Dissociable Forms of Repetition Priming: A Computational Model

Kirill Makukhin
k.makukhin@webage.net.au
University of Queensland, Brisbane QLD 4072, Australia

Scott Bolland
Queensland University of Technology, Brisbane QLD 4059, Australia

Nondeclarative memory and novelty processing in the brain is an actively studied field of neuroscience, and reducing neural activity with repetition of a stimulus (repetition suppression) is a commonly observed phenomenon. Recent findings of an opposite trend—specifically, rising activity for unfamiliar stimuli—question the generality of repetition suppression and stir debate over the underlying neural mechanisms. This letter introduces a theory and computational model that extend existing theories and suggests that both trends are, in principle, the rising and falling parts of an inverted U-shaped dependence of activity with respect to stimulus novelty that may naturally emerge in a neural network with Hebbian learning and lateral inhibition. We further demonstrate that the proposed model is sufficient for the simulation of dissociable forms of repetition priming using real-world stimuli. The results of our simulation also suggest that the novelty of stimuli used in neuroscientific research must be assessed in a particularly cautious way. The potential importance of the inverted-U in stimulus processing and its relationship to the acquisition of knowledge and competencies in humans is also discussed.

1 Introduction

Repetition priming is a well-studied distinct form of the human and animal nondeclarative memory (Squire, 2004) that refers to the changing of the speed and accuracy of the unconscious detection, recognition, and identification of stimuli or their relative combination, depending on previous exposure to the object. The priming phenomenon is tied in with the changing of neural activity in different cortical areas (Gotts, 2003; Schacter, Wig, & Stevens, 2007; Wig, Grafton, Demos, & Kelley 2005; Wiggs & Martin, 1998; Soldan, Habeck, Gazes, & Stern, 2010; Voss & Paller, 2010; Grill-Spector, Henson, & Martin, 2006). In particular, there is much evidence that the repetition of a stimulus (stimulus novelty) triggers changes in neural

activity in the perirhinal and lateral entorhinal area, while novel temporal or special arrangement of objects (associative novelty) affects the activity in the parahippocampal (postrhinal) cortex as well as in the hippocampus (Wan, Aggleton, & Brown, 1999; Eichenbaum, Yonelinas, & Ranganath, 2007; Brown & Aggleton, 2001; Yang, Meckingler, Xu, Zhao, & Weng, 2008).

Despite a large body of the literature published on repetition priming, the generality of this principle and the underlying neural mechanisms involved are still a matter of debate (Soldan et al., 2010; Schacter et al., 2007). In particular, the recent finding of dissociable forms of repetition priming whereby the repetition of unfamiliar items leads to an increase of activation (Henson, Shallice, & Dolan, 2000) seems to defy the general tenets of repetition suppression and its proposed mechanisms.

In this letter, we review existing models and their limitations and introduce a new theory that potentially describes the neural mechanisms involved in dissociable priming. We provide additional support for this model through the use of computational modeling.

1.1 Models of Repetition Suppression. The most commonly observed form of neural activity underlying repetition priming is repetition suppression, whereby neural activity decreases with representation of a stimulus. Repetition suppression has been reported at a variety of scales: from single neuron recordings (Miller, Li, & Desimone, 1991; Sobotka & Ringo, 1994) to EEG (Gruber & Müller, 2002; Voss & Paller, 2010) and fMRI studies (Demb et al., 1995; Naccache, 2001; Gagnepain et al., 2011). The ubiquity of observations of repetition suppression has led to the emergence of a few theories and models that attempt to explain the phenomenon. (For a more comprehensive review of these models, refer to Grill-Spector et al., 2006.)

Facilitation models (Neely, 1977; Ringo, 1996; Sobotka & Ringo, 1996; James & Gauthier, 2006; McMahon & Olson, 2007) suggest that fMRI repetition suppression may occur due to faster processing of well-known stimuli. That is, when a stimulus is repeated, the corresponding neurons fire with a shorter delay, but they also settle to the baseline activity more quickly. As long as fMRI shows an integrated response over a few seconds of activity, a vigorous but significantly shorter duration of activity causes decreased amplitude of the fMRI signal. A rapid response to a stimulus is well correlated with behavioral increases in processing speed.

According to fatigue models (Anderson, 1976; Miller & Desimone, 1994; Grill-Spector & Malach, 2001), the decay of activity is because of an equivalent reduction in neural response to repeated presentations of the same stimulus. That is, the model assumes that there are no spatial or temporal changes in the neural response pattern. However, as Grill-Spector et al. (2006) pointed out, this theory does not provide an explanation of how reduced firing rates can lead to increased speed and accuracy of processing familiar stimuli as a key property of repetition priming. In addition, the fatigue model was criticized more recently (Summerfield, Trittschuh, Monti,

Mesulam, & Egner, 2008) as unable to explain their findings that repetition suppression also may depend on perceptual expectations.

Another class of models suggests that the sharpening of information representation in the cortex is the main reason for observed repetition suppression (Desimone, 1996; Wiggs & Martin, 1998). According to this theory, if a stimulus is novel for the brain, a relatively large number of neurons respond weakly to it, which leads to the high overall activation registered on the population of neurons. However, with repeated exposure, the neural response becomes sharper with fewer neurons firing (but with higher intensity), leading to an overall reduction in activation.

The sharpening theory is supported by several neurological experiments. For example, Rolls, Baylis, Hasselmo, and Nalwa (1989), Li, Miller, and Desimone (1993), and Kobatake, Wang, and Tanaka (1998) have found enhanced stimulus selectivity in the monkey inferotemporal cortex following stimuli repetition, and Rainer and Miller (2000) showed that familiar visual objects activated fewer but more sharply tuned neurons in the prefrontal cortex compared to novel objects.

A few computational models explore and demonstrate the sharpening of representations with repeated exposure to stimuli. For example, the model of complementary learning systems (Atallah, Frank, & O'Reilly, 2004; O'Reilly, Bhattacharyya, Howard, & Ketz, 2011), employs a competitive Hebbian learning algorithm that demonstrates cortical activation sharpening after repeated presentation of a stimulus. This algorithm relies on a simulated lateral inhibition mechanism in which strong neurons actively inhibit weaker ones. In the case of an unfamiliar stimulus, many neurons are allowed to fire due to a lack of lateral inhibition. Similarly, Moldakarimov, Bazhenov, and Sejnowski (2010) presented a computational model based on a simplified but biologically plausible spiking neural network of the early visual processing areas (V1,V2, and V4) that provides evidence that the sharpening effect may occur due to an interplay between Hebbian learning and lateral inhibition.

1.2 Dissociable Forms of Repetition Priming. The theories presented in the previous section are aimed at explaining decreases in neural activity with repetition of a stimulus. However, these theories seem to be unable to account for the phenomenon of repetition enhancement. That is, a number of studies show an opposite trend: in some cases, neural activity rises with the increasing familiarity of a stimulus.

Potentially, the difference between repetition suppression and enhancement (i.e., dissociable priming) may be explained in terms of an inverted U-shape response with respect to stimulus familiarity. That is, converging evidence suggests that the repetition of unfamiliar items will lead to an increase in activation, whereas repetition of familiar items will yield a decrease (i.e., dissociable priming); that is, moderate levels of familiarity will result in the largest neural response, with highly unfamiliar or familiar items

having less of a response. For instance, Henson et al. (2000) has shown that the repetition of truly unfamiliar visual stimuli, such as unfamiliar faces and symbols, could lead to an increase in neural activity in the right fusiform region measured as an fMRI signal, while repeated exposure to familiar stimuli elicits decaying activity. Similarly, Soldan, Gazes, Hilton, and Stern (2008) have registered repetition enhancement in the occipital-temporal cortex related to the repetition of unfamiliar stimuli and decreases in neural activity for familiar stimuli. A study by Fiebach, Gruber, and Supp (2005) presented similar EEG results showing dissociable trends for the repetition of familiar words versus unfamiliar pseudowords.

Numerous other studies also demonstrate that moderate levels of novelty yield the largest neural response, possibly an inverted U-shaped dependence with respect to familiarity. In the hippocampus and surrounding medial temporal lobe (MTL) cortices, this effect is often found related to "expectation violations" in which both novel (i.e., random) and familiar sequences of spatial or temporal patterns yield a low neural response, whereas patterns that explicitly violate known expectations lead to a large response (associative novelty). For example, Yang et al. (2008) provided evidence based on fMRI human research where the right parahippocampal cortex showed increased activation for unexpected stimulus pairs. Schott et al. (2004) and, independently, Düzel, Habib, Guderian, and Heinze (2004) presented results where participants distinguished between familiar and novel configurations of pairs of items. The results show that the activation of the rhinal cortex (as well as of the hippocampus) rises with increasing novelty of stimuli from highly familiar to partially novel. Similarly, Hunkin, Mayes, and Gregory (2002) have demonstrated that the activity of the posterior cingulate cortex rises when subjects hear a novel combination of word triplets. Finally, Kumaran and Maguire (2007) report that hippocampal neurons increase their baseline firing in response to a novel stimuli rearrangement.

In summary, evidence suggests that both neural enhancement and suppression can be expected with stimuli repetition, depending on the familiarity of the stimuli, following an inverted U-shaped trend.

1.3 Modeling the Inverted U-Shaped Response. It seems that to date, there is a lack of models that explicitly attempt to describe dissociable forms of repetition priming. Henson et al. (2000) speculate that repetition enhancement may be due to "the formation of new representations" (p. 1272) that occur with exposure to novel stimuli; however, they suggested no concrete neural mechanisms underlying the phenomenon.

In the following sections, we introduce a theory of how the inverted-U may emerge in cortical layers through an interplay between Hebbian learning and lateral inhibition. We then implement this theory in a spiking neural network, and demonstrate the emergence of this inverted-U with respect to varying levels of familiarity. Then we introduce a more sophisticated model of a cortical layer and replicate the dissociable priming effects that

Henson et al. (2000) found using real-world stimuli. A general discussion
and implications for future research conclude the letter.

2 Cortical Processes Potentially Underlying the Inverted U-Shaped Curve

Henson et al. (2000) have demonstrated two dissociable trends in novelty
processing by the brain. One process is the well-known form of repetition
priming, namely, repetition suppression. It occurs when the brain repeat-
edly observes a familiar stimulus and is characterized by decreased neural
response over sequential representations. We believe that the sharpening
theory (Desimone, 1996; Wiggs & Martin, 1998) accounts for the repetition
suppression, whereby neurons "drop out" of the representation with in-
creased familiarity. Similar to O'Reilly (1996) and Biederman and Vessel
(2006), we believe that lateral inhibition may play a key role in explaining
this sharpening, whereby the strong neurons suppress the weaker.

The second process Henson and colleagues found is repetition enhance-
ment that occurs when a truly unfamiliar stimulus is repeatedly shown.
As neurons can be viewed as pattern detectors, we assert that true novelty
exists when there are no neurons that strongly encode the stimulus itself.
We believe that in this case, the initial neural response will be low, because
there will be a low probability of neurons firing to the small partial match
(see Figure 1A). However, with repetition of the stimulus, its representation
in the brain begins to form due to Hebbian learning. That is, with exposure,
the probability that neurons will fire will increase over time, giving rise to
increased activity (see Figure 1B). This represents the intermediate level of
familiarity condition.

Finally, repeated exposure, Hebbian learning further strengthens neural
connections, making the recipient neurons respond more actively. However,
via lateral connections, this will trigger inhibitory neurons that can suppress
weaker units and reduce overall activation (see Figure 1C), which represents
a well-known stimulus.

Before describing the simulation of dissociable forms of repetition prim-
ing, we introduce a model of a cortical layer based on our theory. We
empirically confirm that the proposed model exhibits both sharpening of
representation and initial rising of network activity due to a process similar
to Hebbian learning, demonstrating an inverted U-shaped dependence of
the network activity with respect to the novelty of stimuli.

**2.1 A Spiking Neural Network with Hebbian Learning and Lateral
Inhibition.** The purpose of the simulation is to empirically examine
whether interplay between Hebbian learning and lateral inhibition can
exhibit an inverted U-shaped dependence of its activity with respect to
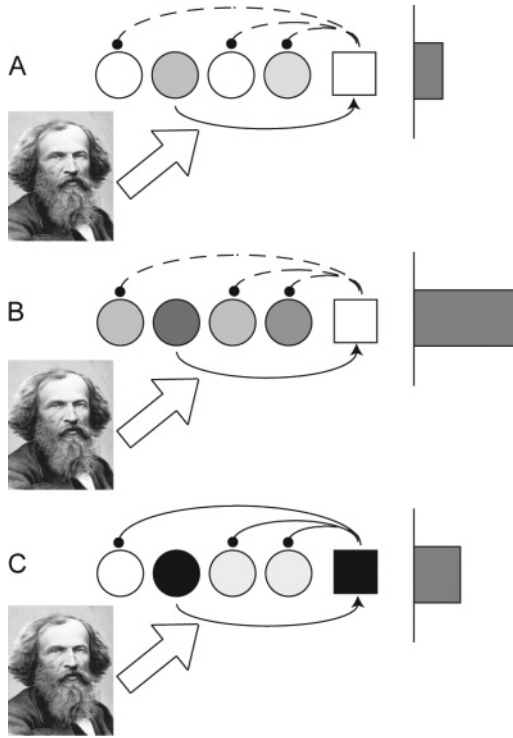stimulus familiarity.

Figure 1: The proposed neural mechanism forms an inverted U-shaped response to novelty. Circles represent excitatory neurons, and squares represent inhibitory neurons. Not all connections are shown for simplicity. The bars at the right show overall layer activity and, thus, expected fMRI response. (A) Unfamiliar stimuli weakly activate a range of neurons. (B) With subsequent repetition, Hebbian learning strengthens these responses, giving rise to a higher activation level. (C) Further learning leads to triggering the inhibitory neurons, suppresses weaker units, leading to sparsification and decreased neural activation.

The simulation network has two sublayers: an input layer that contains a single node and an output layer consisting of 5000 excitatory and 1000 inhibitory units (see Figure 2). The single input node was used to represent a single stimulus, with increased exposure modifying the initially weak connections to the output layer, through Hebbian learning. This simplification is feasible because the firing of the excitatory neurons depends on their net input, making the resulting activity independent of the process of the summation of dendritic signals (meaning that a more complex stimulus pattern is not required in our case).
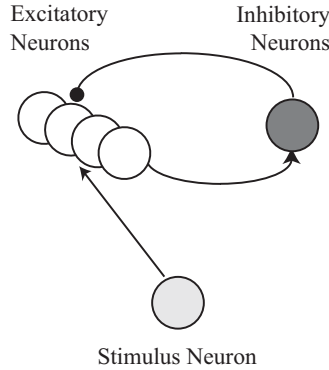
Figure 2: Layout of the spiking neural network with lateral inhibition used in the experiment. The sublayer of the only stimulus neuron is fully connected to the sublayer of the excitatory neurons. The inhibitory neurons are reciprocally fully connected to the excitatory units.

The layout of the network reflects the structure of the cortex, where lateral inhibition is derived through connections with intermediate inhibitory interneurons (Gibson, Beierlein, & Connors, 1999). Full connections exist between the input unit and the excitatory neurons (initialized to small, random weights denoting initially low familiarity of the stimulus), as well as reciprocal connections between the excitatory and inhibitory units that are not plastic and have fixed and relatively strong weights. It is known that about 20% of neurons in the cortex are inhibitory interneurons (Beaulieu, Kisvarday, Somogyi, Cynader, & Cowey, 1992), with our simulation using the same ratio of excitatory to inhibitory neurons.

The stimulus neuron is firing at a constant rate of 10 spikes per second. Over time, Hebbian learning strengthens the connection between the stimulus and excitatory units, representing an increase in familiarity to this pattern. We record the average firing rate of the output sublayer (both inhibitory and excitatory neurons) to measure the activation levels that may be captured by fMRI studies.
'

2.1.1 *Spiking Neurons and Hebbian Learning.* In this experiment we are using a biologically plausible spiking neuron model suggested by O'Reilly and Munakata (2000). The neurons are of the two-compartment integrate-and-fire type, meaning that when a pulse is received as an input signal, the membrane potential increases slowly over time (first flowing through the dendritic compartment and then into the soma), leading to an action potential being fired sometime later if the membrane potential exceeds a threshold.
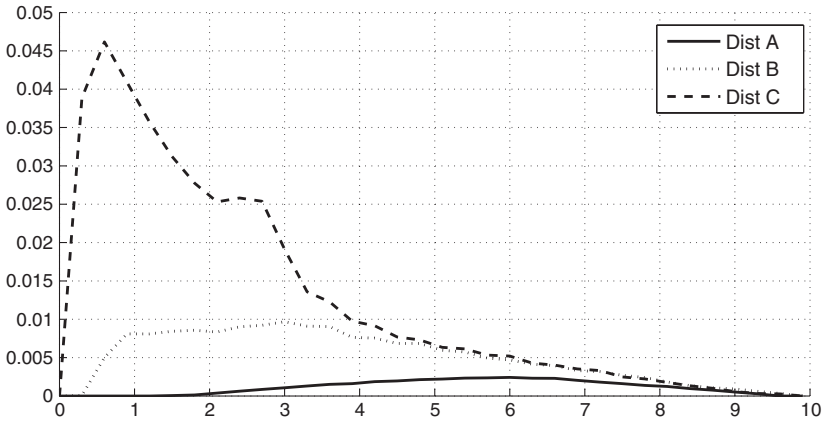
Figure 3: The distribution of neurons' firing with time for three situations: the stimulus is novel (distribution A), partially learned (distribution B), and well learned (distribution C), when the lateral inhibition is turned off. The graph depicts both the number of neurons responding and their latency (temporal distance from the discharge of the input neuron).

Importantly, the inhibitory neurons in the experiment are faster than excitatory neurons, reflecting the dynamics of the brain (Somogyi, Tamás, Lujan, & Buhl, 1998). The connections between the stimulus neuron and excitatory neurons are able to learn, representing the growing familiarity of the input stimulus. The learning is implemented as a form of spike-timing-dependent plasticity, Hebbian learning, following the general theory that neurons that fire together should strengthen their connectivity and otherwise weaken (Gerstner & Kistler, 2002). The appendix provides a detailed description of the models and parameters used in the simulation.

2.1.2 Results. At the beginning of training, when a stimulus is novel for the network, the connection weights between the stimulus and output excitatory neurons are weak. As a result, only a few neurons receive over-threshold net input and fire. Due to a small amount of random noise that is added to the net input of each unit, even units that receive subthreshold activation could occasionally fire with low probability. Thus, a wide range of neurons can be strengthened through Hebbian learning, increasing the population of firing neurons. Over time, more and more neurons start firing, increasing the overall activity. Figure 3 illustrates the changes in the neural activity when the lateral inhibition is temporarily switched off. Distribution A represents the neural activity (the number of neurons firing with respect to their latency behind the stimuli neuron) when a stimulus is new to the network, distribution B when the stimulus is partially learned,
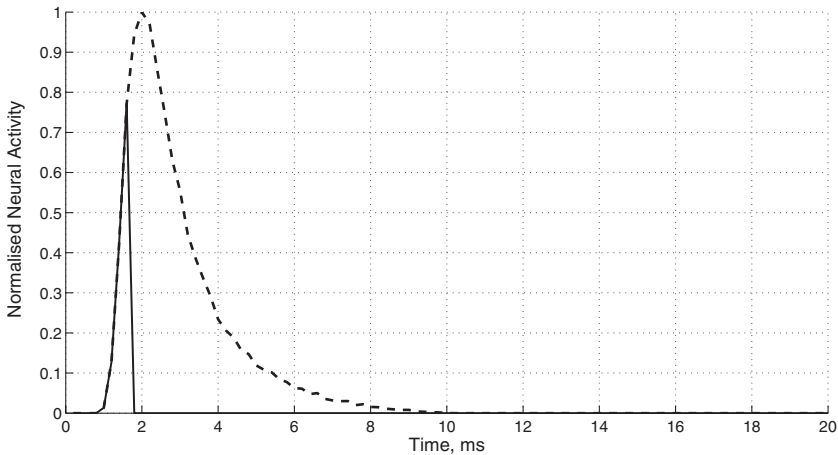
Figure 4: Distribution of the neurons' activity with (solid line) and without (dashed line) lateral inhibition for a well-learned stimulus. Inhibitory neurons cut off the right (latest) part of the distribution, thus reducing overall activation (the effect of sharpening of representation).

and distribution C when the stimulus is well learned by the network. Figure 3 clearly illustrates the increasing nature of activation with respect to familiarity when learning is governed purely by Hebbian mechanisms.

An interesting phenomenon displayed by this model is that not only does the general population increase in activity with learning, but the time delay for neurons to fire also decreases with familiarity. That is, the better a neuron encodes a stimulus, the faster it will be in responding to it. This temporal dependence is well supported by empirical data (Sobotka & Ringo, 1996; James & Gauthier, 2006; Balu, Larimer, & Strowbridge, 2004).

Figure 4 shows the condition when lateral inhibition is implemented in the network. As can be seen, the faster, winning neurons activate the inhibitory neurons, which suppress the slower, losing neurons, cutting off the right part of the distribution and thus resulting in decay of activity.

When the pieces are added together, the results demonstrate that the activity of the network is low for an unfamiliar stimulus because only a few neurons fire, mostly due to their stochastic nature. When a stimulus becomes partially familiar, a large population of excitatory neurons is activated, but because of the weak connections, they fire with a significant time delay. At this time, more activation is spread to the reciprocally connected inhibitory units that also start increasing in population. However, due to a large firing time delay, the inhibitory neurons are unable to affect the overall activity. The situation changes when a stimulus is well known to the network. The strong synaptic connections guarantee a short firing delay, and thus the
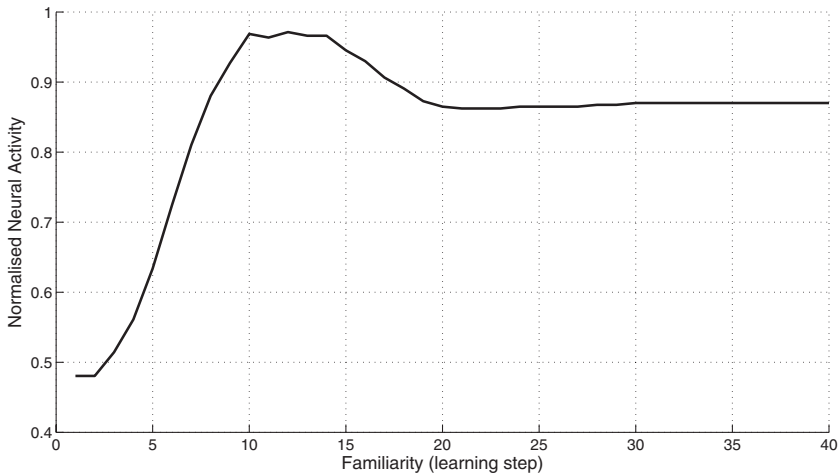
Figure 5: The inverted U-shaped function emerged in the biologically plausible simulation of a spiking neural network with lateral inhibition and Hebbian learning.

inhibitory neurons increase their control over the excitatory units, resulting in decreasing overall activation. Figure 5 shows the resulting inverted U-shaped curve of the relative network activity with learning.

2.1.3 Discussion. The simulation results confirm that the neural activity of a spiking neural network with Hebbian learning and lateral inhibition could naturally exhibit an inverted U-shaped dependence with respect to stimulus novelty. Indeed, the initial increase of the firing neuron population due to the stochastic nature of neurons and Hebbian learning provides the rising activity while forming representations of unfamiliar stimuli. Thereafter, with continuing learning of partially familiar stimuli, the sharpening effect takes over, and the overall layer activity starts to decrease.

2.2 A Restricted Boltzmann Machine with Explicit Lateral Inhibition. In the previous experiment, we showed that the neural activity of a spiking network with Hebbian learning and lateral inhibition could naturally follow an inverted-U shaped function with respect to the novelty of input stimuli, providing both rising activity for unfamiliar stimuli and decaying activity for familiar stimuli. However, replicating neurological experiments requires using complex network structures and real-world input stimuli, thus introducing limitations in the use of spiking networks that are computationally expensive and involve fine-tuning a large number of interrelated parameters. To overcome these limitations, in this section we introduce

a robust and efficient rate-coded neural network model that applies the main characteristics of our theory to a restricted Boltzmann machine (RBM) (Smolensky, 1986; Hinton, 2002, 2010).

RBMs and their variations have been well studied over the past decade and have been shown to be very efficient in many classification, recognition, and dimensionality-reduction tasks (Tieleman, 2008; Nair & Hinton, 2010; Ranzato, Krizhevsky, & Hinton, 2010; Luo, Shen, & Niu, 2010).

Generally an RBM is a network of visible and hidden units that are bidirectionally connected to each other. The units make stochastic decisions about whether to be "on" or "off" depending on their net input. The probability of turning the unit on is:

$$p(h_j = 1) = \frac{1}{1 + e^{-z_j}}, \tag{2.1}$$

$$z_j = b_j + \sum_i v_i \cdot w_{ij}, \tag{2.2}$$

where w_{ij} is the weight of the connection between the ith visible and jth hidden units.

This type of network can be trained to extract statistical regularities from input samples by minimizing the joint energy of visible and hidden units, where the gradient is calculated as

$$\Delta w_{ij} = \left\langle v_i h_j \right\rangle^{data} - \left\langle v_i h_j \right\rangle^{equilibrium}. \tag{2.3}$$

In this equation, $\langle\rangle^{data}$ denotes the expected value of the input multiplied with the inferred hidden states while the input is clamped on the data points, and $\langle\rangle^{equilibrium}$ is the expectation of $v_i h_j$ when the alternating Gibbs sampling of the hidden and visible units was (infinitely) iterated to get samples from the equilibrium distribution (see Hinton, 2002, for details).

For training the model, we use a simple and computationally efficient algorithm that approximates the gradient: the contrastive divergence (CD) algorithm (Hinton, 2002). In brief, the calculation of the weights' gradients consists of two phases. First, at a positive phase, the state of the hidden units h_j^{data} is calculated according to equation 2.1. Then a negative phase starts, and the "fantasy," which is the "believed" input value v_i^{recon}, is reconstructed back given the state of the hidden units and symmetrical weights. This is followed by the calculation of the negative-phase hidden units' state h_j^{recon} given v_i^{recon}. Instead of taking the infinite number of Gibbs sampling iteration, the chain stops after the first loop. The difference between these

two phases is used in the weights' and biases' update:

$$\Delta w_{ij} = \langle v_i h_j \rangle^{data} - \langle v_i h_j \rangle^{recon},$$

$$\Delta b_i = \langle v_i \rangle^{data} - \langle v_i \rangle^{recon},$$

$$\Delta c_j = \langle h_j \rangle^{data} - \langle h_j \rangle^{recon}. \tag{2.4}$$

A more detailed description of the prototypical RBM implementation, including Matlab code examples, can be found in Hinton and Salakhutdinov (2006).

In addition to the prototypical RBM, our implementation explicitly imposes lateral inhibition, which is essential for representation sharpening. This is achieved through incorporating a fuzzy variation of the k-winners-take-all algorithm, suggested by O'Reilly and Munakata (2000), where only a set number of units (k) is permitted to become active:

$$prob(s_i = 1) = \alpha_i \frac{1}{1 + e^{-zi}}, \tag{2.5}$$

where

$$z_i = x_j w_{ij},$$

$$\alpha_i = \begin{cases} 1, & \text{for the first } k_{wta} \text{ units with the strongest net input } z_i \\ 0, & \text{for all other units} \end{cases}.$$

The vector α_i is obtained during a positive phase, and the same vector is also used to mask the activity of hidden units during the negative phase of the CD algorithm.

The proportion of allowed winners k_{wta} is not fixed in our model. Instead, as in the previous simulation, the value of k is determined by the strength of the winning units, with k decreasing as the winners become stronger:

$$k_{wta} = k_0 + k_{add}, \tag{2.6}$$

$$k_{add} = \frac{\varepsilon}{1 + e^{-\gamma C}}, \tag{2.7}$$

$$C = \left(1 - \frac{K_{av}}{K_{exp}}\right) - \beta, \tag{2.8}$$

where $k_0 = 0.1$, $\beta = 0.25$, $\varepsilon = 1 - k_0$, and $\gamma = 10$. The average activation K_{av} is the mean value of the net input of the k winning units among the hidden layer $K_{av} = \|\alpha_i z_i\|$. The expected activation of the winning units K_{exp} is calculated as the running average of K_{av}: $K_{exp}^{t+1} = (1 - \theta) \cdot K_{exp}^t + \theta \cdot$
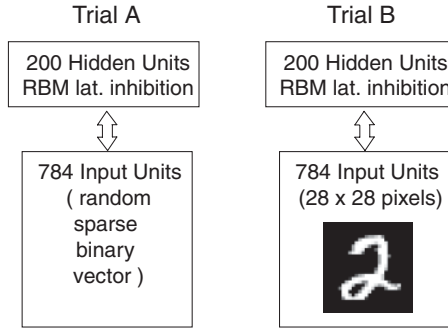
Trial A Trial B

Figure 6: The experiment setup. A modified RBM with explicit lateral inhibition has 784 visible (input) and 200 hidden units. For trial A, randomly generated sparse binary vectors are used as stimuli. For trial B, images from the MNIST data set are presented to the input layer. The activation of the hidden layer is recorded while the learning occurs.

K_{av}, where $\theta = 10^{-3}$. When the net input is low, k_{wta} is high, and the average activation can grow freely. However, when the internal representation of a stimulus has formed, the net input rises and k_{wta} becomes smaller, which leads to reduced the average activity. This mechanism plays the role of lateral inhibition, and the parameter k_{wta} is effectively an analog of the time delay in spiking networks. The average activation of the network in the model is measured as the sum of the probabilities for the k_{wta} winning hidden units to be turned on.

The following section describes an implementation of this model, where we evaluated the effect of familiarity on neural activation given real-world stimuli.

2.2.1 Experiment. The aim of this experiment was to explore the capability of the system to exhibit an inverted-U shape of neural activation with respect to familiarity on complex (and real-world) patterns, in contrast to the single node representation used for the previous simulation.

The network layout is depicted in Figure 6. To explore generality, the experiment consisted of two versions of stimuli. The first trial of the system (A) used randomly generated patterns of 1s and 0s with the sparsity about 0.1,[1] while the second trial (B) used meaningful real-life data, specifically, images of handwritten characters from the MNIST data set (LeCun & Cortes, 1998).

In both trials, the model consisted of a modified RBM with lateral inhibition with 784 input (visible) and 200 hidden units. The model was

---

[1]The sparsity of stimuli must be low to overcome the limitations of a simple binary RBM as discussed by Tang and Sutskever (2011).

pretrained before trials on two patterns that allowed filling the network weights and parameters with initial "knowledge," thus, minimising the influence of random initialization.[2]

During the experiment, the patterns already known by the network were interleaved with a new test pattern (to mimic the more natural occurrence of repetition effects that exist even when the repetitions are temporally segregated), while the average activation of the network (the sum of probabilities) was recorded. The model was trained for 60 iterations by using the contrastive divergence algorithm with a learning rate of 0.002.

2.2.2 Results. Both trials demonstrated similar trends; Figure 7 summarizes the results. Figure 7B shows the average probability of the state among the 10% most active units increases with the learning of an unfamiliar pattern and approaches one, while the reconstruction error is decaying (see Figure 7A). At the same time, the winning probabilities for already familiar patterns stay high—close to one (the large-dashed and fine-dashed lines, mostly overlap each other).

2.2.3 Discussion. Consistent with our expectations, the normalized activation of the hidden units, measured as the sum of all the probabilities divided by the number of units, follows an inverted U-shaped curve with respect to growing familiarity (see Figure 7C). The same trend is demonstrated in both trials.

In the following section, we extend the experiment described with a preprocessing deep network to demonstrate that our model can simulate dissociable forms of repetition priming using real-world data, closely replicating the design and results found in the experiment that Henson et al. (2000) conducted.

3 Dissociable Forms of Repetition Priming

The following section describes a computational experiment that replicates the findings of Henson et al. (2000). The experiment uses the same design as the original study: in two separate trials, either photographs of famous faces (or meaningful symbols) or photographs of nonfamous faces (or meaningless symbols) were repeatedly presented to subjects, and the changes in neural activity were recorded.

3.1 Preprocessing Layers. The real cortex has a hierarchical structure that permits a higher degree of generalization than a single layer. For

---

[2]At trial B, we used five instances of a particular character (chosen randomly at the beginning of test) instead of patterns to evaluate the robustness of the model to small variations of stimuli.
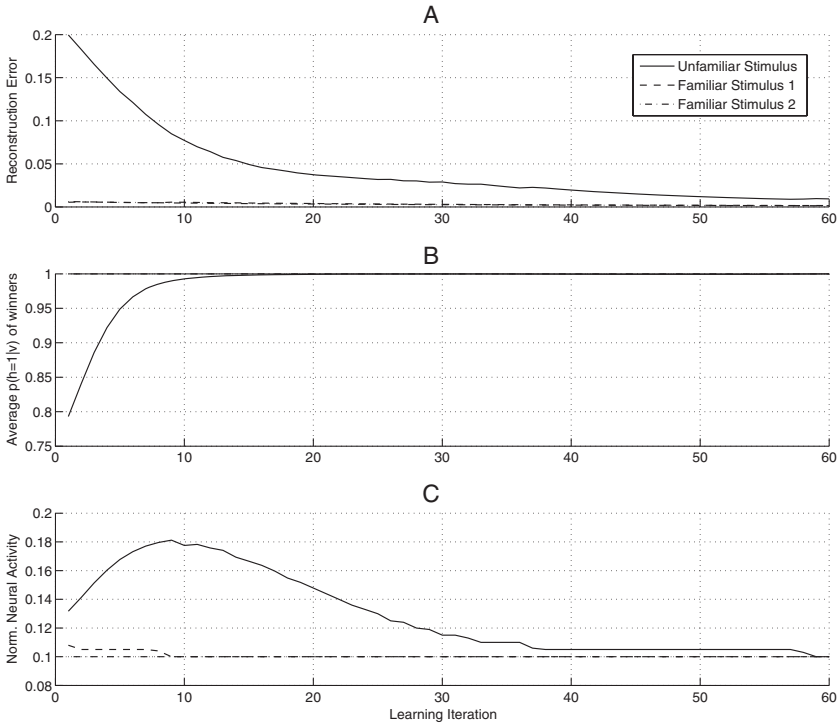
Figure 7: The process of learning a novel stimulus by the proposed model. (A) Reduction of the reconstruction error. (B) Average probabilities to be turned on of 10% winning units given the data. (C) The normalized layer activity of the model network with respect to learning (level of familiarity). The curves for stimulus 1 and 2 mostly overlap.

example, in the ventral visual pathway, superficial cells respond to local specific features, such as bars and edges, whereas in the higher layers, cells respond to complex shapes and objects with a higher degree of invariance (Palmer, 1999). We believe that the ability to generalize is an important property that must be a part of the simulation in order to successfully replicate experiments in which real-world stimuli are used as an input. For example, the repetition suppression effect should exist even when the stimulus is a variant of the original, such as a given celebrity in a different pose.

In order to provide a generalized representation of real-world input stimuli in the simulation, we added a preprocessing module that consists of a deep belief network (DBN) (see Figure 8), identical to the network used in Hinton, Osindero, and Teh (2006). The DBN was separately pretrained before the experiment on a large data set of images (the same data set that is
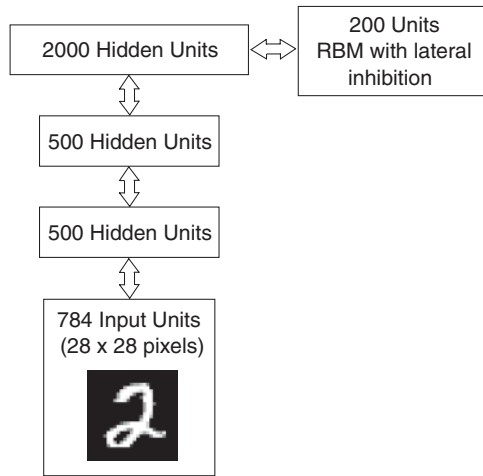
Figure 8: The layout of the network used in the experiment. (Left) A deep belief network (Hinton et al., 2006). (Right) A model of a cortical layer comprising a modified RBM with explicit lateral inhibition.

used in the main experiment—MNIST data set; LeCun & Cortes, 1998) and was kept unchanged during the experiment. We emphasize that the DBN here simply converts realistic visual stimuli (i.e., handwritten characters) to a wide generalized representation (features) that is used as input to the model of a cortical layer. Thus, the DBN might be replaced by another algorithm that is capable of forming such a representation.

The DBN consisted of an input layer of 784 visible units and three hidden layers of 500, 500, and 2000 units. It was trained one layer at a time with the contrastive divergence algorithm (Hinton, 2002) without fine tuning by backpropagation. Features extracted at the last hidden layer are used as an input vector for the model of a cortical layer.

3.2 Stimuli. Like the human brain studies, the simulation uses real-world stimuli analogous to the pictures of symbols, objects, and faces commonly used in fMRI and single-cell recording studies. Specifically, we used 100 random instances of ten handwritten characters randomly chosen from the MNIST data set (LeCun & Cortes, 1998; see the examples in Figure 9).

3.3 Experiment and Results. In order to explore dissociable forms of repetition priming, training of the network on the test character occurred under two conditions: when it had a low and a high level of current familiarity. In the "unfamiliar stimulus" condition, the model of a cortical layer was pretrained on the images of handwritten digits from 0 to 9 except for

Figure 9: An example of the instances of handwritten characters 2 and 6 from the MNIST data set used in the experiment.
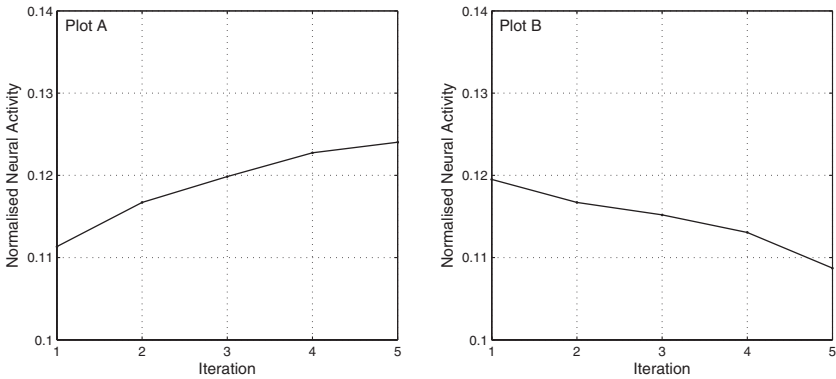

Figure 10: The simulation of dissociable forms of repetition priming. The plots depict changing in the model neural activity with learning (that represents rising stimulus familiarity). (A) The neural activity rises with repetition of an unfamiliar stimulus (repetition enhancement). (B) Decaying activity with repeated exposition to a familiar stimulus (repetition suppression).

one test digit. The layer is then trained for the test digit interleaving with random instances of all other characters for five epochs (exploring the effect of learning on an unfamiliar stimulus), with the dependent variable being the "activity" of the cortical layer. The results are shown in Figure 10A, where the relative neural activity is calculated as the sum of the probabilities of output units in the model of a cortical layer to be "on" divided by the number of the output units.

In the "familiar stimulus" condition, by contrast, the network was trained for a new instance of a familiar character. This condition is analogous to the conditions in fMRI studies when a picture of a familiar symbol or a

famous celebrity face photograph was used: subjects will be able to recognize the content of the image although they may never have seen the exact image before. The results are depicted in Figure 10B. Both experiments are repeated 10 times with using different digit for each simulation for cross-validation.

4 Discussion

There seems to be converging evidence to suggest an inverted-U type function of neural activation with growing stimulus familiarity. Specifically, novel and familiar stimuli induce different types of response in neural activity with repeated stimulus exposure, and intermediate levels of novelty (such as that found in expectation violations) seem to yield the highest neural response. To date, however, in the priming literature, there has been little discussion about the mechanisms underlying this trend.

In this letter, we have examined a theory and associated computational model that could offer an explanation to the neural mechanisms underlying dissociable forms of repetition priming, as well as how the activity of cortical structures generally may depend on novelty. First, we have shown through a computer simulation that a simple model of a spiking neural network with Hebbian learning and lateral inhibition may naturally demonstrate both rising activity with learning unfamiliar stimuli and sharpening of the representation with repetition of familiar stimuli. We then explored how this mechanism may account for dissociable priming effects in experiments that use real-world data.

We have shown that in conditions similar to the neuroscientific study by Henson et al. (2000), the model is able to naturally demonstrate similar results: repetition suppression for familiar stimuli and repetition enhancement for novel stimuli. Specifically we argue that in accordance with the sharpening theory (Desimone, 1996; Wiggs & Martin, 1998), the decline of neural activity with familiar stimuli repetition seems to be caused by lateral inhibition. Indeed, on repeated exposure, neural connections are strengthening due to Hebbian learning, but the strongest neurons inhibit the weaker ones. Hence, the neural representations become sparser, which could be registered as declining activity on an fMRI signal.

A similar process appears able to explain the rising activation for unfamiliar objects. Hebbian learning tends to increase the initially negligible accidental response of the stimulus-specific population of neurons, thus raising the fMRI signal until lateral inhibition starts affecting the population of neurons.

It is widely accepted that there are at least two complementary mechanisms for long-term memory, where neocortical contributions are responsible for memorizing generalized nondeclarative concepts (e.g., "Where is the best place to put my car keys?"), while remembering specific facts and

events (e.g., "Where did I put my car keys today?") is processed by the medial temporal lobe, including the hippocampus and surrounding cortical structures (Squire, 2004; Atallah et al., 2004; O'Reilly et al., 2011). The theory and model presented in this letter are aimed at replicating dissociable forms of repetition priming observed in the fusiform and occipital-temporal cortices, but we speculate that the inverted U-shaped dependence of neural activity with respect to the novelty of observed data could be a more general property of the cortex, including deep structures, such as the rhinal, inferior temporal, parahippocampal cortices, and, arguably, the hippocampus. That is, similar mechanisms may explain raised neural activation in the hippocampus during expectation violations (coinciding with moderate levels of familiarity, where there is only a partial match to preexisting knowledge and expectancies).

The phenomenon of repetition priming seems to be a multifaceted complex process that possibly involves multiple mechanisms. For example, Summerfield et al. (2008) showed that repetition suppression, measured as an fMRI signal in the fusiform face area of the human brain, could be a consequence of a perceptual top-down process that reflects a reduction in prediction error when the brain processes expected stimulus. Recently, Larsson and Smith (2012) provided additional evidence for the existence of different mechanisms of repetition suppression. They showed that in most visual areas, repetition suppression reflects a combination of perceptual expectation and neural adaptation. However, such theories are highly consistent with our own. That is, similar to expectation violations in the hippocampus, top-down expectations may strengthen the activation of winning neurons, increasing the lateral inhibition of the weaker units, leading to sparser representations.

4.1 Further Implications. While our model has some algorithmic similarity to the implementation of the Leabra system (O'Reilly & Munakata, 2000) and its further extension, the complementary learning system (CLS) (O'Reilly et al., 2011), the current research is focused on different goals. In contrast to the CLS, which offers a model of the operation of the MTL memory system, our model explores fundamental neural mechanisms that may be present in various memory areas, giving rise to an inverted-U shape function of neural activation with respect to familiarity.

Interestingly, as explored by Biederman and Vessel (2006), many of the areas that exhibit repetition effects also seem to encode perceived pleasure. That is, the density of μ-opioid receptors is highest in higher levels of the cortex and brain structures such as the parahippocampal cortex. From this perspective, the existence of the inverted-U that we hypothesize in this letter is further supported by behavioral experiments that demonstrate similar trends for perceived pleasure and liking with varying levels of familiarity

(Brennan, Ames, & Moore, 1966; Karmel, 1969; Heyduk, 1975; Williams, 1987; Imamoglu, 2000; Lévy, MacRae, & Köster, 2006).

The link of familiarity to perceived pleasure has implications for theories of intrinsic motivation. Interestingly, in the field of developmental robotics, it has been widely assumed that exploring on the periphery of knowledge is a preferred strategy for the gaining of competence (Oudeyer, Kaplan, & Hafner, 2007; Kaplan & Oudeyer, 2007; Ryan & Deci, 2000; Schmidhuber, 2010). Specifically, if an organism spends time exploring familiar parts of its world, little new knowledge can be gained. Similarly, exploring totally unfamiliar contingencies is problematic in that the contingencies might be purely random (such as trying to predict the next car exiting a tunnel) or may not be easily integrated into current mental models (e.g., integrating a location into a poorly formed map of the world). Instead, exploring new knowledge that extends the old drives the organism to effectively expand its mental models and gain competencies central to survival. Thus, perhaps the inverted-U shape that we have been exploring in this letter and that found in the repetition priming literature is a fundamentally important signal used for shaping effective behavior.

Currently, there is little synergy between these research areas that may be fundamentally exploring the same mechanisms. For example, few papers in the area of repetition suppression explore the potential utility of the signal generated, and few papers in developmental robotics take a close look at the biology involved in generating intrinsic reward signals. It is evident that further exploration of these cross-connections is warranted.

5 Conclusion

In this letter, we have presented a computational model that extends the sharpening theory with concrete algorithms and have demonstrated that lateral inhibition in combination with Hebbian learning could provide a sufficient mechanism that accounts for both repetition suppression for familiar stimuli and repetition enhancement for novel stimuli. Moreover, according to our model, the two forms of priming could be due to the rising and falling parts of an inverted U-shape function with respect to the familiarity of the stimulus.

The preliminary results of this study contribute to our understanding of novelty processing in the brain and raise important questions for further investigation. Specifically, it is suggested that the full inverted U-shape of neural activity could be shown in fMRI study if the novelty of stimuli could be gradually changed from truly novel to familiar. It is also suggested that the neuroscientific community must take special care to choose novel and familiar stimuli for neurological experiments and that a better integration of disciplines that explore the mechanisms and utility of this signal be explored.

Appendix: The Neuron Model and Hebbian Learning Algorithm _____

A.1 Neuron Model. A biologically plausible neuron model (O'Reilly & Munakata, 2000) was used for the spiking network experiment. Specifically, the model assumes that most of the activation work in a neuron is performed by an excitatory synaptic input channel (formed by Na^+ current), inhibitory synaptic input channel (Cl^- current) and always-open leak channel (K^+ current). Thus, the total current for the neuron is defined by

$$I_{net} = I_e + I_i + I_l = g_e(t)\, G_e\left(V_m(t) - E_e\right)$$
$$+ g_i(t)\, G_i\left(V_m(t) - E_i\right) + g_l(t)\left(V_m(t) - E_l\right), \tag{A.1}$$

where $g_*(t)$ is the fraction of the total number of channels that are open at a time t for the excitatory, inhibitory, and leakage channels, and G_* is a maximum conductance of the channel. Thus, the membrane potential is defined by

$$\frac{dV_m}{dt} = D_{vm}I_{net} = D_{vm}[g_e(t)\, G_e\left(V_m(t) - E_e\right)$$
$$+ g_i(t)\, G_i\left(V_m(t) - E_i\right) + g_l(t)\left(V_m(t) - E_l\right)], \tag{A.2}$$

where D_{vm} is a time constant that reflects the capacity of the membrane.
The conductance of the excitatory channel follows the kinetics:

$$\frac{dg_e}{dt} = -D_{net}g_e(t) + D_{net}\left(I_e + I_{rand}\right), \tag{A.3}$$

where I_e is an averaged input excitatory current and I_{rand} is added gaussian noise with mean 0 and standard deviation 2 mV.
Similarly, the conductance of the inhibitory channel is described by

$$\frac{dg_i}{dt} = -D_{net}g_i(t) + D_{net}\left(I_i + I_{rand}\right). \tag{A.4}$$

The leakage current in the model is time independent, so its conductance is given by $g_l(t) = 1$.

The parameters used in the simulation reflect the real neuron dynamics (O'Reilly & Munakata, 2000). In particular, the neurons fire when the membrane potential reaches a threshold of $\theta = -55$ mV. When a neuron fires, the membrane potential resets to a resting potential value $V_{rest} = -70$ mV, and the conductances are reset to zero.

Consistent with neurobiology, the inhibitory neurons in the model are faster to fire compared to the excitatory neurons (Somogyi et al., 1998). This

Table 1: Parameters of the Neuron Model Used in the Simulation (Dimension mV).

Parameter		mV	Normalized Value
Resting potential	V_{rest}	−70	0.15
Firing threshold	θ	−55	0.25
Excitatory (Na^+) channel equilibrium potential	E_e	+55	1.00
Inhibitory (Cl^-) channel equilibrium potential	E_i	−70	0.15
Leakage (K^+, Na^+) channel equilibrium potential	E_l	−70	0.15

Table 2: Parameters of the Neuron Model Used in the Simulation (Dimensionless).

Parameter		Value
Maximum conductance for excitatory channel	G_e	1.0
Maximum conductance for inhibitory channel	G_i	1.0
Maximum conductance for leakage channel	G_l	0.1
Membrane potential time constant, excitatory neuron	D_{vm}	0.14
Membrane potential time constant, inhibitory neuron	D_{vm}	0.02
Conductance time constant, excitatory neuron	D_{net}	0.49
Conductance time constant, inhibitory neuron	D_{net}	0.07

is achieved using a smaller conductance time constant for the inhibitory neurons. The parameters of the neuron model are listed in Tables 1 and 2.

A.2 Hebbian Learning. To simplify our experiment, all excitatory neurons are initially connected to the input stimulus neuron using small random weights to emulate the initial unfamiliarity with the input pattern. The single stimulus neuron is firing at a constant rate of 10 spikes per second, a reasonable spiking rate for regular spiking excitatory neurons found in the cortex (Izhikevich, 2004).

The connections between the input stimulus neuron and excitatory neurons are modified according to an algorithm proposed by Gerstner and Kistler (2002) (Hebbian learning):

$$\Delta w = A_0 + \begin{cases} A_+ e^{-\frac{s}{\tau_+}}, & s \geq 0 \\ -A_- e^{\frac{s}{\tau_-}}, & s < 0 \end{cases}, \tag{A.5}$$

where $s = t_{post} - t_{pre}$ is a time delay between pre- and postsynaptic firings; $A_0 = -0.05$ is an activity independent constant that represents a slow decrease of the weights over time; τ_+, $\tau_- = 20\,ms$, and A_+, $A_- = 1$. The initial

weights are set to a value 4.3, which yields a subthreshold activation of the neurons.

The simulation time step 0.2 ms is chosen to provide a finely grained resolution in order to capture the neuron dynamics. Every learning iteration corresponds to the simulation time frame of 100 ms, which is long enough for the neurons to fire and return to their resting state.

Acknowledgments

References

Anderson, J. (1976). *Language memory thought*. Hillsdale, NJ: Erlbaum.

Atallah, H. E., Frank, M. J., & O'Reilly, R. C. (2004). Hippocampus, cortex, and basal ganglia: Insights from computational models of complementary learning systems. *Neurobiology of Learning and Memory*, *82*, 253–267.

Balu, R., Larimer, P., & Strowbridge, B. W. (2004). Phasic stimuli evoke precisely timed spikes in intermittently discharging mitral cells. *Journal of Neurophysiology*, *92*, 743–753.

Beaulieu, C., Kisvarday, Z., Somogyi, P., Cynader, M., & Cowey, A. (1992). Quantitative distribution of GABA-immunopositive and -immunonegative neurons and synapses in the monkey striate cortex (area 17). *Cerebral Cortex*, *2*, 295–309.

Biederman, I., & Vessel, E. A. (2006). Perceptual pleasure and the brain. *American Scientist*, *94*, 248–255.

Brennan, W. M., Ames, E. W., & Moore, R. W. (1966). Age differences in infants' attention to patterns of different complexities. *Science*, *151*, 354–356.

Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*, 51–61.

Demb, J. B., Desmond, J. E., Wagner, A., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, *15*, 5870–5878.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 13494–13499.

Düzel, E., Habib, R., Guderian, S., & Heinze, H. J. (2004). Four types of novelty-familiarity responses in associative recognition memory of humans. *European Journal of Neuroscience*, *19*, 1408–1416.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123–152.

Fiebach, C. J., Gruber, T., & Supp, G. G. (2005). Neuronal mechanisms of repetition priming in occipitotemporal cortex: Spatiotemporal evidence from functional magnetic resonance imaging and electroencephalography. *Journal of Neuroscience*, *25*, 3414–3422.

Gagnepain, P., Henson, R., Chételat, G., Desgranges, B., Lebreton, K., & Eustache, F. (2011). Is neocortical-hippocampal connectivity a better predictor of subsequent recollection than local increases in hippocampal activity? New insights on the role of priming. *Journal of Cognitive Neuroscience*, *23*, 391–403.

Gerstner, W., & Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge: Cambridge University Press.

Gibson, J. R., Beierlein, M., & Connors, B. W. (1999). Two networks of electrically coupled inhibitory neurons in neocortex. *Nature*, *402*, 75–79.

Gotts, S. J. (2003). *Mechanisms underlying enhanced processing efficiency in neural systems*. PhD dissertation, Carnegie Mellon University.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*, 293–321.

Gruber, T., & Müller, M. M. (2002). Effects of picture repetition on induced gamma band responses, evoked potentials, and phase synchrony in the human EEG. *Cognitive Brain Research*, *13*, 377–392.

Henson, R., Shallice, T., & Dolan, R. J. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, *287*, 1269–1272.

Heyduk, R. G. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. *Attention, Perception, and Psychophysics*, *17*, 84–90.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Computation*, *14*, 1771–1800.

Hinton, G. E. (2010). *A practical guide to training restricted Boltzmann machines*. Toronto: University of Toronto.

Hinton, G. E., Osindero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*, 1527–1554.

Hinton, G. E., & Salakhutdinov, R. (2006). Matlab code for Science paper.

Hunkin, N., Mayes, A., & Gregory, L. (2002). Novelty-related activation within the medial temporal lobes. *Neuropsychologia*, *40*, 1456–1464.

Imamoglu, Ç. (2000). Complexity, liking and familiarity: Architecture and non-architecture Turkish students' assessments of traditional and modern house facades. *Journal of Environmental Psychology*, *20*, 5–16.

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, *15*, 1063–1070.

James, T. W., & Gauthier, I. (2006). Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Human Brain Mapping*, *27*, 37–46.

Kaplan, F., & Oudeyer, P.-Y. (2007). In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience*, *1*, 225–236.

Karmel, B. Z. (1969). Complexity, amounts of contour, and visually dependent behavior in hooded rats, domestic chicks, and human infants. *Journal of Comparative and Physiological Psychology*, *69*, 649–657.

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, *80*, 324–330.

Kumaran, D., & Maguire, E. A. (2007). Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus*, *17*, 735–748.

Larsson, J., & Smith, A. (2012). fMRI repetition suppression: Neuronal adaptation or stimulus expectation? *Cerebral Cortex*, *22*, 567–576.

LeCun, Y., & Cortes, C. (1998). *The MNIST database of handwritten digits.* Retrieved from: http://yann.lecun.com/exdb/mnist/

Lévy, C. M., MacRae, A., & Köster, E. P. (2006). Perceived stimulus complexity and food preference development. *Acta Psychologica*, *123*, 394–413.

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, *69*, 1918–1929.

Luo, H., Shen, R., & Niu, C. (2010). *Sparse group restricted Boltzmann machines*. Arxiv preprint arXiv:1008.4988, 1–9.

McMahon, D. B. T., & Olson, C. R. (2007). Repetition suppression in monkey inferotemporal cortex: Relation to behavioral priming. *Journal of Neurophysiology*, *97*, 3532–3543.

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*, *254*, 1377–1379.

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, *263*, 520–522.

Moldakarimov, S., Bazhenov, M., & Sejnowski, T. J. (2010). Representation sharpening can explain perceptual priming. *Neural Computation*, *22*, 1312–1332.

Naccache, L. (2001). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex*, *11*, 966–974.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proc. 27th International Conference on Machine Learning*. New York: ACM Press.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.

O'Reilly, R. C. (1996). The LEABRA model of neural interactions and learning in the neocortex. Doctoral dissertation, Carnegie Mellon University.

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2011). Complementary learning systems. *Cognitive Science*, 1–20.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, *11*, 265–286.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Rainer, G., & Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, *27*, 179–189.

Ranzato, M., Krizhevsky, A., & Hinton, G. E. (2010). Factored 3-way restricted Boltzmann machines for modeling natural images. In *Proc. of the 13th International Conference on Artificial Intelligence and Statistics*. Cambridge, MA: MIT Press.

Ringo, J. L. (1996). Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. *Behavioural Brain Research*, *76*, 191–197.

Rolls, E., Baylis, G. C., Hasselmo, M., & Nalwa, V. (1989). The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Res*, *76*, 153–164.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54–67.

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology*, *17*, 171–176.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, *2*, 230–247.

Schott, B. H., Sellner, D. B., Lauer, C.-J., Habib, R., Frey, J. U., Guderian, S., . . . Düzel, E. (2004). Activation of midbrain structures by associative novelty and the formation of explicit memory in humans. *Learning and Memory*, *11*, 383–387.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 282–317). Cambridge, MA: MIT Press.

Sobotka, S., & Ringo, J. L. (1994). Stimulus specific adaptation in excited but not in inhibited cells in inferotemporal cortex of macaque. *Brain Research*, *646*, 95–99.

Sobotka, S., & Ringo, J. L. (1996). Mnemonic responses of single units recorded from monkey inferotemporal cortex, accessed via transcommissural versus direct pathways: A dissociation between unit activity and behavior. *Journal of Neuroscience*, *16*, 4222–4230.

Soldan, A., Gazes, Y., Hilton, H. J., & Stern, Y. (2008). Aging does not affect brain patterns of repetition effects associated with perceptual priming of novel objects. *Journal of Cognitive Neuroscience*, *20*, 1762–1776.

Soldan, A., Habeck, C., Gazes, Y., & Stern, Y. (2010). Neural mechanisms of repetition priming of familiar and globally unfamiliar visual objects. *Brain Research*, *1343*, 122–134.

Somogyi, P., Tamás, G., Lujan, R., & Buhl, E. H. (1998). Salient features of synaptic organisation in the cerebral cortex. *Brain Research Reviews*, *26*, 113–135.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*, 171–177.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*, 1004–1006.

Tang, Y., & Sutskever, I. (2011). *Data normalization in the learning of restricted Boltzmann machines*. Toronto: University of Toronto.

Tieleman, T. (2008). Training restricted Boltzmann machines using approximations to the likelihood gradient. In *Proceedings of the 25th International Conference on Machine Learning* (pp. 1064–1071). New York: ACM Press.

Voss, J. L., & Paller, K. A. (2010). Real-time neural signals of perceptual priming with unfamiliar geometric shapes. *Journal of Neuroscience*, *30*, 9181–9188.

Wan, H., Aggleton, J. P., & Brown, M. W. (1999). Different contributions of the hippocampus and perirhinal cortex to recognition memory. *Journal of Neuroscience*, *19*, 1142–1148.

Wig, G. S., Grafton, S. T., Demos, K. E., & Kelley, W. M. (2005). Reductions in neural activity underlie behavioral components of repetition priming. *Nature Neuroscience*, *8*, 1228–1233.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*, 227–233.

Williams, S. M. (1987). Repeated exposure and the attractiveness of synthetic speech: An inverted-U relationship. *Current Psychology*, *6*, 148–154.

Yang, J., Meckingler, A., Xu, M., Zhao, Y., & Weng, X. (2008). Decreased parahippocampal activity in associative priming: Evidence from an event-related fMRI study. *Learning and Memory*, *15*, 703–710.