



# How the brain keeps the eyes still

H. S. SEUNG

Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974

Communicated by John J. Hopfield, California Institute of Technology, Pasadena, CA, June 10, 1996 (received for review January 7, 1996)

**ABSTRACT** The brain can hold the eyes still because it stores a memory of eye position. The brain's memory of horizontal eye position appears to be represented by persistent neural activity in a network known as the neural integrator, which is localized in the brainstem and cerebellum. Existing experimental data are reinterpreted as evidence for an "attractor hypothesis" that the persistent patterns of activity observed in this network form an attractive line of fixed points in its state space. Line attractor dynamics can be produced in linear or nonlinear neural networks by learning mechanisms that precisely tune positive feedback.

The brain moves the eyes with quick saccadic movements. Between saccades, it holds the eyes still, a capability that depends on a "memory" of angular position of the eyes. A person who has lost the ability to store this memory can hold his or her eyes still at only a single null position. After saccades to other positions, the eyes drift back toward the null position, in a disorder known as gaze-evoked nystagmus (1). Localization studies suggest that the memory of eye position is stored in a neural network that extends over several areas in the brainstem and cerebellum (2–19).

When the eyes are still, the pattern of neural activity in the memory network is constant in time. For every position of the eyes, the pattern of activity is different, and can persist much longer than the intrinsic persistence time of a single neuron's activity (20–22). Therefore, the long persistence of network activity appears to be a collective effect that depends on the interactions between the neurons of the memory network.

Based on existing experimental data, it will be argued that the memory of eye position is stored in a neural network with an approximate line attractor dynamics. If synaptic strengths and other parameters are precisely tuned by learning mechanisms, a linear network (23–26) can exactly realize a line attractor dynamics, and a nonlinear network (27, 28) can achieve a good approximation.

## The Memory of Eye Position

The following discussion is restricted to horizontal eye position, and neglects the other two rotational degrees of freedom of the eye (29). Horizontal eye position is chiefly controlled by two extraocular muscles, the lateral and medial recti. When the eye is still, it rests at the mechanical equilibrium point at which the net torque due to the muscles and passive orbital tissues vanishes. The location of this equilibrium point is determined by motor neuron drive, which controls the length-tension relationships of the muscles (30).

The lateral and medial recti are innervated by motor neurons in the abducens and oculomotor nuclei, respectively. When the eyes are still, the firing rates of these neurons are constant in time and are approximately linearly related to eye position (31, 32). Although this suggests that the motor nuclei store the memory of eye position, other evidence is to the contrary. If the oculomotor nerve is electrically stimulated, the eye quickly returns to its

prestimulation position after a transient deflection (33). This stimulation presumably activates neurons in the oculomotor nucleus antidromically, yet evidently does not alter the memory of eye position. Patients with partial damage to the connections between the abducens and oculomotor nuclei suffer from weakness of the medial rectus without loss of the ability to hold the eyes still (1). The most plausible explanation of these empirical observations is that the motor nuclei "read-out" a memory of eye position that is stored elsewhere.

Many premotor neurons in the medial vestibular nucleus (MVN) and prepositus hypoglossi (PH) carry an eye position signal (2–7) and project to the abducens nucleus. Stimulation of the MVN or PH causes persistent changes in eye position (9–11). Electrolytic lesion (12, 13) or pharmacological inactivation (9, 14–19) of these areas causes the disorder gaze-evoked nystagmus. Such experimental results indicate that these premotor neurons are part of the network that stores the memory of eye position.

Feedback connections from the abducens and oculomotor nuclei to the PH and MVN are few (34, 35). Since the eyes can be held still in the dark (36), the contribution of visual feedback in the light seems minor. Proprioceptive feedback also seems unimportant, as there is no evidence of a stretch reflex in the extraocular muscles (37).

To summarize, experiments suggest that neurons carrying an eye position signal can be divided into two populations. One population stores the memory of eye position, and the other reads it out. The memory network includes neurons in the MVN and PH. The read-out network consists of motor neurons in the abducens and oculomotor nuclei. The memory network drives the read-out network, which in turn drives the extraocular muscles. This chain of activation is approximately feedforward.

## State Space Analysis

The preceding account of the memory of eye position can be translated into the language of state space analysis. The firing rates of neurons are denoted by  $v_i$ , where the index  $i$  runs from 1 to  $N$ , the number of neurons in the network.

When the eyes are still, the firing rates are constant and are related to eye position  $E$  by:

$$v_i = v_i^0 + k_i E. \quad [1]$$

The intercept  $v_i^0$  is the firing rate at central gaze  $E = 0$ , and the slope  $k_i$  is called the "position sensitivity." The intercept and slope vary from neuron to neuron.

The state of the network is described by the firing rates  $v_i$  of its neurons. The "state space" of the network is the multidimensional space for which the coordinates are the firing rates  $v_i$ . If the network is placed at a generic point in state space, its state evolves in time, tracing out a trajectory. Different initial conditions lead to different trajectories, which can be depicted in a "state space portrait." States of the form (Eq. 1), observed when the eyes are still, are special points in state space. Because they do not change with time, they are called "fixed points" of the dynamics.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MVN, medial vestibular nucleus; PH, prepositus hypoglossi.

The state space portrait of the memory network has many fixed points, one for each eye position  $E$ . As shown in Fig. 1A, these fixed points lie on a line in state space. This “line of fixed points” is parametrized by the rate-position equations (Eq. 1). The vector  $v_i^0$  is a point on the line. To this vector, multiples of the direction vector  $k_i$  can be added to construct the other points on the line. Eye position  $E$  is coded in the displacement along the line from  $v_i^0$ .

The visual and vestibular inputs to the memory network do not normally vary with static eye position, provided that the head is stationary. As a result, a single state space portrait suffices to describe the dynamics of the memory network when the eyes are still, as depicted in Fig. 1A.

In contrast, the read-out network has a different state space portrait for each static eye position, because it is driven by an

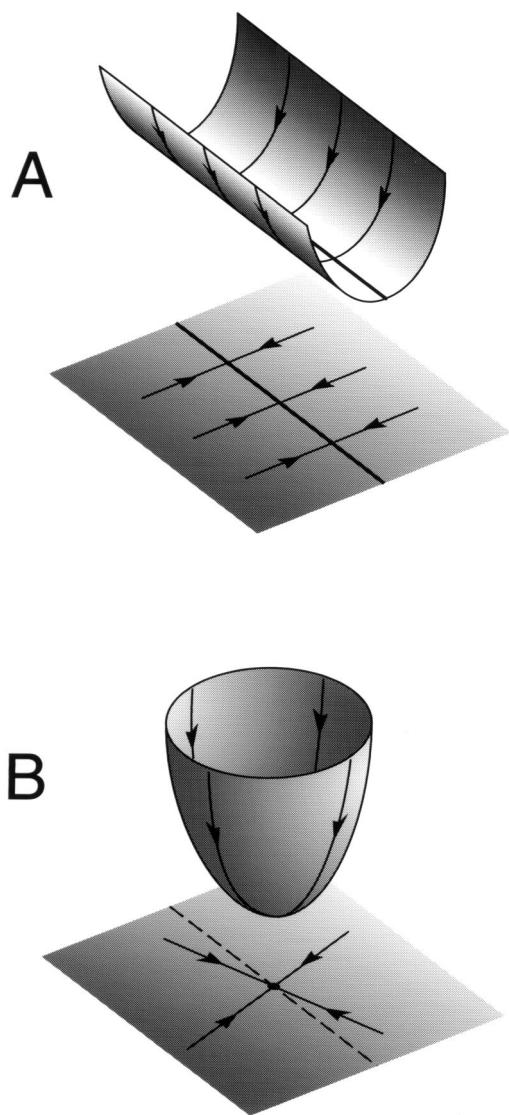


FIG. 1. State space portraits and energy landscapes for the memory and read-out networks. The coordinates of the state space are the firing rates of the neurons in the network. Only two of the dimensions of state space are depicted in the plane, with the height of the energy landscape as the third dimension. The dynamical trajectories in state space correspond to trajectories down the energy landscape. (A) All trajectories in the state space of the memory network flow toward a line attractor (thick line). The corresponding trajectories on the energy landscape flow down the walls of a trough, toward a line of minimum energy at the bottom. (B) All trajectories in the state space of the read-out network flow toward a point attractor. Input to the network controls the location of the point attractor, moving it along the dashed line. A point attractor can be visualized as the minimum of a bowl-shaped energy landscape.

eye position signal from the memory network. Each state space portrait has a “single fixed point” with firing rates  $v_i$  specified by Eq. 1 for a single eye position  $E$ , as shown in Fig. 1B. The inputs from the memory network modulate the state space portrait of the read-out network, determining the location of the fixed point on the line (Eq. 1).

### Perturbations of State

A saccade is a step-like change from one static eye position to another. For a neuron that satisfies the rate-position relationship (Eq. 1), there is a corresponding change from one steady rate of firing to another, but the change does not have the simple time dependence of a step. Some neurons in the memory network burst at high rates or pause in their firing for tens of milliseconds during the saccade. Immediately after the saccade, the firing rates of all the neurons relax slowly over hundreds of milliseconds to new steady levels (2, 3, 5, 8). These “slide-step” postsaccadic responses are the aftermath of the brief pulse input supplied by saccadic command neurons during the saccade (29).

In the language of state space analysis, the pulse input from saccadic command neurons suddenly perturbs the state of the memory network away from the line of fixed points. The ensuing “slide-step” response corresponds to the flow of the network to a new fixed point on the line. The network is kicked in various directions in state space to produce a variety of saccades, yet it always returns to the line. Since all trajectories in state space appear to flow to the line, it is called an “attractor” of the dynamics. It is helpful to visualize this dynamics as a descent on an energy landscape. Then the line attractor is the line of minima at the bottom of a trough with a curved cross-section (see Fig. 1A). Each trajectory flows down the walls of the trough to a point on the bottom. The pulse of input from saccadic command neurons pushes the memory network up the walls of the trough. When the pulse is over, the network relaxes back to the bottom, to a minimum different from the initial one. This new minimum corresponds to a new static eye position. The time scale of the relaxation, which is reflected in the “slide” component of response, is determined by the steepness of the walls of the trough.

To see the dynamics of the read-out network about its single fixed point, its state must be perturbed without an accompanying perturbation of the state of the memory network. This does not happen during saccades, because the command neurons project to both networks. However, such a perturbation can be effected artificially by oculomotor nerve stimulation, as was described earlier (33). This perturbs the state of the oculomotor nucleus via antidromic activation, without perturbing the state of the memory network. As discussed earlier, the eye quickly returns to its prestimulation position. This is indirect evidence that the fixed point is attractive, since it follows from the rate-position relationship (Eq. 1) that the oculomotor nucleus also returns to its prestimulation state. A “point attractor” can be visualized as the minimum of a bowl-shaped energy landscape, as shown in Fig. 1B.

Between saccades, eye position is not exactly constant when measured in darkness, without visual feedback. The eye does a slow random walk (36), which suggests that the state of the memory network does a random walk along the line attractor. The trough visualization makes clear that there is no restoring force opposing perturbations of state along a line attractor. Thus, a memory stored using a line attractor dynamics is susceptible to corruption by noise. Various sources of noise, such as the random fluctuations in the tonic input from vestibular afferents, could be causing the random walk behavior.

### Perturbations of Dynamics

Measurements of eye position in the dark reveal systematic drift, in addition to the random fluctuations described above. The eyes of most normal human subjects slowly drift over a 20–40 s time scale toward the center of the oculomotor range (36). This could

be interpreted as evidence for a state space portrait with a single point attractor, as shown in Fig. 2A. In other subjects, there is no null position, and drift is unidirectional, implying a state space portrait with no fixed points at all, as shown in Fig. 2B.

Both of these cases are best described as "approximations" to the line attractor dynamics of Fig. 1A. There is a trajectory along which motion is slow, and toward which relaxation is fast. On short time scales, this slow trajectory looks like a "line of fixed points." Only on long time scales can the drift along the trajectory be observed.

Perturbations of the dynamics of the memory network can result in much faster eye position drift than is seen in normal subjects. For example, sedatives, anticonvulsants, or alcohol can produce gaze-evoked nystagmus, which is characterized by fast centripetal drift (1). This disorder is presumably due to a memory network with a single point attractor, as in Fig. 2A, but with a longitudinal relaxation time much shorter than the normal tens of seconds. As mentioned earlier, electrolytic lesions of the MVN or PH also produce gaze-evoked nystagmus (12, 13). Vestibular nystagmus (fast unidirectional drift) is often caused by unilateral damage to the vestibular nerve or semicircular canals. Pharmacological inactivation of the MVN or PH can produce centripetal or unidirectional ocular drift (9, 14–19). Some patients with cerebellar disease suffer from centrifugal ocular drift away from a null position, suggesting a memory state space with a saddle point, as shown in Fig. 2C.

Perturbations of dynamics can be visualized as changes in the trough energy landscape. Tilting the trough causes unidirectional drift along its bottom. If the bottom of the trough is uneven, the continuous line of minima breaks up into discrete local minima. Thus, generic perturbations, even very small ones, destroy a line attractor. This sensitivity to infinitesimal perturbations of dynamics is a property known as "structural instability." It suggests that the realization of line attractor dynamics in neural networks requires precise tuning by learning mechanisms.

After about 1 hr of exposure to visual-vestibular conflict stimuli, human subjects develop marked centripetal or centrifugal eye position drift (39). Such deterioration in the dynamics of the memory network is likely due to the same learning mechanisms that normally keep the dynamics well-tuned when presented with natural stimuli. These experiments provide evidence that the time scale of learning (minutes) is much longer than the "computational" time scale of eye movements (seconds).

### Positive Feedback

It was noted above that feedback projections from the read-out to the memory network are sparse or nonexistent. However, there are many feedback loops within the memory network itself. The bilateral MVN and PH nuclei are fully connected by internuclear projections (34, 35). In addition, neurons in the MVN and PH have recurrent collaterals that appear to mediate intranuclear feedback loops. In contrast, motor neurons generally possess no axon collaterals (40).

The feedback loops within the memory network appear important to its function, judging from the fact that damage to the commissural connections of the MVN and PH causes gaze-evoked nystagmus (41, 42). It has been proposed that the role of feedback is to lengthen the persistence time of the network beyond the intrinsic persistence time of a single neuron (23, 28, 43). Positive feedback can oppose the tendency of a pattern of neural activity to decay. If the feedback is tuned to exactly balance the decay, then the activity neither increases nor decreases, but persists without change.

The positive feedback scenario can be expressed in the language of state space analysis. The state space portrait of a group of noninteracting neurons has a single point attractor. If external input perturbs the system, then it relaxes back to the attractor on a time scale determined by the persistence time of an isolated neuron. However, positive feedback in a network of neurons can

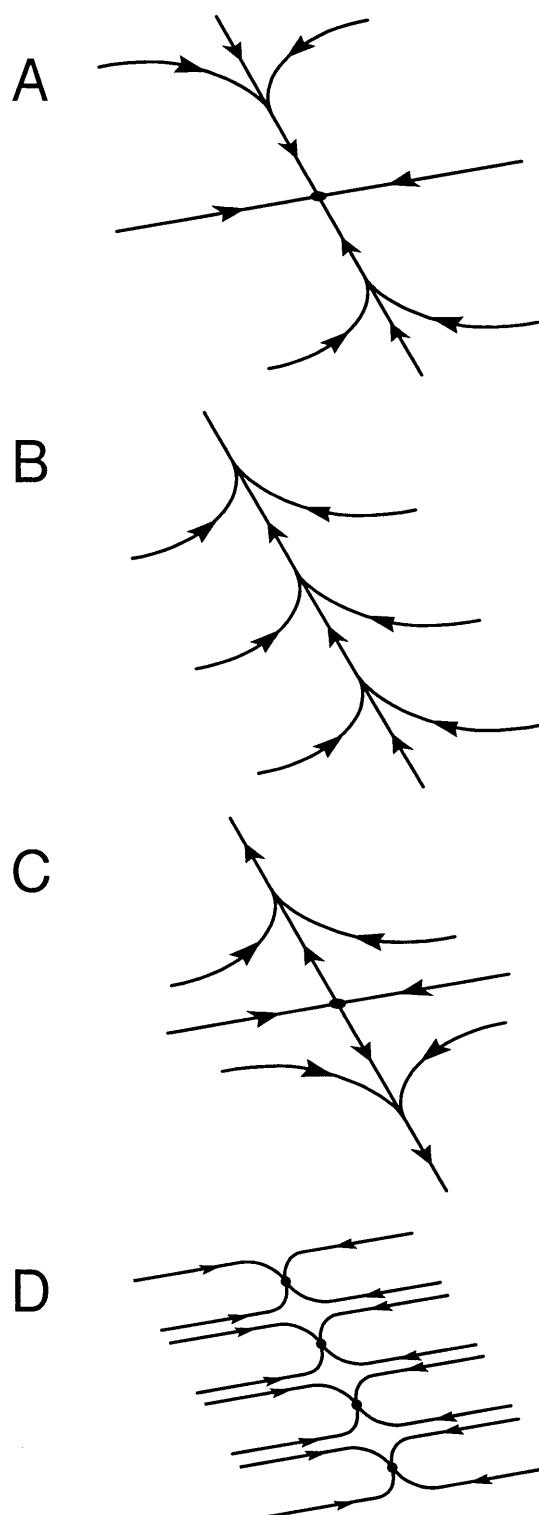


FIG. 2. State space portraits of the memory network. Biological realizations of line attractor dynamics are inevitably imperfect, and are accompanied by various types of systematic eye position drift. In all cases, the trajectory along the oblique axis is very slow, and looks like a line of fixed points on short time scales. (A) Point attractor, corresponding to centripetal ocular drift to a null position. (B) No fixed points, corresponding to unidirectional ocular drift and no null position. (C) Saddle point, corresponding to centrifugal ocular drift. (D) Several point attractors spaced along a line, corresponding to multiple null positions.

lengthen the relaxation time in one of the directions. The feedback can be tuned so that the relaxation time for this direction becomes infinite, producing a line attractor.

The positive feedback scenario assumes a short neuronal persistence time. But the dynamics of ion channels and second messenger pathways can have persistence times of hundreds of milliseconds or even many seconds. Therefore, it is important to consider a scenario in which the persistence time of the memory network is a biophysical time scale (8). State space analysis makes clear the problems with this scenario. Suppose that the intrinsic properties of each neuron endow it with an infinite persistence time, so that it can fire forever at whatever rate to which it is set. Every point in the state space of a noninteracting ensemble of such neurons is a fixed point, which is inconsistent with the experimentally observed line attractor. Consequently, there must be interactions between the neurons that produce transverse relaxation to the line attractor through negative feedback. But this negative feedback scenario is inconsistent with the lesion and inactivation experiments. If it were true, eliminating feedback would increase persistence in the transverse directions, instead of decreasing persistence in the longitudinal direction.

### Linear Network Models

How can positive feedback in a neural network give rise to a line attractor? This question can be investigated at varying levels of detail. Only the simplest class of models is considered here. The total synaptic current  $u_i$  in neuron  $i$  is modeled as a linear superposition of contributions from both recurrent and feedforward connections (44):

$$\tau_s \frac{du_i}{dt} + u_i = \sum_{j=1}^N T_{ij}v_j + h_i. \quad [2]$$

The postsynaptic current in neuron  $i$  due to neuron  $j$  depends on the presynaptic firing rate  $v_j$  and the synaptic strength  $T_{ij}$ . The feedforward term  $h_i$  is constant in time when the head is still, containing only the tonic input from the vestibular afferents. The soma transforms the synaptic current  $u_i$  into a firing rate  $v_i$ , according to some rate-current relationship. According to measurements in MVN slices (20–22), this relationship is roughly linear above threshold:

$$v_i \approx g_i u_i + v_i^p. \quad [3]$$

Even when the current  $u_i = 0$ , there is some spontaneous pacemaker activity  $v_i^p$  (45).

For simplicity, all synaptic currents in Eq. 2 are modeled with a single time scale  $\tau_s = 150$  ms. This important model parameter specifies the persistence time intrinsic to a single neuron. The value  $\tau_s = 150$  ms is consistent with measurements of excitatory synaptic currents in MVN neurons (46). The job of the recurrent feedback connections  $T_{ij}$  is to boost the time constant of the network of  $N$  neurons to a value much longer than  $\tau_s$ .

Fixed-point equations for the rates  $v_i$  can be obtained from Eq. 2 by setting the time derivatives  $du_i/dt$  equal to zero and eliminating  $u_i$  using Eq. 3, yielding:

$$v_i = \sum_{j=1}^N W_{ij}v_j + f_i. \quad [4]$$

The synaptic weight matrix  $W_{ij} = g_i T_{ij}$  is the number of action potentials produced in neuron  $i$  per action potential in neuron  $j$ . The “force”  $f_i = gh_i + v_i^p$  on neuron  $i$  is due to synaptic current  $h_i$  from feedforward inputs and the pacemaker rate  $v_i^p$ , and is the rate at which the neuron would fire if there were no feedback from other neurons.

In general, this set of  $N$  linear equations in the  $N$  unknowns  $v_i$  has a unique solution. In other words, the dynamics has a single fixed point, as in Fig. 2A and C. However, for special choices of  $W_{ij}$  and  $f_i$ , the linear equations can have an infinite number of solutions that lie on a line. These special choices are defined by

two conditions. First, the weight matrix  $W_{ij}$  must have a single unity eigenvalue. This is the condition of tuned positive feedback, and produces a direction in state space along which the energy function has no curvature, as in the trough in Fig. 1A. Second, the force vector  $f_i$  must be orthogonal to the left eigenvector with unity eigenvalue. This condition ensures that spontaneous activity  $v_i^p$  and tonic afferent input  $h_i$  do not push the network along the line attractor. Its effect on the energy function is to make the bottom of the trough perfectly level.

These two conditions for a line of fixed points involve precise tuning. This tuning must be performed by a learning rule (25, 26) that adjusts  $T_{ij}$  and other quantities, such as  $g_i$ ,  $v_i^p$ , and  $h_i$ . If the tuning is perfect, the network time constant is infinite. A more realistic outcome of learning is good but imperfect tuning, producing a slow trajectory that looks like a line of fixed points on short time scales, as in Fig. 2A–C. Motion along the slow trajectory follows an exponential time course with a large but finite network time constant.

The eigenvectors of the weight matrix with unity eigenvalue have interpretations in terms of single-unit properties. For example, the components of the right eigenvector are proportional to the position sensitivities  $k_i$  of the neurons, defined in the rate-position relationship (Eq. 1).

Besides the unity eigenvalue, there are  $N - 1$  other eigenvalues. The real parts of these eigenvalues must be less than unity to make the line of fixed points attractive. Furthermore, the gap separating the  $N - 1$  eigenvalues from the unity eigenvalue must be large. A large gap produces relaxation to the line attractor within time of order  $\tau_s$ , as observed experimentally in the postsaccadic responses of neurons in the memory network.

Cannon *et al.* (23) first proposed the use of linear dynamical equations for modeling the memory network. They interpreted the microscopic time scale  $\tau_s$  in Eq. 2 as a membrane time constant of 5 ms, in contrast to the interpretation given here of  $\tau_s$  as a 150-ms synaptic time constant. They constructed a linear network with a long time constant by tuning an eigenvalue of the weight matrix to unity with an accuracy of 0.025% (the ratio of membrane time constant 5 ms to network time constant 20 s). An obvious criticism of their model was that this degree of precision in tuning seemed biologically implausible. In response, they argued that networks of size  $N$  become more robust to mistuning as  $N$  increases, since the positive feedback becomes distributed over many pathways. This argument was problematic for two reasons.

First of all, although Cannon *et al.* (23) were correct that large networks can be robust to loss of a single neuron, they greatly underestimated the size required. The robustness of their 32-neuron network depended on a feature of the model that is not consistent with the biological data. Namely, several eigenvalues were close to unity, so that there were several directions in state space with long persistence times. This is incompatible with the experimentally observed relaxation to a one-dimensional line of fixed points. To be biologically relevant, a linear network model must have a large gap separating a single unity eigenvalue from the other eigenvalues. In such a network, the sensitivity of the unity eigenvalue to loss of a single neuron is at best on the order of  $1/N$ , as will be explained in more detail elsewhere. By this estimate, roughly 4000 neurons, not 32, are required to keep the eigenvalue tuned to within 0.025%. Accordingly, Arnold and Robinson (25) found their later models to be unrobust to loss of a single neuron, since these models did have large gaps.

Second, the robustness of the Cannon *et al.* (23) model to random fluctuations in synaptic strengths depended on biologically implausible assumptions about correlations between the fluctuations. If the fluctuations are uncorrelated, then their 32-neuron network becomes much less robust. In addition, there are many biologically plausible perturbations of dynamics that have more severe effects than those considered by Cannon *et al.* (23). For example, one can easily imagine a

physiological perturbation that enhances the strengths of all excitatory synapses. A large network is no more robust to this type of perturbation than a small network.

In the description of the neural dynamics (Eqs. 2 and 3), it was argued on biophysical grounds that  $\tau_s$  is most plausibly interpreted as a synaptic time scale on the order of 150 ms, rather than as a membrane time constant on the order of 5 ms. The same conclusion can be reached on the grounds of robustness. Without the biologically inconsistent features of the Cannon *et al.* (23) model, it is difficult for a linear network to achieve the 0.025% tolerance required by the 5-ms membrane time constant. Based on a 150-ms synaptic time scale, a 20-s network time constant can be achieved with only 0.75% tolerance in the tuning of the unity eigenvalue. This is still an impressive degree of tuning, but should be within the reach of biologically plausible learning mechanisms.

This issue could be settled by single-unit recording after lesion or inactivation of a large fraction of the memory network. Disruption of most of the feedback loops in the memory network should make the network time constant fall to a value close to the microscopic time constant  $\tau_s$ . Single-unit recording will reveal whether this time scale is closer to 5 ms, as proposed by Cannon *et al.* (23), or closer to 150 ms, as proposed here. Previous lesion and inactivation experiments have been inconclusive with regard to this issue, because they relied solely on measurements of eye position, which cannot resolve a network time constant shorter than the viscoelastic time constant of the oculomotor plant (150–200 ms).

### Nonlinear Network Models

The linear rate-current relationship (Eq. 3) is only an approximation. Slightly sublinear behavior is evident in the intracellular measurements, especially at high firing rates (20–22). Such deviation from linearity, although small, cannot be neglected in the analysis of a system that depends on precise tuning. Neither can the threshold nonlinearity in Eq. 3 be neglected, because many neurons in the memory network fall below threshold near the center of the oculomotor range (2, 4).

If the rate-current relationship (Eq. 3) takes some nonlinear form  $v_i = g(u_i)$ , the fixed-point equations for  $u_i$  are given by:

$$u_i = \sum_{j=1}^N T_{ij}g(u_j) + h_i. \quad [5]$$

Generally, there is no choice of  $T_{ij}$  and  $h_i$  for which these equations have an infinite number of solutions. This means that a line attractor cannot be perfectly realized, but only approximated. One class of approximations can be found by constraining  $T_{ij}$  to be a matrix of rank one,  $T_{ij} = \xi_i \eta_j$ . Since the matrix-vector product  $\sum_j T_{ij}v_j$  is then a multiple of  $\xi_i$ , it can be shown from Eq. 2 that all dynamical trajectories flow to a line  $u_i = m\xi_i + h_i$  that is parametrized by  $m$ . The full  $N$ -dimensional dynamics is thus reduced to a one-dimensional dynamics in the  $m$  variable, which codes for eye position. A least-squares optimization can be used to tune  $\eta_j$  so that the mean square of  $dm/dt$  is minimized over a range of  $m$ . This produces an approximate line attractor dynamics in which the drift velocity  $dm/dt$  along the line is very slow. The components of the vector  $\xi_i$  are proportional to the position sensitivities  $k_i$  of the neurons, as defined in Eq. 1. The rank-1 construction will be explained in detail elsewhere, along with a biologically plausible learning rule that gives rise to the rank-1 form and performs the least-squares optimization with respect to  $\eta_j$ .

Approximation of line attractor dynamics by a nonlinear network can be qualitatively different than approximation by a linear network. Linear network approximations can have no more than a single fixed point, as shown in Fig. 2 A–C. Nonlinear network approximations can have more than one fixed point, as shown in

Fig. 2D. Some evidence of multiple null positions after partial inactivation of the memory network has been reported (47).

The line attractor dynamics used to store the memory of eye position contrasts with the point attractor dynamics used in other memory networks. In associative memory models (48, 49) and digital memory circuits, the interactions  $T_{ij}$  are chosen to produce strong linear instabilities that are quenched by the nonlinear function  $g$ . The state space portraits of these networks contain discrete point attractors separated by high-energy barriers. The states of these systems are resistant to corruption by noise, and their essentially infinite persistence times do not depend on precise tuning of parameters. Thus, strong nonlinearity can be used to make memory networks that are very robust to perturbations of state and dynamics. In contrast, the analog, graded response of the rate-position relationship (Eq. 1) indicates that the eye position memory network makes weaker use of nonlinear computation. Instead, its robustness appears to be due to the operation of nonlinear learning mechanisms.

### Discussion

The rate-position relationship (Eq. 1) holds in both the memory and read-out networks. In spite of this similarity in neural coding, the two networks have qualitatively different dynamics. The state space of the memory network contains a line attractor. The state space of the read-out network contains a point attractor, the location of which is controlled by input from the memory network.

Point and line attractors can be distinguished by using perturbations of state or dynamics. Perturbations of state can be experimentally induced by electrical stimulation. They also occur naturally when saccadic command neurons provide brief pulses of input. A point attractor dynamics is stable to perturbations of state in any direction. A line attractor dynamics is only stable to perturbations of state in the transverse directions; those in the longitudinal direction are persistent. Perturbations of dynamics can be caused by lesions, pharmacological agents, and many other changes in physiological conditions. A point attractor dynamics is structurally stable, meaning that small perturbations of the dynamics leave the point attractor intact. In contrast, a line attractor is structurally unstable. These issues of stability and structural stability are readily visualized using energy landscapes.

The robustness of a line attractor dynamics is problematic. Its state is corrupted by noise, and perturbations of dynamics shorten its persistence time. The brain deals with this problem by using adaptive mechanisms. Some patients who exhibit severe gaze-holding disorders in the dark can hold their eyes still in the light using visual feedback, demonstrating that sensory feedback can compensate for mistuning in the memory network. On longer time scales, there is behavioral evidence that learning mechanisms tune the line attractor (39). These mechanisms presumably operate by adjusting network parameters such as synaptic strengths or spontaneous firing rates (25, 26).

The long persistence time of the memory network appears to be due to tuned positive feedback (28, 43). If the persistence time intrinsic to a single neuron is a 5-ms membrane time constant, as assumed by Cannon *et al.* (23), the stringent demands on precision of tuning are biologically implausible. Instead, the intrinsic neural persistence time is most likely a synaptic time scale, such as the 150-ms time constant of excitatory synaptic currents observed in the memory network *in vitro* (46). When based on this longer biophysical time scale, models of the memory network become more robust to perturbations of dynamics, making the positive-feedback hypothesis biologically plausible. The opposing hypothesis is that the persistence time of the memory network is purely a biophysical time scale, with little or no contribution from collective interactions (8). This scenario is unlikely because it is difficult to obtain rapid transverse relaxation to the line attractor after saccades.

A line attractor dynamics can be realized exactly by a network of linear neurons (23–26). In the biologically relevant case of nonlinear neurons, a line attractor dynamics can only be approximated. This contrast between exact realization and approximation does not have direct implications for biology, since it exists only in the case of optimal tuning. Biologically plausible learning mechanisms produce good but suboptimal tuning, in which case both linear and nonlinear networks only approximate line attractor dynamics. These approximations can be qualitatively different. Linear approximations have one fixed point, or none. Nonlinear approximations can have more than one fixed point, as in Fig. 2D.

The term “memory network” emphasizes a very limited aspect of its function. In the special case in which the head is still, the feedforward inputs to the network are constant, and the network is an autonomous dynamical system amenable to state space analysis. The autonomous limit was emphasized in this paper because it is easy to understand. During normal behavior, the network receives time-varying vestibular and visual input, and is a driven system rather than an autonomous one. As discovered by Robinson (1), the network integrates these velocity-coded inputs to stabilize eye position with respect to the surrounding environment. Hence, the memory network is more completely described as an “integrator.” The integrator functions as a memory system only in the special situation in which the head is stationary.

In this paper, the memory of horizontal eye position has been treated as separate from that of the vertical and torsional degrees of freedom. This was a simplification made for pedagogical reasons, and may not be accurate. It is known that there are neurons with firing rates that code for more than one degree of freedom of the eye (7). The eyes were also treated as yoked together, when in fact they are under independent control during vergence movements. Thus, a six-dimensional attractive manifold of fixed points may be a more accurate description of state space structure than a one-dimensional line attractor. The dimensionality of this manifold reflects the number of degrees of ocular freedom that must be controlled. The phenomenon of velocity storage (1) has also been neglected here, and involves memory of additional degrees of freedom.

Stimulated by his observations of “reflex after-discharge,” Lorente de Nó (43) formulated his theory of “closed ‘self-reexciting’ chains” in the 1930s. Because the three-neuron vestibuloocular reflex arc was not sufficient to explain persistent neural activity, he was led to the idea of the “reverberating circuit.” The line attractor concept provides a qualitative description of the dynamics of the reverberating circuit that keeps the eyes still, and may also be relevant to the understanding of postural control in other motor systems.

I am indebted to D. Tank for his guidance and encouragement, J. Hopfield for the criticism that stimulated the writing of this paper, K. Drake and D. Lee for help with the figures, and H. Sompolsky for introducing me to line attractors. I am also grateful to R. Baker, K. Delaney, A. Herz, S. Lisberger, P. Mitra, D. A. Robinson, S. Shashy, and B. Shraiman for helpful conversations. This work was supported by Bell Laboratories.

- Leigh, R. J. & Zee, D. S. (1991) *The Neurology of Eye Movements* (Davis, Philadelphia), 2nd Ed.
- McFarland, J. L. & Fuchs, A. F. (1992) *J. Neurophysiol.* **68**, 319–332.
- Scudder, C. A. & Fuchs, A. F. (1992) *J. Neurophysiol.* **68**, 244–264.
- Escudero, M., de la Cruz, R. R. & Delgado-Garcia, J. M. (1992) *J. Physiol. (London)* **458**, 539–560.
- Berthoz, A., Droulez, J., Vidal, P. P. & Yoshida, K. (1989) *J. Physiol. (London)* **419**, 717–751.
- Delgado-Garcia, J. M., Vidal, P. P., Gomez, C. & Berthoz, A. (1989) *Neuroscience* **29**, 291–307.
- Lopez-Barneo, J., Darlot, C., Berthoz, A. & Baker, R. (1982) *J. Neurophysiol.* **47**, 329–352.

- Pastor, A. M., de La Cruz, R. R. & Baker, R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 807–811.
- Cannon, S. C. & Robinson, D. A. (1987) *J. Neurophysiol.* **57**, 1383–1409.
- Godaux, E., Cheron, G. & Gravis, F. (1989) *Exp. Brain Res.* **77**, 94–102.
- Yokota, J.-I., Reisine, H. & Cohen, B. (1992) *Exp. Brain Res.* **92**, 123–138.
- Cheron, G., Gillis, P. & Godaux, E. (1986) *J. Physiol. (London)* **372**, 95–111.
- Cheron, G., Godaux, E., Laune, J. M. & Vanderkelen, B. (1986) *J. Physiol. (London)* **372**, 75–94.
- Cheron, G. & Godaux, E. (1987) *J. Physiol. (London)* **394**, 267–290.
- Cheron, G., Mettens, P. & Godaux, E. (1992) *NeuroReport* **3**, 97–100.
- Mettens, P., Cheron, G. & Godaux, E. (1994) *NeuroReport* **5**, 1333–1336.
- Godaux, E., Mettens, P. & Cheron, G. (1993) *J. Physiol. (London)* **472**, 459–482.
- Mettens, P., Godaux, E., Cheron, G. & Galiana, H. L. (1994) *J. Neurophysiol.* **72**, 785–802.
- Straube, A., Kurzan, R. & Büttner, U. (1991) *Exp. Brain Res.* **86**, 347–358.
- Serafin, M., de Waele, C., Khateb, A., Vidal, P. P. & Muhlethaler, M. (1992) *Exp. Brain Res.* **84**, 417–425.
- Johnston, A. R., MacLeod, N. K. & Dutia, M. B. (1994) *J. Physiol. (London)* **481**, 61–77.
- du Lac, S. & Lisberger, S. G. (1995) *J. Comp. Physiol. A Sens. Neurol. Behav. Physiol.* **176**, 641–651.
- Cannon, S. C., Robinson, D. A. & Shamma, S. (1983) *Biol. Cybern.* **49**, 127–136.
- Cannon, S. C. & Robinson, D. A. (1985) *Biol. Cybern.* **53**, 93–108.
- Arnold, D. B. & Robinson, D. A. (1991) *Biol. Cybern.* **64**, 447–454.
- Arnold, D. B. & Robinson, D. A. (1992) *Philos. Trans. R. Soc. London B* **337**, 327–330.
- Rosen, M. J. (1972) *IEEE Trans. Biomed. Eng.* **19**, 362–367.
- Kamath, B. Y. & Keller, E. L. (1976) *Math. Biosci.* **30**, 341–352.
- Hepp, K., Henn, V., Vilis, T. & Cohen, B. (1989) in *The Neurobiology of Saccadic Eye Movements*, eds. Wurtz, R. H. & Goldberg, M. E. (Elsevier, Amsterdam), pp. 105–212.
- Robinson, D. A. (1964) *J. Physiol. (London)* **174**, 245–264.
- Fuchs, A. F., Scudder, C. A. & Kaneko, C. R. S. (1988) *J. Neurophysiol.* **60**, 1874–1895.
- Pastor, A. M., Torres, B., Delgado-Garcia, J. M. & Baker, R. (1991) *J. Neurophysiol.* **66**, 2125–2140.
- Robinson, D. A. (1968) *Exp. Neurol.* **22**, 130–132.
- McCrea, R. A. (1988) in *Neuroanatomy of the Oculomotor System*, ed. Büttner-Ennever, J. A. (Elsevier, New York), pp. 203–223.
- Highstein, S. M. & McCrea, R. A. (1988) in *Neuroanatomy of the Oculomotor System*, ed. Büttner-Ennever, J. A. (Elsevier, New York), pp. 177–202.
- Hess, K., Reisine, H. & Dürsteler, M. (1985) *Neuroophthalmology* **5**, 247–252.
- Keller, E. L. & Robinson, D. A. (1971) *J. Neurophysiol.* **34**, 908–919.
- Matin, L., Matin, E. & Pearce, D. G. (1970) *Vision Res.* **10**, 837–857.
- Tiliket, C., Shelhamer, M., Roberts, D. & Zee, D. S. (1994) *Exp. Brain Res.* **100**, 316–327.
- Evinger, C. (1988) in *Neuroanatomy of the Oculomotor System*, ed. Büttner-Ennever, J. A. (Elsevier, New York), pp. 81–117.
- Anastasio, T. J. & Robinson, D. A. (1991) *Neurosci. Lett.* **127**, 82–86.
- Godaux, E. & Cheron, G. (1993) *Neurosci. Lett.* **153**, 149–152.
- Lorente de Nó, F. (1933) *Arch. Neurol. Psychol.* **30**, 245–291.
- Hopfield, J. J. & Tank, D. W. (1986) *Science* **233**, 625–633.
- Lin, Y. & Carpenter, D. O. (1993) *Cell. Mol. Neurobiol.* **13**, 601–613.
- Kinney, G. A., Peterson, B. W. & Slater, N. T. (1994) *J. Neurophysiol.* **72**, 1588–1595.
- Crawford, J. D. & Vilis, T. (1993) *Exp. Brain Res.* **96**, 443–456.
- Hopfield, J. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558.
- Amit, D. (1989) *Modeling Brain Function* (Cambridge Univ. Press, Cambridge, U.K.).