Selective networks capable of representative transformations, limited generalizations, and associative memory

(neural net/selection theory/feature detection/parallel circuit/memory)

GERALD M. EDELMAN AND GEORGE N. REEKE, JR.

The Rockefeller University, 1230 York Avenue, New York, New York 10021

Contributed by Gerald M. Edelman, December 7, 1981

ABSTRACT   Two parallel sets of selective networks composed of intercommunicating neuron-like elements have been connected to produce a new kind of automaton capable of limited recognition of two-dimensional patterns. Salient features of this automaton are (*i*) preestablished unchanging connectivity, (*ii*) preassigned connection strengths that are selectively altered according to experience, (*iii*) local feature detection in one network with simultaneous global feature correlation in the other, and (*iv*) reentrant interactions between the two networks to generate a new function, associative memory. No forced learning, explicit semantic rules, or *a priori* instructions are used.

The central problem of neurobiology is to describe how assemblies of neurons connected in networks can mediate a variety of behavioral responses in a fashion adaptive for the organism. The most challenging forms of this problem are posed by higher brain functions such as perception, learning, or problem solving under novel conditions—i.e., situations in which both classification and generalization play a significant role. According to the group-selection theory (1, 2), these higher brain functions are mediated by selection from degenerate repertoires of neural networks that are formed during ontogeny and do not alter in connectivity thereafter. During later experience, selection occurs by differential amplification or diminution of synaptic connection strengths within and between groups of neurons whose preestablished connectivities allow them to respond adaptively to a given input. According to the theory, reentry of outputs from one set of networks to other preceding or parallel networks makes possible spatiotemporal correlations of features from representations of the same object in different networks. Reentry permits the system to deal with both individual and invariant aspects of stimuli drawn from the same or different classes (1).

The group-selection theory makes explicit predictions (1, 2) that can be tested. However, quite independent of the falsifiability of this theory is the question of its self-consistency: Can a prewired network or congeries of networks based on selective principles and reentry respond stably and adaptively to structured inputs to yield both pattern recognition and association without prior instructions, explicit semantic rules, or forced learning? In the present paper, we answer this question affirmatively by describing in a preliminary fashion the structure and performance of a probabilistic automaton ("Darwin II") made up of selective networks. This description is illustrative and does not include a formal assessment of the performance limits of the automaton.

OVERALL SYSTEM DESIGN

For convenience in design and testing, Darwin II was simulated on a digital computer (an IBM 3033). The automaton consists

of an input array on which two-dimensional patterns can be represented, an assembly of interconnected networks or repertoires of recognizing elements that transform input patterns, and an arrangement for coupling these networks to as yet unspecified motor-output functions. Patterns represented on the input array are centrally fixed and scaled to a standard size; mechanisms for detection of motion and for development of translational and scale invariance are essential in real visual systems but have not yet been provided in the automaton. Each recognizing element, called a "group" (Fig. 1), formally represents a connected assembly of neuron-like units, although the connections of such units within a group are not specified. Groups have multiple inputs that may come variously from the input array or from the outputs of groups in the same or different repertoires. The state of each group is characterized by a single time-dependent scalar variable, *s*, that is determined from the inputs and past history of that group according to a nonlinear response function.

$$s_i(t) = \sum_j c_{ij}(s_{l_{ij}} - \theta_E) - \sum_k \beta(s_k - \theta_I) + N + \omega s_i(t - 1), \quad [1]$$

where $s_i(t)$ = state of the *i*th group at time *t*; c_{ij} = connection strength of *j*th input to group *i* ($c_{ij} > 0$, excitatory; $c_{ij} < 0$, inhibitory); $s_{l_{ij}}$ = state of the group specified by l_{ij} (i.e., of the group connected to the *j*th input of group *i*); θ_E = excitatory input threshold (only inputs with $s_{l_{ij}} \geq \theta_E$ are included); β = a fixed inhibition coefficient; s_k = state of group defined by *k*, which ranges over all groups within a specified inhibitory neighborhood around group *i*; θ_I = inhibitory input threshold (only inputs with $s_k \geq \theta_I$ are included); N = noise drawn from a normal distribution with chosen mean and SD; and ω = persistence parameter ($\omega = e^{-1/\tau}$, where τ is a characteristic time constant). The first and second term of the response function are ignored unless their sum exceeds a positive firing threshold (θ_P) or is less than a negative inhibitory threshold (θ_N). (Provision is also made for groups to have a refractory period following suprathreshold excitation.) The number of groups in each repertoire and the number of connections to each group can be varied at will; a maximum of $\approx 10^6$ connections, distributed in any way among the various repertoires, is permitted by the available computer memory.

The amplification function, which is designed to alter the "synaptic strength", c_{ij}, of a connection according to the activity of the pre- and post-synaptic groups, is

$$c_{ij}(t + 1) = c_{ij}(t) + \delta \cdot \phi(c_{ij}) \cdot (s_i - \theta_{M_I}) \cdot (s_j - \theta_{M_J}), \quad [2]$$

where δ = amplification factor ($0 \leq \delta < 1$); $\phi(c)$ = saturation factor to prevent $|c_{ij}|$ from becoming larger than 1 [$\phi(c) = 1 + 2c^2 - c^4$ if $c \cdot (s_i - \theta_{M_I}) \cdot (s_j - \theta_{M_J}) > 0$; $\phi(c) = 1$ if $c \cdot (s_i -$
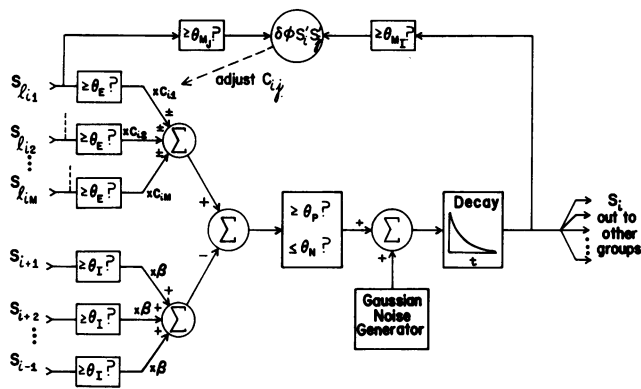
FIG. 1. Structure and function of a group. Inputs, shown at the left, are specified by a connectivity matrix that can range over any groups in the system (upper left) or by a geometric rule for lateral inhibition by neighboring groups (lower left). These inputs are combined with spontaneous activity or "noise" and with persisting activity from earlier times according to Eq. 1.

$\theta_{M_I} \cdot (s_j - \theta_{M_J}) \leq 0$; and θ_{M_I}, θ_{M_J} = amplification thresholds for postsynaptic groups, i, and for presynaptic groups, j]. A total of $3^4 = 81$ amplification rules can be constructed by taking δ to be positive, negative, or zero according to whether s_i and s_j are greater than or less than the thresholds θ_{M_I} and θ_{M_J}, respectively. (In terms of real neuronal function, only a few of these would be sensible. In most of the examples in this paper, δ was zero if $s_i \leq \theta_{M_I}$ and $s_j \leq \theta_{M_J}$; otherwise δ was positive.)

Darwin II is modeled on the group-selection theory (1). Accordingly, the connectivity and connection strengths of each group, and hence the range of input patterns to which it can respond, are established when the group is constructed, prior to the presentation of any inputs. The connectivity remains unchanged throughout a particular instantiation of the model and is not used to embody predetermined information for recognizing *particular* stimulus objects; the connection strengths change continually in accord with an amplification function (Eq. 2) that ensures group selection. When a group responds to a particular stimulus, that response is differentially strengthened so that the same group is more likely to respond to future presentations of the same or similar stimuli. More than one group can respond at some level to a given input and certain groups may respond above threshold to two or more different inputs. This overlapping degeneracy (1) of groups in a repertoire ensures that there will be at least some response to *any* stimulus.

Repertoire Construction and Connectivity. In its current form, Darwin II has two parallel sets of networks, named "Darwin" and "Wallace" for convenience, that comprise six repertoires in all (Fig. 2). Darwin (Fig. 2 *Left*) transforms each pattern presented on the input array into a distinct and stable selected output representation. It consists of **R**, a local feature-recognizing network that extracts salient aspects of the stimulus object, of **R-of-R** ("recognizer of recognizers"), a randomly organized abstractor, and of **R_rep**, an output or reporter network designed to sample the activity of **R-of-R**. Wallace (Fig. 2 *Right*) is concerned with limited generalization and produces relatively invariant output patterns for all members of a given class of input objects. It consists of (*i*) a computer-simulated visual prosthesis that accomplishes global feature correlation by contour tracing (*ii*) two similarly constructed recognizing networks, **R_M** (features) and **R_M** (relations), onto which the output of the prosthesis is mapped; and (*iii*) an output network **R_Mrep** similar in principle to **R_rep**. Locally reentrant connections are provided within each repertoire except **R**. More globally reentrant connections are provided in both directions between the **R-of-R**
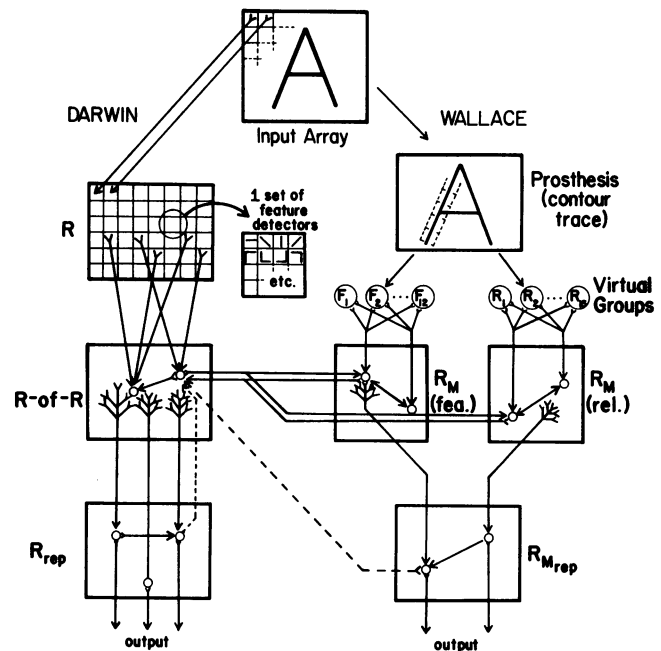

FIG. 2. Overall system design. Stimulus objects are presented on an input array. The **R** repertoire consists of a series of local feature detectors connected topographically to visual areas in the input array. **R-of-R** groups are connected randomly to multiple groups in **R** and may also receive input from any or all of the other repertoires. **R_rep** groups summarize the activity of collections of neighboring **R-of-R** groups. The visual prosthesis traces objects in the input array to obtain nonlocal characteristics of lines and the junctions between them. Virtual groups are excited according to these characteristics, which correspond, for example, to the numbers, lengths, curvatures, and orientations of lines and, for junctions, to the number of lines, their relative lengths, and whether or not they terminate at that junction. **R_M (features)** [**R_M (fea.)**] and **R_M (relations)** [**R_M (rel.)**] groups are connected to the virtual groups as described in the text. These repertoires serve to correlate physically separated object features to give a global representation that is relatively insensitive to local alterations in the stimulus patterns. **R_Mrep** groups summarize the activity of **R_M** groups. (Connections shown as dashed lines are usually omitted.)

network on the one hand and **R_rep**, **R_M** (features), **R_M** (relations), and **R_Mrep** on the other (Fig. 2).

The connectivity of a group is determined by a list of the groups connected to its inputs. Connections are different for each group but are constructed according to a common rule within each repertoire. Different repertoires follow different connection rules, as follows:

R Repertoire. The construction of the **R** repertoire is entirely systematic. Each **R** group is designed, by assignment of appropriate connection strengths, to respond to a particular local pattern feature, such as a horizontal line segment. The inputs to the group (usually 16) are connected to a compact set of grid points on the input array designated a "visual area." Each visual area has one group for each feature-detecting pattern used (up to 40) and the visual areas are mapped topographically onto the **R** sheet.

Prior to presentation with objects and amplification, all repertoires other than **R** have random connection strengths drawn from a normal distribution with specified mean and SD. Both the numbers of connections and the proportions of excitatory and inhibitory connections of each type can be set at will.

R-of-R Repertoire. **R-of-R** groups may receive input connections from groups in any repertoire or set of repertoires, including **R-of-R** itself. The origins of these connections may be distributed over an entire source repertoire or over some spec-

Neurobiology: Edelman and Reeke

Proc. Natl. Acad. Sci. USA 79 (1982) 2093

ified local subrepertoire. In either case, the exact groups to be connected to a given **R-of-R** group are selected at random. The connections from **R** constitute the basic patterned input to **R-of-R**; the other connections are used to form associations as described below.

R_{rep} *Repertoire.* Each reporter group is connected to groups within a localized area on the **R-of-R** sheet. The centers of these areas are selected randomly and uniformly on **R-of-R**. Individual connections to the reporter groups are then distributed around these center points. This allows convergence onto a relatively small number of output groups while maintaining the distinctness of each **R-of-R** representation.

Visual Prosthesis and R_M Networks. R_M groups are connected to a visual prosthesis that traces over the input array, extracting certain features (e.g., lines and points) and relationships between them (e.g., junctions). The trace program uses an exploratory algorithm to search along contours of a pattern in a series of trapezoids. Various information about features and relationships (the exact items to be included can be varied) is encoded as a list of numbers. This information is used to specify the states of a set of "virtual groups" that are mapped to a particular subrepertoire of R_M. Virtual groups do not have the ordinary input connections or response functions but do have an output, s, that is computed according to the trace results. The trace routine is a prosthesis in that its functions *could* be carried out by a network of standard groups. However, the design complexity of such a network and the irrelevance of the details to our present aims dictated the prosthetic approach.*

Each R_M group is connected to all virtual groups of the appropriate kind: R_M **(features)** groups are connected to virtual groups encoding feature characteristics and R_M **(relations)** groups are connected to virtual groups encoding relationship characteristics. The decay time of the response of an R_M group is long compared with the trace time. Therefore, as the virtual groups pass through various states in the course of a trace, each R_M group responds essentially to the sum of these states for all features traced, ensuring independence of the order in which a trace happens to be made.

R_{Mrep} *Repertoire.* R_M reporter groups are arranged in the same fashion as **R-of-R** reporter groups, but the R_{Mrep} repertoire is divided into two subrepertoires—one with input from R_M **(features)** and one with input from R_M **(relations)**. In addition, R_M and R_{Mrep} groups may receive reciprocal back connections from **R-of-R**. The need for reentry has been discussed elsewhere (1, 2) as a fundamental functional feature of selective networks; here, the purpose of these connections is to provide a basis for an associative memory in Darwin based on the interaction between Wallace and Darwin[†].

Event Cycle. Darwin II operates phasically. In each unit time interval, new states are calculated for all groups in all repertoires in turn, and connection strengths are then modified in accordance with the amplification rule. A number of such cycles may be carried out before a new stimulus is presented. Performance statistics are gathered during the last cycle for each stimulus and are reported when all stimuli have been processed. A few tens of cycles are usually sufficient for the connection strengths to reach convergence.

RESULTS

The performance of Darwin II is evaluated by a number of statistical criteria. Typically, network parameters are tuned so that
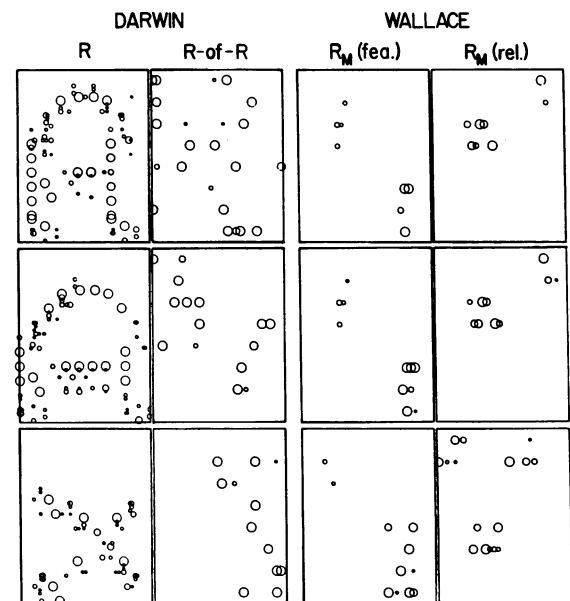

FIG. 3. Typical response patterns of different repertoires to various stimuli. For illustration, the groups of each repertoire are placed on a square lattice and a circle is drawn at the position of each group responding with $0.5 \leq s \leq 1.0$. The diameter of the circle is proportional to s. No circles are drawn for groups with $0 \leq s < 0.5$. Reentrant connections between Darwin and Wallace are not active. Stimuli presented were an A (*Top*), a different, broader, A with a lower crossbar (*Middle*), and an X (*Bottom*).

1–5% of the groups in each repertoire respond above the θ_P threshold for a typical stimulus. The progress of amplification can then be judged by examining such statistics as the number of groups responding above θ_P ("hits"), the maximum excitation received by any group in the repertoire, and the sharpness of the distribution of s. In addition, the statistical routine (but not the automaton itself) is informed of the classes to which stimulus objects have been assigned by the outside observer. It is then able to report counts of the numbers of groups that respond to pairs of stimuli in the same or different classes. These numbers are compared with those that would be obtained by chance assuming a uniform distribution of hits over a whole repertoire. We present here only a few examples of the broad range of responses exhibited by the system. They do not test the general performance of Darwin II nor do they give a secure indication of the formal limitations of such networks.[‡]

Typical response patterns obtained under conditions in which the Darwin and Wallace networks were decoupled from each other are illustrated in Fig. 3. As expected, the **R** responses resemble the stimulus patterns because of the topographical mapping from the input array to **R**. The **R-of-R** responses bear no geometrical relationship to the stimuli and are nearly as different for two different A stimuli as for an A and an X. R_M responses, on the other hand, are very similar for the two A stimuli, but different for the X. The mapping of particular lines and junctions to different subrepertoires of R_M can be seen by the clustering of the response circles. Table 1 contains statistical data for several such runs. The ratios of the numbers of groups responding to pairs of stimuli in the same class to the numbers of groups responding to pairs of stimuli in different classes are much larger for R_M than for **R** or **R-of-R**. These ratios reflect

---

Table 1. Behavior of Darwin and Wallace networks in parallel but in the absence of reciprocal reentry

	Darwin		Wallace	
	R	R-of-R	R_M (features)	R_M (relations)
Before amplification				
Fraction of groups above θ_P threshold	0.02	0.08	0.01	0.02
Disjunction statistic*	0.70	0.71	0.64	0.85
Intraclass cross-response ratio[†]	3.68	1.45	43.86	43.14
Interclass cross-response ratio[‡]	1.88	1.20	0.48	9.04
Intraclass ratio/interclass ratio	1.96	1.21	90.93	4.77
After amplification[§]				
Fraction of groups above θ_P threshold	0.02	0.07	0.01	0.02
Disjunction statistic*	0.71	0.76	0.79	0.96
Intraclass cross-response ratio[†]	3.91	1.06	53.09	51.64
Interclass cross-response ratio[‡]	1.91	0.75	0.22	12.22
Intraclass ratio/interclass ratio	2.05	1.41	241.3	4.23

Numbers of groups used were **R**, 3840; other repertoires, 4096 each; total connections, 368,640. Connections in both directions between **R-of-R** and **R_M** were inactive. Stimuli included four examples of each of four letters (A, E, Y, Ξ). In eight separate runs, either four letters of the same class (four runs) or four letters, one from each class (four runs), were presented. Averages over appropriate sets of these runs are tabulated. Mean responses ($\bar{s}$) of groups in different repertoires ranged from 0.05 to 0.10.

* Defined as the mean response ($\bar{s}$) of groups for which $s \geq \theta_P$ minus the mean response of groups for which $s < \theta_P$; the statistic increases as the number of groups having responses near $s = 0$ or $s = 1$ increases.

† Number of groups responding to both members of a pair of stimuli of the same class, summed over the six possible pairings of the four stimuli/number of such groups expected if the observed number of hits were distributed uniformly over the repertoire.

‡ Same as above except for pairs of stimuli belonging to different classes.

§ Amplification consisted of three series of presentations of the four stimuli in turn, with eight basic cycles for each stimulus (96 cycles in all). $\delta = 0.20$, $\theta_{M_I} = 0.5$, $\theta_{M_J} = 0.5$. R groups were not modified.

the limited but definite ability of **R_M** to make classifications imitating those made by the outside observer. [Here **R_M** (relations) has more groups responding to stimuli in different classes, and hence a smaller intraclass/interclass ratio, than **R_M** (features) because the stimuli presented had some types of junctions in common.] Amplification improves these classification ratios. It also changes the distribution of s, yielding more groups with very high or very low responses, while producing only small changes in the mean response and numbers of groups excited above the hit level.

Table 2 outlines one kind of protocol under which associative recall can be demonstrated. Connections in both directions between **R_M** and **R-of-R** are initially established with low mean synaptic strength. During early experience ("training"), those connections that happen to link groups in the two repertoires responding to the same stimulus are selectively strengthened (or "validated") by the standard amplification procedure. After training, the visual prosthesis is inactivated and the validated connections from **R-of-R** to **R_M** are used to activate an image in **R_M** of the trace that was there during training (when the prosthesis was on). The reciprocal connections from **R_M** back to **R-of-R** then provide a pathway by which these traces in **R_M** can activate patterns of response in **R-of-R** corresponding to other stimuli in the same class, patterns that were active during training at the same time as the original **R_M** trace. Depending on the time constants chosen for decay and refractory periods and on the amount of lateral inhibition used, these associated responses in **R-of-R** can occur simultaneously or in sequence. Association is demonstrated in the example shown by the fact

that, after training, 38 rather than 8 groups responded in test to both of two stimuli presented successively under conditions in which **R-of-R** responses were not dominated by input from **R**. The 30 additional groups that responded to both stimuli when each was presented in successive tests were among the groups that had responded to only one of the two stimuli during training.

DISCUSSION

The construction of Darwin II was prompted by three related aims: (*i*) to test the self-consistency of the notion (1, 2) of group selection in nonlinear degenerate repertoires made up of neuron-like groups that have preassigned connectivity and connection strengths; (*ii*) to design networks capable of making unique representations of individual objects by means of local feature detection, as well as networks for making relatively invariant representations of objects within a class by means of global feature correlation; and (*iii*) to arrange reentrant interactions of the two networks carrying out these separate operations in parallel so that an associative memory linking individual representations within a class is developed. The preliminary design and performance described here suggest that, at least in part, these goals have been achieved.

While there are certain aspects of Darwin II that resemble perceptrons (4) or self-organizing neural networks (5), our model differs from them in several respects that enable it to circumvent certain limitations of such systems (6). Darwin II is selective rather than instructive and embodies parallel networks for local and global feature detection; these produce a

Neurobiology: Edelman and Reeke

Proc. Natl. Acad. Sci. USA 79 (1982)   2095

Table 2. Associative recall reflected in **R-of-R** repertoire responses

Condition	No. of groups of 256 in repertoire that responded above θ_P		No. of those that responded above θ_P to both stimuli
	To stimulus 1	To stimulus 2	
Related stimuli*			
Training	32	27	8
Test	38	39	38
Unrelated stimuli†			
Training	17	34	4
Test	13	31	4

All repertoires had 256 groups except **R**, which had 3840. Amplification parameters were $\delta = 0.20$, $\theta_{M_I} = 0.4$, $\theta_{M_J} = 0.35$. Initially, reciprocal connections between **R-of-R** and **R$_M$** were present with a low mean c_{ij} (0.0333) and a high θ_E (0.6). Training consisted of four alternating presentations (eight cycles per presentation) of each of two stimuli. Reciprocal **R-of-R** to **R$_M$** connections (see Fig. 2) were allowed to be amplified after two of these presentations. After training, θ_E for these reciprocal connections was decreased to 0.2 and the visual prosthesis input to **R$_M$** was inactivated. One of the stimuli in each pair was then presented during test for four cycles to the Darwin network alone. All input from the input array was then removed and the automaton was allowed to run for eight more cycles.
* An X and a +.
† An A and an X.

patterned correlation of responding groups rather than a single binary output for each stimulus. It also contains reentrant pathways that enable it to deal with a wider range of individual and class characteristics.

Neither the representative transformations nor the limited generalizations performed by Darwin II require the direct intervention of a classifying observer, either through instruction or forced learning. These capabilities arise instead from the selective principle embodied in the network by means of feature detection, reentry, and differential amplification. Selection permits the system to make certain discriminations based on internally generated criteria. If the system is allowed to evolve according to its experience, these criteria become correlated with the relevant physical properties of the outside world. The resulting behavioral primitives provide the necessary substrates for more sophisticated behavior, including association, learning, and problem solving. These will require modifications to the machine not described here.

The first of these, association, has, however, in part been realized by activating the reentrant "commissural" connections between Darwin and Wallace. This reciprocal mapping between the two major subnetworks enables associations to be made by amplifying connections that link representations of individual objects of a class already present in Darwin. The association depends on the parallel functioning of the two subnetworks and on the validation of **R$_M$** to **R-of-R** connections against patterns of activity stimulated by input from **R** and sus-

tained by reentry within **R-of-R**. Wallace thus allows associative connections to be established in Darwin that could not arise spontaneously. In these associations, representations of individual identity and class membership are both largely preserved.

There are some limitations inherent in the design of Darwin II. Although unstable behavior has been observed in other complex nonlinear systems (7), we have found it possible to adjust the system parameters to avoid explosive responses and oscillatory behavior. Other limitations reflect the deliberate restriction of semantic input, the decision not to include forced learning, and the omission of defined survival rules based on output behavior. Until these functions are explicitly introduced, Darwin II cannot, for example, develop associations based on convention (such as upper case A with lower case a) and cannot show evidence of problem solving. For these reasons, present performance levels only give an indication of possible performance and are not comparable with those of pattern-recognizing machines based on principles of artificial intelligence. Our goal was to achieve a minimal level of performance in selective networks; further experiments and formal analysis will be required to determine maximal performance.

Despite present limitations, it is significant that two parallel concatenations of selective networks, each carrying out different functions on the same input patterns, can generate a new function not achievable by either alone. Other such dyadic systems having composite functions can be envisioned, as can higher order interactions among multiple networks.

It should be emphasized that Darwin II is not an explicit model of either the whole or part of any nervous sytem. Nevertheless, it was designed heuristically with nervous systems in mind and it would not be surprising if it reflected some aspects of their performance. In any case, it serves to demonstrate the capabilities of group selection among repertoires of neuron-like groups and even its present limited performance establishes the self-consistency of the ideas of group selection and reentry in multineuron networks. Demonstration of their reality obviously will depend on direct experimentation in real nervous systems.

1. Edelman, G. M. & Mountcastle, V. B. (1978) *The Mindful Brain* (MIT Press, Cambridge, MA), pp. 51–100.
2. Edelman, G. M. (1981) in *The Organization of the Cerebral Cortex*, eds. Schmitt, F. O., Worden, F. G., Adelman, G. & Dennis, S.-G. (MIT Press, Cambridge, MA), pp. 535–563.
3. Van Ryzin, J. (1977) *Classification and Clustering* (Academic, New York).
4. Rosenblatt, F. (1958) *The Perceptron. A Theory of Statistical Separability in Cognitive Systems* (Cornell Aeronautical Laboratory, Ithaca, NY), Report no. VG-1196-G-1.
5. Von der Malsburg, C. (1973) *Kybernetik* **14**, 85–100.
6. Minsky, M. & Papert, S. (1969) *Perceptrons. An Introduction to Computational Geometry* (MIT Press, Cambridge, MA).
7. Ashby, W. R., von Foerster, H. & Walker, C. C. (1962) *Nature (London)* **196**, 561–562.