

Theoretical Note

Parsimony in neural representations: Generalization of a model of spatial orientation ability

ROBERT B. GLASSMAN
Lake Forest College, Lake Forest, Illinois

Brains must be adequate to process complex information, but they must also have a simple enough underlying organization to have evolved by natural selection. Beginning with a network model, originally developed to show how spatial orientational behavior might be organized with a modest number of neural connections, the following hypothesis is offered about a pattern of connections recurring within the brain: All significant activation patterns of a large number of input elements are transformed to output patterns by small numbers of associative subsystems, or modules, which do not require computerlike algorithms. They are single neurons or other functional units, which individually merely summate inputs; together, they discriminate among probable inputs without requiring a complex representation to do so. They may be thought of as registering (1) values of input dimensions, (2) combinations of activated input elements, or (3) numerical labels for distinct inputs. The informational capacity of a set of modules is a function of its number of modules and their dynamic range. The set works most efficiently if all its modules have the same range. Elementary combinatorial considerations suggest that besides receiving patterns of connections appropriate to the information being processed, sets of associative modules probably receive inputs systematically restricted in more general ways, for example by lateral inhibition or by connections from a small or large, but not intermediate, proportion of the set of input modules.

Efforts to develop neural network theories include computer simulations, vector concepts, and other sophisticated techniques of multidimensional analysis (e.g., Feldman & Ballard, 1982; Fukushima, 1984; Grossberg, 1978; Hinton & Anderson, 1981; Kohonen, 1978; Pellionisz & Llinas, 1982; Willshaw, 1981; and Levine's, 1983, extensive review), but some ideas about neural networks lend themselves to more elementary mathematical techniques. This paper offers a new hypothesis about numbers of connections required by associative systems and about the dynamic ranges of their components. The idea began with a surprising observation in cats recovering from somatosensory cortical damage.

Cutaneous Orientation-Localization in Cats

When blindfolded cats orient the head towards a tactile stimulus, muscles of the limbs, trunk, and neck must all adjust to bring the mouth into contact with the point of the body that was touched, the pattern of adjustment de-

pending on the location of the stimulus (Glassman, 1970). This behavior has an interesting degree of complexity, greater than in local reflexes, which simply involve motor output to the same part of the body from which input is received.

While cats were behaviorally recovering from unilateral damage to SII cortex and adjacent areas, many were unexpectedly observed to pass through a stage in which the vertical and horizontal orientational movements were largely dissociated. For example, when the cutaneous stimulus was applied to the left forepaw, in some cases the blindfolded cat pitched its head downward to paw level before initiating any leftward movement; in other cases, the lateral movement occurred before the vertical component (Glassman, 1983).

These observations are reminiscent of an aspect of the organization of industrial robot arms; that is, they require only three sets of motors and controllers to reach any point in a volume (Engelberger, 1980). The observations suggest that the cat's brain may somehow behave analogously in moving the mouth to any point in the three-dimensional space occupied by its body surface. What are the critical factors that might underlie a parsimonious, three-module neural system? First, consider a "straw system," which is unlikely to exist for orientation-localization in higher animals.

I thank Roger J. Faber and Robert J. Troyer for their helpful comments about the mathematics. This paper includes some of the ideas and references kindly offered by J. A. Anderson, S. Clarke, R. L. Curtis, H. Szechtman, and G. Yuen during my Society for Neuroscience poster presentation.

The author's mailing address is: Department of Psychology, Lake Forest College, Sheridan and College Rds., Lake Forest, IL 60045.

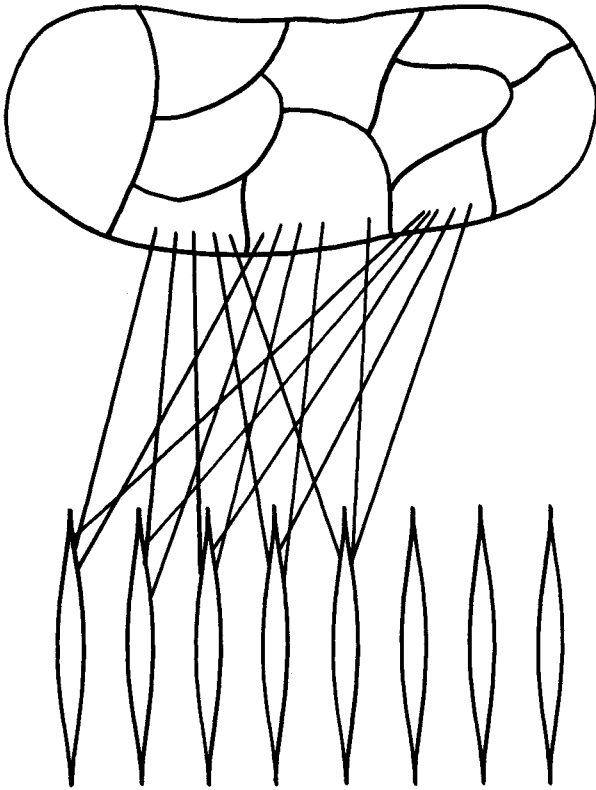


Figure 1. Scheme of a hypothetical system for controlling orientation of the head toward parts of the body; each discriminable skin area is directly connected to each muscle. Only some of the connections are sketched in.

A Two-Level Input-Output Model

In the neural model for orientation-localization schematized in Figure 1, a simplifying assumption is that the skin surface comprises discrete, discriminable units; each unit has an appropriately weighted output line to each muscle. (This model and the following one emphasize spatial organization and disregard temporal sequencing and feedbacks. Their purpose is to show how the contraction of the muscles may be set for the head to reach a point on the way to the target.)

The scheme of exhaustive direct input-output connections might describe a hypothetical primitive creature with a set of sensory receptor cells connected directly to a set of motor effector cells; alternatively, it describes the net effect of multisynaptic pathways through the brain of a more complicated organism. More generally, the arguments in this paper pertain to the organization between two connected layers of a multilevel system. When parts of a similar model were ablated in computer simulations, brainlike effects of apparent mass action or redundancy were observed (Anderson, Silverstein, Ritz, & Jones, 1977; Wood, 1982).

When there are many elements in each of two sequential levels, the scheme of direct connections between all elements in each of the layers requires a large number of connections. In the orientation-localization model, if there are s discriminable skin areas and r muscles, then

a total of sr weighted connections is needed. Thus, the evolutionary and ontogenetic problem of maintaining orderly input-output relationships grows rapidly with the size of organism or its discriminative capacity.

A Three-Level Orientation-Localization Model with Associative Subsystems

Figure 2 illustrates a system that can mediate orientation-localization behavior while requiring many fewer connections. Roughly speaking, each of the three associative subsystems or modules represents one of the three dimensions of space, that is, one for proximodistal localization, one for left versus right, and one for localization along the anterior-posterior axis. Each discriminable skin area connects to each associative module with appropriate weighting; the connection from each associative module to each muscle sets the degree to which that muscle participates in a localizing motion of the head in the "dimension" represented by the module.

These associative modules are plausible biologically because they do not require computerlike algorithms for their functioning; each works simply on the basis of diffuse excitation. Thus, they may be single neurons, mass-acting groups of neurons, neurotransmitter pools, or perhaps

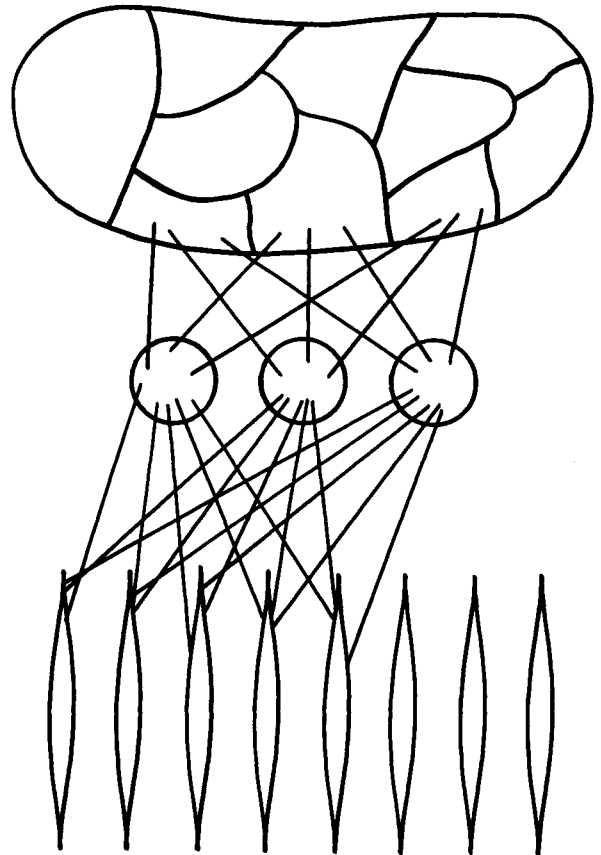


Figure 2. Input and output elements communicate via a small number of associative modules. Each discriminable skin area is connected to each associative module, and each associative module is connected to each muscle. Fewer total connections are needed than in the system shown in Figure 1. Not all connections are sketched in.

other functional units. The discriminative and motor capabilities of the system as a whole derive from the pattern of combined activation of the associative modules, and the pattern of skeletal moorings of the muscles. The total number of connections is $3s+3r = 3(s+r)$. In comparison with the model of direct input-output connections, the savings are great, particularly if s and r are large. For example, if $s = 1,000$ and $r = 100$, then the first model requires 100,000 connections, whereas the one with associative modules requires only 3,300 connections.

Although the empirical dissociation data suggested this model, it is not obvious how to relate it to neuroanatomy. The mappings of visual and tactile space to cortex and tectum do not involve a radical convergence onto simple modules which represent spatial dimensions by their dynamic properties; rather, these central neurons are arrayed in a two-dimensional sheet. Output control of orienting is distributed over the tectal brain maps in frogs (Arbib, 1982) and cats (Stein, Magalhaes-Castro, & Kruger, 1976). Nevertheless, topographic brain maps must do something besides simply transfer stimulus space to brain space, or there would hardly be any need for a brain. Indeed, in cats, orientation-localization depends on the topographic mappings in SII and subjacent orbital-anterior sylvian association cortex; a good deal of convergence occurs to these cortical areas (Glassman & Glassman, 1977).

An anatomical entity might act as a unitary summator of inputs for one function (hence, a single module) while comprising an internally differentiated set of modules for some other function. For example, the left- versus right-turning function in orientation-localization may be skimmed from neurons having another main function. The brain has fairly symmetrical hemispheres and primarily contralateral sensory input. Therefore, left versus right turning could simply follow the overall central neural asymmetry of diffusely summed cutaneous input activity, regardless of the fine spatial distribution of inputs to the topographic maps of either hemisphere. Behavioral evidence is consistent with this hypothesis (Glassman, 1983).

Generalization of the Three-Level Model

More generally, not 3 but "a" associative subsystems categorize input-output relations in various brain systems; the s skin areas and r muscles of the orientation-localization model are replaced in the general case by s input modules and r output modules. Connections are saved whenever $a(s+r)$ in a three-level system is less than sr in a two-level system. This occurs when the number of associative modules is much less than both the number of input and output modules. [Specifically, $a(s+r) < sr$ whenever $a < sr/(s+r)$.]

However, the model for orientation-localization is too simple for other functions, because it contends only with binary (on vs. off) and punctiform inputs (one input module on at a time). A more general model must handle spatially distributed patterns of numerous input modules si-

multaneously active in various degrees. Even so, the associative codes may be highly impoverished, lacking counterparts for much of the richness of inputs and outputs; they need only a sufficient number of distinct representations to differentiate probable inputs. The remainder of the structure required for adaptive behavior must be in other levels of the system, including the sensory interface and motor apparatus. (Also see Arbib, 1975, on "action-oriented perception" and Michaels & Carello, 1981, on "affordances.")

The problem of mapping distinct input patterns to distinct outputs can be understood by considering the information in associative subsystems in three ways: (1) as values on dimensions of a graph, (2) as elements in a combinatorial process, or (3) as digits in a counting process. These views will clarify what it means to consider the informational capacity of a set of modules, but first, one more concept needs to be made explicit.

Associative modules must have a *range* of possible degrees of activation; as a simplification, think of these degrees as discretely different. One possible example is the firing rate of a neuron. There is also a need to consider resolution in this example because there may be random fluctuations of firing rates, with slightly different values not eliciting distinct outputs. The maximal neuronal firing rate of about 1,000/sec sets an upper limit on what we might imagine as the range of a single-neuron module.

No matter how many possible input patterns there are, one could logically hypothesize a single associative module with a sufficient range to uniquely represent each input. However, such a drastic convergence to a large-range module might require a complicated algorithm. By contrast, a categorization using three diffusely summing modules was appropriate in the orientation-localization model.

Associative Modules Representing Dimensions—Handling Nonbinary Inputs

Figure 3 illustrates a possible organization of inputs, converging onto associative modules that are connected in the manner of any two of the three modules of Figure 2. The input weights are illustrated as ordered pairs of numbers (x,y) labeling the graphic cells in Figure 3. If s is the number of binary inputs and range is represented by g , then $(g_x)(g_y) = s$. The most efficient use of modules' ranges occurs when the ranges are equal. As informal proof, referring to Figure 3, recall that a square encompasses a given area with a shorter perimeter than any other rectangle.

One way to handle input modules with a range greater than binary "ons" versus "offs" would be to add a third associative module to register input amplitude, which could be represented as a z -axis in Figure 3. This is conceptually simplest, but there are many ways to map input elements' ranges into the ranges of any number of associative modules; in the future, these need to be considered with regard to physiological and evolutionary plausibility.

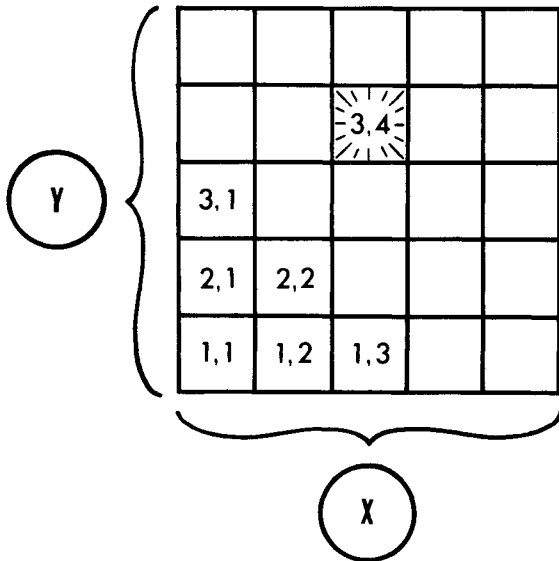


Figure 3. Each associative module can take one value at a time on a "dimension." In this example, when the x-axis module is active at a value of 3 and the y-axis module is active at a value of 4, this is represented at graphic cell (3,4).

Associative Modules Representing Combinations or Counts—Handling Both Nonbinary and Distributed Inputs

In the orientation-localization model, combined patterns of activity in several associative modules were allowed as representations, whereas inputs were restricted to one point at a time. The punctiform restriction is an extreme case; more generally, the brain's associative systems must cope with spatially distributed input patterns. In such a pattern, a combination of input modules is active, with each module at some level of its range. If the range of input module *i* is represented as *g_i*, then the total number of possible activation patterns of *n* input modules is the product of all their ranges, which can be expressed as

$$\prod_{i=1}^n g_i$$

If the range of all input modules is the same (call it *g*), this formula reduces to *gⁿ*.

With this last formula, each module also may be considered as containing a digit in a counting process. Thus, if each module were binary (*g*=2), a system of *n* modules as a whole could represent 2^{*n*} distinct patterns; this is analogous to the way an 8-bit byte in a computer can take any of 2⁸ = 256 unique states, counting from 0 to 255. If all the modules have a range of 10, then *g* = 10; so, for example, if *n* = 6, the six modules would be sufficient to represent 1,000,000 distinct patterns, or to "count" from 0 to 999,999. Of course, even a neural associative system whose modules happen to have a range of 10 is not likely to count in the decimal method used

by a whole human; this analysis merely suggests minimal requirements for the discriminative capacity of an associative system.

Input Restrictions

The immediately preceding comments imply that saving connections requires restricting an associative system's inputs. Input restrictions are familiar in sensory physiology and perceptual psychology. For instance, lateral inhibition may greatly reduce the number of possible input combinations to deeper levels. If neurons are arrayed in a thin layer, and if activity in a row of neurons inhibits an equal number of neurons in rows on either side, then the associative system to which they connect needs to contend with only one-third of its input neurons' being active at once or, at most, with half, if they happened to be active in a tightly striped pattern.

As an additional example, tendencies of Gestalt perception such as perceived closure of nearly closed curves (Hochberg, 1972) may depend upon early signal conditioning that has evolved in reducing the multiplicity with which deeper associative systems must cope. Because different analyses occur at different places in the brain, inputs may often be restricted not at the source, but someplace between the source and the various sets of associative modules to which they distribute. Empirically there are some great expansions in going from one to another brain level (e.g., retina to visual cortex); the present model suggests looking at these expanded representations as comprising numerous functional segments, each of which is a small set of associative modules. As a whole, an expansive feature map may serve as a staging ground for subsequent parsimonious convergences further downstream.

Associative systems would generally require few modules if these modules received convergences either from few or from most members of the input set. This is because the quantities represented by the combinatorial formula for *n* things (e.g., input modules) taken *r* at a time,

$${}^n C_r = \frac{n!}{(n-r)!r!}$$

are smallest for combinations that are either much smaller than half of *n* or are close to *n*. This interesting empirical implication of the model may help in understanding evidence that the various visual cortical maps have local or global specializations (Sprague, Hughes, & Berlucchi, 1981).

REFERENCES

ANDERSON, J. A., SILVERSTEIN, J. W., RITZ, S. A., & JONES, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413-451.
 ARBIB, M. A. (1975). Artificial intelligence and brain theory: Unities and diversities. *Annals of Biomedical Engineering*, *3*, 238-274.
 ARBIB, M. A. (1982). Modelling neuronal mechanisms of visuomotor

- coordination in frog and toad. In S. Amari & M. A. Arbib (Eds.), *Competition and cooperation in neural nets. Lecture notes in bio-mathematics* (Vol. 45). New York: Springer-Verlag.
- ENGELBERGER, J. F. (1980). Robotics in practice: Management and applications of industrial robots. London: Kogan Page.
- FELDMAN, J. A., & BALLARD, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, **6**, 205-254.
- FUKUSHIMA, K. (1984). A hierarchical neural network model for associative memory. *Biological Cybernetics*, **50**, 105-113.
- GLASSMAN, R. B. (1970). Cutaneous discrimination and motor control following somatosensory cortical ablations. *Physiology & Behavior*, **5**, 1009-1019.
- GLASSMAN, R. B. (1983). Dissociation of vertical and horizontal components of somesthetic orientation-localization during recovery from cortical damage: Implication regarding central associative functions. *Physiological Psychology*, **11**, 47-53.
- GLASSMAN, R. B., & GLASSMAN, H. N. (1977). Distribution of somatosensory and motor behavioral function in cat's frontal cortex. *Physiology & Behavior*, **18**, 1127-1152.
- GROSSBERG, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. *Progress in Theoretical Biology*, **5**, 233-374.
- HINTON, G. E., & ANDERSON, J. A. (EDS.) (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.
- HOCHBERG, J. (1972). Perception. I. Color and shape. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth & Schlosberg's experimental psychology* (3rd ed.): Vol. 1: *Sensation and perception*. New York: Holt, Rinehart & Winston.
- KOHONEN, T. (1978). *Associative memory: A system-theoretical approach*. New York: Springer-Verlag.
- LEVINE, D. S. (1983). Neuronal population modeling and psychology: A review. *Mathematical Biosciences*, **66**, 1-86.
- MICHAELS, C. F., & CARELLO, C. (1981). *Direct perception*. Englewood Cliffs, NJ: Prentice-Hall.
- PELLIONISZ, A., & LLINAS, R. (1982). Space-time representation in the brain. The cerebellum as a predictive space-time metric tensor. *Neuroscience*, **7**, 2949-2970.
- SPRAGUE, J. M., HUGHES, H. C., & BERLUCCHI, G. (1981). Cortical mechanisms in pattern and form perception. In O. Pompeiano & C. Ajmone-Marsan (Eds.), *Brain mechanisms and perceptual awareness*. New York: Raven Press.
- STEIN, B., MAGALHAES-CASTRO, B., & KRUGER, L. (1976). Relationship between visual and tactile representations in cat superior colliculus. *Journal of Neurophysiology*, **39**, 401-419.
- WILLSHAW, D. (1981). Holography, associative memory, and inductive generalization. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.
- WOOD, C. C. (1982). Implications of simulated lesion experiments for the interpretation of lesions in real nervous systems. In M. A. Arbib, D. Caplan, & J. C. Marshall (Eds.), *Neural models of language processes*. New York: Academic Press.

(Manuscript received December 27, 1983;
revision accepted for publication March 20, 1985.)