Selective Enhancement of Recall through Plasticity Modulation in an Autoassociative Memory

A. Sandberg, A. Lansner, K. M. Petersson

January 26, 2000

Abstract

The strength of a memory trace is modulated by a variety of factors such as arousal, attention, context, type of processing during encoding, salience and novelty of the experience. Some of these factors can be modeled as a variable plasticity level in the memory system, controlled by arousal or relevance-estimating systems. We demonstrate that a Bayesian confidence propagation neural network with learning time constant modulated in this way exhibits enhanced recall of an item tagged as salient, proactive and retroactive inhibition of other items is also demonstrated as well as an inverted U-curve response curve to overall plasticity.

1 Introduction

Long-term memory (LTM) formation in everyday life often occurs incidentally without explicit intention to remember the information processed. It has been suggested that memory formation is the dynamic consequence of information processing and system plasticity [1, 2]. Research indicates that specific kinds of information processing contributes to LTM-formation, including meaning-based, context and relational processing and factors like emotional significance and attentional allocation (for recent reviews see e.g. [3, 4]).

Endogenous processes activated by experience can modulate memory strength in terms of recall probability [5]. For example, emotionally arousing [6] or humorous [7] experiences are generally better remembered than less affective experiences, and hormones and neuromodulators can affect how strongly experiences are retained [8]. The novelty of a stimulus also plays an important role. The von Restorff effect consists of improved recall or recognition of an item (the isolate) in a set that is distinct or different from the others, while the other items are less well recalled (retroactive and proactive inhibition) [9]. While this has mainly been studied in human list recall, a similar effect has been observed in rats [10] and monkeys [11].

Some of these factors can be interpreted in the framework of memory consolidation as a relevance modulation of the "print-now" signal by regulating memory encoding and synaptic plasticity. During ordinary events the plasticity is lower than when salient, arousing or motivational stimuli increase the learning rate, for example through a modulating mechanism. Such a mechanism relates closely to neuromodulation in the brain, e.g. the effect of dopamine [12] and acetylcholine in synaptic plasticity [13, 14].

We have previously studied the properties of an autoassociative Bayesian confidence propagation neural network (BCPNN) [15, 16, 17, 18] with incremental learning [19]. The BCPNN is based on a probabilistic view of learning and retrieval, with input and output unit activities roughly representing confidence of feature detection and posterior probabilities of outcomes, respectively. Besides in ANN models, it has also been used for simulation of cortical dynamics and cell assemblies [20, 21].

This learning rule has a learning time constant of learning that determines how quickly it will adapt to new information. By modulating this time constant we can model the modulatory regulation of the print-now signal on associative encoding of information into for example long term memory.

This paper describes a simple model of an autoassociative network with plasticity modulation for one item, and shows that it can produce the enhanced recall of the isolate, proactive and retroactive inhibition and an inverted U-curve response curve to overall plasticity similar to the one commonly observed in arousal-performance or dose-response plots.

2 Network

We have used the BCPNN with incremental learning and hypercolumns [19, 22]. The BCPNN (Bayesian Confidence Propagation Neural Network) is based on a Hebbian learning rule derived from Bayes rule [15, 16]. It is an autoassociative network with a learning rule where weights are updated depending on estimates of unit activity regulated by a learning time constant. Units are combined into hypercolumns with normalized output activities (corresponding to mutually exclusive or combinations of otherwise dependent features [18]).

$$
\begin{aligned}
\frac{dh_{ii'}(t)}{dt} &= \beta_{ii'}(t) + \sum_{j}^{N} \log\left(\sum_{j'}^{M_i} w_{ii'jj'}(t) o_{jj'}(t)\right) - h_{ii'}(t) \\
o_{ii'}(t) &= \frac{e^{h_{ii'}}}{\sum_j e^{h_{ij}}} \\
\frac{d\Lambda_{ii'}(t)}{dt} &= \kappa(t)\alpha([(1-\lambda_0)o_{ii'}(t) + \lambda_0] - \Lambda_{ii'}(t)) \\
\frac{d\Lambda_{ii'jj'}(t)}{dt} &= \kappa(t)\alpha([(1-\lambda_0^2)o_{ii'}(t)o_{jj'}(t) + \lambda_0^2] - \Lambda_{ii'jj'}(t)) \\
\beta_{ii'}(t) &= \log(\Lambda_{ii'}(t)) \\
w_{ii'jj'}(t) &= \frac{\Lambda_{ii'jj'}(t)}{\Lambda_{ii'}(t)\Lambda_{jj'}(t)}
\end{aligned}
$$

where $h_{ii'}$ the potential of the i'th unit in hypercolumn i, $o_{ii'}$ the output of the unit, β the bias, w the weights, Λ rate estimates, λ_0 a low background activity (set to 10^{-4} throughout this paper), N the number of hypercolumns, M_i the number of neurons in hypercolumn i and α the inverse of the learning time constant. Low α correspond to slow learning and forgetting, high α fast learning and forgetting.

$\kappa(t)$ is the relevance signal, assumed to be sent from a modulator system as a response to the current experience. In the following experiments it was kept at $\kappa(t) = 1$ except for one pattern in the training set, the isolate, where $\kappa(t)$ was set to κ_i during training.

3 Results

A 100 neuron BCPNN network with 10 hypercolumns of 10 neurons each was trained by clamping unit activity to each training pattern $\boldsymbol{\xi}^p$ for one unit of time, allowing the weights to adapt. Retrieval was tested by activating the neurons with a trained pattern where the activity in three hypercolumns of ten had been randomized, and then allowed to relax for one unit of time (no learning was used during testing). Performance was measured by the overlap $\boldsymbol{\xi}^p \cdot \mathbf{x}/|\boldsymbol{\xi}^p||\mathbf{x}|$.

Figure 1 shows the mean overlap for all patterns as a function of the learning time constant α, with no isolate element ($\kappa_i = 1$). At low α the network learns too slowly to learn the patterns, while at high α the network learns quickly but forgets the oldest patterns.

Figure 2 shows the selective enhancing effect on recall when an isolate pattern occurs. Note the inhibition of recall of other patterns in the high κ_i condition. α was set to 10^{-8} in this and the following plots.

Figure 3 shows the mean overlap as a function of the modulation strength for the isolate and for the normal items. As κ_i increases the recall of the isolate becomes better and better, while there is an inhibition effect on the other items.

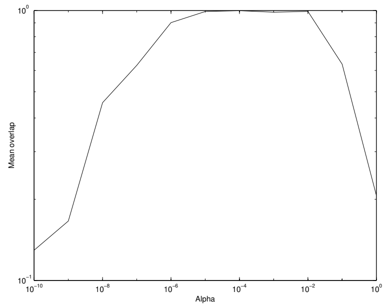
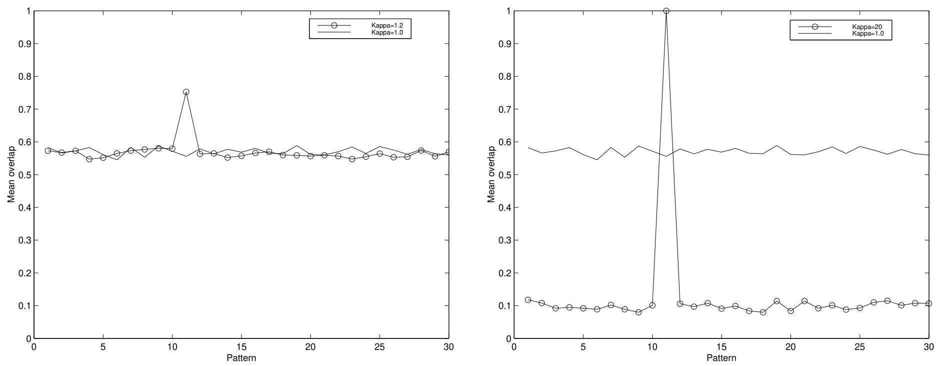


Figure 1: Mean overlap after convergence from noisy cue as a function of the inverse learning time constant α. Below $\alpha = 10^{-6}$ the network learns too slowly to store the shown patterns, and above $\alpha = 10^{-2}$ it forgets too quickly to store all patterns.



(a) $\kappa_i = 1.2$

(b) $\kappa_i = 20$

Figure 2: Mean overlap after convergence from noisy cue with or without an isolate pattern. In the dotted run, κ_i was set to 1.2 (left) or 20 (right) for pattern 11.

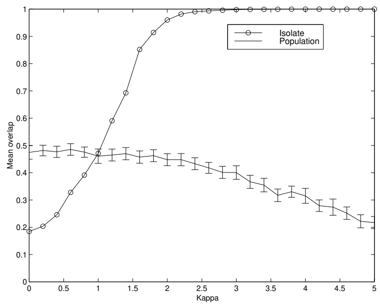


Figure 3: Mean overlap after convergence from noisy cue of the isolate pattern and the other patterns as a function of the increase in plasticity κ_i.

However, it is possible to avoid inhibition in this model for low levels of the relevance signal while still observing a recall enhancement for the isolate.

When the network is stimulated by a random pattern it will converge to a given attractor state with a probability depending on the relative volume of the basin of attraction to the volume of the state space. In figure 4 we can see that the isolate is more likely to be recalled, suggesting that for these parameters the increased plasticity has enlarged its basin of attraction relative to the other attractors.

If the network has a shorter learning time constant as in the right picture, there will be a memory gradient due to fast forgetting. The increase in plasticity caused by a strong modulatory input prevented encoding of the isolate and retroactively interfered with the early patterns without affecting the subsequent storage of the network.

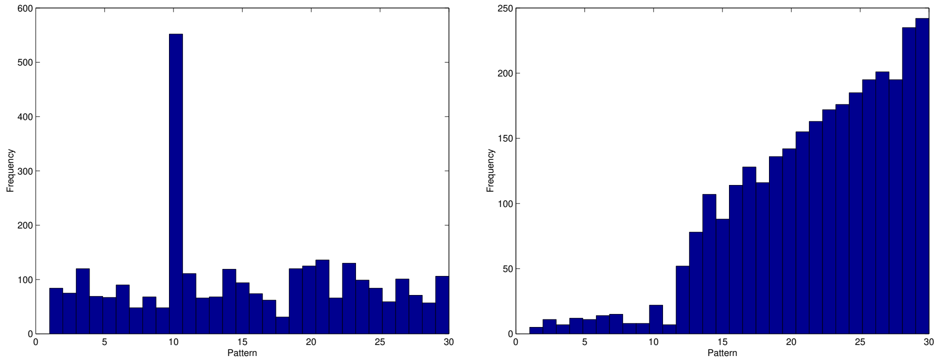
4 Discussion

We have described a simple model of an autoassociative memory where the learning rate is modulated by a presumably multi-dimensionally determined relevance signal and shown that it can exhibit selective enhancement of items associated with a relevance signal.

We have shown that a model of an autoassociative memory with modulated learning rate can exhibit selective enhancement of items associated with a relevance signal. When the signal is strong inhibition of other items occurs as well as a relative expansion of the isolate item basin of attraction.

The inverted-U curve in figure 1 has similar properties as the Yerkes-Dodson law [23] or the observed inverted-U dose-response relationship in many memory enhancing drugs [8, 24, 25].

If there is a need to encode relevant experiences at different levels of trace strength, for example to distinguish between everyday and important stimuli, the baseline plasticity needs to be in a range where the increase in plasticity



(a) $\kappa_i = 1.5$, $\alpha = 10^{-8}$

(b) $\kappa_i = 20$, $\alpha = 10^{-1.5}$

Figure 4: Frequency of ending up in the attractors corresponding to the learned patterns when activating the network with a random pattern (one randomly activated unit per hypercolumn) for a low and high level of plasticity and modulation.

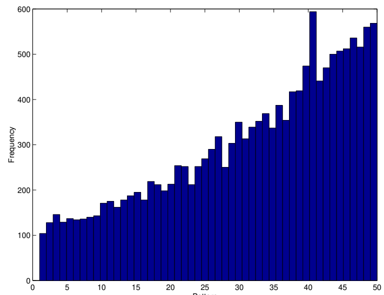


Figure 5: Frequency of attractors with medium stimulation and plastic network. $\kappa_i = 2$, $\alpha = 10^{-2}$. Pattern 41 is the isolate.

due to the relevance signal produces a positive change in encoding success. If the memory exhibits a U-shaped response to plasticity, this suggests that the baseline plasticity is not close to the maximum of the overlap curve but below it, and that an increased level of baseline arousal would decrease the distinction between relevant and irrelevant memories.

Interestingly, a strong modulatory input increases the plasticity in this model to the extent that the overall effect becomes a memory impairment rather than enhancement.

References

[1] K.M. Petersson, C. Elfgren, and M. Ingvar. A dynamic role of the medial temporal lobe during retrieval of declarative memory in man. *Neuroimage*, 6:1–11, 1997.

[2] K.M. Petersson, C. Elfgren, and M. Ingvar. Dynamic changes in the functional anatomy of the human brain during recall of abstract designs related to practice. *Neuropsychologia*, 37:567–587, 1999.

[3] R.L. Buckner, W.M. Kelley, and S.E. Petersen. Frontal cortex contributes to human memory formation. *Nature Neuroscience*, 2:311–314, 1999.

[4] A.D. Wagner, W. Koutstaal, and D. Schacter. When encoding yields remembering: insights from event-related neuroimaging. *Phil. Trans. R. Soc. Lond. B*, 121:1307–1324, 1999.

[5] James L. McGaugh. Memory – a century of consolidation. *Science*, 287:248–251, 2000.

[6] S.-A. Christianson, editor. *Handbook of Emotion and Memory: Current Research and Theory*. Erlbaum, Hillsdale, NJ, 1992.

[7] S.R. Schmidt. The effects of humor on sentence memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20:953–967, 1994.

[8] Joe L. Martinez, Gery Schulteis, and Susan B. Weinberger. *Learning and Memory: A Biological View*, chapter 4, pages 149–198. Academic Press, second edition, 1991.

[9] Hedwig von Restorff. Analyse von vorgängen in spurenfeld. *Psychologice Forschung*, 18:299–342, 1933.

[10] Phil Reed and Angela Richards. The von restorff effect in rats (rattus norvegius). *Journal och Comparative Psychology*, 110(2):193–198, 1996.

[11] Amanda Parker, Edward Wilding, and Colin Akerman. The von restorff effect in visual object recognition memory in humans and monkeys: The role of frontal/perirhinal interaction. *Journal of Cognitive Neuroscience*, 10(6):691–703, 1998.

[12] J. Wickens and R. Kötter. Cellular models of reinforcement. In James C. Houk, Joel L. Davis, and David G. Beiser, editors, *Models of Information Processing in the Basal Ganglia*, pages 187–214. MIT Press, 1995.

[13] N. J. Woolf. The critical role of cholinergic basal forebrain neurons in morphological change and memory encoding: a hypothesis. *Neurobiol Learn Mem*, 66(3):258–66, November 1996.

[14] M. E. Hasselmo, B. P. Wyble, and G. V. Wallenstein. Encoding and retrieval of episodic memories: role of cholinergic and gabaergic modulation in the hippocampus. *Hippocampus*, 6(6):693–708, 1996.

[15] Anders Lansner and Örjan Ekeberg. An associative network solving the "4-Bit ADDER problem". In Maureen Caudill and Charles Butler, editors, *IEEE First International Conference on Neural Networks*, pages II–549, San Diego, CA, June 21–24 1987.

[16] Anders Lansner and Örjan Ekeberg. A one-layer feedback artificial neural network with a bayesian learning rule. *Int. J. Neural Systems*, 1(1):77–87, 1989.

[17] Anders Lansner and Anders Holst. A higher order Bayesian neural network with spiking units. *Int. J. Neural Systems*, 7(2):115–128, May 1996.

[18] Anders Holst. *The Use of a Bayesian Neural Network Model for Classification Tasks*. PhD thesis, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, Sweden, September 1997. TRITA-NA-P9708.

[19] Anders Sandberg, Anders Lansner, Karl-Magnus Petersson, and Örjan Ekeberg. A palimpsest memory based on an incremental bayesian learning rule. In *CNS'99, Computational Neuroscience*, Pittsburgh, 1999. To be published in Neurocomputing.

[20] Erik Fransén and Anders Lansner. A model of cortical associative memory based on a horizontal network of conneted columns. *Network*, 9:235–264, 1998.

[21] Erik Fransén and Anders Lansner. Low spiking rates in a population of mutually exciting pyramidal cells. *Network*, 6(2):271–288, 1995.

[22] A. Sandberg, A. Lansner, K.M. Petersson, and Ö. Ekeberg. Bayesian attractor networks with incremental learning. Submitted to Network, 1999.

[23] R.M. Yerkes and J.D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurological Psychology*, 18:459–482, 1908.

[24] M. M. Boccia, S. R. Kopf, and C. M. Baratti. Phlorizin, a competitive inhibitor of glucose transport, facilitates memory storage in mice. *Neurobiol Learn Mem*, 71(1):104–12, January 1999.

[25] M.W. Parsons and P.E. Gold. Glucose enhancement of memory in elderly humans: an inverted-u dose-response curve. *Neurobiol Aging*, 13(3):401–4, May-Jun 1992.