Author: David Vogel

Word Counts:

Abstract: 229

Main Text: 10,219

References: 758

Entire Text: 11,349

A NEURAL NETWORK MODEL OF MEMORY AND HIGHER COGNITIVE FUNCTIONS IN THE CEREBRUM

David Vogel
Ross University
Portsmouth
Commonwealth of Dominica
Windward Islands
dvogel@rossmed.edu.dm

Short Abstract

I first review a biologically plausible, neural network model of a small region of the brain that produces associative memory. In new work, identical copies of this local network are connected to form a network that produces serial memory, operant conditioning, refabrication of memory, and fabrication of possible future events. The network distinguishes between real and imagined events, and can respond to the absence as well as the presence of stimuli. Some of these properties are achieved in ways that seem to provide instances of self-awareness and imagination suggesting that consciousness may emerge as an epiphenomenon in simple brains.

Long Abstract

I first review a neural network model of a small region of the brain that produces associative memory. The model depends, unconventionally, on disinhibition of inhibitory links rather than long-term potentiation of excitatory projections. The model may be shown to have advantages over traditional models both in terms information storage capacity and biological plausibility. The simple learning and recall algorithms are independent of network architecture, and require no thresholds, finely graded synaptic strengths, or implausible stochastic smoothness. In new work, several identical copies of this local network are connected by means of massive, weak, reciprocal, excitatory projections that appropriately control the recall of information in one region by activity in another (but reduce the information storage capacities of local networks). The result is a simple network with properties that seem analogous to serial memory, classical and operant conditioning, refabrication of memory, and fabrication of possible future events. The network distinguishes between real and imagined events, and can predicate its response on the absence as well as the presence of stimuli. Some of these properties are achieved in ways that seem to provide instances of self-awareness and imagination suggesting that consciousness may emerge as an epiphenomenon in simple brains. The network further suggests mechanisms by which memory consolidation might occur in real brains. Given a tentative definition of consciousness, mechanisms are demonstrated by which well rehearsed behaviors might become unconscious.

Key words

associative; cognition; conditioning; memory; neural network; P-net; projective network; refabrication; R-net; sparse.

1. INTRODUCTION

Prima facie, trying to model the cerebrum at the neural network level is a fool′s game. The architecture and cellular function are only partially known. Moreover, any evolved control system is not only likely to be complex, it is almost certain to be absurdly complex. Any model that represents such systems with methodological economy is unlikely to represent them accurately.

On the other hand, there are sufficient reasons for the attempt. Even models that are in fundamental respects inconsistent with the known cerebrum have provided insights into mechanisms that may contribute to actual cortical functioning. Marr′s (1971) model of the hippocampus has influenced research and our beliefs concerning that structure for 3 decades.

Moreover, it is extremely difficult for people to imagine how the workings of a machine can explain the existence of a mind. The allegedly transcendent mind is a correspondingly powerful object of the sort of mystification and special pleading that has long been inimical to scientific inquiry. Writing in the Encyclopedia of Ignorance shortly after the introduction of Marr's model, Gregory (1977) observed that there are many "gaps" in science, but the "mind-brain gap" seems, somehow, much greater than all the others. Any intellectually satisfying progress toward filling this gap serves the interests of reason.

Many neural network models of psychological processes replicate some aspect of the input-output relationships of real brains, but they are often “neural network” models in the engineering usage of the term more than they are biological models (e.g., Jani & Levine 2000; Salum et al. 2000). Some models have large scale block designs that are roughly analogous to structures in the brain but do not specify biologically realistic connection matrices (e.g., Taylor & Taylor 2000). Models that are more representative of the real cerebrum are likely to be restricted to special aspects of lower level problems (e.g., Fukai, 1999; Körner et al. 1999). Most attempts at realistically representing significant aspects of the cerebrum have been models of associative memory derived from Marr′s (1971) model of archicortex.

I believe this paper details an unusually plausible neural network that is both simple and able to account for a compelling variety of psychological phenomena including serial memory, classical and operant conditioning, predictive fabrication, and refabrication of incomplete memories. Random excitation of the network allows it to discover sensory consequences of motor activity and to subsequently repeat the motor activity by recalling the sensory consequence. The network distinguishes between imagination and reality, and is able to predicate behaviors, not only on the presence of stimuli, but on their absence. It may provide insight into processes that lead to consolidation of memory. Some of these properties are achieved in ways that seem to provide instances of self-awareness and imagination and suggest that consciousness may emerge as an epiphenomenon in relatively simple brains. The model of consciousness further accommodates the conversion with practice of conscious behavior into unconscious behavior.

I begin, in Section 2, with a brief description of a partial model of a small region of the brain that I have developed more fully in previous papers (Vogel 1998; 2000). The small model describes a region that might amount to a few square millimeters of archicortex (or other structure) and provides a biologically plausible account of associative memory. The completed model does not support certain presuppositions about information storage. Memory, for example, does not depend on the strengthening of excitatory synapses as in long-term potentiation (LTP). Rather, even though only a small fraction of cortical neurons are inhibitory, memory depends on the weakening of inhibitory links (disinhibition). LTP and the sea of excitation actually degrade memory though they provide essential control of the coupling of activities in separate local regions.

I then describe, in Section 3, new work in which I have linked 2 of these small regions together into a larger network by means of large numbers of relatively weak, reciprocal excitatory projections. The resulting model of serial memory seems more satisfying than those I have described previously (Vogel 1995).

Finally, in Section 4, I investigate the role of excitatory connections in gating the flow of information from one region to another by modeling a conditioned response in a network consisting of several local regions, each with the same architecture and information processing rules but variously regarded as sensory, motor, or integrative on the basis of their positions in the network. A similarly constructed “reinforcement” region is also present that responds to a small subset of “sensory” inputs by initiating synaptic training and inhibiting activity in certain other local regions.

This model exhibits enough complexity to represent a miniature mind. It does not incorporate a number of conspicuous features of neocortex including its layered structure and cortical columns. These features are not needed in the present paper though I regard work that incorporates these structures into models of such phenomena as rapid sensory perception (Körner et al. 1999) or spatial working memory (Tanaka 1999) as developing important, complementary ideas about neocortex.

As with all sparsely connected neural networks, information stored in these models must be sparsely coded (i.e., a “sensory” event must be represented by the firing of a small fraction of the neurons in the network). In real brains, information at the level of primary sensory neurons is densely coded. The process of feature extraction presumably generates sparse coding in “higher” regions of real brains. All of the networks described here use information that has already been sparsely coded. Near relatives of these networks are able to convert dense to sparse coding, and I do not imagine that real brains make an exclusive distinction between mnemonic and encoding processes. However, a discussion of network designs that optimize encoding as opposed to mnemonic processes is not required for this paper.

2. A MODEL OF A LOCAL REGION

2.1 Traditional Neural Networks

In models that follow the work of Marr (1971), excitatory neurons, analogous to principal cells, are randomly connected to one another as in Figure 1. The artificial neural networks are sparsely connected with a probability of one neuron projecting onto another comparable to that of the actual cerebrum.

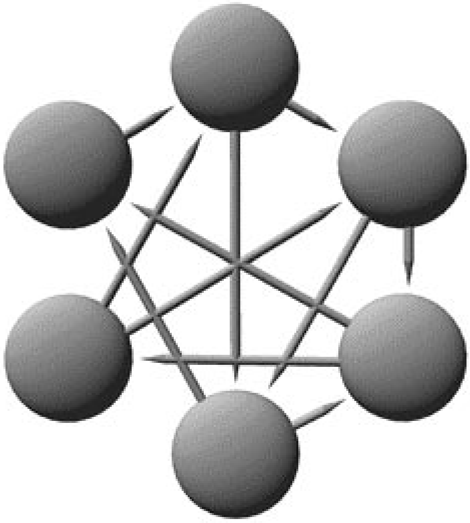

Figure 1. A representation of the standard model of cortical memory in which neurons have random, excitatory projections onto one another.

In operation, a subset of the neurons in the network is activated. I will refer to this subset as a "training set." Synapses between active neurons are then trained according to a "Hebbian" learning rule which specifies that the strength of excitatory synapses is increased when both the pre- and postsynaptic neurons are active at the same time. After training on some number of sets, a subset of a training set (a "recall set") may be activated with the objective of reactivating the original training set (the "target set"). Because of previous synaptic training, members of a target set are likely to receive more excitatory activation from a recall set than non-target neurons. An activation threshold for firing may be computed stochastically such that most target neurons are above the threshold and fire while most non-target neurons are below the threshold and do not. Recall of a target set when a recall set fires is referred to as "auto-associative memory." "Hetero-associative memory" in which one training set recalls another is developed along similar lines and appears in the larger networks discussed below.

As the number of stored training sets increases, the number of recall errors also increases. Errors may consist of active neurons that are not members of the target set ("spurious neurons") or inactive members of the target set ("missing neurons"). In general, the relationship between the number of training sets stored in a network and the number of errors is steeply concave upward so that the number of training sets that may be stored with an error rate of 1% is not greatly different from the number that may be stored with an error rate of 10%.

The required threshold depends on the size of the recall set. As the size of the recall set increases, the amount of activation of all neurons increases and the threshold that separates target from non-target neurons increases. These networks are made more robust with respect to recall set size by positing a diffuse inhibitory feedback. Increasing this inhibition is equivalent to increasing the threshold, and the inhibition is made proportional to the number of active neurons in a recall set (or the number of active neurons on the previous cycle of firing if the network is allowed to run through more than one cycle).

These traditional networks provide some insight into principles by which real brains may store information. However, they have several features that are biologically implausible.

First, the connectivity of the real cortex is at the lower edge of the connectivity required for successful information storage by these neural networks. Even with the assumption of perfectly uniform inhibitory feedback, the networks are brittle with respect to training and recall set sizes. The relatively sophisticated network of Bennett, Gibson, & Robinson (1994) does not effectively store training sets much larger than 330 neurons because the number of synapses trained increases with the square of the number of neurons per training set. It does not store training sets much smaller than 330 neurons because recall is not stochastically smooth.

Second, the introduction of stochasticity in the inhibitory feedback destroys the already tenuous storage capacity of the networks.

Third, the mechanism of uniform inhibitory feedback is not made explicit, and no such mechanism seems to exist in the real brain. Single active inhibitory synapses appear to be extremely effective at silencing postsynaptic neurons (Conors, Malenka & Silva 1988; . & Schwartzkroin 1995; Miles et al. 1996). No finely graduated inhibitory cloud is present.

Fourth, the strengths of synapses must be closely regulated and are dependent on architectural parameters such as connectivity. This seems to make traditional networks evolutionarily implausible because any change in architecture deoptimizes synaptic strength.

Fifth, the brain is not randomly connected, as are the networks in these models. While non-random connection matrices exist that are stochastically as smooth as random matrices (Vogel & Boos 1997), they are relatively rare objects that are not likely to be discovered by evolutionary means.

Sixth, the analysis of the storage capacities of large versions of these networks is suspect. The problem arises in the approximation of a binomial distribution by a normal distribution (Marr 1971). For small networks, equations for the number of errors during recall closely approximate the actual number of errors found in simulations. It seems natural to assume that as network size increases the normal approximation of a binomial distribution will get better and the equations more accurate. However, it is not the network size that is the number of trials in the binomial distribution, it is the recall set size, and the recall set size typically remains constant in moving from simulations to analyses of large networks. It sometimes appears (Gibson & Robinson 1992) that in simulations of small networks, the number of errors is .slightly larger than analysis predicts. If the number of neurons in a recall set remains constant, the number of spurious neurons on the first cycle of operation may be expected to increase linearly with network size, and a small under-estimate of the number of errors in small networks, becomes large in large networks, and large networks may become swamped with noise. It is possible that this analytical problem is less severe in networks in which modestly fewer training sets are stored. However, I am unaware of further work on this problem.

2.2 The Local Network Model

While inhibitory synapses constitute only a small fraction of the synapses in the cerebrum, I have previously demonstrated (Vogel 1998; 2000) that disinhibition of inhibitory pathways provides enormous storage capacity while overcoming all of the objections to traditional networks raised above. In the local network model developed in this section, direct projections between excitatory neurons are ignored. (Their function will be explored in subsequent sections.) Rather, I consider only random projections of excitatory neurons onto inhibitory neurons, and of inhibitory neurons back onto excitatory neurons. In previous papers I have referred to this class of networks as “R-nets” to distinguish their random architectures from those of mathematically more tractable “P-nets” whose connection matrices form projective planes. A more complete discussion of biological plausibility may be found in previous work (Vogel 2000).

In this artificial neural network model of a local region, excitatory neurons are called “primary neurons” and are analogous to the principal cells of the cerebrum. These primary neurons project onto inhibitory “secondary neurons” that are analogous to the interneurons and project back onto the primary neurons (Figure 2). Pairs of primary neurons are then said to be “linked” by inhibitory pathways of length 2. Of the various kinds of interneurons present in the cortex, the basket cells are most common and most appropriate for the construction of these networks.

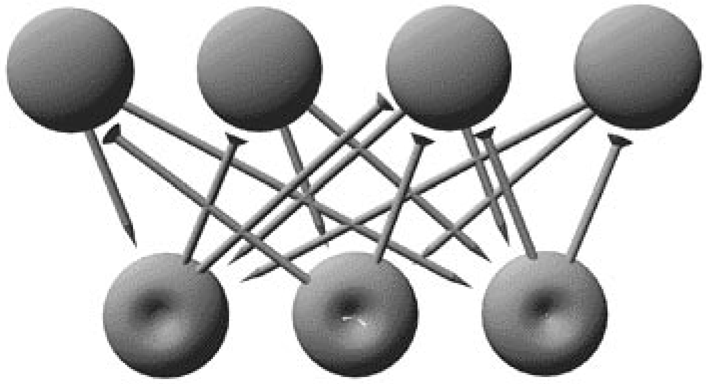

Figure 2. A representation of an R-net. Axons with points indicate excitatory projections. Axons with cups indicate inhibitory connections. Excitatory neurons project only onto inhibitory neurons and vice versa. The numbers of excitatory and inhibitory neurons and synapses represented are not proportional to the numbers in either simulated or biological networks.

A conservative estimate of the size of a biologically relevant network may be made as follows. Each principal cell synapses on at least 1500 interneurons (Freund & Buzsáki 1996); the number of synapses of interneurons on principal cells is at least 320 (Sik, Tamamaki, & Freund 1993); and the ratio of interneurons to principal cells is roughly 0.2. The results of Miles (1990) indicate that about 30% of principal cell pairs have inhibitory links. If another 10% have inhibitory links that have been disinhibited, as in the fully trained models discussed below, then an assumption of uniform synaptic density leads to a rough estimate of the number of principal cells in a mammalian network of 10^6. This is probably a considerable underestimate because the single pyramidal cell studied by Sik, Tamamaki, and Freund (1993) had a total of only 15,295 synapses, and only synapses on parvalbumin staining cells were counted as synapses on interneurons. Perhaps as many as half the basket cells in the hippocampus do not stain with parvalbumin (DeFelipe 1997), and Primate (Ribak, Seress, & Leranth 1993) and especially human (Seress et al. 1993) cortex has larger numbers of inhibitory neurons than the rat, and some of these interneurons appear to have much more extensive connections.

It is, of course, impractical to simulate networks with 10^6 primary neurons. The smaller simulated networks discussed here are constructed with the ratios of primary to secondary neurons, and excitatory to inhibitory synapses given above. The total number of synapses is adjusted so that 40% of primary neuron pairs are linked though at least 1 secondary neuron.

I refer to a collection of primary and secondary neurons with the architecture described above as a “region”. For this paper, all regions simulated on a serial computer consist of 4000 primary neurons, each with 20 projections onto secondary neurons, and 800 secondary neurons, each with 100 projections onto primary neurons. This provides a model with a ratio of primary to secondary neurons comparable to that of primate archicortex. The connectivity is more like that of the rat archicortex.

Learning occurs by activating training sets of primary neurons which in turn activate secondary neurons. When an active secondary neuron links two active primary neurons, both synapses in the link are trained. The training of both synapses results in disinhibition of the link. There are at least two possible interpretations of how training leads to disinhibition.

Disinhibition of inhibitory pathways under conditions comparable to those that produce LTP is well established. Manisha et al. (1994) suggest that disinhibition requires changes in both synapses in a link. They suppose that disinhibition involves weakening of synapses in the pathway. However, inhibitory synapses that employ gamma-aminobutyric acid (GABA) become excitatory at high rates of GABA release. In fact, Staley, Brandi, and Proctor (1995) have concluded from the data of Buzsáki et al. (1992) that the firing rates of most interneurons are sufficiently high to be excitatory. Accordingly, the process of disinhibition may involve potentiation of transmitter release at both synapses. This interpretation is consistent with the data of Manisha et al. (1994) who measured postsynaptic potentials rather than postsynaptic currents.

Neural network models may be constructed that employ either interpretation of disinhibition. The models discussed here assume strengthening of both synapses and conversion of inhibitory to excitatory synapses. This interpretation is also consistent with the following findings: (1) Tetanic stimulation of feed-forward projections of the perforant pathway onto inhibitory neurons in CA1 produces LTP of excitatory synapses on the inhibitory neurons of CA1 (Buzsáki & Eidelberg 1982; Xie & Lewis 1995). (2) Tetanic stimulation of presynaptic axons in the stratum radiatum of guinea-pig CA1 neurons produces LTP of inhibitory synapses (Xie et al. 1995). (3) However, short trains of action potentials in the Schaffer collaterals from CA3 to CA1 combined with depolarization of the CA1 pyramidal cells produces disinhibition (. et al. 1995). In the first two protocols, only one synapse in a potential inhibitory link is tetanized, and potentiation of the synapse results. It is only in the last of these protocols that the 2-synapse link is investigated under conditions comparable to those in the model, and disinhibition results. Of course, in these protocols the three pathways into CA1 are different and may well have different properties.

All synapses in these R-nets are binary, being either trained or untrained. The primary neurons are also binary, being active or not. When a set of primary neurons is activated, the secondary neurons linearly sum the projections of primary neurons onto them with trained synapses having 5 times the weight of untrained synapses. The resulting "activity" of a secondary neuron may be thought of as its action potential firing rate. The ratio of trained to untrained synaptic weights is not critical. A value of 5 was chosen to represent the relative synaptic currents of biological synapses before and after LTP (Malinow 1991).

Following activation of a recall set, an active synapse of a secondary neuron on a primary neuron is inhibitory unless two conditions are met. (1) The secondary neuron has an activity at least as great as that produced by a single trained excitatory synapse, and (2) the synapse of the secondary neuron on the primary neuron was previously trained. Primary neurons fire during the *n*th cycle if and only if they receive no inhibition from neurons active during the (*n-1*)th cycle.

Except for extremely rare events in which five untrained synapses give a secondary neuron the activation produced by a trained synapse, these simple rules produce the following important result: A primary neuron fires on the *n*th cycle if and only if all of its links to primary neurons that were active on the *(n-1)*th cycle are trained. (By a "trained link" I mean one in which both synapses are trained.)

This result leads to large information storage capacities. Mathematical analysis (Vogel and Boos 1997) of carefully structured P-nets demonstrates that brain-sized versions of these networks will store 0.25 bits per synapse. I have further demonstrated that R-nets with somewhat greater synaptic densities than those discussed here must scale up as efficiently as P-nets so that simulations of both P- and R-nets lead to the rigorous conclusion that R-nets with synaptic densities sufficient to link nearly all neuron pairs have large storage capacities. I have not rigorously demonstrated similar storage capacities for the somewhat less densely connected networks employed here. However, in simulations of networks with up to 5000 primary neurons, the information storage capacities per synapse of these less densely connected networks are greater than those of more densely connected networks. Assuming less densely connected R-nets scale up with storage capacities per synapse as large as those of more densely connected networks, an R-net with 10^6 primary neurons and brain-like connectivity will store at least $2x10^8$ bits of information.

The disinhibition algorithm requires no threshold setting, and all it requires of the architecture is that sufficiently many pairs of primary neurons have inhibitory links. As far as I can tell, the model requires no properties that are not properties of real brains. Moreover, no parameters of the training and recall algorithms depend on network architecture, and the model seems both evolutionarily plausible and appropriate to a variety of neural structures. As implemented, here, the architecture most closely resembles archicortex.

Extremely noisy recall sets may be used with only a small modification of the recall algorithm. In order to keep noise from inhibiting target neurons, neurons are allowed to accumulate inhibition (up to some maximum). This inhibition decays linearly with time. In general, noise in a recall set will turn off all primary neurons. However, following excitation with a noisy recall set, target neurons will receive less inhibition than nontarget neurons, will fire again before nontarget neurons, and prevent nontarget neurons from firing.

In all of the networks considered here, primary neurons are allowed to accumulate up to 20 units of inhibition, and if inhibition decays at a rate of one unit per cycle, complete recall requires roughly 20 cycles of activity.

All training sets imposed on the networks are 20-neuron training sets. However, training sets evolved by spontaneous activity of the networks described below may contain less than 20 neurons.

3. MULTI-REGION NETWORKS

3.1 General Considerations for a 2-Region Network.

As in the real brain, two or more regions may be connected by excitatory projections of the primary neurons of one region onto the secondary neurons of another. (For now, I neglect projections onto excitatory neurons.) The training and recall rules remain the same as those for local connections. In all of the simulations discussed below, each primary neuron has 5 projections to each non-local region to which it projects, and unless otherwise specified, regions are reciprocally connected (reciprocity being common though not unfailing in the cortex).

For convenience, I will treat any network with more than one region as having sensory and motor regions. I do not mean to imply that the principles developed here apply only to pairs of regions that have a sensory-motor relationship. I now describe a two-region model with the following properties.

1. Regions are able to maintain active sets independently of one another.

2. A region in which no training set is active fires neurons at random. These randomly firing sets may be trained to form local representations of activity in other parts of the network.

3. Under appropriate conditions an active set in one region is able to activate a set in another

region to which it has been trained.

The simple two-region network with primary projections only onto secondary neurons does not have properties 1 or 2. Rather, if sets that are not trained together are activated in different regions, the inhibitory influence of one region will eventually silence the other.

The intended properties may be obtained by adding a diffuse cloud of excitatory projections from the primary neurons of one layer onto those of another. In the simulations that follow, each primary neuron has 500 random projections onto the primary neurons of any other region to which it projects. A stricter model of the cerebrum would have more such projections, but 500 projections produces sufficient stochastic smoothness to permit proper functioning of the network without slowing serial computer simulations unacceptably.

The state of a primary neuron depends on a weighted, linear sum of the active inhibitory and excitatory synapses. If the total excitation is greater than or equal to the total inhibition, a primary neuron fires. The relative weights of the excitatory and inhibitory synapses are not critical. If the amount of inhibition produced by one inhibitory synapse is taken as 1, I have been able to obtain the above properties with excitatory synaptic weights ranging from 0.02 to 0.4.

In the simulations that follow, it is sometimes desired that activation of a set in the sensory region lead to activation of a motor set to which the sensory set has been trained. In order to speed up this process, I have commonly made trained inhibitory links somewhat excitatory and employed the equivalent of LTP of excitatory synapses. Again, the excitatory synapses are binary and potentiation of excitatory synapses is governed by a simple Hebbian rule. I emphasize that training of excitatory synapses actually decreases the storage capacity of the networks. LTP contributes to gating the flow of information from region to region rather than to the storage of information.

Unless otherwise specified, for all of the networks described here untrained excitatory synapses acting on primary neurons have a weight of 0.05, trained excitatory synapses have a weight of 0.1, and the excitations of trained inhibitory synapses have a weight of 0.3.

I have made no particular effort to optimize the above synaptic values. The effort needed to do so, empirically, could not produce results of proportional interest. Instead, I emphasize the robustness of the unoptimized networks that follow.

3.2 Formation of Training Sets

If a training set is active in a region, its activity suppresses the activity of other neurons. Otherwise, a region fires small numbers of neurons spontaneously. Suppose that some activities in the motor region result in corresponding activities in the sensory region as, for example, motor activity produces corresponding proprioceptive responses. Suppose, further, that some mechanism exists (discussed in Section 4) that causes training to occur either because of the sensory activity, or as the result of repetition of correlated motor-sensory activity. Then random activity in the motor region that produces sensory feedback will result in the conjoint training of sets in both regions. Later activation of a sensory recall set will result in activation of the corresponding motor target neurons.

In the networks discussed here, regions with no active training set spontaneously fire about 6 neurons per cycle. The particular neurons vary in a pattern that appears random to the casual observer even if the network has been trained, previously, on many training sets. (Given sufficiently many hundreds of cycles a region will fall into one of those training sets.) A 6-neuron training set is not sufficiently large to reliably suppress random activity. However, larger training sets may be produced by at least four means.

First, neurons that have fired most recently may be regarded as currently firing. If the network trains neurons that have fired over the last 5 to 10 cycles, training sets of roughly 20 neurons per region will be formed.

Second, larger networks may be expected to fire larger numbers of neurons spontaneously. The number of neurons fired per cycle is not directly proportional to the number of neurons in the network. (As the size of the network increases, the number of randomly firing neurons increases which decreases the probability of any one neuron firing.) If the average number of spontaneously firing neurons increases with roughly the square root of the size of the network, a brain-sized network should form spontaneous training sets of about 100 neurons.

Third, if training sessions involving the same motor neurons are repeated, randomly firing neurons will be added to the training set until the set is sufficiently large to inhibit random firing.

Fourth, because real brains exhibit behaviors that are not learned, there are instances in which it is appropriate to construct a network with some pathways that are already trained.

In the conditioned learning paradigm considered in Subsection 4.3, representations of events in motor and sensory regions are created in an integrative region. The representations consist of sets of roughly 20 randomly chosen neurons. In some simulations the representations have been created by the first of these methods. In others, I have simply chosen 20 neurons at random.

3.3 Serial Memory

A small modification of the above training rules will allow a 2-region network to learn sets in series. Consider a series of training sets each of which spans sensory and motor regions. Suppose that, in addition to training links between currently active neurons, links from previously active motor neurons to currently active sensory neurons are also trained. Finally, suppose that either (1) there is a periodic event that causes strong inhibition of currently active neurons in the sensory region, or (2) sensory neurons are burst neurons that spontaneously decrease their firing rates.

After training the series, suppose recall set 1 is activated in the sensory region. Activation of the sensory elements of set 1 permits activation of the motor elements. Links from motor set 1 to sensory set 2 have been trained; so motor set 1 would permit firing of sensory set 2. However, sensory set 2 is inhibited by sensory set 1 until the event occurs that slows the firing rate of the active neurons in the sensory region. After this event, sensory set 2 becomes active, inhibiting motor set 1 and permitting the firing of motor set 2. After sensory set 2 neurons cease to fire rapidly, motor set 2 permits activation of sensory set 3, and so forth. The rate of read-out of the series depends on the frequency with which active neurons in the sensory region are inhibited (or on the length of bursting).

This two region serial network cannot reliably recall two series that share a common training set. Similarly, it has difficulty recalling a series that repeats the same training set. These difficulties are overcome in the next section.

4. A MORE COMPLEX MULTI-REGION NETWORK

4.1 Architecture and Rules

There are many interesting ways in which regions with similar or identical properties may be brought together to perform functions similar to those of the network described in this section. Consider the specific architecture shown in Figure 3.

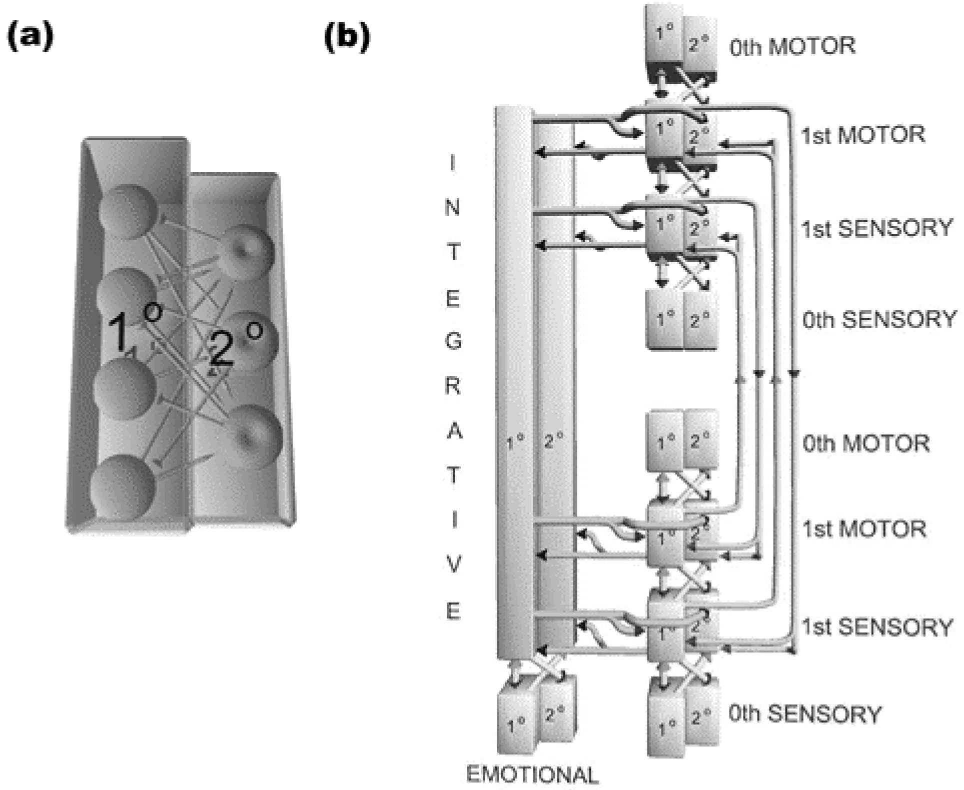

Figure 3. (a) A block representation of an R-net. Each block represents one region containing both primary (excitatory) and secondary (inhibitory) neurons. (b) A block diagram of a network with two channels. No inhibitory projections extend outside a block. Each arrow represents all projections of primary neurons in a region onto primary and secondary neurons of some other region.

Direct activation of neurons in the sensory layer can force states on networks that they would otherwise resist. Accordingly, I have added “0th sensory regions” that are reciprocally connected to the “1st sensory regions” in exactly the way the 1st sensory regions are connected to motor regions. All sensory information is forced on the 0th sensory regions and may or may not control activity in the corresponding 1st sensory regions.

Similarly, each motor region connects to a corresponding 0th motor region. When modeling conditioned responses, no motor action is regarded as actually occurring until the training set that represents it is active in the 0th motor region. In general, 0 denotes the sensory or motor region nearest the periphery.

I call a series of regions from a 0th sensory to a 0th motor region a “channel.” In networks with more than one channel, the channels are weakly connected between 1st sensory and 1st motor regions. Each primary neuron in a 1st sensory (motor) region has 2 projections onto secondary and 200 projections onto primary neurons in the 1st motor (sensory) regions of other channels. This arrangement is just a chunked version of the familiar cortical architecture in which neurons project most densely to the centers of their target areas.

In addition, the architecture has an integrative region to which all 1st sensory and 1st motor regions project reciprocally just as 1st sensory and 1st motor regions project to one another. As before, I do not mean the words "sensory", "motor", and "integrative" to be taken too literally.

Finally, there is a "reinforcement" sensory region that connects only to the integrative region. In it are two pretrained sets I call the "reward" and "punishment" sets. By a mechanism that I have not made explicit in the model, activation of either the reward or punishment set causes (1) training to occur throughout the network according to the previously described rules, (2) inhibition of active neurons in the 1st sensory regions, and (3) self-inhibition of active neurons in the reinforcement region. In addition, activation of the punishment set causes inhibition of active neurons in the 0th motor regions (i.e., it prevents response). I will refer to this combination of events as a "training-inhibitory" or "T-I" event.

When the reward or punishment sets are activated, links between currently active neurons are trained. Additionally, projections of neurons in the 1st motor regions onto neurons in the 1st sensory regions are trained for those motor neurons that were active during the previous 100 cycles. 100 cycles was chosen because, when a new set is activated in the 0th sensory regions, the network will converge on the new set within about 50 cycles. Training backward for 100 cycles insures that the current set is trained to the previous set. Inhibition induced by the reward or punishment sets persists for at least 50 cycles.

In these studies a T-I event only produces a decrease in firing rate of active sensory neurons. Following a T-I event a previously active neuron with no current inhibitory feedback fires on every twentieth cycle (giving the neuron bursting behavior). There is no compelling reason to tie this inhibition to activation of the reward or punishment sets and training. Bursting may as well be made an intrinsic property of the neurons. I have linked the two phenomena chiefly because, in the simulations that follow, it allows the network to reject punished behavior and seek rewarded behavior more rapidly. It also seems to provide a step toward a mechanism for allowing a series to be read out at a rate determined by some other region. (The rate of readout depends on the time between inhibitory events).

Trained excitatory projections from 0th sensory onto 1st sensory and from 1st sensory onto 1st motor regions have weights of 1.5. This rather large weight could be made correspondingly smaller by increasing the strength of trained inhibitory synapses. This weight is necessary for downstream information to control the activity in upstream regions under the conditions specified in the next section.

This network produces the following behavior in the integrative region. If a trained set is activated within the 1st sensory region at a time when there is an active trained set in the integrative region, activation of the set has little effect on the integrative region. If activation of the set in the 1st sensory region is followed by activation of elements of the set in the 1st motor region, their combined activities will have one of two effects. If there is a representation of this state that has been trained in the integrative region, it will become active. Otherwise, the integrative region will appear to fire at random.

The properties of this network seem most easily described by giving examples of learning and recall paradigms. I have simulated several versions of each of these scenarios on a serial computer. Each simulation has been repeated in networks with from 1 to 4 channels, channels being added in a strictly modular fashion with no change in synaptic parameters or local architecture.

4.2 Serial Learning

Suppose one of the methods described in Subsection 3.2 has been used to produce numerous training sets that span channels. Consider an event in which recall sets are activated in the 0th sensory regions of each channel. (None of these recall sets are in their original contexts.) Firing these recall sets will produce one of two possible outcomes. (1) If there are no active training sets in the network, the recall sets will progressively recall target neurons in the 1st sensory, 1st motor, and 0th motor regions of their respective channels. I may write that the channels "converge" on their target sets. (2) If there are active trained sets in any of the 1st sensory regions, they will prevent activation of target neurons in the corresponding channel. Instead, the recall set will remain active until a T-I event occurs after which target neurons may or may not become active in the rest of the channel depending on conditions described below.

We begin to create a series by activating a recall set, set 1, in the 0th sensory region of each channel. In less than 100 cycles the channels will converge on their respective sets denoted 1. Suppose the response of some 0th motor region results in a sensory event in the reinforcement region and produces a T-I event. Synapses will be trained that link active neurons of the various 1st sensory (motor) regions to active neurons in the various 1st motor (sensory) regions. (Links within channels have already been trained.) Active neurons in both 1st sensory and 1st motor regions will also be trained to a randomly selected set of roughly 20 neurons in the integrative region. These neurons in the integrative region form a representation of the activity in the 1st sensory and 1st motor regions.

Next activate a recall set 2 in each channel. If a set 1 is firing slowly (in some channel) following a T-I event, the set 1 is unable to inhibit set 2 and the channel converges on set 2. (The delay between the T-I event following convergence on set 1 and the activation of set 2 may be as long as one wishes because continued slow firing in the 1st sensory region will maintain the network in a steady state even though it does not prevent penetration of set 2.) If set 1 is still firing rapidly at the time set 2 is activated, set 2 will remain active in the 0th sensory region until a T-I event occurs at which time the channel will converge on set 2.

Once the sets denoted 2 are firing in both sensory and motor regions, their combined effects will cause the integrative region to fire at random. Again, let some response result in a T-I event. The sets denoted 2 are trained to a representation of the state of the network in the integrative region. In addition, because the network has converged on set 2 in less than 100 cycles, links from 1st motor elements of sets denoted 1 onto 1st sensory elements of sets denoted 2 are trained.

The serial training procedure may now be repeated for as many sets as the network can store. Recall may be initiated in any of several ways. An obvious choice is activation of a recall set in a 0th sensory region. Alternatively, one may reactivate the representation of set 1 in the integrative region.

If one begins by activating a recall set of sets denoted 1, the network will converge on set 1. When target neurons in both 1st sensory and 1st motor regions become active, the integrative representation of set 1 will become active and activate the appropriate set in the reinforcement region either inhibiting the response or permitting it. Activation of the reinforcement region inhibits set 1 in the 1st sensory region and permits activation of set 2 in the 1st sensory region. Activation of set 2 inhibits set 1 in the 1st motor region and permits activation of set 2. The combined action of 1st sensory and 1st motor regions forces activation of the corresponding representation in the integrative layer and the appropriate set in the reinforcement region. The network converges on set 2. After the next T-I event, the network converges on set 3. In this way the entire series is read out.

This procedure works reliably only if set *x* is turned off in the 0th sensory region at some time between activation of set *x* and the T-I event following convergence on set *x+1*. If it is not turned off, after the network converges on set *x+1*, recall may run backward to set *x* rather than on to set *x+2*. One may either extend the T-I event to include inhibition the 0th sensory region or simply regard the 0th sensory region as having gone on to reflect other inputs during this time.

The network is able to store different series with common sets provided the context of the two series is different. The small number of projections of each primary neuron of a 1st sensory region onto 1st motor regions of other channels are sufficient to insure recall of the next set in the correct series. In multi-channel networks, if out-of-context recall sets from different series are activated in different channels at the same time, the channels can recall their series independently. However, if both series have representations in the integrative network with one trained to the reward set and the other to the punishment set, the behavior of the integrative and reinforcement networks is unpredictable.

Note that, in this example, the 1st motor sets of various channels are serving as context for the corresponding 1st sensory sets. It does not follow that every 1st motor set corresponds to a behavior. Rather, the 1st motor sets may only be representations of the activity in the sensory region at the time of training.

4.3 Conditioned Learning

It is now easy to understand how the network learns a conditioned response. Suppose, for example, a recall set 1 is activated, and the network converges on set 1 which produces a motor response. Suppose, further, that the motor response leads to a sensory event that includes activation of the punishment set. The set 1 response is inhibited. Next, suppose a recall set 2 is activated, and the network converges on set 2 which produces a motor response that leads to activation of the reward set.

If the recall set 1 is repeated, the network will again converge on set 1. Activation of both 1st sensory and 1st motor elements of set 1 permits activation of the representation of set 1 in the integrative region which in turn permits activation of the punishment set. This activation inhibits the set 1 response and permits the network to converge on set 2 which activates the reward set and permits the set 2 response. The network responds to the first (punished) stimulus with the response originally associated with the second (rewarded) stimulus.

4.4 Prediction and Refabrication

Predictive fabrication and refabrication of memory depend on the same principle. If set *x* is trained to set *y* during event 1, and set *x* is trained to set *z* during event 2, then if sets *y* and *z* are active during event 3, set *x* will be activated in preference to any other set to which *y* or *z* may be independently trained. The principle is illustrated in more detail in the following simulation:

Suppose, on some occasion such as my birthday, represented by set 1 in channel 1, I have traveled to Toucari Bay, set 1 in channel 2, and my birthday and Toucari Bay become associated with some other event, set 1 in channel 3, and a dock, set 1 in channel 4. On some other occasion, set 2 in channel 1, some other event occurs, set 2 in channel 2, which is associated with swimming, set 2 in channel 3, and again, a dock, set 1-2 in channel 4. Now, on some still further occasion, set 3 in channel 1, I go swimming in Toucari Bay. Both Toucari Bay, set 1 in channel 2, and swimming, set 2 in channel 3, are associated with the dock, set 1-2 in channel 4. Activating Toucari Bay, set 1 in channel 2, and swimming, set 2 in channel 3 along with any randomly chosen sets in channels 1 and 4 will activate target neurons in the 1st sensory, and then 1st motor regions. These sets are sufficient to activate my representation of the dock, set 1 in channel 4 at the next T-I event. In other words, swimming in Toucari Bay will cause me to think of the dock, and if a more complete network has been appropriately trained, I may think of swimming toward the dock though I have never swum in Toucari Bay before.

It is worth noting that the activation of my representation of the dock may not occur immediately. If the randomly chosen sets in channels 1 and 4 are trained to a response, that response may occur first, but some later T-I event will result in activation of the representation of the dock.

In this way, the network may fabricate imaginary future events. Refabrication of complete memories from fragments is a familiar psychological process that does not appear to be significantly different from that which results in predictive fabrication.

4.5 The Null Stimulus and Reality

The network described so far is unable to respond to the absence of a particular stimulus, and cannot distinguish between a stimulus from the real world and a memory. Both problems may be solved if direct, feed forward projections from the 0th sensory region are provided to the integrative region so that real input may serve as part of the context of a response. (Feedback projections are omitted to prevent the integrative region from controlling activity in the 0th sensory region producing what might be thought of as hallucination.)

Suppose we wish to teach the network to run to the mouse hole and think about the mouse without pouncing until the real mouse appears. In more formal terms, suppose we wish the network to recall a series ending with a training set that represents “mouse pounce,” but to refrain from executing “mouse pounce” unless the mouse is present. The required training requires at least two episodes of punishment.

The network is trained by exciting the series and rewarding “mouse pounce” in the presence of the mouse. It is then trained at least twice on events in which “mouse pounce” is punished in the absence of a mouse.

During training, each time the network pounces in a new context, a new representation is formed in the integrative region. On recall, anytime “mouse pounce” becomes active in a context that has not previously been trained, the representation in the integrative region will include a mixture of the representations for each of the previous appearances of “mouse pounce.” (Please ignore the contradiction until the end of the next paragraph.) If a majority of such appearances have been punished, the mixed representation will recall the punishment set in the reinforcement region. If, however, “mouse pounce” appears in a context that was rewarded, the integrative representation associated with that context will be recalled, the reward set will be activated, and “mouse pounce” will be executed. If training on 2 “mouse pounce” events with no mouse does not produce reliable inhibition of the “mouse pounce” response, then training on a third or fourth such event will do so.

As described so far, the procedure actually fails for the following reason. After training the network with examples in which the mouse is present, the network must be trained on the examples with the mouse absent. When “mouse pounce” is activated without the presence of a mouse, some integrative neurons from the representations with the mouse present are activated, and these are incorporated into the representation with the mouse absent. Subsequently, the network makes frequent errors. Of various simple solutions to this problem, the one that seems biologically most appropriate consists of separating the reward and punishment channels so that there are separate integrative and reinforcement regions for reward and punishment with reward training only the corresponding integrative region. In this way it is not possible for the representations for rewarded and punished contexts to become mixed. It is then straight-forward to construct the networks so that, in a novel context, the most frequently excited reinforcement network turns on first and inhibits the other through lateral inhibition. In these simulations, I have simply randomized activity in the integrative layer each time a training event occurs, ceasing to train further once the network has learned to solve the problem. This procedure requires biologically unrealistic supervision but demonstrates the basic mechanism.

In the example, I have trained the network to respond only in the presence of the real stimulus. It is, of course, possible to reverse the sets being rewarded and punished so that the network always responds in the absence of the real stimulus.

4.6 Quantitative Simulations

I have simulated the network described in Subsection 4.1 on a conventional serial computer to obtain the results described qualitatively in Sections 4.2 through 4.5. Each of the following quantitative results is from 10 random trials.

Associative storage capacity was studied in the following way. The network was first trained on some number of training sets each of which included elements in all regions of all channels. A recall set consisting of all the elements of a particular set in all the 0^{th} sensory regions was then activated, and the network was allowed to run for 100 cycles. If the network had converged on the target set with no more than 4 errors per region (both spurious and missing neurons), the training set was regarded as successfully stored and recalled. In a network with 1 channel, the average number of training sets that could be stored and recalled was 46.1 (range 42 to 51). Not surprisingly, increasing the number of channels, which produces only minor collateral effects, had no discernable effect on the storage capacity per region. With 2 channels, 47.5 sets were stored (range 41 to 50). Increasing the tolerance for errors from 4 to 5 increased the average number of sets stored to 49.5. As is usual with such networks, changing the error tolerance has little effect on storage capacity.

Serial memory was tested by activating set 1 of a series in the 0th sensory regions. If all sets were recovered in the correct order, the series was lengthened until the sets were not recovered correctly. Most commonly, failure consisted of activating sets in the wrong order rather than with excessive numbers of errors. Frequently, the network simply skipped a set in the series. On other occasions, the network repeated the same set indefinitely. In a network with 1 channel the average greatest length for a series was 22 sets (range 20 to 26). The average numbers of missing and spurious neurons per region per set were 0.46 and 2.0, respectively. In networks with 2 or 3 channels, the average greatest length for a series was 25.4 and 25.6 sets, respectively.

For serial recall, the network is modestly sensitive to synaptic parameters. Increasing the weight of trained excitatory synapses that project from the 1st motor region to the 1st sensory region from 0.1 to 0.2 produced a 9% increase in the storage capacity of the network with 1 channel, and a comparable decrease in the storage capacities of the networks with 2 or 3 channels. Increasing this weight tends to shift the cause of failure from repeating a set to skipping a set, and the network with 1 channel is more likely to repeat sets.

4.7 Consolidation

The network could easily be made to show a kind of memory consolidation that reminds one of the role of the hippocampus, which is needed for the establishment of certain kinds of memory, and for their immediate recall, but not for their later recall. In the conditioning example of Subsection 4.3, after training is complete, reactivation from memory of the stimulus that leads to punishment will lead to convergence on the rewarded response in less than 200 cycles, many fewer cycles than would be required for the imagined events to play out in real time. One could easily program the network to train the sensory set with a punished unconditioned response directly to the motor set that represents the conditioned response as though they were simultaneous events. Subsequent activation of the sensory set would activate the motor output without intermediate activation of representations in the integrative region. What is not included in the model is (1) an internal mechanism for automatic rehearsal of the sequence, and (2) a mechanism (as is seen in sleep) that prevents activation of the 0th motor region even though the reward set is active.

5. DISCUSSION

5.1 Memory and Conditioned Responses

This partial model depends chiefly on familiar properties of the brain. It depends most conspicuously on (1) local, inhibitory links that are more effective than excitatory synapses, (2) distant excitatory projections onto both excitatory and inhibitory neurons, (3) a distribution of such distant excitatory projections that is more dense in the central region of the projections, (4) training of inhibitory links that results in disinhibition, (5) conventional LTP, and (6) an internal standard of success which, by inhibiting or permitting responses, provides an impoverished version of the kinds of emotional functions commonly attributed to the rather vaguely defined "limbic system." In addition to these properties, which are clearly properties of real brains, the model depends on (1) the presence of an integrative region in which representations are formed of combinations of sensory and motor states, (2) a mechanism by which active neurons in the sensory and reinforcement regions show bursting behavior, and (3) training of links from certain recently active neurons to certain currently active neurons. (The training could have been between any two regions such as the integrative and sensory or motor regions.)

Given these properties, the network is robust requiring no critical regulation of synaptic weights or careful construction of a connection matrix. It appears to store sufficient information to account for mammalian memory, although rigorous proof of this claim remains beyond my ability.

The model is clearly incomplete. It does not make use of such obvious organizational features as cerebral columns. However, I know of no reason why this model could not be integrated with models that use columnar structure to account for other aspects of cerebral function (e.g., Körner et al. 1999; Tanaka 1999) .

Aside from built-in, stimulus-response functions, random excitation of the motor regions allows the network to discover motor-stimulus relationships. The network is subsequently able to execute motor activity merely by (in some sense) "thinking" of the sensory consequence of the activity.

In the single example given in Subsection 4.4, the network demonstrates classical, operant, and backward conditioning. More limited examples are easily constructed that demonstrate these effects independently.

As described here, series learning requires that each set in the series produces sensory feedback that results in reward or punishment. However, it is not difficult to imagine that coincidence firing induces a state in a link that may lead to training if reward or punishment occurs in the near future. Such an arrangement would permit a series to be learned in response to a single reward. The problem would lie in determining the start of the series. The start might, for example, be defined by the onset of novelty. As I am more interested in demonstrating the behavioral complexity of simple networks than in building airy castles, I have not included a novelty detector in these models. I note, however, that information about novelty is present in the network in terms of the number of neurons active on any one cycle in the integrative region, the number being large when familiar events occur and small when novel events occur.

I have emphasized simple conditioned behaviors in a small network. I think more elaborate versions of this network may reasonably be expected to produce more complex psychological results. For example, in a network that is capable of marking a series for training by means of a criterion such as novelty, a series fabricated from memory might be trained by the recall of reinforcement rather than actual sensory input.

5.2 The Null Stimulus

The simulation described in Subsection 4.5 is roughly analogous to teaching the network to run to the mouse hole and wait for the real mouse before pouncing. It is not especially disturbing that this is more difficult to learn than a statement such as “pounce if you think of a mouse.” Any mother cat can tell you it takes more than one training episode to teach a kitten when to pounce.

If, in the simulation of Subsection 4.5, the states that are rewarded or punished are reversed, the learned program is more nearly analogous to the statement, “Run to the mouse hole. If there is no cat, run to the kitchen”. That this requires more than one training episode is also not surprising. The distinction between “If there is no cat, run to the kitchen”, and the easily learned “If there is a cat, do not run to the kitchen” is subtle even to a human brain.

In general, it appears to be possible for this simple network to be trained to follow almost any if...then... rule with almost any combination of conjunctions (with the possible exception of XOR), negations, or existential imperatives in the subclauses.

5.3 Consciousness

The first and most obvious problem in discussing consciousness lies in having some notion of what it is one intends to discuss. Requests to colleagues for characterizations of consciousness lead to long lists of inconclusive adjectives. Discussions of the subject come to cluster around two notions, “self-awareness” and, less commonly, “imagination” - whatever these are.

“Self-awareness” seems to indicate something self-referential, and physiologists rather commonly speak in some way of the brain looking at itself. For example, Crick (1979) writes, “There is more than a suspicion that such phenomena result from the computation pathways acting in some way on themselves.…” Allusions to imagination are less frequent, but Gregory (1977) seems to allude to something like imagination when he remarks that “Perceptual reality - consciousness - may be a feature of the simulated world or the brain on which we act predictively.”

Consider the following tentative definition of consciousness: “Consciousness” is the ability of an information processing system to (1) use information about how the system is processing information adaptively (self-awareness), and (2) provide parts of the system that are capable of processing information about events outside the system with information about events that might occur (imagination).

Compare this definition with the behavior of the network during the simple conditioned learning described in Subsection 4.3. During recall, set *x* becomes activated progressively in the sensory and motor regions. When it is active in both regions the representation of this activity in the integrative region becomes active. This activity in the integrative region contains information about how the network is processing information. Specifically it contains the information that set *x* has become active in the sensory region and information has been processed so that set *x* is active in the motor region, and the corresponding response is about to be made in the 0th motor region. This information is acted on adaptively in a process that involves recall in both the sensory and reinforcement regions of training sets that represent events that might occur.

The network may be further used to model how conscious actions become unconscious with practice. The process is the same as that described in Subsection 4.7. Consolidation of the memory directly associates the sensory event having a punished unconditioned response to the rewarded motor event, thus removing the need for representations in the integrative network of sensory-motor states that have lead to punishment. The difference between consolidation and conversion of practiced actions into unconscious responses lies in whether the process is the result of some internal mechanism for rehearsal, or of actual repeated execution of the actions.

A mind-brain dualist reading this paper might well say, “Yes, there are correlates of the mind in the brain that appear to be discoverable, but when I look at these networks I do not see in them what I experience as consciousness.”

A reductionist might reply, “Examining these networks is not the same as being them. Humans cannot experience what it is to be a network by examining it. Indeed, such experiences are probably among the many thoughts of which humans *a priori* are not capable. You cannot demand of reductionism that looking at a network will make you feel what it is to be the network.”

At another time, a reductionist might be able to claim that the network has all the known properties readily attributable to consciousness by dualists or reductionists. The problem for the reductionist lies in words like “known.” There may always be, there may have to be, properties of consciousness that remain undescribed. So reductionists cannot really understand whether they have “really” understood consciousness. By the same token, however, dualists should not demand that an observed network “feel like consciousness.” On the account given here, which involves a conjecture that the limits of given sorts of thoughts may be indefinable to thinkers of such thoughts, the issue lying between the reductionist and the dualist will never be resolved. Put as an antinomy in this form, it would be a waste of time to pursue the definitive resolution.

Aside from consciousness, it could be argued that any aspect of the “psychological complexity” of this simple network is the result of naming computational features of the network with complex names. As the properties of networks (like the behaviors of other individuals) are observed more or less empirically while such entities as our own emotions are observed introspectively, the antinomy just described pervades every attempt to integrate the psychology of the mind into the rest of science. Because looking at the “memory” exhibited by these networks does not feel like my memory, there may always be unknown properties of memory that prevent any reductionist understanding of memory.

However, granting every objection to the view that the network described here harbors a nanospark of consciousness, if there are correlates of mind in the activities of the brain, then finding properties of the network that behave like correlates of consciousness still argues for the point of view that some magnitude of consciousness may be harbored by small, simple brains. The famous test for consciousness, perhaps not quite accurately attributed to Turing, in which an observer communicating with a Teletype machine is unable to distinguish human from machine, is both unnecessary and insufficient. Consciousness is not so much a matter of what a black box can do, but of what mechanisms of self-awareness and imagination adapt the black box to its environment.

6 REFERENCES

Bennett, M.R., Gibson, W.G., & Robinson, J. (1994) Dynamics of the pyramidal neuron autoassociative network in the hippocampus. *Philosophical Transactions of the Royal Society of London,* B 343: 167-187.

Buckmaster, P.S. & Schwartzkroin, P.A. (1995) Interneurons and inhibition in the dentate gyrus of the rat in vivo. *Journal of Neuroscience* 15: 774-789.

Buzsáki, G. & Eidelberg, E. (1982) Direct afferent excitation and long-term potentiation of hippocampal interneurons. *Journal of Neurophysiology* 48: 597-607.

Buzsáki, G., Horvath, Z., Urioste, R., Hetke, J., & Wise, K. (1992) High-frequency network oscillation in the hippocampus. *Science* 256: 1025 -1030.

Collin C., Devane W.A., Dahl D., Lee C-J., Axelrod J., Alkon D.L. (1995) Long-term synaptic transformation of hippocampal CA1 gammaminobutyric acid synapses and the effect of anandamide. Proceedings of the National Academy of Science USA 92: 10167-10171.

Conors, B.W., Malenka, R.C. & Silva, L.R. (1988) Two inhibitory postsynaptic potentials, and $GABA_A$ and $GABA_B$ receptor-mediated responses in neocortex of rat. *Journal of Physiology London* 406: 443-468.

Crick, H.F.C. (1979) Thinking about the brain. *Scientific American* Sept.: 130-137.

DeFelipe, J. (1997) Types of neurons, synaptic connections and chemical characteristics of cells immunoreactive for calbindin-D28K, parvalbumin and calretinin in the neocortex. *Journal of Chemical Neuroanatomy* 14: 1-19.

Freund, T.F. & Buzsáki, G. (1996) Interneurons of the Hippocampus. *Hippocampus* 6: 347-470.

Fukai, T. (1999) Sequence generation in arbitrary temporal patterns from theta-nested gamma oscillations: a model of the basal ganglia - thalamo-cortical loops. *Neural Networks* 12: 975-987.

Gibson, W.G. & Robinson, J. (1992) Statistical analysis of the dynamics of a sparse associative memory. *Neural Networks* 5: 645-661.

Gregory, R.L. (1977) Consciousness. In: *The Encyclopedia of Ignorance*, ed. R. Duncan & M. Weston-Smith, Pergamon Press Inc.

Jani, N.G. & Levine, D.S. (2000) A neural network theory of proportional analogy-making. *Neural Networks* 13: 149-183.

Körner, E., Gewaltig, M.-O., Körner, U., Richter, A., & Rodemann, T. (1999) A model of computation in neocortical architecture. *Neural Networks* 12: 989-1005.

Malinow, R. (1991) Transmission between pairs of hippocampal slice neurons: quantal levels, oscillations, and LTP. *Science* 252: 722-724.

Manisha, A.S., McBain, C.J., Jauer, J.A., & Conn, P.J. (1994) Metabotropic glutamate receptor-induced disinhibition is mediated by reduced transmission at excitatory synapses onto interneurons and inhibitory synapses onto pyramidal cells. *Neuroscience Letters* 181: 78-82.

Marr, D. (1971) Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London*, B 262: 23-81.

Miles, R. (1990) Synaptic excitation of inhibitory cells by single CA3 hippocampal pyramidal cells of the guinea-pig in vitro. *Journal of Physiology* (London) 428: 61-77.

Miles, R., Katalin, T., Gulyás, A.I., Hajos, N., & Freund, T.F. (1996) Differences between somatic and dendritic inhibition in the hippocampus. *Neuron* 16: 815-823.

Ribak, C.E., Seress, L., & Leranth, C. (1993) Electron microscopic immunocytochemical study of the distribution of parvalbumin-containing neurons and axon terminals in the primate dentate gyrus and Ammon's Horn. *Journal of Comparative Neurology* 327: 298-321.

Salum, C., Morato, A.C., & Roque-da-Silva, A.C. (2000) Anxiety-like behaviour in rats: a computational model. *Neural Networks* 13: 21-29.

Seress, L., Gulyás, A.I., Ferrer, I., Tunon, T., Soriano, E., & Freund, T. (1993) Distribution, morphological features, and synaptic connections of parvalbumin- and calbindin D_{28k}-Immunoreactive neurons in the human hippocampal formation. *Journal of Comparative Neurology* 337: 208-230.

Sik, A., Tamamaki, N., & Freund, T.F. (1993) Complete axon arborization of a single CA3 pyramidal cell in the rat hippocampus, and its relationship with postsynaptic parvalbumin-containing interneurons. *European Journal of Neuroscience* 5: 1719-1728.

Staley, K.J., Brandi, L.S., & Proctor, P.W. (1995) Ionic mechanisms of neuronal excitation by inhibitory $GABA_A$ receptors. *Science* 269: 977-981.

Tanaka, S. (1999) Architecture and dynamics of the primate prefrontal cortical circuit for spatial working memory. *Neural Networks* 12: 1007-1020.

Taylor, N.R. & Taylor, J.G. (2000) Hard-wired models of working memory and temporal sequence storage and generation. *Neural Networks* 13: 201-224.

Vogel, D. (1995) Burst neurons and lateral inhibition produce serial memory in a projective network. *World Congress on Neural Networks '95,* I 462-465.

Vogel, D. (1998) Auto-associative memory produced by disinhibition in a sparsely connected network. *Neural Networks* 11: 897-908.

Vogel, D. (2000) A biologically plausible model of associative memory which uses disinhibition rather than long term potentiation. *Brain and Cognition* 45: 212-228.

Vogel, D. & Boos, W. (1997) Sparsely connected, Hebbian networks with strikingly large storage capacities. *Neural Networks* 10: 671-682.

Xie, C.W. & Lewis, D.V. (1995) Endogenous opioids regulate long-term potentiation of synaptic inhibition in the dentate gyrus of rat hippocampus. *Journal of Neuroscience* 15: 3788-3794.

Xie, Z., Yip, S., Morishita, W. & Sastry, B.R. (1995) Tetanus-induced potentiation of inhibitory postsynaptic potentials in hippocampal CA1 neurons. *Canadian Journal of Physiology and Pharmacology* 73: 1706-1713.