331
2. Gazzaley, Adam, and Anna C. Nobre. “Top-down Modulation: Bridging Selective Attention and Working Memory.” *Trends in cognitive sciences* 16.2 (2012): 129–135.
3. Dayan, Peter, and L. F. Abbott. “Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems”. Cambridge, Mass.: MIT, 2001. Print.
4. Moldakarimov S, Bazhenov M, Sejnowski TJ. “Top-Down Inputs Enhance Orientation Selectivity in Neurons of the Primary Visual Cortex during Perceptual Learning.” PLOS Computational Biology 10(8) (2014).