(1991) *J. Neurosci.* 11, 3227–3238
7 Hartenstein, V. and Jan, Y. N. (1992) *Roux's Arch. Dev. Biol.* 201, 194–220
8 Nose, A., Mahajan, V. B. and Goodman, C. S. (1992) *Cell* 70, 553–567
9 Bourgouin, C., Lundgren, S. E. and Thomas, J. B. (1992) *Neuron* 9, 549–561
10 Bate, M. (1990) *Development* 110, 791–804
11 Bate, M., Rushton, E. and Currie, D. A. (1991) *Development* 113, 79–89
12 Jan, L. Y. and Jan, Y. N. (1976) *J. Physiol.* 262, 189–214
13 Jan, L. Y. and Jan, Y. N. (1976) *J. Physiol.* 262, 215–236
14 Anderson, M. S., Halpern, M. E. and Keshishian, H. (1988) *J. Neurosci.* 8, 242–255
15 Cantera, R. and Nässel, D. R. (1992) *Cell. Tiss. Res.* 269, 459–471
16 Budnik, V. and Gorczyca, M. (1992) *Soc. Neurosci. Abstr.* 18, 39
17 Budnik, V., Zhong, Y. and Wu, C-F. (1990) *J. Neurosci.* 10, 3754–3768
18 Zhong, Y., Budnik, V. and Wu, C-F. (1992) *J. Neurosci.* 12, 644–651
19 Sink, H. and Whitington, P. M. (1991) *J. Neurobiol.* 22, 298–311
20 Cash, S., Chiba, A. and Keshishian, H. (1992) *J. Neurosci.* 12, 2051–2064
21 Patel, N. H., Snow, P. M. and Goodman, C. S. (1987) *Cell* 48, 975–988
22 Sink, H. and Whitington, P. M. (1991) *Development* 112, 307–316
23 Jacobs, J. R. and Goodman, C. S. (1989) *J. Neurosci.* 9, 412–424
24 Broadie, K. S. and Bate, M. (1993) *J. Neurosci.* 13, 144–166
25 Broadie, K. S. and Bate, M. (1993) *J. Neurosci.* 13, 167–180
26 Broadie, K. S. and Bate, M. (1993) *Nature* 361, 350–353
27 Schuster, C. M. *et al.* (1992) *Science* 254, 112–114
28 Chiba, A., Hing, H., Cash, S. and Keshishian, H. (1993) *J. Neurosci.* 13, 714–732
29 Chang, T. N. and Keshishian, H. (1992) *Soc. Neurosci. Abstr.* 18, 580
30 Bier, E., Jan, L. Y. and Jan, Y. N. (1990) *Genes Dev.* 4, 190–203
31 Uemura, T., Shepherd, S., Ackerman, L., Jan, L. Y. and Jan, Y. N. (1989) *Cell* 58, 349–360
32 Sink, H. and Whitington, P. M. (1991) *Development* 113, 701–707
33 Tosney, K. W. and Landmesser, L. T. (1984) *J. Neurosci.* 4, 2518–2527
34 Tosney, K. W. and Landmesser, L. T. (1985) *J. Neurosci.* 5, 2336–2344
35 Whitelaw, V. and Hollyday, M. (1983) *J. Neurosci.* 3, 1199–1215
36 Whitelaw, V. and Hollyday, M. (1983) *J. Neurosci.* 3, 1216–1225
37 Patel, N. H., Ball, E. E. and Goodman, C. S. (1992) *Nature* 357, 339–342
38 Goodman, C. S. and Doe, C. in *The Development of* Drosophila melanogaster (Bate, M. and Martinez-Arias, A., eds), Cold Spring Harbor Laboratory Press (in press)
39 Snow, P. M., Bieber, A. J. and Goodman, C. S. (1989) *Cell* 59, 313–323
40 Grenningloh, G., Rehm, E. J. and Goodman, C. S. (1991) *Cell* 67, 45–57
41 Bieber, A. J. *et al.* (1989) *Cell* 59, 447–460
42 Chiba, A., Chang, T. N. and Keshishian, H. (1992) *Soc. Neurosci. Abstr.* 18, 1273
43 Landmesser, L., Dahm, L., Tang, J. and Rutishauser, U. (1990) *Neuron* 4, 655–667
44 Tomaselli, K. J., Neugebauer, K. M., Bixby, J. L., Lilien, J. and Reichardt, L. F. (1988) *Neuron* 1, 33–43
45 Bixby, J. L., Pratt, R. S., Lilien, J. and Reichardt, L. F. (1987) *Proc. Natl Acad. Sci. USA* 84, 2555–2559
46 Elkins, T., Zinn, K., McAllister, L., Hoffman, F. M. and Goodman, C. S. (1990) *Cell* 60, 565–575
47 Lin, D., Van Vactor, D., Tiemeyer, D. and Goodman, C. S. (1992) *Soc. Neurosci. Abstr.* 18, 36
48 Thomas, J. B. and Wyman, R. J. (1982) *Nature* 298, 650–651
49 Van Vactor, D. and Goodman, C. S. (1992) *Soc. Neurosci. Abstr.* 18, 1273
50 Eisen, J. S. and Pike, S. H. (1992) in *The Nerve Growth Cone* (Kater, S. B. and Macagno, E. R., eds), pp. 323–335, Raven
51 Lance-Jones, C. and Landmesser, L. (1981) *Proc. R. Soc. London Ser. B* 214, 19–52
52 Roth, J., Kempf, A., Reuter, G., Schauer, R. and Gehring, W. J. (1992) *Science* 256, 673–675

The computational leech

Shawn R. Lockery and Terrence J. Sejnowski

The local bending reflex of the leech computes a well-defined sensorimotor input–output function in which each of several unique patterns of sensory input elicits a unique pattern of motoneuron activity. Interneurons in the reflex respond to most input patterns and contribute to most motor patterns, suggesting a distributed processing mechanism for the reflex. This suggestion is supported by models in which connection strengths are adjusted by a neural network optimization algorithm to reproduce the local bending input–output function. In addition, computational parallels between the local bending network and the perceptron, a major class of artificial neural networks, brings the functional role of local bending interneurons into question and suggests new physiological experiments.

Shawn R. Lockery is at the Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA, and Terrence J. Sejnowski is at the Computational Neurobiology Laboratory, The Salk Institute, La Jolla, CA 92037, USA.

A decade of research on parallel distributed processing (PDP) networks has demonstrated the impressive capacity of artificial networks to perform complex tasks. Using simplified neuron-like processing elements and a variety of computer algorithms for connecting them, networks have been created that do everything from reading zipcodes to predicting the stock market. In the study of biological neural systems, the PDP perspective provides theoretical tools for the analysis of computational strategies and a range of powerful computer algorithms, such as backpropagation, for constructing working models of large, highly interconnected networks. Although these algorithms started as models of learning, there is now wide agreement that they are poor representations of the mechanisms of learning and memory. Rather, they have proven to be useful as biologically neutral ways to fit complex models to physiological data.

To illustrate the utility of the PDP modeling approach, we describe our recent theoretical efforts to understand a simple, well-defined network of repeatably identifiable interneurons in the leech. A realistic model of the system was constructed by forcing backpropagation to operate within stringent physiological and anatomical constraints. The model has elucidated the functional role of interneurons in the network and the connectivity of interneurons still to be identified. In addition, the theoretical framework provided by neural network theory focuses future experimental work in new directions.

Networks of idealized neurons

Perhaps the most widely studied artificial PDP network is the two-layered perceptron[1]. Figure 1A shows a simple perceptron with two input lines (X1

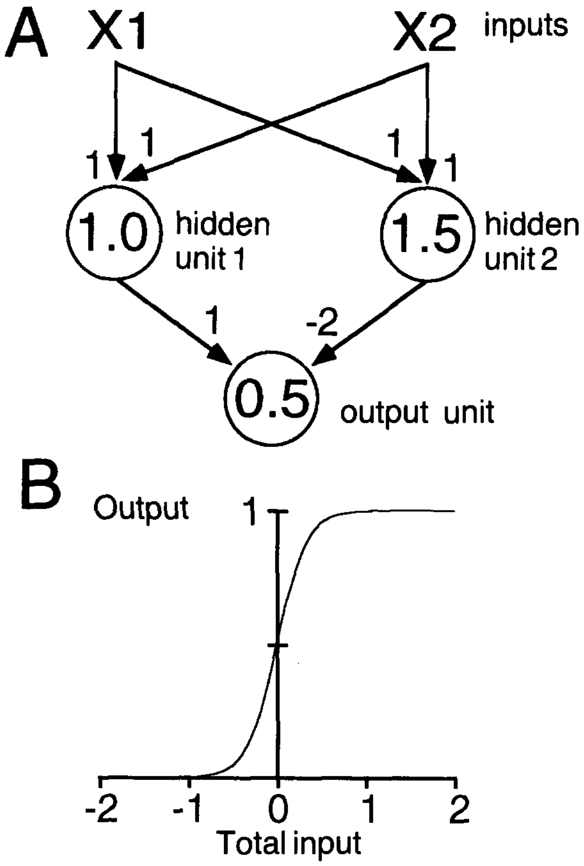

Fig. 1. (A) *A two-layered perceptron with two hidden units and one output unit. The number inside each unit is the threshold, the level of synaptic input at which the output equals 0.5. Net synaptic input is the sum of the outputs of each presynaptic unit weighted by the connection strengths indicated next to the arrowheads.* **(B)** *The sigmoidal function relating total synaptic input to the output of a unit.*

and X2), two interneuron-like hidden units and one output unit. Each connection in the network is assigned a numerical strength called a weight and the output of each unit (Fig. 1B) is a sigmoidal function of the net synaptic input, namely, the weighted sum of the outputs of the presynaptic units. The sigmoid curve idealizes the relation between synaptic input and firing frequency in real neurons. The level of net synaptic input at which the output of the unit equals 0.5 is called the threshold of the unit because it determines where the transition from inactive (off) to active (on) occurs. Processing in a PDP network is parallel in the sense that all the units transform input into output simultaneously; processing is distributed in the sense that information from each input is spread out among many hidden units.

A common task artificial networks perform is to classify particular combinations of features (e.g. pitch, color, shape, etc.). In the simple example of Fig. 1A, activity along input lines X1 and X2 could be used to represent the degree to which two independent features are present. The region enclosed by the X1 and X2 axes defines the input space of the network. The receptive field of the output unit is visualized by plotting its activity as a function of all possible combinations of input unit activity (Fig. 2). By virtue of its threshold, each hidden unit cuts the input space into an 'on' and 'off' region. A variety of output unit receptive fields can be constructed by judiciously aligning the cuts made by each hidden unit. To illustrate, the weights in Fig. 1A were chosen to produce the diagonally oriented receptive field in Fig. 2A.

PDP networks owe their computational power to both the distributed nature of connections from one layer to the next, and to the nonlinearity of the units themselves. Selective alterations of the network in Fig. 1A provide a convenient demonstration of this point. For example, if one 'de-distributes' the input by deleting the crossed connections in Fig. 1A, one limits the possible cuts to those parallel to the X1 and X2 axes, seriously limiting the range of receptive field shapes (Fig. 2B). If instead one replaces the sigmoidal input–output function with a linear relationship, no cuts are possible, and the receptive field becomes a plane (Fig. 2C). The output unit now responds to the entire input space, albeit with excitation in some regions and inhibition in others. Deleting the crossed connections in Fig. 2C produces a planar receptive field with a different slope. In fact, it can be shown that such a planar receptive field can be produced by an even simpler 'network': a single linear output unit together with its inputs[2].

The computational power of PDP networks can be increased by additional circuit elements. For example, adding units to the hidden-unit layer in Fig. 1A would further subdivide the input space, and thereby sharpen the receptive field of the output unit. Indeed, by adding whole layers of hidden units, any receptive field is possible[3], though the number of hidden units needed may become impractically large. Another problem that must be overcome in multi-layered networks is how to choose the connection strengths between units. Indeed, research into PDP networks was hindered by this problem for many years[4], but interest in them was re-ignited by the introduction of a variety of network optimization or 'training' algorithms that could adjust the weights (and other network parameters) to achieve almost any input–output function[5]. These include: 'supervised' procedures in which the weight adjustments are calculated by taking the difference between actual and desired output for each unit[6]; 'reinforcement' procedures that involve only a good–bad assessment of network output[7]; and 'unsupervised' procedures in which weights are adjusted according to local learning rules that depend on things like the correlation between presynaptic and postsynaptic activity[8–10].

Using these algorithms, it has been shown that expanded versions of the simple network in Fig. 1A can perform sophisticated information-processing tasks that have captured the imagination of the specialist and layman alike. These include many biologically relevant tasks such as pattern recognition[11,12], data compression[13], visually guided locomotion[14], interpolation[15], signal detection[16], prediction of future events[17] and speech synthesis[18]. Often, the network performs as well as or better than humans or specially designed computer programs[11,16,19]. Thus, a variety of successful applications show the perceptron to be a surprisingly powerful and quite general computational strategy.

Back to biology

Biological PDP networks, to which artificial networks owe their inspiration, are both common[20] and notoriously difficult to study[21]. This has led to the hope that artificial PDP networks may one day return the favor, so to speak, by providing insights and new tools for the investigation of distributed processing in biological networks. Two immediate possibilities come to mind. First, the idealized representation of neurons embodied in artificial networks fosters theoretical analyses[10,22] that can point experimentalists to new experiments that reveal essential computational details. Second, network optimization algorithms can be applied to realistic models of biological networks[23,24]. This exciting development may provide a way out of the double-bind that characterizes the study of biological PDP networks: models are essential tools in the design and interpretation of experiments, yet the size and complexity of most distributed systems far exceeds the physiological data available to construct the model in the first place. There is an urgent need, therefore, for what we might loosely call a 'physiologist's assistant': an automated procedure that provides educated guesses as to the value of unknown model parameters in anticipation of our making the necessary physiological measurements. By sharpening our experimental insights, these working models could greatly accelerate collection of relevant physiological data, hastening the day when working models are supplanted by models fully constrained by physiological measurements[25].

Local bending behavior

We have been exploring this possibility using the local bending reflex of the leech as a test case in the analysis of biological PDP networks (Fig. 3A). In response to a moderate mechanical stimulus, the leech withdraws from the point of contact by contracting longitudinal muscles beneath the stimulus and relaxing those on the opposite side of the body, resulting in a U-shaped local bend[26]. For example, a dorsal stimulus causes dorsal longitudinal muscles to contract and ventral ones to relax. Analogous patterns of contraction and relaxation underlie ventral and lateral bends.

The local bending network

The circuitry underlying the local bending reflex invites comparison to a two-layered perceptron (Fig. 3B). The nervous system of the leech consists of a chain of segmental ganglia, each of which contains circuitry sufficient to produce the behavior. This has made possible a detailed description of the local bending circuit[26–31]. Major input to the reflex is provided by dorsal and ventral pressure-sensitive mechanoreceptors, the P cells. Contraction and relaxation of longitudinal muscles is controlled by eight types of motoneurons, an excitatory (DE or VE) type and an inhibitory (DI or VI) type for the dorsal and ventral quadrants, respectively, on the left and right side of each body segment. Motoneurons are connected by numerous chemical and electrical synapses introducing feedback within the motor layer. Input from sensory neurons to motoneurons is mediated by a layer of interneurons. Synaptic transmission from interneurons to motoneurons, and among motoneurons, is a graded function of presynaptic voltage (Fig. 3C)[32,33]. Pairwise intracellular recordings showed that the synaptic transfer function is approximately linear at low presynaptic voltage but saturates (flattens out) at higher voltages.

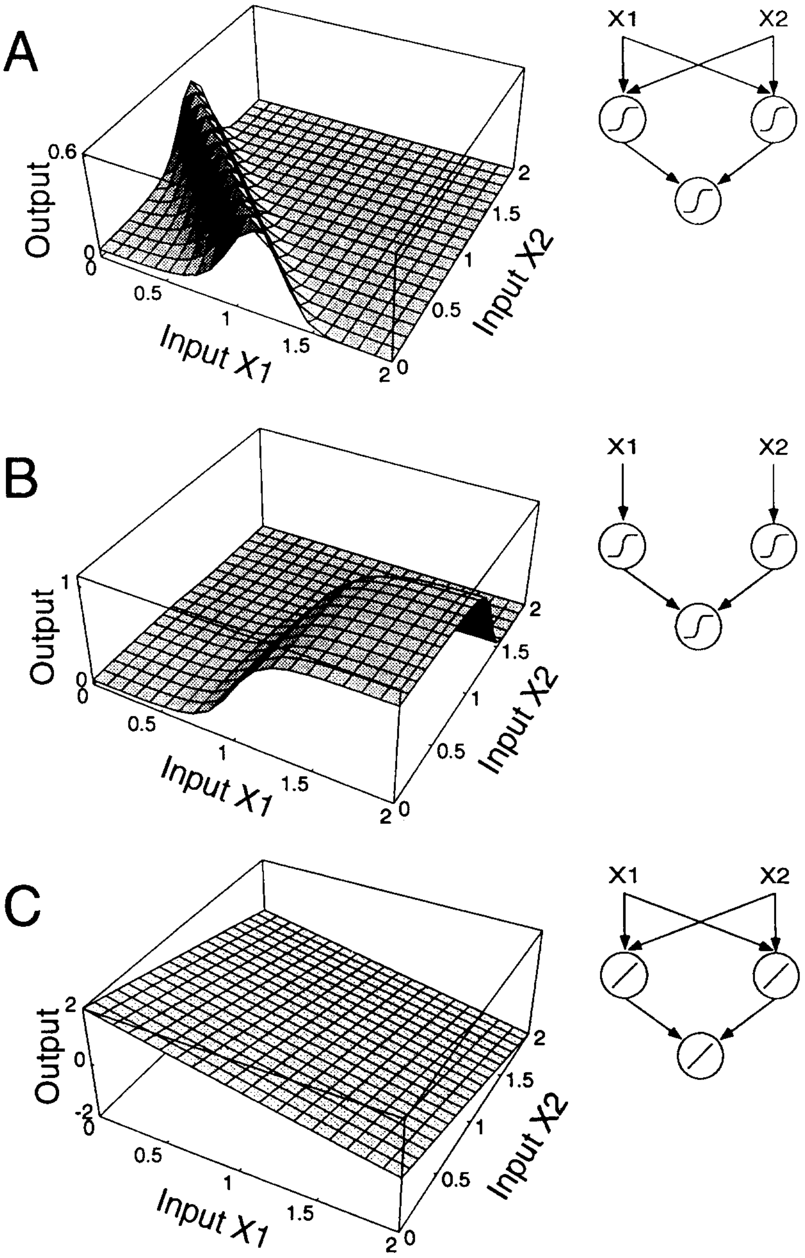

Fig. 2. **(A)** *The receptive field of the output unit in Fig. 1A. The input space of the network is the region between the X1 and X2 axes. Activity of the output unit is plotted on the vertical axis. Each hidden unit divides the input space into an 'on' and 'off' region separated by diagonal lines in the X1–X2 plane. The output unit subtracts the activity of hidden unit 2 from hidden unit 1, resulting in a diagonally oriented receptive field.* **(B)** *The receptive field of the output unit after deleting the crossed connections in* (A). *The lines separating the on and off regions must now be parallel to the X1 or X2 axis because each hidden unit is sensitive only to one input.* **(C)** *The receptive field of the output unit when the sigmoidal function is replaced by a linear input–output relationship. The receptive field becomes a plane with excitation near the origin and inhibition elsewhere. Linearization of the network precludes spatially restricted receptive fields. Moreover, the hidden units now serve no function, because the same receptive field can be produced by a single unit receiving inhibitory connections with a strength of −1 from each input and a resting activation level of 2.5.*

Local bending and neural computation

The primary function of interneurons in the local bending network is to associate with each sensory stimulus that pattern of motoneuron excitation and inhibition sufficient to withdraw from the site of contact. In computational terms, the reflex computes a function that maps a four-dimensional input vector,

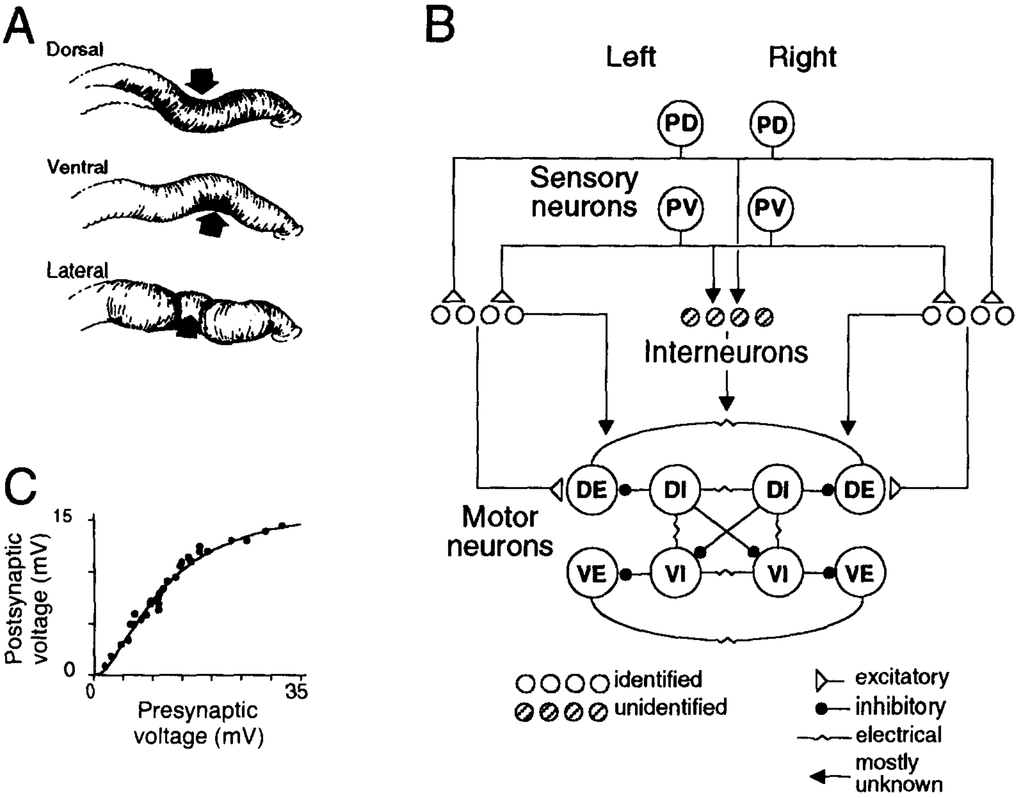

Fig. 3. *The local bending reflex.* **(A)** *Behavior: dorsal, ventral and lateral stimuli produce local U-shaped bends.* **(B)** *Simplified neural circuit: the main input to the reflex is provided by the dorsal and ventral P cells (*PD *and* PV*). Control of local bending movements is largely provided by motoneurons whose projective fields are restricted to one quadrant (left or right, dorsal or ventral) of the body. Dorsal and ventral quadrants are innervated by both excitatory (*DE *and* VE*) and inhibitory (*DI *and* VI*) motoneurons. There are two to four representatives of each motoneuron type in a midbody leech ganglion. Motoneurons are connected by numerous chemical and electrical synapses. A subpopulation of interneurons has been identified (small hatched circles) that receive excitatory input from dorsal P cells and excite the dorsal excitatory motoneurons. Other interneurons, including those that excite the ventral excitors, remain to be identified (small open circles).* **(C)** *The synaptic transfer function for the DI to DE and VI to VE connections. Data points are from pairwise intracellular recordings of steady-state postsynaptic voltage following prolonged current injection in the presynaptic neuron[33]. The smooth curve is an empirical fit to the data points[34]. This function is believed to be typical of the transfer function at other synapses in the local bending network.*

encoding stimulus location (the activity of the four P cells), onto an eight-dimensional output vector, encoding the associated movement (the activity of the eight types of motoneurons). One of the strengths of the local bending reflex as an experimental system is that the input–output function can be measured precisely by recording the motoneuron synaptic potentials produced by stimulation of P cells in various combinations[30]. To date, we know the patterns of motoneuron excitation and inhibition produced by eight different combinations of single or paired P cell stimulation (e.g. Fig. 4). Each of the patterns is consistent with the withdrawal behavior observed in response to mechanical stimuli that would activate the same P cells. A major focus of our physiological studies has been to determine how this input–output function is computed.

Identification of local bending interneurons

In perhaps the simplest conceptual model of how the computation is performed, local bends are produced by dedicated interneurons specific for dorsal, ventral or lateral forms of the response (Fig. 5A). To determine how the interneurons in the reflex actually compute the local bending input–output function, a subpopulation of local bending interneurons contributing to dorsal bending was identified using physiological and morphological criteria[31]. Nine types of dorsal bending interneurons, which have excitatory connections to the dorsal excitatory motoneurons and receive excitatory connections from the dorsal P cells, have been identified (Fig. 5B). All but one of these types occur as left–right bilateral homologues. This means there are at least 17 dorsal bending interneurons per ganglion. Consistent with the number of interneurons in each ganglion, removing single interneurons from the circuit produced small but detectable decrements in motoneuron responses during behavior. Other types of local bending interneurons, e.g. those that inhibit the dorsal excitors, remain to be identified.

Two aspects of the connections made by the subpopulation of identified interneurons are inconsistent with a commitment to only dorsal bending, and thus with the simplest model. First, all but one type of dorsal bending interneuron receives substantial excitatory input from one or more ventral P cells, indicating that those neurons previously considered to be dorsal bending interneurons are also active in the ventral and lateral bending behaviors. Second, the effect of an interneuron on an inhibitory motoneuron is not always opposite in sign to its effect on the excitatory motoneuron controlling the same body quadrant (Fig. 5B, arrows). Thus, the connections of the local bending interneurons suggest a distributed processing strategy in which each interneuron is active in some or all forms of local bending and has output connections that are not completely consistent with any single form of the response.

A dynamic neural network model of the local bending reflex

Modeling the reflex was prompted by the need to show that a network of interneurons with distributed sensory inputs could produce the physiological responses seen in the motoneurons during the reflex. The possibility remained that other interneurons, perhaps like the dedicated neurons of Fig. 5A, are the ones actually responsible for the reflex and that these had been missed in the original search.

The basic model we studied had four sensory neurons, eight motoneurons and 40 interneurons, and thus 480 connections, representing the actual local bending circuit (Fig. 3B)[34]. The number of interneurons was based on an upper estimate of the number of local bending interneurons that remain to be identified in the biological network. Each neuron in the model was represented as an isopotential cellular element called an electrical compartment[35] with a physiologically determined input resistance and time constant. The strengths of connections between motoneurons were adjusted by trial and error to reproduce the results of pairwise motoneuron recordings[33]. However, because of the larger number of connections between sensory neurons and interneurons, and between interneurons and motoneurons, it was not practical to adjust these by trial and error. We therefore used backpropagation[6] as an optimization procedure to select a set of weights that could reproduce the input–output function of the reflex. Our model differs from typical backpropagation networks, however, because the neurons are dynamic (the response of each neuron evolves in time), due to the resistance and capacitance of the cellular compartments. This makes the model more realistic, but necessitates the use of a variation of the backpropagation algorithm for dynamic neurons[36].

At the start of optimization, the weights to be optimized are randomly assigned small initial values. Each input–output association is then presented and the total discrepancy between the network's output and the desired output, called the error, is calculated. Weights that increase the error are decreased and those that decrease the error are increased. This procedure is repeated many (typically 1000–100 000) times until the total error is small. Because it must calculate the effect on the error of each weight, backpropagation is not intended as a model for learning or development. It is used here as a curve-fitting technique, where the input–output relationships of the network are the data points to be fit and the weights are the adjustable coefficients or parameters. In the local bending model, weights were optimized to reproduce the amplitude and timecourse of synaptic potentials recorded in the motoneurons in response to each pattern of sensory input (Fig. 4). Thus, after optimization, the model reproduced almost exactly the input–output behavior seen in physiological experiments.

To make the model more realistic, the optimization algorithm was forced to operate under five additional physiological constraints. (1) Only excitatory connections were allowed from sensory neurons to interneurons in the model, because only excitatory connections have so far been found at this layer in the biological network[31]. (2) The sigmoidal function commonly used in artificial networks (Fig. 1B) was replaced by the physiologically determined relationship between presynaptic voltage and postsynaptic response (Fig. 3C). (3) Each interneuron on the left of the ganglion was paired with one on the right to maintain homologous input and output connections, reflecting the bilateral symmetry of the leech nervous system. (4) No connections between interneurons were allowed because none have so far been found. (5) The model included all of the known chemical and electrical connections between the motoneurons. Thus, the model differs from artificial networks designed by computer scientists and engineers who are under no obligation to remain true to any particular biological network.

After optimization, the input and output connections of interneurons in the model network qualitatively resemble the connections of identified local bending interneurons (Figs 5B,C). In particular, all interneurons receive inputs from ventral as well as dorsal P cells, most have connections to all motoneurons, and the connections to the inhibitory motoneurons are not always opposite in sign to those onto the excitatory motoneurons controlling the same body quadrant (Fig. 5C, arrows).

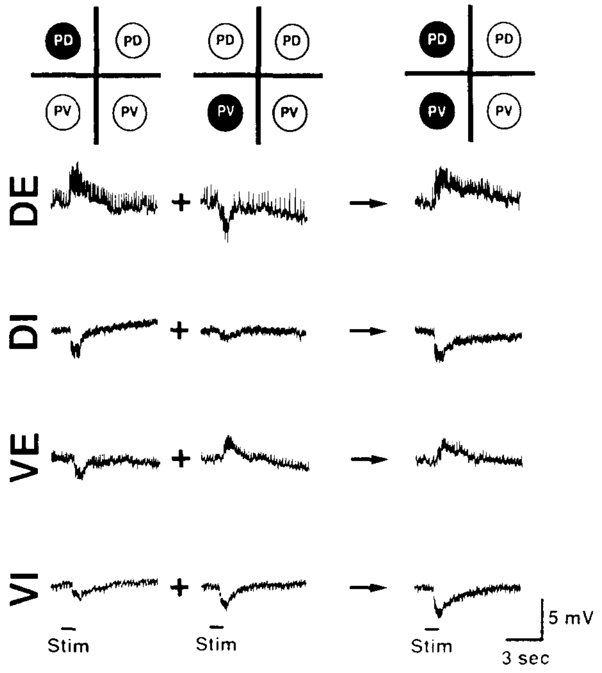

Fig. 4. *The local bending input–output function. Intracellular recordings from four motoneurons in response to stimulation of one or two P cells (filled circles). The motoneurons have projective fields ipsilateral to the stimulated P cell(s). Co-stimulation of ipsilateral dorsal and ventral P cells (right column) activates excitatory and inhibits inhibitory motoneurons on the stimulated side. This pattern is consistent with withdrawal from a lateral stimulus. Similar recordings were obtained with other patterns of P cell stimulation and from contralateral motoneurons.* (Reproduced, with permission, from Ref. 30.)

The model led to several new insights that could not have been reached without optimization. First, the similarity between model interneurons and interneurons in the biological network shows that additional interneurons with receptive or projective fields (defined by the postsynaptic motoneurons) that differ radically from those of the subpopulation of identified interneurons are not required. Second, in hundreds of optimization runs from different randomly chosen initial connection strengths, a different final network was reached each time. Thus, there are many different networks, with different sets of connections, that produce a physiologically accurate local bending input–output function. The multiplicity of networks raises the intriguing possibility that, as a result of differences in genetics or experience, different leeches may achieve the same reflex behavior using very different local bending networks. At present, however, this possibility has not been tested experimentally.

Variations on the local bending model

Varying the constraints placed on the algorithm provided additional insights[34]. For instance, we

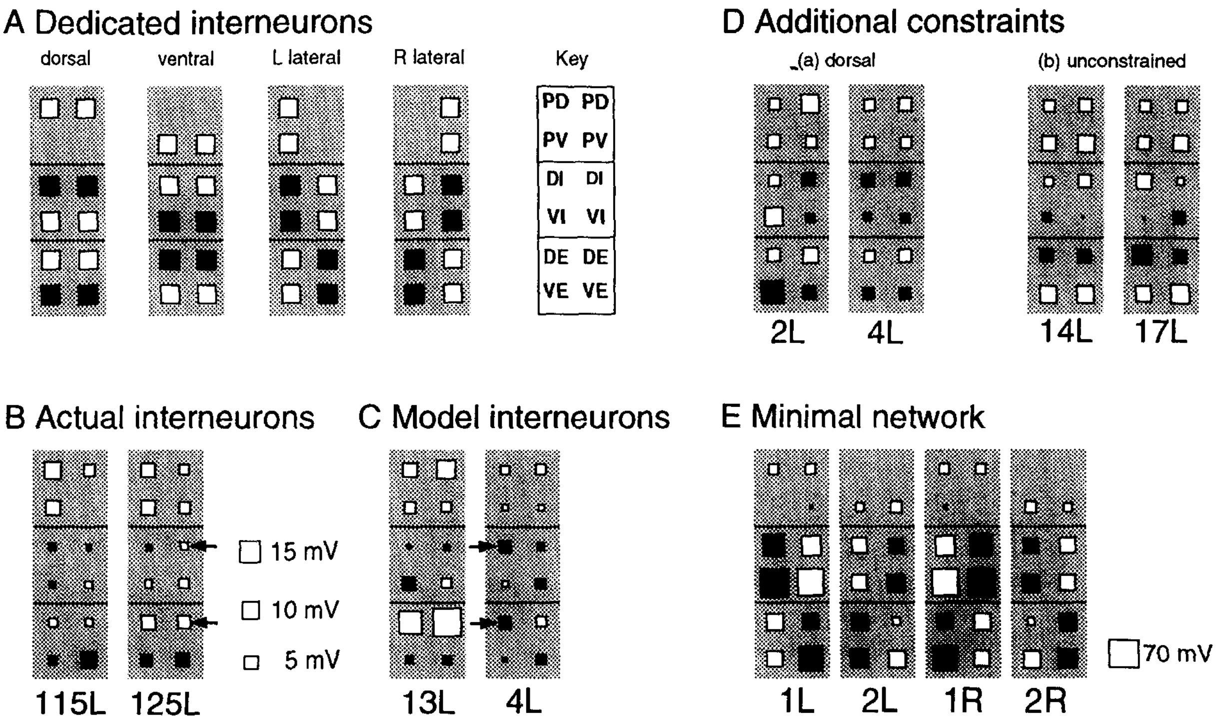

Fig. 5. *Comparison of input and output connections of model and actual local bending interneurons. In each gray box, the upper panel shows input connections from sensory neurons, the middle panel shows output connections to inhibitory motoneurons, and the lower panel shows output connections to excitatory motoneurons (see key). Open boxes are excitatory connections and filled boxes are inhibitory connections. Box area is proportional to synaptic strength (see scale). Panels* (B)–(D) *are shown at the same scale. Each interneuron is one of a bilaterally symmetrical pair.* **(A)** *A hypothetical network of dedicated interneurons in which each interneuron has inputs and outputs specific for dorsal, ventral or lateral bends.* **(B)** *Connections of identified local-bending interneurons* 115L *and* 125L. *Output connections to excitatory motoneurons are as predicted for a dedicated dorsal bending interneuron [see* (A)*], but input connections and conflicting effects on excitatory and inhibitory motoneurons of the same body quadrant (arrows) suggest a distributed processing strategy for the reflex.* **(C)** *Connections of two interneurons in a model network after optimization by backpropagation. Model interneurons, like the real ones [see* (B)*] received dorsal and ventral P cell inputs, had output effects on all motoneurons, and exhibited conflicting effects on excitatory and inhibitory motoneurons of the same body quadrant (arrows).* **(D)** *Connections of interneurons in a model network with two subpopulations of interneurons. The subpopulation* (a) *was constrained to have output effects on excitatory motoneurons that were the same sign as those of the identified neurons [see* (B)*]. No such constraint was placed on* (b)*, the unconstrained subpopulation. The model demonstrated that all three forms of local bending can be achieved by a model having just two basic types of interneurons.* **(E)** *Connections of interneurons in a minimal model network with just two pairs of interneurons. These interneurons have input connections specific for dorsal or ventral inputs, but outputs consistent with lateral bending. Optimization thus revealed a previously unanticipated solution involving dedicated interneurons with a dissociation between input and output specificities.* (Data are reproduced, with permission, from Figs 5, 12 and 14 of Ref. 34.)

divided the 40 interneurons into two populations, (a) and (b). Population (a) represented the identified interneurons in being constrained to excite the dorsal excitatory motoneurons and inhibit the ventral excitatory motoneurons. Population (b) was under no additional constraints. After optimization, the connectivity of population (b) represents the possible connections of interneurons yet to be identified in the biological network (Fig. 5D). Most of the interneurons had outputs consistent with a major contribution to ventral bending, suggesting that all three forms of local bending could be produced by just two basic types of interneuron: one specialized for dorsal bending, the other for ventral bending. This notion was supported by reducing the number of unconstrained interneurons and re-optimizing the network. All the constrained interneurons then had outputs consistent with ventral bending. In a different modeling experiment, we reduced the number of interneurons in the network and found that recognizable local bending motor patterns can be produced by networks with as few as four interneurons (Fig. 5E). Interneurons in these networks resembled the dedicated interneurons in Fig. 5A, except they were specific for either dorsal or ventral inputs, yet had outputs consistent with lateral bending. Before the modeling, we had not anticipated such a possibility.

The local bending network as a biological perceptron

The essential computational aspects of the biological local bending network are captured in a subnetwork containing just two sensory inputs (e.g. ipsilateral PV and PD neurons) and a single motoneuron (DE) (Fig. 6A). This is because of the bilateral symmetry of both the local bending input–output function and the interneurons that compute it. Such a sub-network is like the simple two-layered perceptron in Fig. 1A, except that there are more interneurons. Thus we can compare the performance of the subnetwork and the perceptron by plotting the motoneuron's receptive field, so to speak, i.e. the synaptic potentials actually recorded from motoneuron DE in response to either single or simultaneous activation of the two P cells (Fig. 6B). This demonstrates that the output of the biological network is nearly planar, like that of the linearized perceptron of Fig. 1D. In theory then, the interneurons in the local bending network could be replaced by direct connections from sensory neurons to motoneurons without loss of function. This is because, as noted above, a one-layer network is sufficient to produce a planar receptive field (Fig. 2C). Similar results are obtained for the other motoneurons in the biological network.

That the output of the network appears to be linear suggests the interneurons in the local bending network are not there to divide the input space into localized receptive fields as we saw for the perceptron of Fig. 2A. A wide range of alternative functional roles are conceivable. These include increasing the gain from sensory neurons to motoneurons[37], serving as control points for dynamical gain modulation[37,38], increasing the stability of the output in the event of noise in (or loss of) individual interneurons or synapses, and effecting a compromise between local bending and other behaviors[39]. One might also suggest, albeit with some disappointment, that the local bending interneurons are merely vestiges of evolutionary tinkering[40].

It is worth noting, however, that the reflex could turn out to be nonlinear, and interneurons essential, after all. This is because we have not yet recorded the response of motoneurons to stronger P cell stimuli, which might be capable of driving the interneurons into the nonlinear (saturating) region of the synaptic transfer function (Fig. 3C). In light of the previous discussion of simple perceptrons, it is interesting to speculate on what the functional consequences of this could be. One possibility is that the saturating nonlinearity could partition the receptive field of individual motoneurons into restricted on and off regions as is the case for the output unit in Fig. 1B. For example, if the effects of excitatory and inhibitory interneurons saturated at different presynaptic voltages (Fig. 6C), the non-zero output of the motor neuron (labeled 'sum') would be confined to a small region of the input space near the origin. Complex receptive fields could be constructed out of the simple building blocks provided by this mechanism. For example, the receptive field in Fig. 6D was produced by linking the dorsal P cell to two different sets of interneurons. The net effect on the motoneuron of the first set is as shown in Fig. 6C. Thus, the profile of the receptive field along the PD axis is equivalent to the line marked 'sum' in Fig. 6C. The net effect on the motoneuron of the second set is opposite in sign. This can be seen by examining the profile along the PV axis. When all combinations of PV and PD activity levels are examined, the result is a receptive field with intersecting excitatory and inhibitory troughs. Note that this receptive field has an approximately planar region (dark shaded zone) near the origin that is consistent with physiological data gathered to date (Fig. 6B).

Concluding remarks

The close parallels between simple perceptrons and many biological networks make it natural to model and analyse these systems using theoretical contributions from artificial neural networks. In our study of the local bending reflex, optimization algorithms enabled us to construct a working model in advance of a complete set of physiological measurements. This demonstrated that our conceptual model of the reflex could be made to work in practice. Moreover, analysis of the reflex in light of neural network theory has raised interesting new experimental questions regarding the role of interneurons in the reflex when stronger P cell stimuli are considered.

Although backpropagation is a powerful and efficient means of adjusting connections in model networks, we do not believe the leech uses anything like backpropagation to set the weights in the biological network. Connection weights in the biological network are probably set by some combination of genetic and epigenetic factors, although at this point one can only speculate. We use backpropagation, therefore, not as a model for development or learning, but as a curve-fitting technique, where the input–output relationships of the network are the data points to be

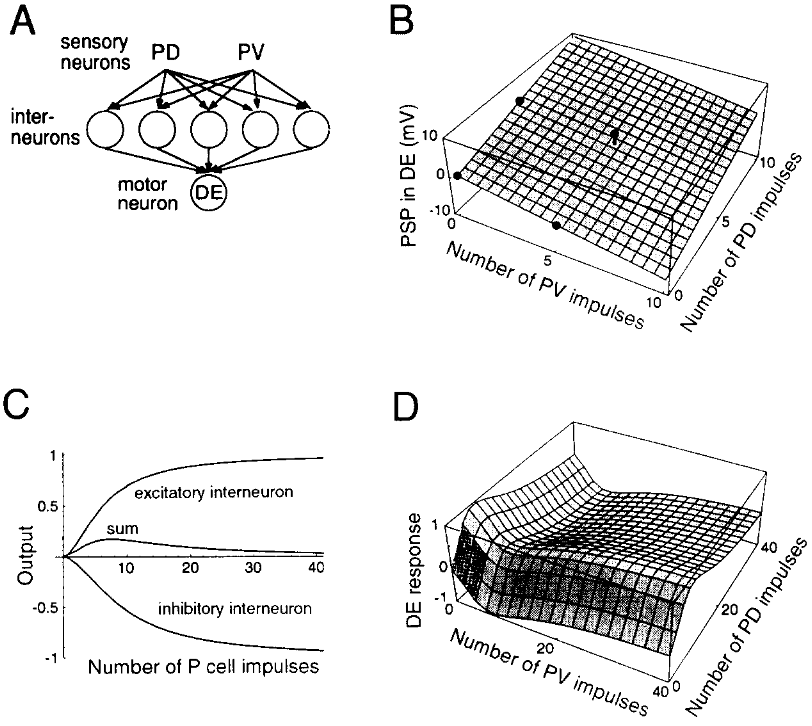

Fig. 6. The local bending network as a perceptron. **(A)** *The local bending network reduced to ipsilateral dorsal and ventral P cells and the dorsal excitor motoneuron* (DE). *Each of the nine or more interneurons is excited by both P cells and has either an excitatory or inhibitory effect on the motoneuron.* **(B)** *The actual receptive field of the dorsal excitor in the input space defined by the activation level of the two P cells (number of impulses per stimulus). The points indicate average peak synaptic potentials of the motoneuron in experiments like the one illustrated in Fig. 4, in which P cells are stimulated (10 Hz for 0.5 s) alone or in pairs. The receptive field is nearly planar, with the point for paired stimulation* (PV + PD) *lying just above the plane. Thus, in the range of low-intensity stimuli studied to date, the network behaves as a linearized perceptron (Fig. 2C).* **(C)** *Hypothetical effect on a motoneuron of excitatory and inhibitory interneurons that saturate at different presynaptic voltages. In future experiments, the P cells will be stimulated at higher frequencies expected to drive interneurons into the saturating region of their synaptic transfer function (Fig. 3C). In the example illustrated, an excitatory motoneuron saturates at a lower voltage than an inhibitory neuron, resulting in a net excitation of the motoneuron at low P cell stimulus intensities and no response at high intensities.* **(D)** *Hypothetical receptive field of a motoneuron in response to paired P cell stimulation. The first P cell,* PD, *is assumed to affect the motoneuron as shown in* (C). *A second P cell,* PV, *is assumed to have the opposite effect. These two effects sum linearly to produce a complex receptive field in which an excitatory ridge intersects an inhibitory trough.*

fit and the weights are the adjustable coefficients or parameters. This is justified because we are primarily concerned with how the fully mature system operates, not how it is built. The same strategy has elucidated the functional role of interneurons in the shortening reflex of the leech[41] and provided a variety of working models of larger systems that are less tractable physiologically. These include sensorimotor integration in the vestibulo–ocular reflex[42,43], spatial localization of visual stimuli[23,44], shape from shading[24] and cortical motor control[45]. Optimization in these models, as in the local bending model, provides novel interpretations of existing physiological data and the impetus for new experiments.

Optimization is thus emerging as a general tool for a wide range of problems in systems neuroscience. For example, we have used optimization to study possible sites of synaptic plasticity underlying nonassociative learning in the local bending reflex[46,47]. Optimization can also be applied to the networks that do not have perceptron-like circuitry, including networks with recurrent or feedback connections like invertebrate central pattern generators[48–50] and possible cortical circuits for short-term memory[51]. The range of applications can be greatly increased by using neural network units to represent quantities other than firing rate or voltage. For example, one can optimize models in which units represent subcellular processes such as Hodgkin–Huxley variables[52], the biochemical mechanisms of the genetic control of development[53] or even environmental quantities like force and torque in models of animal orientation[50,54]. In the future, we can expect to see a variety of models thought impossible until recently, and for each model, many new challenges for experimentalists.

Acknowledgements

We thank William Kristan, Jr for discussion and comments on the manuscript. Support was provided by The Bank of America–Giannini Foundation for Medical Research, an NIH Postdoctoral Fellowship, The Howard Hughes Medical Institute and the Mathers Foundation.

Selected references

1 Rosenblatt, F. (1962) *Principles of Neurodynamica*, Spartan Books
2 Rumelhart, D. E. and McClelland, J. L. (1986) *Parallel Distributed Processing*, MIT Press
3 Minsky, M. L. and Papert, S. A. (1988) *Perceptrons*, MIT Press
4 Cowan, J. D. and Sharp, D. H. (1988) *Q. Rev. Biophys.* 21, 365–427
5 Hinton, G. E. (1989) *Artif. Intell.* 40, 185–234
6 Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) *Nature* 323, 533–536
7 Sutton, R. S. and Barto, A. G. (1981) *Psychol. Rev.* 88, 135–170
8 von der Malsburg, C. (1973) *Kybernetik* 15, 85–100
9 Kohonen, T. (1984) *Self-Organization and Associative Memory*, Springer-Verlag
10 Linsker, R. (1986) *Proc. Natl Acad. Sci. USA* 83, 7508–7512
11 Le Cun, Y. *et al.* (1990) *Advances in Neural Information Processing Systems 2* (Touretzky, D. S., ed.), pp. 461–473, Morgan Kaufmann
12 Qian, N. and Sejnowski, T. J. (1988) *J. Mol. Biol.* 202, 865–884
13 Cottrell, G. W., Munro, P. and Zipser, D. (1987) in *Ninth Annual Conference of the Cognitive Science Society*, pp. 462–473, Erlbaum
14 Pomerleau, D. A. (1989) *Neural Information Processing Systems I (Denver 1989)* (Touretzky, D. S., ed.), pp. 305–313, Morgan Kaufmann
15 Ahmad, S. and Tesauro, G. (1989) *Advances in Neural Information Processing Systems I* (Touretzky, D. S., ed.), pp. 160–168, Morgan Kaufmann
16 Gorman, R. P. and Sejnowski, T. J. (1988) *Neur. Networks* 1, 75–89
17 Lapedes, A. and Farber, R. (1988) *Neural Information Processing Systems (Denver 1987)* (Anderson, D. Z., ed.), pp. 442–450, American Institute of Physics
18 Sejnowski, T. J. and Rosenberg, C. R. (1987) *Complex Syst.* 1, 145–168
19 Tesauro, G. (1990) *Neur. Comp.* 1, 321–323
20 Goldman-Rakic, P. S. (1988) *Annu. Rev. Neurosci.* 11, 137–156
21 Larimer, J. L. and Pease, C. M. (1988) *J. Exp. Zool.* 247, 45–55
22 Miller, K. D., Keller, J. B. and Stryker, M. P. (1989) *Science* 245, 605–615
23 Zipser, D. and Andersen, R. A. (1988) *Nature* 331, 679–684
24 Lehky, S. R. and Sejnowski, T. J. (1988) *Nature* 333, 452–454
25 Sejnowski, T. J., Koch, C. and Churchland, P. S. (1988) *Science* 241, 1299–1306
26 Kristan, W. B., Jr (1982) *J. Exp. Biol.* 96, 161–180
27 Ort, C. A., Kristan, W. B., Jr and Stent, G. S. (1974) *J. Comp. Physiol. A* 94, 121–154
28 Stuart, A. E. (1970) *J. Physiol.* 209, 627–646
29 Nicholls, J. G. and Baylor, D. A. (1968) *J. Neurophysiol.* 31, 740–756
30 Lockery, S. R. and Kristan, W. B., Jr (1990) *J. Neurosci.* 10, 1811–1815
31 Lockery, S. R. and Kristan, W. B., Jr (1990) *J. Neurosci.* 10, 1816–1829
32 Angstadt, J. D. and Calabrese, R. L. (1991) *J. Neurosci.* 11, 746–759
33 Granzow, B., Friesen, W. O. and Kristan, W. B., Jr (1985) *J. Neurosci.* 5, 2035–2050
34 Lockery, S. R. and Sejnowski, T. J. (1992) *J. Neurosci.* 12, 3877–3895
35 Segev, I., Fleshman, J. W. and Burke, R. E. (1989) *Methods in Neuronal Modeling* (Koch, C. and Segev, I., eds), pp. 63–96, MIT Press
36 Pearlmutter, B. A. (1989) *Neur. Comp.* 1, 263–269
37 Burrows, M. (1992) *Trends Neurosci.* 15, 226–232
38 Nagayama, T. and Hisada, M. (1987) *J. Comp. Neurol.* 257, 347–358
39 Kristan, W. B., Jr, Lockery, S. R., Wittenberg, G. and Cottrell, G. (1989) *The Computing Neuron* (Durbin, R., Miall, C. and Mitchison, G., eds), pp. 180–204, Addison-Wesley
40 Jacob, F. (1977) *Science* 196, 1161–1166
41 Wittenberg, G. and Kristan, W. B., Jr (1992) *J. Neurophys.* 68, 1693–1707
42 Anastasio, T. J. and Robinson, D. A. (1990) *Biol. Cybern.* 63, 161–167
43 Arnold, D. B. and Robinson, D. A. (1991) *Biol. Cybern.* 64, 447–454
44 Mazzoni, P., Andersen, R. A. and Jordan, M. I. (1991) *Cereb. Cortex* 1, 293–307
45 Fetz, E. E. *The Neurobiology of Neural Networks* (Gardner, D., ed.), MIT Press (in press)
46 Lockery, S. R. and Kristan, W. B., Jr (1991) *J. Comp. Physiol. A* 168, 165–177
47 Lockery, S. R. and Sejnowski, T. J. *Neurobiol. Behav.* (in press)
48 Doya, K. and Yoshizawa, S. (1989) *Neur. Networks* 2, 375–385
49 Rowat, P. F. and Selverston, A. I. (1991) *Network* 2, 17–42
50 Beer, R. D. and Gallagher, J. C. (1992) *Adapt. Behav.* 1, 91–122
51 Zipser, D. (1991) *Neur. Comp.* 3, 179–193
52 Doya, K. and Selverston, A. I. *Neur. Comp.* (in press)
53 Mjolsness, E., Sharp, D. H. and Reinitz, J. (1991) *J. Theor. Biol.* 152, 429–453
54 Lockery, S. R. and Sejnowski, T. J. (1991) *Soc. Neurosci. Abstr.* 18, 1490