

Competitive Learning in Biological and Artificial Neural Computation

Nathan Intrator
School of Mathematical Sciences
Tel Aviv University, Tel Aviv 69978, Israel
nin@math.tau.ac.il

Shimon Edelman
Dept. of Brain & Cognitive Science
MIT E25-201, Cambridge, MA 02142, USA
edelman@ai.mit.edu

July, 1997

Keywords: Competitive Learning, Resource Allocation, Memory, Striatum

Summary

Following a brief discussion of 'classical' competitive learning and of approaches that involve mixtures of experts, we concentrate on competitive learning guided by temporal structure present in the stimuli. In this context, we suggest a general principle for resource allocation and memory management, which may be able to account for a range of psychophysical and neurophysiological findings.

1 Biological evidence for competition in representation

Learning – modification of performance precipitated by prior experience – is commonplace in biological information processing, and, in particular, in human perception [1, 2]. Although mere repeated exposure normally leads to a degradation of the perceptual capacity (e.g., the detection threshold in grating adaptation [3]), the performance invariably improves when the subject is required to discriminate among several stimuli. In this note, we concentrate on a particularly important aspect of the learning process, which underlies many perceptual and cognitive phenomena: *competition* among various functional units that contribute to the solution of the task at hand.

In biological perceptual systems, the tuned response properties and the distributed nature of the representational substrate instigate competition among the participating functional units. One kind of behavioral evidence in support of the existence of tuned units is provided by the specificity of improvement in typical learning situations to the given task and stimulus. The specificity of learning is pervasive: in vision, it occurs in the entire range of tasks from intensity grating discrimination [4] to face recognition [5]. In some cases, stimulus-specific learning has been shown to lead to a degradation of performance on a different range of stimuli in the same task. For example, the lowering of vertical vernier detection threshold with practice transfers neither to horizontal verniers,

nor to other acuity-related tasks [6]; moreover, subjects trained on verniers of one orientation exhibit initially a threshold that is elevated relative to the baseline when tested on verniers of the orthogonal orientation [7]. These findings suggest that the perceptual system contains tuned functional units, and that such units are allocated and their response properties are modified conservatively and competitively.

Because stimuli are typically represented by *ensembles* of functional units, each such unit may be expected to have to “justify” its existence by vying for an ever larger share of the representational burden.¹ As a result, percepts formed by a representational system are superpositions of the activities of a number of functional units, each tuned to a somewhat different aspect of the feature space spanned by the stimuli. A beautiful illustration of this principle can be found in the early work of J. J. Gibson [9] who showed that an objectively curved line is perceived as nearly straight following adaptation to curves of the same sense; a straight line is then perceived as curved in the opposite direction (note that we choose not to distinguish here between perceptual learning and adaptation). In auditory perception, analogously, the pitch of a tone appears higher than it is, following adaptation to a somewhat lower tone; the reverse effect is obtained when the adapting tone is higher than the test one [10]. Likewise, in two-way classification of parametrically-defined stimuli arranged along a line in some feature space, the performance on unseen stimuli situated progressively farther from the class boundary first gets better, then worse than the performance on trained stimuli [11]. In all these examples, the peculiarities of observed performance can be explained by positing a pool of functional units that (1) are differentially tuned to line orientation, or tone pitch, or whatever feature dimension is used in the definition of the classification task, (2) are adaptable, and (3) whose joint pattern of activation determines the subjective quality of a percept (curvature of a line, pitch of a sound, category of a shape).

The notion of a limited pool of individually adaptable tuned units is more than a functional abstraction useful for psychological modeling. The reality of a distributed representational substrate whose members compete for a share in the representation of the stimulus has been demonstrated in a variety of electro-physiological studies [12]. These studies can be divided into two major groups. In the first group, competition among members of a pool of tuned units has been enhanced by the withdrawal of stimulation from some of the units. For instance, the induction of a retinal scotoma leads to the invasion of the visual space of affected cells by receptive fields of other, neighboring cells [13]; an analogous effect, albeit on a much longer time scale, is observed in the somatosensory modality [14]. In comparison, in the second group, the increased prominence of a sub-population of functional units (whose emergence may in itself be a product of learning) is precipitated merely by the prevalence of the preferred stimuli of those units in the sensory repertoire of the system. In the monkey, this phenomenon has been glimpsed in relation with face representation [15], as well as the representation of general natural [16] and artificial [17] objects.

The cursory survey of the perceptual learning literature presented above suggests that competition among functional units is a widespread phenomenon, whose understanding may contribute to the identification of cortical learning mechanisms common to various subsystems and modalities. In all cases, competitive learning amounts to a dynamic redistribution of “responsibilities” of various units over parts of the representation space (which may refer to the regular retinal space, or to some more abstract feature space; cf. [18]). In the next section, we survey some of the computational algorithms that have been developed in response to the need to model competitive learning, and to employ it in the search for the solutions to some common neural computation tasks.

¹The involvement of inhibition among tuned units has been extensively discussed in the literature, following early findings such as [8].

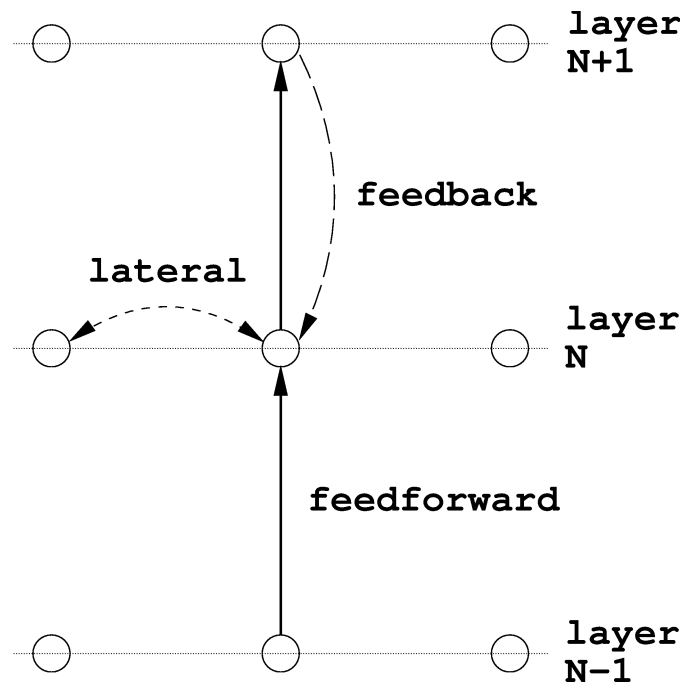


Figure 1: A competitive learning model commonly consists of several layers of “neurons” (three such layers are illustrated here). Units in each layer are connected via excitatory connections to higher layers; inhibitory lateral links to adjacent neurons are a common way to model competition. In addition to those, some models call for top-down feedback connections to steer and regulate the competition (note that both the lateral and the feedback connections can be made to operate on a different time course, e.g., to persist after the feed-forward signal decays). A model can be also purely feed-forward, yet contain an element of competition. For example, the (necessarily limited) degree of activation of a cell shared by a number of converging connections may constitute a target resource for competition, leading to the emergence of inequalities among the units that connect to it (cf. [23], figs. 1 and 8).

2 Families of competitive learning algorithms

The central computational appeal of competitive learning lies in its ability to generate sparse representations. Such representations, in turn, economize on resources (a consideration mainly of interest in biological modeling), and facilitate further processing such as detection of coincidences [19, 20].

2.1 Competition over entire problem space

By definition, competitive learning algorithms for artificial neural networks employ some sort of competition between neurons, via “lateral” (as contrasted with feed-forward connections; see Figure 1.) Early algorithms of this kind involved *hard* competition, in which only a single neuron (the initially strongest one) remains active at the end of the learning cycle [21]. In more recent algorithms, the competition does not necessarily drive all neurons but one to zero [22]. Such *soft* competition has desirable mathematical properties, as it fits a mixture of Gaussians formulation.

Current competitive learning algorithms may be distinguished on the basis of the learning rule they employ (determined by the desired objective function for the learning process), and the form

and the role of the competition during learning. In some cases, a competitive learning algorithm amounts to an application of a demonstrably successful single-neuron learning rule in a multi-unit network with lateral connections. The latter are needed to ensure that each neuron will extract a different feature from the data. More generally, competition may occur between subnetworks, and not single units, exposed to a given learning task. This will be discussed in section 2.2.

Another family of algorithms is characterized by the indispensability of the lateral connections; in such an algorithm, turning these off results in an abrupt loss of the ability to extract useful features. The importance of competition in this case suggests that pitching units against each other may have an intrinsic appeal (over and above that of the single-neuron learning rules), and that algorithms involving competition may succeed where other forms of learning run into difficulties. A simple example of such a situation is clustering — the process of grouping together data points based on some measure of inter-point distance. In the absence of inhibition, all units in a typical clustering network are likely to converge to the mean of the data distribution. Under a simple Hebbian update rule for synaptic weights [24], this happens because learning drives the preferred stimulus of a tuned unit to the center of the distribution, where the correlation between input and output activities is maximized. A smarter algorithm, which looks instead for tight clusters (cf. [25]) will likely find those that occur with the highest probability, or those that are the closest to the initial synaptic weights. In comparison, when inhibition is turned on, the clusters move to new locations and become sharper and more distinct [26].²

2.2 Division of the problem space: local experts

The idea of having experts (entire modules) rather than units compete with each other [28, 29] is a direct extension of the regular competitive learning approach. For instance, each hidden unit of a radial basis function (RBF) clustering network can be replaced by an expert module, while the linear summation output unit becomes a gating network that receives inputs not only from the hidden layer but also directly from the input layer. Such a gating network is taught to combine expert outputs in a different manner for the different regions in the input space, effectively dividing it among the (local) experts.

The competition among neurons is replaced here by a competition among experts, all striving to perform well on the data in various parts of the input space. The assumption behind this model is that the data have been generated by several processes, each with its own set of parameters. Therefore, the best performance is achieved if an expert is trained, and performs best, on the output of a single generator, rather than on the entire data space. The splitting of this space is based on the performance of the experts, with the hard competition corresponding to the assumption that each data point is due to a single generator; the soft competition assumes instead a probabilistic mixture of the involvement of the different generators. It should be noted that each of the experts and the gating network may differ in their architecture and complexity; for instance, [29] used an RBF module and a sigmoidal feed-forward network as experts in the same model.

During training, the gating network learns to direct to each expert those patterns on which it performs the best. Consequently, each expert is trained on a sub-population of the training data. The soft competition version directs inputs to all experts, but weighs the output of each expert by its relative previous performance. Similarly, the error used for training the experts is affected by the amount of “belief” in each expert as determined by the gating network. This approach can be given a Bayesian formulation [30].

²There is evidence that inhibition plays a similar role in various brain processes, e.g., in the creation of orientation columns in the visual cortex [27] and in the sharpening of receptive fields in the somatosensory cortex [21, 14]

2.3 Hierarchical mixture of experts

Analogies can be drawn between the various ways in which one can construct gating networks and the various statistical approaches to classification. For instance, one can either construct a wide shallow network in which input patterns are split only once on their way to the experts, or a deep tree with many splits in which the regression or classification task is performed only at the leaf nodes. This latter approach has gained considerable attention with the appearance of the Classification and Regression Trees [31] methodology, which constructs a tree leading to different decisions in different regions of pattern space.³

A direct extension of competitive learning motivated by such recursive partitioning methods is the hierarchy of experts [34]. In this tree-structured network, each expert can be, in turn, a competitive mixture of experts. Learning in this model is more complicated, and therefore the different mixture components are simplified to be generalized linear models. The use of the Expectation-Maximization (EM) algorithm accelerates the learning.

3 Competitive learning and resource allocation

We conclude this review by focusing on resource allocation — an issue that is merely a technical detail in computational learning algorithms, yet may be of a crucial importance in biological instantiations of competitive learning. As we saw in section 1, competition over limited resources and the allocation of such resources are important characteristics of learning in biological systems. How do such systems manage and distribute their resources? It seems plausible to assume that a sophisticated system would choose the allocation strategy in response to the characteristics of the task. Consider, for instance, a situation that involves two consecutive learning tasks; in this case, one may distinguish between two possibilities:

1. The second task involves the same data as the first one;
2. The second task involves new data (albeit presented in the same sensory modality).

Let us assume the existence of a reduced (i.e., low-dimensional; cf. [35]) internal representation of the data, whose computation incurs a cost in terms of time and effort. It is then clear that in case (1) the system would be better off if it retains the same internal representation in both tasks. In comparison, in case (2) the existing representation may be modified, or a new one may be created. An intriguing glimpse into the approach taken by the brain in such a situation is provided by the recent experimental results concerning the representation of extra-corporal space in rat hippocampus [36].

Wilson and McNaughton studied place cells in the hippocampus using chronic multi-electrode implantation techniques that allow long-term recording of cell activities. They recorded more than a hundred cells from free-moving rats that were trained to become familiar with a part of a simple maze. In these rats, certain hippocampal cells fire in relation to their place in the maze; when the rats are suddenly exposed to a new place, they undergo a short (about 5 minutes) phase of learning in which cells that have been silent in the previous location are recruited to encode the new location. This process is consistent with a possible effective reduction in inhibition to cells that have not been recently active. Such a reduction of inhibition may enable them to modify their

³Experimental evidence from the olfactory cortex suggests that it may be performing hierarchical clustering [32]. This evidence led to the development of a biologically inspired model for hierarchical clustering [33], which revealed the potential of recursive partitioning in complex pattern classification.

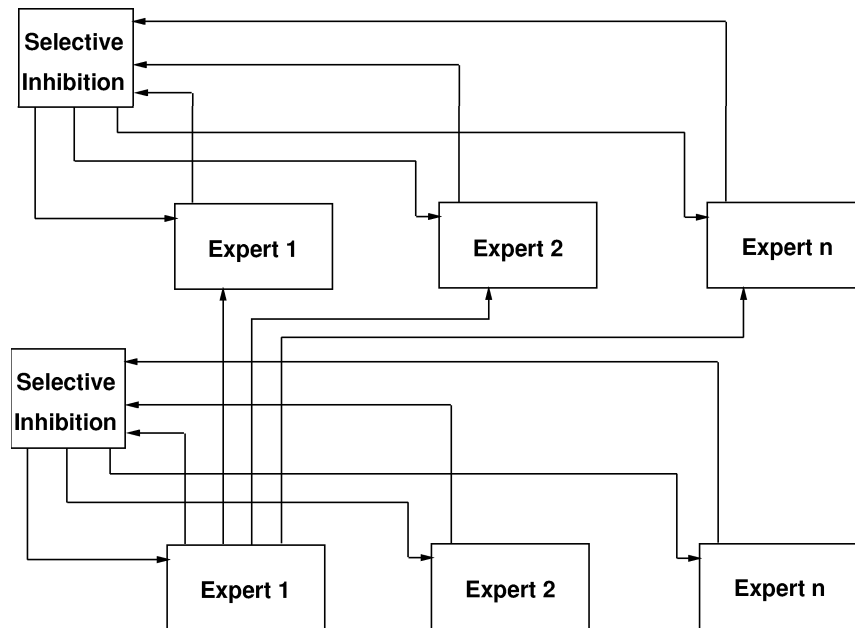


Figure 2: Resource allocation can be achieved by an effective reduction in inhibition to experts (sub-networks) that were not recently active, via a selective inhibition mechanism. Thus, when new inputs arrive which require some learning, only experts with low background inhibition are able to adapt. For clarity, only some of the connections between the first and the second layer are drawn.

response patterns rapidly and to become coherently active in the new location. Selective inhibitory distributions have been discussed in [37].

These results are consistent with a very simple mechanism for resource management that is based on short-term memory (STM) (Figure 2). According to the classical functional notion of short-term memory [38, 39], a transient memory trace can be transferred into long-term storage, subject to certain conditions. The memory management algorithm (which, in computer systems, is usually built around a table that holds pointers to the currently active memory pages) that oversees STM seems to operate on the basis of *recency*: a “flag” that indicates whether or not each particular clique of neurons has been active recently. The time span regarded as recent is probably different from one area of the brain to another; for the hippocampal place cells it is estimated to be around one minute. There, a unit that is not active for some time *enables the reduction of its surround inhibition*. When a global indication of the initiation of learning arrives at such a unit, its inhibition is selectively reduced, making it amenable for (re)allocation (learning). The plausibility of this scheme stems from its reliance on a global learning initiation signal (possibly carried by a striatal projection system; cf. Chevalier et al., 1985), [40]) which may be easier to implement than a general-purpose computer-like pointer-based memory management that requires a dedicated page-tracking subsystem and the ability to manipulate pointers to a pool of free units.⁴

Recently, a recurrent extension of the mixture of experts framework was suggested [41] in which the gating network has feedback connections that allow it to take temporal context into account. An extension that has similarities to hidden Markov models was also proposed [42]. Analogous developments may be identified in studies of the neurophysiological substrate of learning.

⁴The same mechanism may be useful for steering attention or managing short-term memory, rather than allocating long-term memory, if it is constrained to operate on a shorter time scale.

Carandini and Ferster recently reported the existence of a factor that slowly changes the membrane potential of a cortical neuron [43]. Barlow ties this result with a mechanism for the increase in the cortical response that occurs when an animal is attending to a particular stimulus [44]. It would be interesting to test whether such a mechanism is involved in the selective reduction of inhibition to inactive cortical regions, as proposed above.

In summary, utilization of the temporal structure of the stimulus — a major feature in the neurobiological accounts of learning — seems to be worth considering in the context of computational models of competitive learning as well. The natural role for the temporal information in such models is to shape the gating signal that controls the processing of the current stimulus by the competing functional units. By concentrating on a separate gating mechanism [28, 29], this approach (and, in particular, the simple method for recency-based resource allocation outlined above) augments other recent proposals for a general framework for competitive learning, such as the distinction between competition over pre-synaptic vs. post-synaptic resources [45].

An integration of these approaches, which may be accomplished by bridging the computational, the psychophysical and the neurophysiological characterizations of competitive learning, should lead to a better understanding of some of the more exciting issues in cognition, such as the determination of the role of stimulus-driven processes in forming the representations of higher-level sensory entities, such as entire objects and scenes in vision. Specifically, it would be interesting to characterize psychophysically the degree of competition among tuned functional units, in tasks other than vernier acuity (e.g., in categorization, as in the works of Mackintosh and his colleagues). In addition one would like to distinguish (experimentally and theoretically) between competition based on lateral inhibition and competition driven by a separate (e.g., top-down) flow of information.

References

- [1] R. D. Walk. Perceptual learning. In E. C. Carterette and M. P. Friedman, editors, *Handbook of Perception*, volume IX, pages 257–298. Academic Press, New York, NY, 1978.
- [2] D. Sagi and D. Tanne. Perceptual learning: learning to see. *Current opinion in neurobiology*, 4:195–199, April 1994.
- [3] C. Blakemore and F. W. Campbell. Adaptation to spatial stimuli. *J. Physiol.*, 200:11–13, 1968.
- [4] A. Fiorentini and N. Berardi. Perceptual learning specific for orientation and spatial frequency. *Nature*, 287:453–454, 1981.
- [5] Y. Moses, S. Ullman, and S. Edelman. Generalization to novel images in upright and inverted faces. *Perception*, 25:443–462, 1996.
- [6] T. Poggio, M. Fahle, and S. Edelman. Fast perceptual learning in visual hyperacuity. *Science*, 256:1018–1021, 1992.
- [7] M. W. Fahle, S. Edelman, and T. Poggio. Fast perceptual learning in hyperacuity. *Vision Research*, 35:3003–3013, 1995.
- [8] C. Blakemore and E. A. Tobin. Lateral inhibition between orientation detectors in the cat’s visual cortex. *Expl Brain Res.*, 15:439–440, 1972.
- [9] J. J. Gibson. Adaptation, after-effect, and contrast in the perception of curved lines. *J. Exp. Psychol.*, 16:1–31, 1933.
- [10] G. von Békésy. Zur theorie des Hörens. *Physikalische Zeitschrift*, 30:115–125, 1929. see K. R. Boff, L. Kaufman and J. P. Thomas, eds., *Handbook of perception and human performance*, vol. I, ch. 15, p. 22 (Wiley, 1986: New York).

- [11] N. J. Mackintosh. Categorization by people and pigeons: The twenty-second Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, 48B:193–210, 1995.
- [12] C. D. Gilbert. Neuronal dynamics and perceptual learning. *Current Biology*, 4:627–629, 1994.
- [13] M. W. Pettet and C. D. Gilbert. Dynamic changes in receptive-field size in cat primary visual cortex. *Proceedings of the National Academy of Science*, 89:8366–8370, 1992.
- [14] M. M. Merzenich, G. Recanzone, W. M. Jenkins, T. T. Allard, and R. J. Nudo. Cortical representation plasticity. In P. Rakic and W. Singer, editors, *Neurobiology of Neocortex*, pages 41–68. Wiley, New York, NY, 1988.
- [15] E. T. Rolls, G. C. Baylis, M. E. Hasselmo, and V. Nalwa. The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Exp. Brain Res.*, 76:153–164, 1989.
- [16] E. Kobatake, K. Tanaka, and Y. Tamori. Long-term learning changes the stimulus selectivity of cells in the inferotemporal cortex of adult monkeys. *Neuroscience Research*, S17:237, 1992.
- [17] N. K. Logothetis and D. L. Scheinberg. Visual object recognition. *Annual Review of Neuroscience*, 19:577–621, 1996.
- [18] R. N. Shepard. Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323, 1987.
- [19] H. B. Barlow. Conditions for versatile learning, Helmholtz’s unconscious inference, and the task of perception. *Vision Research*, 30:1561–1571, 1990.
- [20] D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6:559–601, 1994.
- [21] C. von der Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100, 1973.
- [22] D.E. Rumelhart and D. Zipser. Feature discovery by competitive learning. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, volume 1,2. MIT Press, Cambridge, MA, 1986.
- [23] Y. Weiss, S. Edelman, and M. Fahle. Models of perceptual learning in vernier hyperacuity. *Neural Computation*, 5:695–718, 1993.
- [24] T. H. Brown, E. W. Kairiss, and C. L. Keenan. Hebbian synapses: biophysical mechanisms and algorithms. *Ann. Rev. Neurosci.*, 13:475–511, 1990.
- [25] R. O. Duda and P. E. Hart. *Pattern classification and scene analysis*. Wiley, New York, 1973.
- [26] E. Oja. Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, 1:61–68, 1989.
- [27] A. S. Ramoa, M. A. Paradiso, and R. D. Freeman. Blockade of intracortical inhibition in kitten striate cortex: Effects on receptive field properties and associated loss of ocular dominance plasticity. *Experimental Brain Research*, 73:285–296, 1988.
- [28] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991.
- [29] S. J. Nowlan. Soft competitive adaptation: Neural network learning algorithms based on fitting statistical mixtures. Ph.D. dissertation, Carnegie Mellon University, 1991.
- [30] F. Peng, R. A. Jacobs, and M. A. Tanner. Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition. *Journal of the American Statistical Association*, 91:953–960, 1996.
- [31] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. The Wadsworth Statistics/Probability Series, Belmont, CA, 1984.
- [32] G. Lynch. *Synapses, Circuits and the Beginings of Memory*. MIT Press, Boston, 1986.

- [33] J. Ambros-Ingerson, R. Granger, and G. Lynch. Simulation of paleocortex performs hierarchical clustering. *Science*, 247:1344–1348, 1990.
- [34] M. I. Jordan and R. A. Jacobs. Hierarchical mixture of experts and the EM algorithm. *Neural Computation*, 1993.
- [35] S. Edelman and N. Intrator. Learning as extraction of low-dimensional representations. In D. Medin, R. Goldstone, and P. Schyns, editors, *Mechanisms of Perceptual Learning*. Academic Press, 1997. in press.
- [36] M. A. Wilson and B. L. McNaughton. Dynamics of hippocampal ensemble code for space. *Science*, 261:1055–1058, 1993.
- [37] P. O. Otterson and J. Storm-Mathisen. Distribution of Inhibitory Amino Acid Neurons in the Cerebellum With Some Observations on the Spinal Cord: An Immunocytochemical Study With Antisera Against Fixed GABA, Glycine, Taurine, and β -Alanine. *The Journal of Mind and Behavior*, 8(4):503–518, 1987.
- [38] R. C. Atkinson and R. M. Shiffrin. Human memory: a proposed system and its control processes. In K. W. Spence, editor, *Psychology of learning and motivation: advances in research and theory*, volume 2, pages 89–195. Academic Press, New York, 1968.
- [39] E. Tulving and D. L. Schacter. Priming and human memory systems. *Science*, 247:301–306, 1990.
- [40] G. Chevalier, S. Vacher, J. M. Deniau, and M. Desban. Disinhibition as a basic process in the expression of striatal function. I. The striatonigral influence on tecto-spinal/tecto-diencephalic neurons. *Brain Research*, 334:215–226, 1985.
- [41] T. W. Cacciatore and S. J. Nowlan. Mixtures of controllers for jump linear and non-linear plants. In J. D. Cowan, G. Tesauero, and J. Alspector, editors, *Neural Information Processing Systems 6*. Morgan Kaufman, San Mateo, CA, 1994.
- [42] Y. Bengio and P. Frasconi. An input output hmm architecture. In G. Tesauero, D. S. Touretzky, and T. K. Leen, editors, *Neural Information Processing Systems 7*, pages 427–434. MIT Press, Cambridge, MA, 1995.
- [43] M. Carandini and D. Ferster. A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science*, 276:949–952, 1997.
- [44] H. B. Barlow. Adaptation by hyperpolarization. *Science*, 276:913–914, 1997.
- [45] D. Willshaw, J. Hallam, S. Gingell, and S. L. Lau. Marr’s theory of the neocortex as a self-organizing neural network. *Neural Computation*, 9(4):911–936, 1997.