Hamilton, *J. Cardiovasc. Pharmacol.* **12**, 535 (1988); T. C. Hamilton and A. H. Weston, *Gen. Pharmacol.* **20**, 1 (1989).

22. We thank F. Altmayr, A. Dischner, L. Müller, and W. Zeitz for technical support, and J. Dudel, W. Baedeker, and H. Blömer for help and advice. Supported by the Deutsche Forschungsgemeinschaft (Da 177/2-4).

27 October 1989; accepted 24 January 1990

Simulation of Paleocortex Performs Hierarchical Clustering

JOSÉ AMBROS-INGERSON, RICHARD GRANGER,* GARY LYNCH

Simulations were performed of layers I and II of olfactory paleocortex, as connected to its primary input structure, olfactory bulb. Induction of synaptic long-term potentiation by means of repetitive sampling of inputs caused the simulation to organize encodings of learned cues into a hierarchical memory that uncovered statistical relationships in the cue environment, corresponding to the performance of hierarchical clustering by the biological network. Simplification led to characterization of those parts of the network responsible for the mechanism, resulting in a novel, efficient algorithm for hierarchical clustering. The hypothesis is put forward that these corticobulbar networks and circuitry of similar design in other brain regions contain computational elements sufficient to construct perceptual hierarchies for use in recognizing environmental cues.

HOW VARIOUS PROPERTIES OF memory might emerge from design features of circuits in cerebral cortex is a major problem area for neural network research (*1, 2*). In previous studies, we addressed this in models of the superficial layers of the olfactory cortex (*3*) by incorporating several of the characteristics of the synaptic long-term potentiation (LTP) effect (*4*). Implementation of a repetitive sampling feature meant to represent the cyclic sniffing behavior of mammals (*5*) produced a system that exhibited a kind of dual encoding of learned cues: early cycles (sniffs) generated response patterns that were common to a subset of cues that resembled each other, whereas later responses were specific to an individual member of the subset. This could mean that the cortical model simply constructs two types of representations (category and individual) or that it discovers hierarchical structure in the cue world and stores memory in this highly structured form. Human subjects in perceptual studies robustly recognize objects first at categorical levels and subsequently at successively subordinate levels (*6*), suggesting the presence of structured memories that are organized and searched hierarchically during recognition. Here we show that the olfactory cortex–olfactory bulb model, during learning, generates a multilevel hierarchical memory that uncovers statistical relationships inherent in collections of learned cues, and, during retrieval, sequentially traverses this hierarchical recognition memory. Moreover, simplification of the network results in an algorithm that provides a novel and efficient solution to the computationally difficult problem of hierarchical clustering.

The elements and circuitry simulated are shown in Fig. 1. Two networks, bulb and cortex, consisting of distinct architectures and physiologies, are extensively connected by both feedforward and feedback projections (*7*). The entire system works in synchrony with a 4- to 7-Hz (theta) sampling pattern that is characteristic of small mammals (*5*). Bulb mitral cells (those neurons innervated by the peripheral receptors and that project to cortex) receive inputs presented repetitively for brief periods. Inputs to the cortical network arise from the resultant synchronous bursting in a subset of mitral cells, yielding cyclic activity in relatively discrete "operation cycles" time-locked to the sampling rhythm. Sparse random connectivity in the simulation selectively activates those cortical cells whose dendrites are most connected to the input lines that are active. Learning increments active synapses on sufficiently depolarized cells via a rule based on LTP (*3*), which has been shown to produce a measurable increment in synaptic strength during even a single 50-ms burst of activity (*4*), that is, within a single operation cycle in the model. Learning requires only a few training trials per

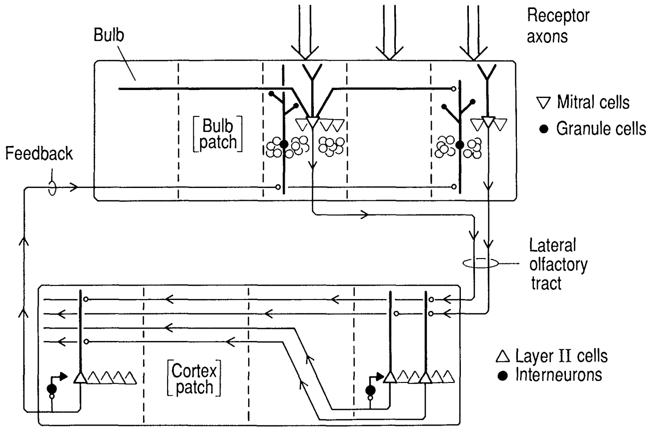


Fig. 1. Anatomical architecture of the bulbar-cortical simulation. The bulb simulation contains 400 projection (mitral) units (simulated neurons), divided into 40 separate groups, each of which receives an input from one group of peripheral receptor axons (*22*). The intensity of an input is reflected in the number of cells within the appropriate group (or groups in the case of multicomponent cues) that it activates. The (excitatory) mitral projection cells in bulb have been shown to have extremely long oblique dendrites that form dendrodendritic contacts with a dense granule cell inhibitory network (*22*). We adopt an assumption made by others (*12*) that this excitatory-inhibitory arrangement serves to normalize the output of the bulb (that is, the total number of mitral cells that are activated is reasonably constant across cues with different intensities and compositions). The inhibitory neurons (granule and probably periglomerular) are represented in the model by a single layer of cells and are innervated by randomly organized excitatory feedback from cortex. The strength of the simulated feedback contacts is set during a "development" period in which hundreds of cues are presented and the strength of feedback synapses allowed to vary according to a correlational (Hebb) rule. The mitral cells of bulb project sparsely and nontropographically to the outermost layer of olfactory cortex via the LOT, both biologically (*7*) and in the simulation. The cortex is simulated as a layer of 1000 excitatory layer II cells that are assumed in the model to be arranged into patches of 20 cells each by the radial axonal arborizations of local (feedback) inhibitory interneurons. The modeled neurons sum the voltages from their active synaptic inputs and require different amounts of depolarization for discharges, bursts of discharges, and for induction of synaptic change via LTP. A more detailed simulation of individual patches has shown that active cells that trigger inhibitory interneurons can suppress firing by other cells in a patch; because of this, typically only one or two cells in a patch will discharge in response to bulbar inputs, making each patch into a competitive (or modified winners-take-all) arrangement of the type discussed by many authors (*9, 10*). Such an arrangement is assumed in the present simulation. Each cortical cell receives input from the LOT and from a feedforward associational system generated by the cortical neurons themselves, in both cortex (*7, 23*) and the simulation. The operating rules for the model are based on physiological data reported in the literature (*5, 24*).

Center for the Neurobiology of Learning and Memory, University of California, Irvine, CA 92717.

*To whom correspondence should be addressed.

cue, as in rapid olfactory learning in mammals (*8*). Input lines shared across many similar input cues, and thus participating in many learning episodes, will strengthen their target synapses more than lines that participate in relatively fewer episodes. The result is that cortical dendrites (which can be viewed as vectors being moved by synaptic learning) become increasingly well tuned to those inputs containing the shared subset; that is, those inputs that are sufficiently similar to constitute members of a cluster. We have shown that this circuit will generate cell-firing responses that group learned cues by similarity. For a given threshold of input similarity among a set of cues, outputs are identical for all of the cues, whereas below that similarity threshold, outputs are much less similar than corresponding inputs (*3*). This form of unsupervised learning (*9, 10*) is to be distinguished from supervised learning, in which categorization information is provided to the learner.

Feedback from cortex to the bulb inhibitory layer in the model (Fig. 1) is trained by means of a correlational rule during an earlier "developmental" period via a Hebb rule coarsely correlating activity in cortex with activity in bulb. The feedback then selectively inhibits the mitral cells in those bulb patches (Fig. 1) that are most responsible for the cortical output response, via relatively long-lasting inhibition that has been shown to exist in olfactory bulb (*11*). Resulting renormalization of bulb activity (*12*) maintains the total number of firing mitral cells at a roughly constant level. This renormalization thus recruits additional mitral cells in remaining (uninhibited) bulb patches to fire to compensate for those selectively inhibited by the feedback. Thus, the pattern of mitral cell firing on the next operation cycle (roughly 200 ms later) is distinct spatially from the previous pattern. This in turn activates a distinct set of cortical cells. This sampling cycle (bulb activation → cortical activation → inhibitory feedback → renormalization) can be repeated until bulb is sufficiently inhibited to be largely quiescent. The sequence of cortical responses after the initial (first-sample) response becomes progressively more different for different cues (*3*), increasingly approximating a given cue, thus producing unique encodings for individuals.

Until now it had not been determined whether the network had the ability to discover secondary or intermediate structure in hierarchically organized input cues. Table 1 gives responses of the simulation after training to such a structured environment, one consisting of clusters, subclusters within those clusters, and individuals within the subclusters. On its first cycle, network responses to members of a given cluster of input cues are nearly identical, thus grouping those cues together, whereas between-group overlap is extremely low. Second-cycle responses are nearly identical for members of subclusters, but not for members of the containing clusters. Third-cycle network responses are nearly unique for each individual cue. The network thus discovers intermediate structure, exhibiting an ability corresponding to hierarchical clustering (*13*), which identifies multilevel statistical structure in novel data.

Empirical results of this type do not by themselves elucidate the mechanism by which the hierarchical clustering operation is accomplished. Of primary interest is the identification and characterization of the essential design features of the network underlying its hierarchical clustering ability. Analysis led to such a characterization, and controlled testing revealed that the resulting simplified formulation of the network provides a novel and efficient method of hierarchical clustering.

The simplified formulation of the cortex contains a weight matrix W [corresponding to the connections between the lateral olfactory tract (LOT) and piriform layer II cell dendrites] (Fig. 1), divided into H nonoverlapping "winners-take-all" or "competitive" (*9, 10*) subnets of cells, each competing to "win" within their subnet. Each (densely connected) subnet $S_i \in \{S_1, S_2, \ldots, S_H\}$ contains weight vectors $\mathbf{C}$ (that is, columns of W) such that $W = \cup_i S_i$ (*14*). Each subnet S_h corresponds to the network response at hierarchical level h. The simplified bulb receives N-dimensional real-valued vectors (corresponding to frequency and spatial patterns of activation input to bulb) and passes these to cortex, relaxing the requirement in the biological model that inputs to cortex must be binary. The cortical network is thus trained on a set of N-dimensional real-valued vectors via an extension of a correlational (Hebbian) learning algorithm (a simplification of the LTP rules for synaptic modification). An input vector $\mathbf{X}$ is presented first to the highest subnet in the hierarchy, S_1. The column vectors (dendrites) $\mathbf{C}$ in this subnet that win the (winners-take-all) competition on $\mathbf{X}$ (corresponding to target cells that are most depolarized by this input) are identified. The synaptic contacts on these winning vectors are then trained, mov-

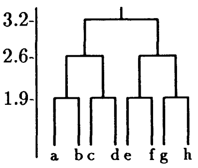
Table 1. Overlaps among responding vectors in cortical-bulbar simulation to eight hierarchically organized input cue vectors. The input cues can be visualized as the hierarchical cluster dendrogram at the left side of the table (see the legend to Fig. 2b), breaking the input space into two superordinate clusters, which are divided into four intermediate subclusters (two subclusters in each cluster), which in turn are subdivided into eight individual cues (a to h), two individuals to a subcluster. Average Euclidean distance between the two cues in a subcluster is 1.9; between the means of the subclusters with a cluster is 2.6; and between the means of the two clusters is 3.2. Response overlaps are defined as follows: let $\zeta = \{\zeta_1, \zeta_2, \ldots, \zeta_k\}$ be the set of clusters at one level (superordinate, intermediate, or subordinate), where each cluster consists of a set of cues. For $\mathbf{X} \in \zeta_i$ let $R_h(\mathbf{X})$ be the set of cortical cells responding to $\mathbf{X}$ at operation cycle h. The average percent within-cluster overlap at level ζ for cycle h is:

$$\mathcal{W}_{\zeta,h} = \frac{100}{A|\zeta|} \sum_{\zeta_i \in \zeta} \left| \bigcap_{\mathbf{X} \in \zeta_i} R_h(\mathbf{X}) \right| \tag{5}$$

where A is the number of patches (assuming one winner per patch) and $|\zeta|$ is the cardinality of set ζ. Analogously, average percent between-cluster overlap is

$$\mathfrak{B}_{\zeta,h} = \frac{100}{A\binom{|\zeta|}{2}} \sum_{\substack{\zeta_i,\zeta_j \in \zeta \\ i<j}} \left[\sum_{\substack{\mathrm{X}_i \in \zeta_i \\ \mathrm{X}_j \in \zeta_j}} \frac{|R_h(\mathrm{X}_i) \cap R_h(\mathrm{X}_j)|}{|\zeta_i||\zeta_j|} \right] \tag{6}$$

First-cycle responses were nearly identical (average overlap among responses was 91%) for all objects within either of the two (superordinate) clusters, indicating simply membership of a cue in a cluster, whereas between-cluster overlap between the superordinate clusters (right-hand column) was extremely low (8% or less) for all trials. Second-cycle responses were very similar (average within-group overlap 88%) for subcluster members, but were distinct for other members of the containing cluster that were not members of the subcluster (average within-group overlap 31%). Third-cycle responses were distinct for each individual cue: little or no overlap exists among responses to individual members of clusters (0%) or subclusters (17%).



	$\mathcal{W}_{\zeta,h}$*			$\mathcal{B}_{\zeta,h}$*
Cycle	Superordinate (clusters) (%)	Intermediate (subclusters) (%)	Subordinate (individuals) (%)	Superordinate (clusters) (%)
1	91	99	100	8
2	31	88	100	5
3	0	17	100	2

*$\mathcal{W}_{\zeta,h}$, Within-group overlap among all responses; $\mathfrak{B}_{\zeta,h}$, between-group overlap between superordinates.

ing the vectors closer to **X** by an increment γ_c. Feedback from the just trained vectors then partially inhibits or "masks" the input. The remainder of the input is presented to the next lower subnet of cells in the hierarchy, until all hierarchical subnets $S_1, S_2, \ldots, S_H$ have been trained, over H operation cycles. At any given hierarchical level S_h, the **C**s in that subnet can be shown to converge to the means of the clusters of cues on which they are trained, as in related "competitive learning" algorithms (*3, 9, 10*). The feedback inhibition step enables vectors in W assigned to the subordinate hierarchical levels to converge to means of subclusters of the data, allowing secondary (and H-ary) structure to be identified (for H divisions of the weight matrix into subnets). Formally

Step 1. Do steps 2 to 5 for each input **X** to be learned.

Step 2. Do steps 3 to 5 for each hierarchical level $h \epsilon \{1, 2, \ldots, H\}$.

Step 3. Identify winning cells (column vectors) in subnet S_h for input **X**: win($\mathbf{X}$, S_h).

Step 4. Train each of the winning cells $\mathbf{C} \epsilon$ win $(\mathbf{X}, S_h)$ identified in step 3: $\mathbf{C} \leftarrow \mathbf{C} + \gamma_c(\mathbf{X} - \mathbf{C})$.

Step 5. Subtract winners from input: $\mathbf{X} \leftarrow \mathbf{X} - \text{mean}[\text{win}(\mathbf{X}, S_h)]$.

where H is the depth of the hierarchy; $\text{win}(\mathbf{X}, S_h) = \{\mathbf{C} \epsilon S_h : [\mathbf{X} \cdot \mathbf{C} = \max_{\mathbf{C}_i \epsilon S_h}(\mathbf{X} \cdot \mathbf{C}_i)] \wedge \mathbf{X} \cdot \mathbf{C} > 0\}$ is the set of weight vectors within a subnet S_h that wins the competition on the input **X**; and γ_c is the learning rate (*15*).

To enable controlled testing of the ability of the simplified formulation to identify hierarchical structure, we created cue environments with known hierarchical structure. The cues each consist of the sum of a sequence of orthogonal multidimensional vectors with noise, forming a hierarchy of subclusters within clusters:

$$\text{cue} = M + \sum_{i \epsilon \text{ path to cue}} (\mathbf{B}_i + K_i) \qquad (1)$$

where each $\mathbf{B}_i$ is an orthogonal vector in the summation path to the cue (Fig. 2); M is noise and

$$K_i = N(\mu_i, \sigma_i) \frac{\mathbf{B}_i}{\|\mathbf{B}_i\|} \qquad (2)$$

that is, each K_i is unidimensional Gaussian noise in the direction of component i. The result is a set of vector sums that correspond to groups of vectors that are naturally clustered at each of the i levels of the hierarchy (Fig. 2, a and b) (*16*).

The dendrograms produced by the simplified formulation (*17*) are shown in Fig. 3a. Single-level competitive learning partitions the input space in a piecewise linear fashion (*9*); for the present algorithm, the input space at level h is itself a partition generated at level $h - 1$ and is recursively subpartitioned in the same piecewise linear manner. Like most probabilistic algorithms, the one presented here can fail to identify the full hierarchical structure in some circumstances (for example, degenerate initial conditions, initial skew bias in the data); empirically however, failure tends to be graceful in that correct structure is identified, although intermediate structure may be either missed or interposed. The breadth of the categories created is dependent on the number and distribution of units in a subnet.

The simplified formulation of the network, besides representing selected characteristics of the interacting biological systems (olfactory bulb and cortex) in the larger simulation, can be treated as a proposed novel algorithm for hierarchical clustering. As such, its performance on structured data can be compared directly against standard algorithms in the hierarchical clustering literature (Fig. 3b). Moreover, the space and time complexity of the algorithm can be evaluated. The weight matrix W contains H layers or distinct sets (subnets) of units. For complete separability of n cues, the bottom of the hierarchy (S_H) must contain at least n units, so the complete hierarchy will contain roughly $b(b - 1)^{-1}n$ units, organized into a tree consisting of $\log_b n$ hierarchical layers, where b is the average branching factor at each level. Thus the number of units required grows linearly with the number of cues to be learned: the space complexity of the algorithm is $O(nN)$, where N is the dimensionality of **X** [a function $g(n)$ is "order" $f(n)$—denoted $O(f(n))$—if for large n, $g(n) \leq kf(n)$ for some constant k].

The three time costs of the algorithm for each input vector **X** at level h are (i) summation of inputs on subnets S_h; (ii) computation of subnet winners **C**; and (iii) weight modifications on **C**. On a serial processor, after processing all levels (i) is $O(nN)$, (ii) is $O(n)$, and (iii) is $O(N \log n)$. Because of the inherent parallelism of the algorithm, on a suitable parallel processor, (i) is $O(\log N)$, (ii) is $O(\log n)$, and (iii) is constant. Thus training time per presentation, assuming $O(n)$ units in the net, is $O(nN)$ in serial and $O(\log n \cdot \log nN)$ in parallel. We have not determined analytically the number of instance presentations per cue required for convergence, though empirically a small number (~10) has sufficed (see legend to Fig. 3). Hence, we conjecture that training time to process a collection of n objects to convergence is $O(n^2N)$ in serial and $O(n \cdot$

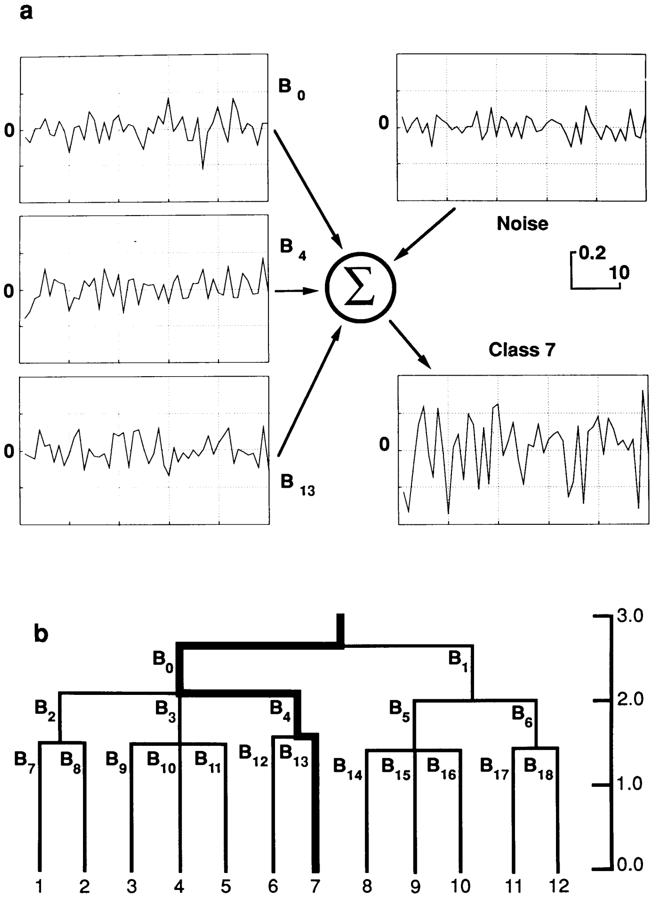


Fig. 2. (**a**) A sample member of one group of a created hierarchy of cues. The 50-dimensional vector, an example of one group (class 7) of 12 [see (b)], is the sum of three orthogonal vectors plus a noise vector. Positions on the *x*-axes correspond to each of the 50 dimensions of the vectors; the *y*-axes denote the values of each of the dimensions. This particular vector sum corresponds to the end point of the path indicated by darkened lines in the hierarchical tree in (b). (**b**) A hierarchy of 50-dimensional real-valued vectors like the instance in (a), created such that they are naturally clustered at each of three levels, via an algorithm described in the text. Each vector is the sum of three orthogonal component vectors plus a noise vector. All vectors under node $\mathbf{B}_0$ contain the same initial component ($\mathbf{B}_0$); all under node $\mathbf{B}_1$ contain initial component $\mathbf{B}_1$; of the vectors under node $\mathbf{B}_0$, those under node $\mathbf{B}_3$ all contain initial component $\mathbf{B}_0$ and second component $\mathbf{B}_3$; and so on. The height of each node corresponds to the average distance among the means of the data below the node.

log $n \cdot \log nN$) in parallel. The inherent parallelism and uniformity of the steps permits efficient hardware implementation (*18*).

The simplified network formulation aforementioned consists of specific operations of winning a competitive subnet, training winning vectors, and masking the input by the trained vectors. Admitting a range of different mechanisms for competition, training, and masking yields formulation of a general algorithm of which the simplified network is a special case

Step 1. Do steps 2 to 5 for each input **X** to be learned.

Step 2. Do steps 3 to 5 for each hierarchical level $h\epsilon\{1, 2, \ldots, H\}$.

Step 3. Identify winning cells (column vectors) in subnet S_h for input **X**.

Step 4. Train each winning cell identified in step 3.

Step 5. Mask structure identified by winners in step 3 from input **X** to produce remainder of input **X** for further processing.

The generalization bears some relation to algorithms identifying principal components of data (*19, 20*); indeed, the algorithms set forth by Oja *et al.* and Sanger (*19*) can be cast as distinct special cases of the generalized formulation of the algorithm. We conjecture that the new general algorithm characterizes a class of repetitive sampling algorithms that successively approximate statistical aspects of data, including eigenvectors and clusters.

Our results provide an instance in which a novel and efficient algorithm for a well-studied computational problem is developed from a simulation of a specific cortical network. Reflecting the system from which it is derived, the algorithm is inherently parallel and hence lends itself to efficient implementation in hardware. The present findings also point to the hypothesis that approximate hierarchical clustering will emerge as a fundamental property of memories based (at least in part) on LTP-like synaptic modifications in damped oscillatory networks of the type found in the bulbar-cortical system. The model is sufficiently detailed to make testable predictions at both behavioral and physiological levels (for example, different cortical cells should discharge over successive sampling cycles with progressively more selective tuning), but relevant experimental data is not yet available. Finally, the general architectural plan of the bulbar-cortical system finds parallels in certain aspects of thalamo-cortical relations [that is, secondary thalamic projection nuclei → superficial neocortex → deep layers of cortex → nucleus reticularis → projection nuclei (*21*)]. This similarity suggests a possible connection with attentional functions proposed for thalamocortical interactions (*2*), and raises the possibility that this biologically generated mechanism for hierarchical clustering may be a routine part of perceptual recognition memory behavior in animals and humans (*6*).

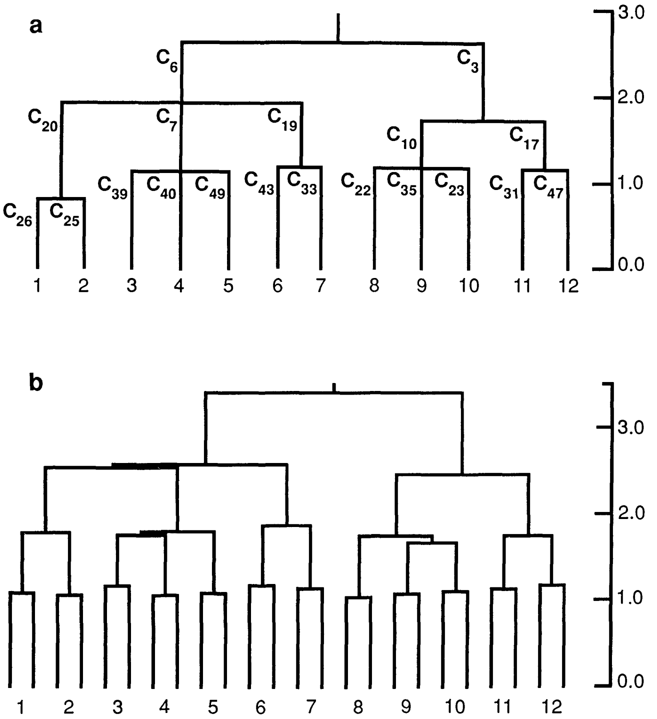


Fig. 3. (**a**) Dendrogram structure created by simplified network after training. The network was trained on three passes over a training set consisting of a sample of 120 cues in random order (each of 12 categories represented by approximately ten instances). After training, the network was tested on a distinct testing set consisting of ten novel instances of each of the 12 categories, and a record was kept of which cells won on each cue presentation. Analysis of this record showed that cells become tuned to groups of cues that correspond to categories; cell responses are indicated in the dendrogram. For instance, cell $\mathbf{C}_6$ wins on the first sniff for all instances of cues from categories 1 through 7 and not on any other (that is, $\mathbf{C}_6$ wins if and only if the cue is from category 1 through 7); cell $\mathbf{C}_{19}$ wins on the second sniff for cues from categories 6 and 7 and no other; cell $\mathbf{C}_{33}$ wins on the third sniff if and only if the cue comes from category 7. The height of each node is given by the average Euclidean distance among the weight vectors represented by each cell at that level (for example, the average distance between weight vectors $\mathbf{C}_{39}$, $\mathbf{C}_{40}$, and $\mathbf{C}_{49}$ is 1.11). (**b**) Dendrogram created by agglomerative hierarchical clustering algorithm using Euclidean group–mean distance metric (*13*) on the same testing set data as in (a). Height of each node denotes distance between the two immediately subordinate constituents of the node.

REFERENCES AND NOTES

1. M. A. Gluck and R. F. Thompson, *Psychol. Rev.* **94** (2), 176 (1986); W. J. Freeman, *Mass Action in the Nervous System* (Academic Press, New York, 1975); L. B. Haberly and J. M. Bower, *Trends Neurosci.* **12**, 258 (1989); K. D. Miller, J. B. Keller, M. P. Stryker, *Science* **245**, 605 (1989); C. von der Malsburg and D. J. Willshaw, *Exp. Brain Res. Suppl.* **1**, 463 (1976); G. M. Edelman and G. N. Reeke, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2091 (1982); M. F. Bear, L. N. Cooper, F. F. Ebner, *Science* **237**, 42 (1987); D. H. Ballard, G. E. Hinton, T. J. Sejnowski, *Nature* **306**, 21 (1983).
2. F. Crick, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4586 (1984).
3. R. Granger, J. Ambros-Ingerson, H. Henry, G. Lynch, in *Neural Information Processing Systems*, D. Anderson, Ed. (American Institute of Physics Press, New York, 1988); R. Granger, J. Ambros-Ingerson, G. Lynch, *J. Cog. Neurosci.* **1**, 61 (1989); G. Lynch and R. Granger, *Psychol. Learn. Motiv.* **23**, 205 (1989); R. Granger, J. Ambros-Ingerson, U. Staubli, G. Lynch, in *Neuroscience and Connectionist Theory*, M. A. Gluck and D. E. Rumelhart, Eds. (Erlbaum, Hillsdale, NJ, 1989); R. Granger, J. Ambros-Ingerson, P. Anton, G. Lynch, in *Connectionist Modeling and Brain Function: The Developing Interface*, S. J. Hanson and C. R. Olson, Eds. (MIT Press, Cambridge, MA, 1989).
4. G. Lynch, *Synapses, Circuits and the Beginnings of Memory* (MIT Press, Cambridge, 1986); J. Larson and G. Lynch, *Science* **232**, 985 (1986); *Brain Res.* **441**, 111 (1988); D. Muller, M. Joly, G. Lynch, *Science* **242**, 1694 (1988).
5. F. Macrides, *Behav. Biol.* **14**, 295 (1975); ——, H. Eichenbaum, B. Forbes, *J. Neurosci.* **2**, 1705 (1982); B. R. Komisaruk, *Neurosciences Research Program Bulletin* **11**, 376 (1973); D. v. Holst and H. Kolb, *J. Comp. Physiol.* **105**, 243 (1976).
6. I. Biederman, *Science* **177**, 77 (1972); M. C. Potter and B. A. Faulconer, *Nature* **253**, 437 (1975); A. M. Collins and E. F. Loftus, *Psychol. Rev.* **82**, 407 (1975); E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, P. Boyes-Braem, *Cog. Psychol.* **8**, 382 (1976); E. Rosch and B. B. Lloyd, *Cognition and Categorization* (Erlbaum, Hillsdale, NJ, 1978); E. E. Smith and D. L. Medin, *Categories and Concepts* (Harvard Univ. Press, Cambridge, MA, 1981); J. Hoffmann and C. Ziessler, *Z. Psychol.* **194**, 135 (1983); P. Jolicoeur, M. A. Gluck, S. M. Kosslyn, *Cog. Psychol.* **16**, 243 (1984); J. Corter, M. A. Gluck, G. H. Bower, *Proc. 10th Annu. Conf. Cog. Sci. Soc.* (Erlbaum, Hillsdale, NJ, 1988); M. A. Gluck and G. H. Bower, *J. Exper. Psych.* **117**(3), 227 (1988).
7. J. L. Price, *J. Comp. Neurol.* **150**, 87 (1973); L. B. Haberly and J. L. Price, *Brain Res.* **129**, 152 (1977); L. B. Haberly, *Chem. Senses* **10** (2), 219 (1985).
8. J. W. Jennings and L. H. Keefer, *Psychol. Reports* **24**, 3 (1969); B. M. Slotnick and H. M. Katz, *Science* **185**, 796 (1974); B. M. Slotnick and N. Kaneko, *ibid.* **214**, 91 (1981); H. Eichenbaum, A. Fagan, N. J. Cohen, *J. Neurosci.* **6**, 1876 (1986); U. Staubli, D. Fraser, R. Faraday, G. Lynch, *Behav. Neurosci.* **101**, 757 (1987).
9. S. Grossberg, *Biol. Cybernetics* **23**, 121 (1976); T. Kohonen, *Self-Organization and Associative Memory* (Springer-Verlag, New York, 1984).
10. C. von der Malsburg, *Kybernetik* **14**, 85 (1973); D. E. Rumelhart and D. Zipser, in *Parallel Distributed Processing*, D. Rumelhart and J. McClelland, Eds. (MIT Press, Cambridge, MA, 1986).
11. R. A. Nicoll, *Brain Res.* **14**, 157 (1969); K. Mori and G. M. Shepherd, *ibid.* **172**, 155 (1979); K. Mori, M. C. Nowycky, G. M. Shepherd, *J. Physiol.* **314**, 311 (1981).

12. G. M. Shepherd, in *The Neurosciences: Second Study Program*, F. O. Schmitt, Ed. (Rockefeller Univ. Press, New York, 1970); G. M. Shepherd and R. K. Brayton, *Brain Res.* **175**, 377 (1979); J. S. Kauer, in *The Neurobiology of Taste and Smell*, T. E. Finger, Ed. (Wiley, New York, 1987).
13. Agglomerative (bottom-up) hierarchical clustering algorithms are of the following form: (i) calculate (or update) a distance matrix (all pairwise distances among cues); (ii) identify "closest" pair of cues in the data according to a chosen distance metric, and group those into a cluster; (iii) repeat steps (i) and (ii) using newly formed clusters in place of their members, until all elements are members of a cluster. Divisive (top-down) hierarchical clustering methods typically subdivide each cluster (initially a single one comprising all the data) into k subclusters for each hierarchical level (k is typically chosen by the user). This is done by iteratively attempting to locally optimize a given objective function (for example, mean-squared distance from cluster members to cluster mean) by refining an initial, often arbitrary, assignment of data points into k clusters [R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy* (Freeman, San Francisco, 1963); G. N. Lance and W. T. Williams, *Nature* **212**, 218 (1966); R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973); B. Everitt, *Cluster Analysis* (Wiley, New York, 1980)].
14. Each subnet S_h in the simplified formulation consists of those cells that respond at particular operation cycle h of the network. The subnet structure is constructed from the anatomical network via two steps: (i) the sparsely connected anatomical patches in the network are combined into a single network with dense connectivity; (ii) the cells in this single network are then divided into H subnets and the cells in each subnet are allowed to respond during only one prespecified operation cycle. That is, each subnet at level h contains only those cells that respond on operation cycle h; these correspond to the responses at hierarchical level h. Even without this simplification, the network still identifies hierarchical structure, as shown earlier (Table 1); the simplification enables formal distinction among cells responding to a given operation cycle, enabling a tractable algorithmic formulation.
15. The learning rate, γ_c, is chosen to satisfy $\gamma_c(t) \rightarrow 0$ and $\Sigma_t \gamma_c(t) = \infty$ where t is the number of times **C** has been trained [L. Ljung, *IEEE Trans. Autom. Control* **AC-22**, 551 (1977)].
16. Parameters are set as follows: $\mu_{0,1} = 1.6$; $\mu_{2,\ldots,6} = 1.3$; $\mu_{7,\ldots,18} = 1.0$; $\sigma_i = 0.1$, $\|\mathbf{B}_i\| = 1$, $i = 0, \ldots, 18$; and M is uncorrelated Gaussian noise on each input line with $\mu = 0.0$ and $\sigma = 0.1$.
17. $H = 3$; $\gamma_c = 0.2t^{-1/2}$ (satisfying the specified constraints for γ); and the cardinalities of S_i's are set so that $|S_{i+1}| = b|S_i|$ is maintained as closely as possible (b, the average branching factor of the tree, is a small number in the range 2 to 5). We use $|S_1| = 7$, $|S_2| = 14$, and $|S_3| = 29$. W was constructed by selecting pseudorandom vectors from a uniform distribution over the hypersphere of radius $r = 0.5$ [G. S. Watson, *Statistics on Spheres* (Wiley, New York, 1983)].
18. J. Bailey, D. Hammerstrom, J. Mates, M. Rudnick, in *An Introduction to Neural and Electronic Networks*, S. F. Zornetzer, J. L. Davis, C. Lau, Eds. (Academic Press, New York, 1989); D. Hammerstrom, in *Neural Information Processing Systems*, D. Anderson, Ed. (American Institute of Physics Press, New York, 1988).
19. E. Oja, *J. Math. Biol.* **15**, 276 (1982); ——— and J. Karhunen, *J. Math. Anal. Appl.* **106**, 69 (1985); T. D. Sanger, *Neural Network* **2**, 459 (1989).
20. R. Linsker, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7508 (1986); *ibid.*, p. 8390; *ibid.*, p. 8779; *ibid.*, p. 8783; *IEEE Computer* **21**, 105 (1988).
21. M. E. Scheibel and A. B. Scheibel, *Brain Res.* **1**, 43 (1966); E. G. Jones, *J. Comp. Neurol.* **162**, 285 (1975); M. Herkenham, in *Cerebral Cortex*, E. Jones and A. Peters, Eds. (Plenum, New York, 1986), vol. 5, pp. 403–445.
22. G. M. Shepherd, *Physiol. Rev.* **52**, 864 (1972); F. R. Sharp *et al.*, *J. Neurophysiol.* **40**, 800 (1977); K. Mori, *Prog. Neurobiology* **29**, 275 (1987).
23. J. E. Schwob and J. L. Price, *Brain Res.* **151**, 369 (1978).
24. W. J. Freeman and W. Schneider, *Psychophysiology* **19**, 44 (1978); L. B. Haberly and J. M. Bower, *J. Neurophysiol.* **51**, 90 (1984); G. Tseng and L. B. Haberly, *ibid.* **59**, 1352 (1988); F. Roman, U. Staubli, G. Lynch, *Brain Res.* **418**, 221 (1987).

25. We thank J. W. Whitson, Jr., and P. Antón for their assistance with this research. We also thank the helpful comments of the reviewers of this paper. Supported in part by the Office of Naval Research under grants N00014-89-J-1255 and N00014-89-J-3179.

19 October 1989; accepted 19 January 1990
