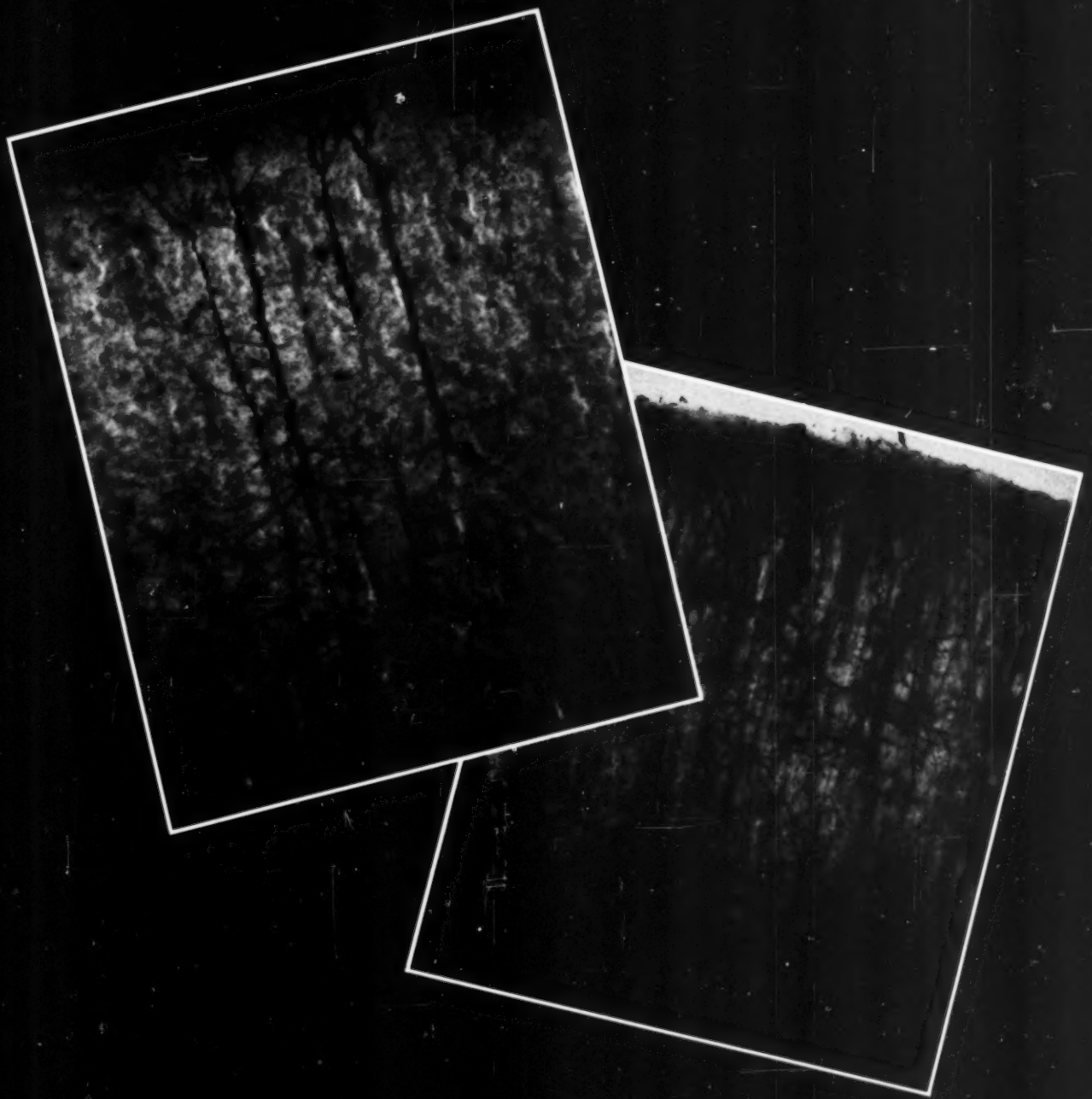
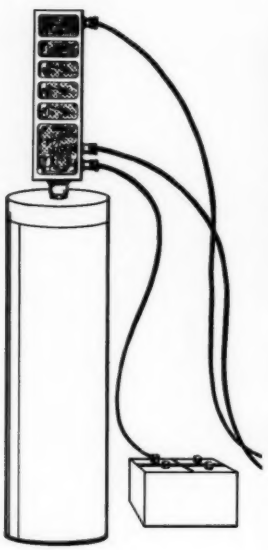
Naval Research Reviews

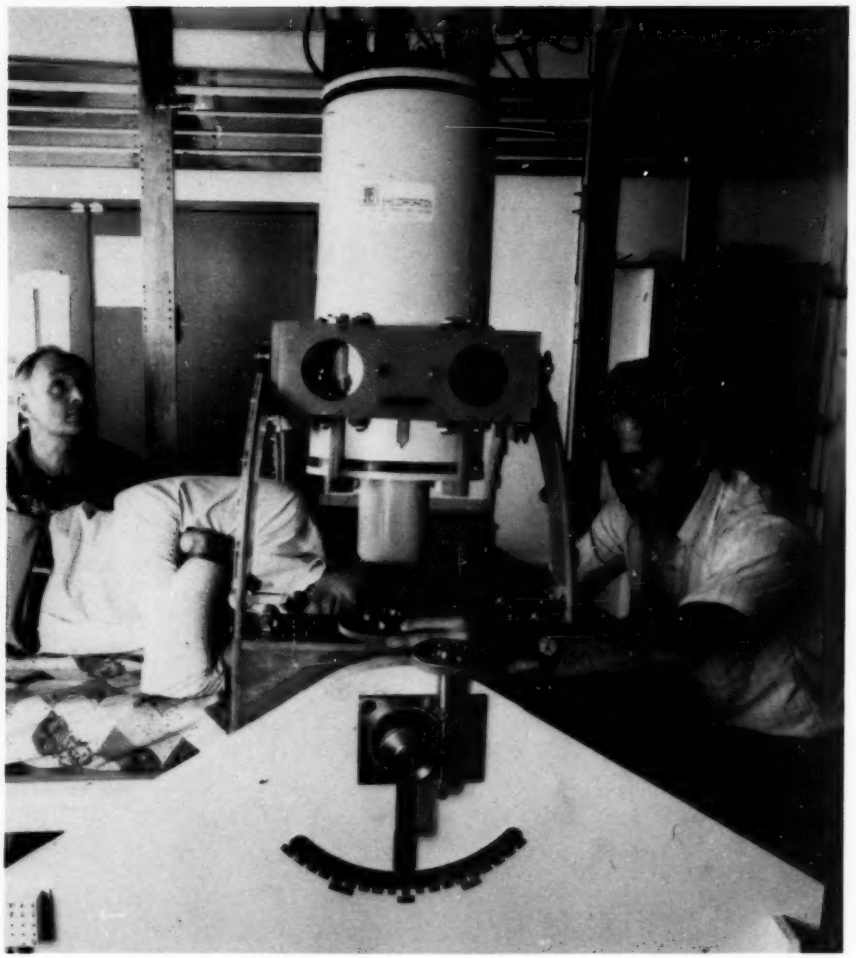
Office of Naval Research
Four / 1985
Vol XXXVII



Noninvasive Detection of the Brain's Magnetic Fields



One SQUID Device



FREDDY

The Office of Naval Research has been supporting superconducting research for measuring magnetic fields since the mid-1950's. The early research was directed towards devices and techniques for possible submarine detection. Later, this research was applied to detecting the magnetic fields eminating from the human body. The first measurements made in 1969 by ONR supported scientists were of the magnetic fields associated with the human heart; an event which opened a new era for biomagnetic research. Over the years, ONR support has led the way in developing supersensitive devices for measuring the magnetic fields of the brain. Professor S. L. Williamson and L. Kaufman of New York University detected magnetic field activity in the brain with a one SQUID (Superconducting Quantum Interference Device) gradiometer in an unshielded room in 1974. Recently, Williamson and Kaufman developed FREDDY, a multisensor SQUID gradiometer. FREDDY can map the circuitry of the brain and may produce measurements which may lead in the future to an understanding of how brain cells communicate.

Naval Research Reviews

Office of Naval Research
Four / 1985
Vol XXXVII

Articles

Departments

About the Cover

Shown on the front cover are photomicrographs (X500) of two regions of the neocortex from an immature rat after treating the tissue with an immunochemical technique that labels a protein named "fodrin." Neocortex is thought by most neuroscientists to be the locus of much of the brain's memory system and according to a recent hypothesis (see page 22), fodrin is a critical element in the chemistry responsible for storage. As can be seen in the pictures, the protein is densely concentrated throughout the dendritic tree that emanates from the pyramid-shaped cell bodies of the cortical neurons; fodrin is postulated to stabilize the surface chemistry and micro-anatomical organization of the many branches that form the tree. The neurons also contain an enzyme that breaks up the fodrin molecules in specific regions of the dendrites; and thereby irreversibly changes the connections between the dendrite and its inputs from other neurons. This produces a stable alteration in the level of communication between the affected cells. To the extent this hypothesis is correct, these photomicrographs provide a view of one portion of the machinery used by brain to encode information.

(Prepared by Drs. Christine Gall, Gwen Ivy, and Robert Siman at the University of California, Irvine)

Naval Research Reviews publishes articles about research conducted by the laboratories and contractors of the Office of Naval Research and describes important naval experimental activities. Manuscripts submitted for publication, correspondence concerning prospective articles, and changes of address, should be directed to Code OPARI, Office of Naval Research, Arlington, Va. 22217–5000. Requests for subscriptions should be directed to the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. *Naval Research Reviews* is published from appropriated funds by authority of the Office of Naval Research in accordance with Navy Publications and Printing Regulations NAVSO P-35.

Introduction

A great deal of structural and biochemical information about the nervous system has accumulated within the last twenty-five years. Although all the details of the neuroanatomic and neurophysiologic mechanisms have not been specified, many scientists believe that our understanding of information flow through neural circuits is advanced enough to begin a serious attempt to explain complex cognitive processes such as learning and memory in biological terms. We are at the point were realistic and powerful models of cellular "information processing" networks underlying learning and memory can be developed. It is one of the underlying assumptions of the Physiology/Neurobiology Group at the Office of Naval Research (ONR) that increased understanding of the "original" information processing system, the brain, will lead to advances in the design of information processing systems. Entirely new approaches to large scale information storage, retrieval and decision making systems may well be the result.

The three articles in this issue of *Naval Research Reviews* have been written by contractors representing programs in the current ONR Learning and Memory Program. The program was begun in 1982 and will continue through 1987. Although acknowledged to be at the forefront in their respective fields, the investigators who participate in this program have all made a commitment to broaden their experimental expertise and to begin to investigate learning and memory with a more multidisciplinary approach.

Dr. Richard Thompson's contribution deals with the identification, localization, and analysis of essential memory trace circuits in the mammalian brain. He has taken a very discrete, highly reproducible type of behavior (that has been described in most mammals, including humans) and traced the brain circuits that acquire and store the information needed to reproduce the behavior. Dr. Thompson's work is primarily empirical but is strongly influenced by theoretical computational modelling. His article reaffirms the importance of the cerebellum in the learning and memory process and suggests that memory storage may be truly localized.

Dr. James McGaugh leads a group that has approached the problem of understanding the neurobiology of learning and memory in different yet complementary ways. Dr. McGaugh's group has long been a leading force in investigating the chemistry of the brain and peripheral hormonal systems which facilitate and impair memory storage. Dr. Gary Lynch's work has centered around the identification of an enzymatic process that may be the critical step in forming certain types of memory. Dr. Norman Weinberger's interests include how specific cells in the auditory system exhibit learning-like behavior prior to the appearance of behavioral manifestations of memory. Lastly, Dr. Richard Granger's group has developed an innovative and powerful process model of learning and memory as the basis for artificial intelligence approaches to the field. These four authors demonstrate that these issues are highly interrelated, and that findings at one level of analysis are necessary for advances at other levels.

Dr. Stephen Kosslyn's contribution to this issue is representative of a more cognitive approach to learning and memory, one that Dr. Kosslyn describes as "computational neuropsychology." Much of this work is directed at a general theory of high level vision. This theory consists of those aspects of visual processing where one's knowledge and beliefs are utilized in carrying out information processing. This may include aspects of recognition, navigation, tracking, and mental imagery. Some of the empirical data for this work are derived from "split-brain" human subjects whose unique lateralized condition allows the study of human visual representation in two separate cerebral hemispheres. Dr. Kosslyn's theory of high level vision makes use of a hybrid functional architecture, one component consisting of "connectionist" parallel processing networks and one component consisting of a more standard von Neumann-like architecture. This work, like that of the other contributors to this issue, may have profound implications for a new generation of artificial-intelligence devices.

The human brain and its functional abilities represent the ultimate in evolutionary processes that began millions of years ago. Understanding how learning and memory takes place is one of the most important basic questions still to be answered in the life sciences. I hope this issue of Naval Research Reviews conveys some of the current excitiment in this field.

JOEL L. DAVIS
Program Manager
Physiology/Neurobiology Group
Biological Sciences Division, ONR

LOCALIZATION OF MEMORY TRACES IN THE MAMMALIAN BRAIN

Richard F. Thompson,
Stanford University

Perhaps the ultimate goal of research in the area of brain function and behavior is an understanding of the human mind. A key aspect of the mind is the nature of our thoughts or cognitions—how are they assembled and maintained? Simply put—how do we learn and how do we remember? Without memory, there can be no mind.

How it is that the brain codes, stores and retrives memories is among the most important and baffling questions in science. At the cellular level, there are two fundamental types of information coding. One of these is the familiar genetic code, shared by organisms from virus to man. In higher organisms, literally millions of bits of information are coded in the DNA of the cell nucleus. Over the course of evolution, a quite different kind of information coding has developed—the cellular encoding of acquired information in the brain. This coding is no less remarkable than the genetic code. It has been estimated that a well-educated adult human has millions of bits of acquired information stored in the brain.

The generic term for acquired information coding in the brain is the "memory trace" or "engram." The fundamental difference between the genetic code and the engram code is, of course, that each individual human's engram store is acquired through experience and learning. It is the biological substrate for the growth of knowledge and civilization. The individual uniqueness of each human being is due largely to the engram store—the biological residue of memory from a lifetime of experience.

The development, elaboration and growth in size of the human brain over the past three million years is unprecedented in evolution. Indeed, some biologists now feel that a new principle or mechanism of evolution may have come into play, a mechanism that involved the rapidly growing ability to learn, with its consequent development of culture and transmitted knowledge. The learning that occurred in each individual created a more complex environment that required still better learning, and hence a larger brain to survive and be fit. The cellular roots of this ability to learn can be traced to simpler organisms. It is best developed in mammals, where basic phenomena of learning, like Pavlovian and instrumental conditioning, exhibit the same fundamental properties from rat and rabbit to human. The complexity of information processing and learning increases in an orderly fashion over the class of mammals.

The ability to learn is an emergent property of cellular tissue. The ability, as such, has a clear genetic base; it is dependent on the structural and functional organization of the brain and on the elaboration of cellular storage processes. It would not be entirely surprising if the genetic material itself plays a role in the process of learning.

In the past generation, understanding of the biological basis of learning and memory has undergone a revolution. The earlier notion that memory is diffusely distributed in the brain has been largely discounted. It is now clear that various forms and aspects of learning and memory involve particular systems and circuits in the brain. The realization means that it is now both theoretically and technically possible to define these circuits, localize the sites of memory storage and analyze the cellular and molecular mechanisms of memory.

The roots of this new understanding lie in several different disciplines. From psychology has come a clear characterization of the behavioral properties of learning—what learning and memory are—and a developing theoretical analysis of the nature of the associative processes that form the basis of learning and memory. The work of such psychologists as Robert Rescorla at the University of Pennsylvania and Alan Wagner at Yale University provide examples.[1] The term "associative", incidentally, describes the basic property of learning, the fact that stimulus or event becomes associated with another stimulus, event or response, as in learning a foreign language.

From the broad field of psychobiology or behavioral neuroscience has come the recognition that identifiable neural memory systems and circuits in the brain can be characterized and analyzed. One example of this is our own current work, to be described later, on the neuronal circuits that code basic associative learning in the mammalian brain. Another example is current work where the human disorder of amnesia is being modelled in the primate brain by such scientists as Mortimer Mishkin at the National Institute of Mental Health[2] and Larry Squire at the University of California at San Diego.[3] Still another example concerns how chemical and hormonal processes can modulate the storage and retrieval of memories, as shown in the work of such scientists as James McGaugh at the University of California at Irvine[4] and Stephen Zornetzer[5] and Joel Davis[6] at the Office of Naval Research. From the field of Neurobiology, we are learning about the cellular, biophysical and molecular mechanisms that underlie elementary forms of associative learning in simple neural circuits in invertebrates in the work of such scientists as Eric Kandel at Columbia University,[7] and Daniel Alkon at Woods Hole.[8]

The success of this collective approach has been the source of great optimism in the field; fundamental insights into the physical basis of memory will continue to be achieved over the next few years. Indeed, many of us feel that the field is now in the critical "breakthrough" stage—the present would seem to be the most exciting phase in the history of the field.

The Search For The Memory Trace

The idea that memory traces can be localized to particular circuits in the brain is not new. Karl Lashley began the search for the memory trace early in the century. He trained rats to learn mazes and then removed portions of the "highest" region of the brain, the cerebral cortex. But he was unable to find any one region of the cortex that seemed to be particularly involved in the memories for the maze habit.

Because of his (and others) inability to find localized memory traces in the cerebral cortex, Lashley drew the following rather pessimistic conclusion at the end of his career:

> "This series of experiments has yielded a good bit of information about what and where the memory trace is not. It has discovered nothing directly of the real nature of the engram. I sometimes feel, in reviewing the evidence on the localization of the memory trace, that the necessary conclusion is that learning just is not possible. It is difficult to conceive of a mechanism which can satisfy the conditions set for it. Nevertheless, in spite of such evidence against it, learning does sometimes occur"[9]

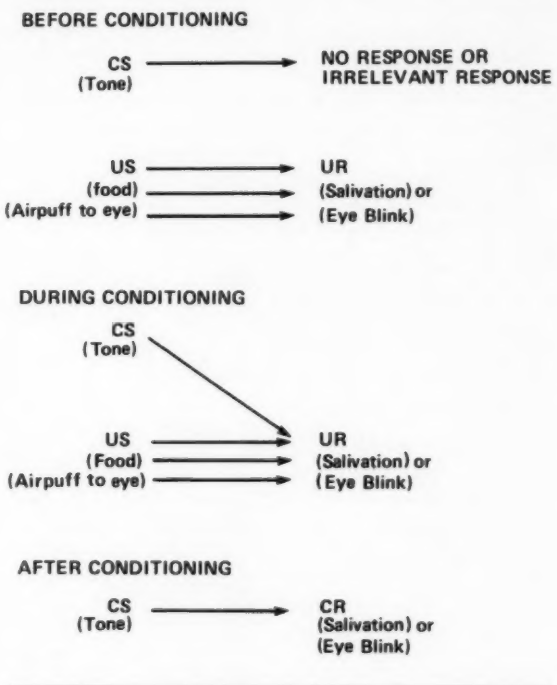
With the advantage of hindsight it now appears that certain brain structures below the level of the cerebral cortex play a more important role in maze learning in the rat. One such structure is the hippocampus, a part of the so-called limbic forebrain system that also seems to be involved in amnesia in humans.

As Lashley early recognized, the overriding problem for analysis of memory traces in the mammalian brain has been that of localization. In order to analyze neuronal/synaptic mechanisms of memory storage and retrieval, it is first necessary to identify and localize the brain circuits, structures and regions that are critically involved. The general approach to localization of memory traces that has proved most productive is to use what are termed "model biological systems"—to utilize a form of basic associative learning in an animal preparation where the memory circuits can be identified and analyzed.

Ivan Pavlov was perhaps the first to develop and use the model biological system approach to learning and memory. From the time he discovered the conditioned reflex, he saw it as a tool to investigate higher functions of the brain. Lashley, influenced by Pavlov, Bechterev and Watson, was the first Western scientist to state explicitly the model system approach, namely to use simply conditioned reflexes as models of associative learning and to localize memory traces by tracking the essential conditioned response circuitry through the brain.

Figure 1:

Diagram of classical conditioning. Initially, the conditioned stimulus (CS) does not elicit the response to be conditioned (unconditioned response, UR). On the other hand, the unconditioned stimulus (US) or "reinforcement" does elicit the unconditioned response. If the conditioned stimulus (CS) is then paired with the unconditioned stimulus (US) such that the CS precedes the US by an appropriate time interval, then a predictive association is formed between the two and the CS comes to elicit the conditioned response (CR). If the CS is then presented repeatedly without the US the CR is gradually forgotten or extinguished.



Pavlovian or classical conditioning has come to be viewed as the basic process of associative learning. All readers are probably familiar with Pavlov's discovery of the conditioned reflex—if a tone is sounded just before food is given to a dog, the dog learns to salivate to the tone—the tone signals or predicts the presentation of food. It is a basic way that all animals, including humans, learn about predictability and cause and effect in the world. The tone is termed a conditioned stimulus (CS)—initially it does not elicit salivation. But when it has been paired with food, a biologically meaningful simulus that elicits a response directly, the unconditioned salivation response (UR), then the tone comes to elicit salivation, the conditioned response (CR). The food is termed the unconditioned stimulus (US)—it directly elicits the salivary response. This is schematized in Figure 1. In this and many other examples of Pavlovian conditioning, the learned response of salivation (or eyeblink) to the tone CS is very similar to the unlearned response of salivation to the food US (or eyeblink to the corneal airpuff US).

In the Western world, discrete behavioral responses like leg flexion or eyelid closure have been more widely used than autonomic responses like salivation to study the properties of associative learning. But they all show the same basic properties and phenomena of Pavlovian learning in all mammals, including humans, and in invertebrate preparations as well. The major breakthroughs that have occurred in the past few years in terms of identifying critical memory trace circuits and processes in both invertebrates and vertebrates have used Pavlovian conditioning procedures. These basic procedures, completely under the control of the experimenter, are an important part of the model system approach.

Some years ago, we adopted a particularly clear cut and robust form of associative learning in the intact mammal as a model system: Pavlovian conditioning of the rabbit eyelid response to an acoustic or visual CS using a corneal airpuff US. Thus, a tone CS is sounded and about a quarter of a second later a puff of air is delivered to the cornea of the eye. Initially, the tone does not evoke any response of the eye and the airpuff US evokes eyelid closure, a simple defensive reflex of the eye. Be repeating such pairings the tone comes to elicit a full blown closure of the eye, a conditioned response, that closely resembles the reflex eyeblink initially elicited by the airpuff. It is a highly adaptive learned response. Over a wide range of time intervals from tone onset to airpuff onset, e.g., from 1/10 second to more than a second, the animal (and human) learns to close the eye maximally at the time the airpuff to the cornea occurrs, thus providing maximal protection of the eye.

Classical conditioning of the eyelid response was first done in humans. Ernest Hilgard, now emeritus professor of psychology at Stanford, was the first to study eyelid conditioning in infrahuman animals in his classical work with Donald Marquis on dogs and monkeys.[10] He and Marquis, then both at Yale, recognized immediately that eyelid conditioning provided an excellent model system for analysis of brain substrates of memory and they undertook a series of lesion studies with the assistance of John Fulton. They used a visual CS and showed that the visual cortex was not essential. Their results pointed to a subcortical memory trace. Isadore Gormezano, then at Indiana, was the first to publish eyelid conditioning studies in the rabbit, an animal that is docile and tolerates restraint well.[11] The conditioned eyelid response in the rabbit has been used to very good effect in analysis of basic theoretical issues in learning.[12]

Classical conditioning of the rabbit eyelid response has a number of advantages for analysis of brain substrates of learning and memory, which have been detailed elsewhere.[13] A particular advantage of the conditioned eyelid response is the fact that eyelid conditioning has become perhaps the most widely used paradigm for the study of basic properties of classical or Pavlovian conditioning of striated muscle responses in both humans and infrahuman subjects. It displays the same basic laws of learning in humans and other animals. Consequently, it seems highly likely that neuronal mechanisms found to underlie conditioning of the eyelid response in rabbits will hold for all mammals, including humans. We view the conditioned eyelid response as an instance of the general class of discrete, adaptive behavioral responses learned to deal with aversive stimuli and adopt the working assumption that neuronal mechanisms underlying associative learning of the eyelid response will in fact be general for all such learning.

In science, it is generally easier to rule out possibilities rather than rule them in, as Sherlock Holmes was fond of pointing out. We completed a long series of studies ruling out various brain structures and circuits as a part of the essential memory trace circuit or locus of the essential memory trace itself. In brief, our results and work from other laboratories argued against essential participation of such higher brain structures as the cerebral cortex and hippocampus. Interestingly, memory traces do develop in the hippocampus as a result of such basic associative learning,[14] but this hippocampal memory trace circuit is not essential for learning. The development of such higher order memory systems in basic associative learning, incidentally, may provide simplified models for the study of neuronal substrates of the more complex or cognitive functions of higher regions of the brain.

We also ruled out certain circuits in the brain stem as the locus of the memory trace, e.g., the primary auditory nuclei that relay information from the ear about the tone CS to other brain structures and the motor nuclei that generate the behavioral response, although portions of them are of course a part of the memory trace circuit. But we were still left with much of the brain stem, midbrain, and cerebellum as possible sites. Since there was no *a priori* way of determining which of these regions and structures are involved in the memory trace, we undertook, beginning some years ago, to map these structures by systematically recording neuronal unit activity in already trained animals.[15] For this purpose we developed a chronic micromanipulator system that permits recording of nerve cell activity in a substantial number of neural loci per animal. Increases in unit activity that form a temporal model within a trial of the learned behavioral response were prominent in certain highly localized regions of the cerebellum, in the cortex and in the interpositus nucleus. The results of the mapping studies pointed to substantial engagement of the cerebellar system in the generation of the conditioned response. An example is shown in Figure 2 with unit recordings from the cerebellar interpositus nucleus on the same side as the trained eye. This animal was given unpaired training before acquisition began. Average histograms reveal that the unit activity showed only minimal responses to the tone and airpuff during the unpaired day of training. However, during acquisition, as the animal learned, the unit activity developed a "model" of the conditioned response but no clear model of the unconditioned response. The cerebellar neuronal unit model of the learned response precedes the behavioral response significantly in time and predicts the actual amplitude-time course of the learned behavioral CR. The course of development of the conditioned behavioral response and the concomitant growth in the neuronal unit "model" of the conditioned response in the interpositus nuclear region show very high correlations (e.g., $r = .90$).[16]

The neural unit recordings displayed in Figure 2 are from small groups or clusters of neurons, so-called multiple unit recording. In current work in which we record the action potentials from single neurons, we find two patterns of activity in neurons that respond in relation to the behavioral learned response: (1) those that show increases in frequency of discharge that precede and predict the form of the learned behavioral response, as in Figure 2, and (2) those that show mirror images of such responding (inhibition) but are equally predictive of the learned response. Neurons showing these positive or negative models of the learned response are found in very localized regions of the interpositus nucleus and in very localized regions of the cerebellar cortex, where the neurons have been identified as Purkinje cells, the only neurons that send information out from the cerebellar cortex, in this case to the interpositus nucleus.

Figure 2:

Histograms of unit recordings obtained from a chronic electrode implanted at the lateral border of the interpositus nucleus. The animal was first given random, unpaired presentations of the tone and airpuff (104 trials of each stimulus) and then trained with two days of paired training (117 trials each day). Each histogram is an average over the entire day of training indicated. The upper trace represents movement of the NM. The first vertical line represents the onset of the tone CS, while the second line represents the onset of the corneal airpuff US. Each histogram bar is 9 milliseconds in duration. Notice that these neurons develop a model of the conditioned but not unconditioned response during learning, and that this neuronal model precedes the learned behavioral response substantially in time (see reference 16).

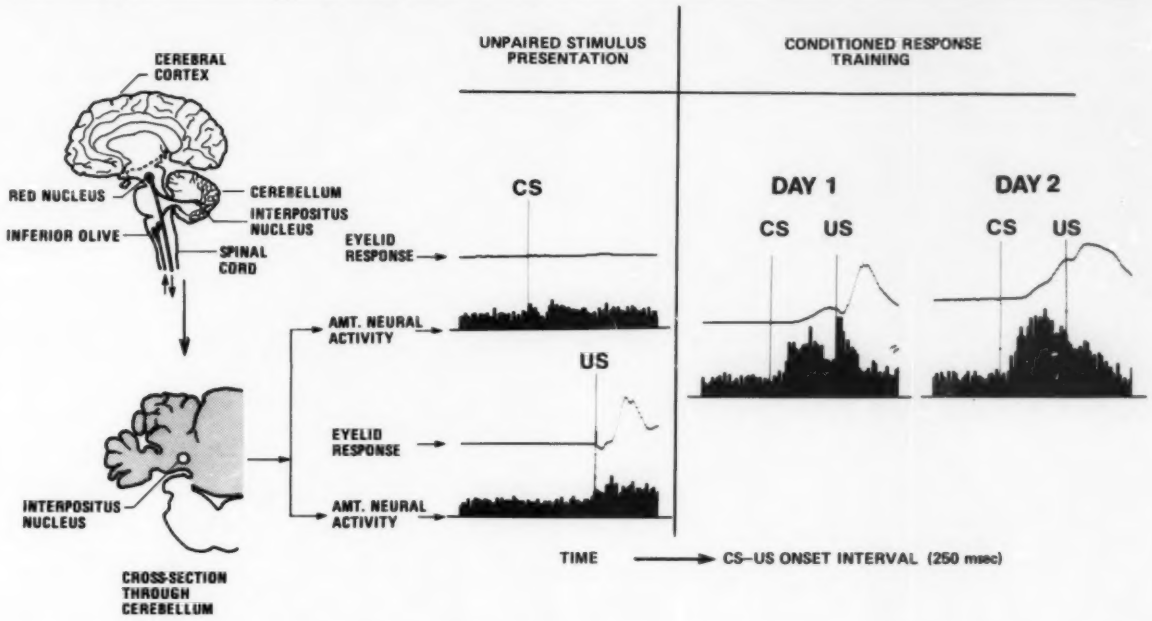
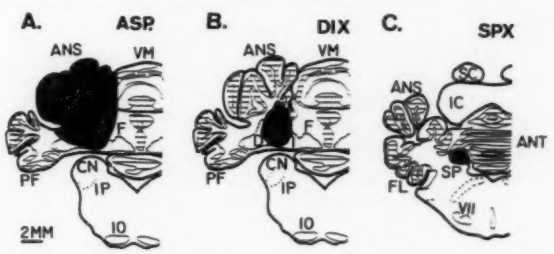


Figure 3:

Examples of cerebellar lesions that abolish the conditioned eyelid response: A, large lesion of both cortex and interpositus nucleus; B, smaller lesion of interpositus; C, lesion of the output pathway from the interpositus. Lesions limited to the same side of the cerebellum as the trained eye are sufficient to abolish the learned response (and prevent learning if made before training). Drawings are cross sections through the cerebellum and labels refer to various structures of the cerebellum and brain stem.



In current work, we have found that lesions on the same side as the trained eye in the cerebellum (Figures 3 and 4) permanently abolish the CR but have no effect on the UR and do not prevent subsequent learning by the other eye.[17] If the lesion is made before training, no learning occurs. The critical region is the lateral interpositus nucleus. Perhaps, most dramatic is a recent study involving kainic acid (it destroys nerve cell bodies but not nerve fibers or terminals); destruction of neuron cell bodies in a region as small as a cubic millimeter in the lateral portion of the interpositus causes complete and permanent loss of the learned eyelid response.[18]

Electrical stimulation through recording microelectrodes in the critical lateral interpositus nuclear region of the cerebellum elicits a discrete eyelid closure response prior to training. In fact, a wide range of discrete behavioral responses—eyelid closure, leg flexion, head turning—can be elicited by microstimulation of the interpositus, the type of response depending upon the exact location of the stimulating electrode.

Figure 4:

Effects of ablation of left lateral cerebellum on the learned NM/eyelid response (six animals). Solid triangles, amplitude of conditioned response (CR); open diamonds, amplitude of unconditioned response (UCR). All training was to left eye (ipsilateral to lesion) except where labelled "R" (for right eye). The cerebellar lesion completely and permanently abolished the CR of the ipsilateral eye but had no effect on the UCR. P_1 and P_2 indicate initial learning on the two days prior to the lesion. L_1-L_4 are four days of post-operative training to the left eye. The right eye was then trained and learned rapidly (R), thus controlling for nonspecific lesion effects. The left eye was again trained and showed no learning. Numbers on abscissa indicate 40 trial periods, except for "right eye" which are 24 trial periods (see reference 17).

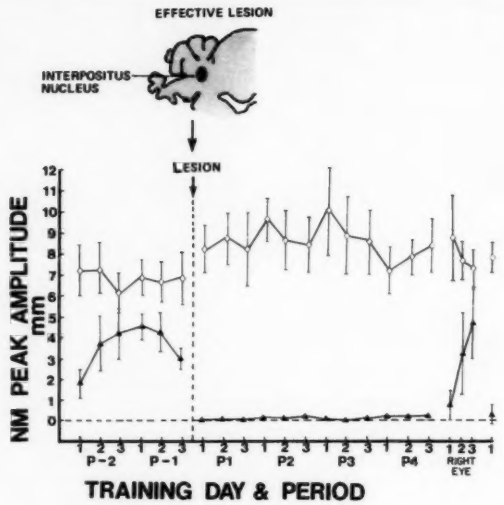
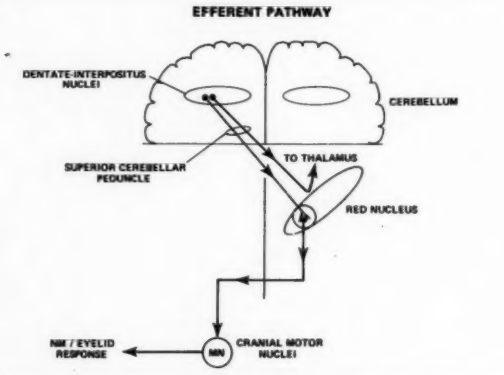


Figure 5:

The efferent (output) pathway from the cerebellar memory trace system to motor nuclei.



These results indicate that the essential memory trace circuits for these learned responses are extremely localized in the brain. How general is this finding? Most of our work has been on the learned eyelid response, but we have shown that a different part of the interpositus is essential for hindlimb flexion conditioning.[19] We and our associated laboratories are now exploring the generality of our results across learning procedures and species. Particularly important are current results in collaboration with Michael Patterson at Ohio University. He developed a procedure for eyelid conditioning in the cat, and has now found that the interpositus nucleus on the same side as the trained eye is essential for the learned response in this species, as well as in rabbit. It would seem that our results may be generalizable to all mammals. In other work, Patterson developed a procedure for "instrumental" avoidance learning of the eyelid closure response in the rabbit, where eyelid closure to tone CS before onset of corneal airpuff results in the airpuff not being delivered. He found that lesions of the ipsilateral interpositus nucleus abolish this instrumental avoidance response, as well as the classically conditioned response. Hence, we feel we can generalize our essential cerebellar circuit to discrete, adaptive responses learned to deal with aversive stimuli in both classical and instrumental avoidance learning paradigms in mammals.

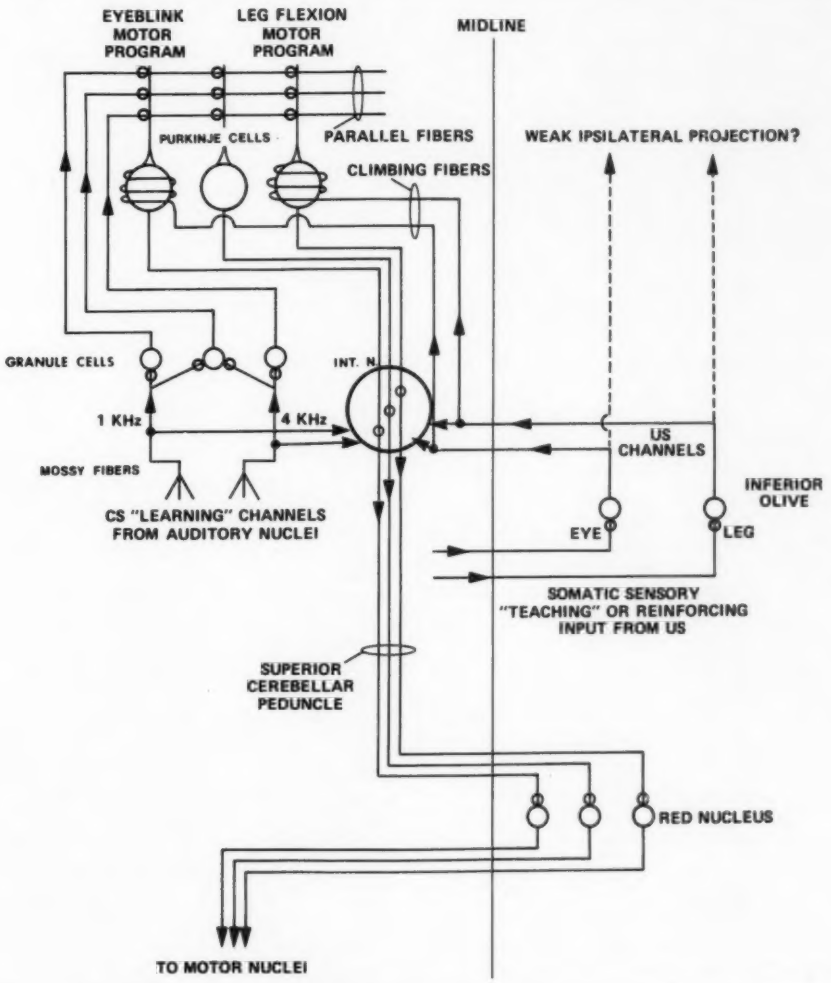
In other work, we have identified much of the efferent or output pathway from interpositus nucleus to motor nuclei (Figure 5). Our results indicate that it courses out the superior cerebellar peduncle, crosses to the contralateral side in the peduncle, relays in the magnocellular division of the red nucleus, crosses back to the ipsilateral side and projects to lower brain stem as a part of the descending rubral pathway.[20] The essential CR circuit we have so far defined could be called the "efferent" limb in that destruction of any part of the circuit abolishes the CR to any CS (e.g., light or tone), but the association between CS and US/UR appears to be formed in the cerebellum, as will be seen below.

A Hypothetical Model

In late 1983, we developed a hypothetical schema or model of how the cerebellar memory trace circuit might work, based on the well-known anatomy of the system, on our results described above, and on theories of how the cerebellum might function as a learning machine (Figure 6—see Figure 8 for a more realistic schematic of the cerebellar circuitry). The cerebellar cortex has great appeal to theoreticians and modellers because of its elegant uniformity and simplicity and because of the striking fact of two quite different inputs to each Purkinje cell—a mossy fiber-granule cell-parallel fiber input that is widely distributed and a climbing fiber input that is highly localized. All models, both verbal-logical and computational, have stressed this point, and those that focus on learning and memory have universally hypothesized that the mossy fiber-granule cell-parallel fiber system is the learning input, and the climbing fiber input is the "teaching" input.[21]

Figure 6:

Scheme of hypothetical memory trace system for associative learning of discrete, adaptive, somatic-motor responses to deal with aversive unconditioned stimuli. Most interneurons are ommitted. It is assumed that the site of the memory trace is at the Purkinje cells shown in the upper left under "motor programs" and/or at associated interneurons. The locus of the memory trace is shown as cerebellar cortex and the basic notion is very similar to earlier theories of cerebellar plasticity (Albus, Marr, Eccles, Ito), but traces could also or alternatively be formed in the interpositus nucleus (int. n.), we assume by an analogous circuitry. A given CS (1 KHz) activates a subset of parallel fibers that in turn activate weakly all Purkinje cells shown. A different tone also activates all Purkinje cells but by a partially different group of parallel fibers. The US pathway is assumed to be via the inferior olive and climbing fibers. A given US is assumed to activate only a limited group of Purkinje cells coding the motor program for the defensive response that is specific for the US (eyelid closure, leg flexion). When parallel fiber activation occurs at the appropriate time just prior to climbing fiber activation, the connections of the parallel fibers to the Purkinje cells activated by the particular US are strengthened or weakened. The efferent pathway from Purkinje cells to motor neurons is by way of the interpositus nucleus, superior cerebellar peduncle, and red nucleus. The scheme accounts for stimulus specificity, i.e., the fact that CRs shown a stimulus generalization gradient, for response specificity of learned responses, transfer effects and is consistent with all evidence to date. Although hypothetical, insofar as being the substrate for memory traces, each aspect and assumption is amenable to experimental test.



The details of our schema need not concern us here.[22] In brief, it is assumed that the locus of the memory trace is at the Purkinje cells in cerebellar cortex (and/or in analogous cells in the interpositus nucleus). A given CS (tone, light) is assumed to activate a subset of mossy fibers-granule cells-parallel fibers that in turn activate weakly a number of Purkinje cells. The US (corneal airpuff) is assumed to activate a limited number of climbing fibers from the inferior olive that in turn activate strongly only a few of the Purkinje cells also activated by the CS, the few that result in activation of the appropriate motor program via the interpositus nucleus, e.g., eyelid closure for a corneal airpuff or leg flexion for a paw shock. When parallel fiber activation occurs at the appropriate time just prior to climbing fiber activation, the connections of the parallel fibers to the Purkinje cells activated by the particular US are strengthened. The schema accounts for stimulus specificity, e.g., the fact that CRs show a stimulus generalization gradient—if a different stimulus is used as a test CS the learned response will be weaker or nonexistent, the more different the test CS is from the training CS. It also accounts for response specificity of learned responses and is consistent with all our evidence to date. Although much of this circuit was hypothetical, insofar as it's being a substrate for the formation of memories is concerned, each aspect and assumption is testable. Indeed, current work in our laboratory is providing strong new evidence favoring such a schema.

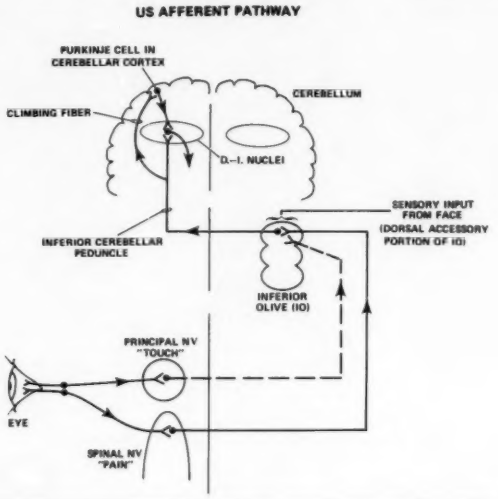
The Inferior Olive–The Necessary and Sufficient Teaching Input for the Learning of Discrete Behavioral Responses

A major system that projects information to the cerebellum is the inferior olive (IO)-climbing fiber system (Figures 6, 7 and 8). In recent work, we have found that lesions of the appropriate region of the IO (rostromedial DAO) do not abolish the CR but instead lead to relatively normal *extinction* with continued paired CS-US training.[23] Lesions of all other regions of the IO do not affect the CR. The dorsal accessory olive (DAO) appears to be the essential afferent limb for the reinforcing or "teaching" input from the US. The fact that lesions of the DAO do not immediately abolish the CR but instead lead to its eventual extinction argues that the essential memory trace cannot be there.

In current work, we find that electrical microstimulation of the DAO (see Figure 8) can elicit a variety of behavioral responses, including eyelid closure, the nature of the threshold response being determined by the exact location of the stimulating electrode. If this is now used as the US/UR and paired with a tone CS, the *exact response* elicited by DAO stimulation is learned to the tone as a CR rapidly and with all the properties of a normal CR.[24] Lesion of the critical interpositus region abolishes this IO-established CR, *and* abolishes the response elicited by IO stimulation. Control stimulation 1-2 mm dorsal to the DAO in the reticular formation can also elicit movements, presumably by activation of descending reticular pathways, but these elicited movements cannot be trained to a CS.

Figure 7:

Our hypothesis of the effective US reinforcing or teaching input circuit from the cornea of the eye to the cerebellum. The critical region of the inferior olive is the dorsal accessory olive.



These IO results strengthen the argument that the IO and its climbing fiber input to the cerebellum is the essential US teaching input and that the trace is localized to the cerebellum. These are also the first clear empirical evidence supporting the pioneering hypothesis and network models of Albus, Eccles, Ito, Marr and others (noted in reference 21) that the IO-climbing fiber system is the "teaching" input for behavioral learning in the cerebellum. (Ito has developed analogous findings in the context of plasticity of the vestibulo-ocular reflex, and Llinás reports a similar role for the inferior olive-climbing fiber system in recovery from postural abnormalities following vestibular damage.[25])

Creation of a Known CS Pathway–Mossy Fiber Projections to Cerebellum

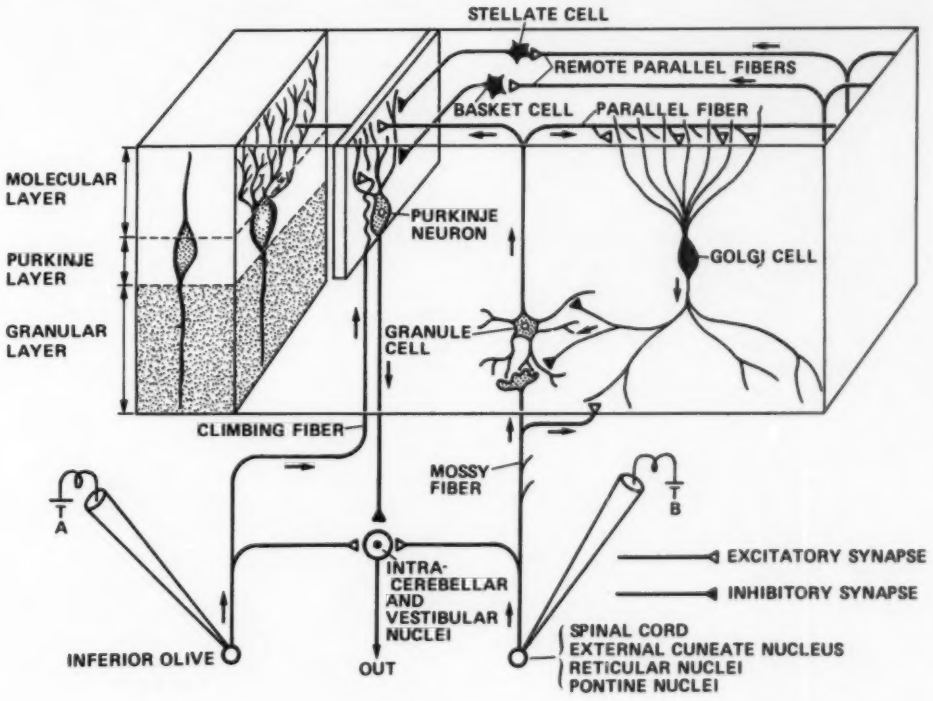
We have succeeded in creating a known CS pathway by using electrical microstimulation of mossy fiber projections to the cerebellum as the CS (see Figure 8). Animals rapidly learn normal behavioral conditioned responses to this CS (e.g., eyeblink CR with corneal airpuff US). To date we have successfully used stimulation of mossy fibers from the dorsolateral pontine nucleus, the lateral reticular nucleus and the mossy fibers themselves, as CSs.[26]

Finally, in current pilot work we find that *normal behavioral CRs* are learned with electrical stimulation of mossy fibers as the CS and of DAO-climbing fibers as the US, thus creating a "reduced" preparation within the intact, behaving animal (Figure 8). This preparation promises much in terms of fine-grained localization of the memory traces and analysis of mechanisms.

Figure 8:

Synaptic organization of the cerebellar cortex. Purkinje cells are excited directly by climbing fibers and indirectly (via parallel fibers from the granule cells) by the mossy fibers. Stellate and basket cells, which are excited by parallel fibers, act as inhibitory interneurons. The Golgi cells act on the granule cells with feedback inhibition (when excited by parallel fibers) and feedforward inhibition (when excited by climbing and mossy fiber collaterals). The output of the Purkinje cell is inhibitory upon the cells of the intracerebellar and vestibular nuclei.

*In normal behavioral training, a tone is used as the CS and a corneal airpuff, that elicits an eyeblink, is used a the US. However, electrical microstimulation of a portion of the inferior olive-climbing fiber projection to the cerebellum (see A in Figure 8) can also elicit an eyeblink response and serves as a very effective US in place of corneal airpuff. Similarly, electrical microstimulation of the mossy fiber projection to the cerebellum (see B in Figure 8) can serve as a very effective CS in place of tone or light. Finally, joint stimulation of climbing fibers as the US and mossy fibers as the CS produces normal learning of discrete behavioral responses. See text for details. (Modified from Kandel, G.R. and Schwartz, J.H., **Principles of Neuroscience**, New York, Elsevier, 1984; Figure 30.3, p. 338.)*



Conclusions

In more general terms, our results have demonstrated quite clearly that the essential memory trace circuits in the brain for the basic category of associative learning we have studied are highly localized; and all our evidence to date points to discrete regions of the cerebellum as the locus of memory storage for discrete, adaptive behavioral responses learned associatively to deal with aversive events.

The effects of cerebellar damage in humans are to impair movements, particularly skilled movements. Most movements that humans make are to a significant degree learned, i.e., skilled. In this context, Eccles has proposed the following:

"We can say that normally our most complex muscle movements are carried out subconsciously and with consummate skill. The more subconscious you are in a golf stroke, the better it is, and the same with tennis, skiing, skating, or any other skill. In all these performances we do not have any appreciation of the complexity of muscle contractions and joint movements. All that we are conscious of is a general directive given by what we may call our voluntary command system. It is my thesis that the cerebellum is concerned in this enormously complex organization and control of movement, and that throughout life, particularly in the earlier years, we are engaged in an incessant teaching program for the cerebellum. As a consequence, it can carry out

all of these remarkable tasks that we set it to do in the whole repertoire of our skilled movements in games, in techniques, in musical performances, in speech, dance, song, and so on)."[27]

One of the most striking features of the cerebellum is the high degree of regularity in the anatomical organization of the cortex over its extent and over species. It is highly and regularly organized; indeed, hard-wired. If a major function of this structure is to code the associative learning of discrete, adaptive movements and more generally skilled movements, then the basic neural circuitry for such learning preexists. No new pathways are formed. Instead, the memory traces must involve changes in the excitability of preexisting circuits. In this sense, at least, our results are closely consistent with studies of invertebrate preparations that exhibit learning, where the circuits are hardwired and the mechanisms appear to involve changes in membrane properties.[28]

The general possibility that learning circuits are hard-wired in the mammalian brain is consistent with all we know about the high degree of anatomical organization of the brain and not inconsistent with all we know about learning and memory. Even the most complex form of human learning and memory, language, appears to have differentiated anatomical substrates in the cerebral cortex and uniformities in the "deep structure" of language itself across all languages.[29] In terms of mechanisms, the possibility that memory circuits are hard-wired does not exclude anatomical substrates for memories in terms of the growth or alteration of the microstructure, i.e., change in the size and properties of synapses or even the development of new synapses.

The work that we have reported here provides a particularly clear example where theoretical computational models derived from such fields as artificial intelligence and cognitive psychology can interact with and guide empirical neurobiological research. When we set out to find the essential US teaching pathway, we had two major choices. Information about the corneal airpuff and other face skin sensations is conducted into the brain by the 5th cranial nerve to the 5th nucleus. From here there are two major ways this information can get to the cerebellum: directly by way of mossy fibers and indirectly by way of the inferior olive and climbing fiber. The computational theories of cerebellar learning argued eloquently for the climbing fiber system as the teaching input because of its unique anatomical organization, one climbing fiber to one Purkinje cell, as opposed to the distributed projections of the mossy fiber—granule cell—parallel fiber system to many Purkinje cells (see above). Following these models, we focused on the climbing fiber system and found that it was indeed the necessary and sufficient teaching input for the learning of discrete, adaptive behavioral responses.

The distributed nature of the mossy fiber-granule cell-parallel fiber system, on the other hand, makes it ideally suited as the "learning" input, a point also stressed by the computational models. Again, following the models, we have shown that activation of mossy fibers is indeed a very effective CS learning pathway.

As we define the essential cerebellar memory trace circuit in more detail, it will permit us to develop more precise and powerful computational models that could well have more general and practical applicability, i.e., perhaps for control and guidance systems and for robotics. In terms of this last possibility, the manner in which our animals learn to make very precise and discrete movements to a "command" (tone CS) when electrical microstimulation of the inferior olive is used as the teaching US input is strikingly machine-like. As noted earlier, animals learn to make the *exact* movement evoked by the electrical stimulus and show no sign of emotion or affect when doing so. The cerebellum does indeed appear to be an extremely efficient biological machine for the learning of precise movements. Perhaps when we understand in sufficient detail how it accomplishes this, we can make use of the same principles in non-biological machines.

In general terms, it seems clear that as the neuronal circuits in the brain are defined that are essential for learning and other aspects of behavior, it will be necessary to develop theoretical computational models of these circuits. Even with very simple circuits, it is not possible to predict with any precision what in fact the circuit does or is capable of doing without quantitative modeling of the circuit.[30]

Biography

The author is the subject of the "Profiles in Science" on page 14.

References

1. Rescorla, R. A. & Wagner, A. R., "A Theory of Pavlovian Conditioning; Variations in the Effectiveness of Reinforcement and Nonreinforcement," A. H. Black & W. A. Prokasy (Eds.), *Classical Conditioning II: Current Theory and Research*, (1972) New York, Appelton-Century-Crofts.

2. Mishkin, M., A Memory System in the Monkey, *Philosophical Transactions of the Royal Society of London B*, (1982) 85-95.

3. Squire, L. R. & Zola-Morgan, S., "The Neurology of Memory: the Case for Correspondence between the Findings for Man and Non-human Primates," J. A. Duetsch (Ed.), *The Physiological Basis of Memory* (2nd edition), (1982) New York, Academic Press.

4. McGaugh, J. L., "Hormonal Influences on Memory Storage," *Journal of the American Psychological Association*, 38:2 (1983a) 161-174.

5. Zornetzer, S. F., "Neurotransmitter Modulation and Memory: A New Neuropharmacological Phrenology?" M. A. Lipton, A. DiMascio, & K. F. Killam (Eds.), *Psycho-Pharmacology: A Generation of Progress*, New York, Raven Press (1978) 637-649.

6. Davis, J. L., Pico, R. M., & Cherkin, A., "Arginine Vasopressin Enhances Memory Retroactively in

Chicks," *Behav. Neural Biol.*, 35 (1982) 242-250.

7. Kandel, E. R. & Schwartz, J. H., "Molecular Biology of Learning: Modulation of Transmitter Release," *Science*, 218 (1982) 433-443.

8. Alkon, D. L., "Cellular Analysis of a Gastropod (*Hermissenda Crassicornis*) Model of Associative Learning," *The Biological Bulletin*, 159 (1980) 505-506.

9. Lashley, K. S., "In Search of the Engram," *Society of Experimental Biology*, Symposium 4 (1950) 454-482.

10. Hilgard, E. R. & Marquis, D. G., (1936). "Conditioned Eyelid Responses in Monkeys, with a Comparison of Dog, Monkey, and Man. *Psychol. Mongr.*, 47:212 (1936) 185-198.

11. Gormezano, I., "Investigations of Defense and Reward Conditioning in the Rabbit," A. H. Black and W. F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory*. New York, Appleton-Century-Crofts (1972) 151-181.

12. Wagner, A. R., "SPO: A Model of Automatic Memory Processing in Animal Behavior," N. E. Spear and R. R. Miller (Eds.), *Information Processing in Animals: Memory Mechanisms*. Hillsdale, N.J., Lawrence Erlbaum Associates (1981).

13. Thompson, R. F., Berger, T. W., Cegavske, C. F., Patterson, M. M., Roemer, R. A., Teyler, T. J., & Young, R. A., "The Search for the Engram," *Am. Psychol.*, 31 (1976) 209-227.

14. Berger, T. W., & Thompson, R. F. (1978c). Identification of pyramidal cells as the critical elements in hippocampal neuronal plasticity during learning. *Proc. nat. Acad. Sci. (Wash.)*, 75, 1572-1576.

15. McCormick, D. A., Lavond, D. G., & Thompson, R. F., "Neuronal Responses of the Rabbit Brainstem During Performance of the Classically Conditioned Nictitating Membrane (NM)/Eyelid Response," *Brain Research*, 271 (1983) 73-88.

16. McCormick, D. A. & Thompson, R. F., "Neuronal Responses of the Rabbit Cerebellum During Acquisition and Performance of a Classically Conditioned Nictitating Membrane-eyelid Response," *Journal of Neuroscience*, 4:11 (1984b) 2811-2822.

17. McCormick, D. A., Clark, G. A., Lavond, D. G., & Thompson, R. F., "Initial Localization of the Memory Trace for a Basic Form of Learning," *Proceedings of the National Academy of Sciences*, 79:8 (1982b) 2731-2742.

18. Lavond, D. G., Hembree, T. L., & Thompson, R. F., "Effect of Kainic Acide Lesions of the Cerebellar Interpositus Nucleus on Eyelid Conditioning in the Rabbit," *Brain Research*, 326 (1985) 179-182.

19. Donegan, N. H., Lowry, R. W., & Thompson, R. F., "Effects of Lesioning Cerebellar Nuclei on Conditioned Leg-flexion Responses," *Neuroscience Abstracts*, 9:331 (No. 100.7)(1983).

20. Thompson, R. F., Clark, G. A., Donegan, N. H., Lavond, D. G., Madden, J. IV, Mamounas, L. A., Mauk, M. D., & McCormick, D. A., "Neuronal Substrates of Basic Associative Learning," L. Squire and N. Butters (Eds.), Neuropsychology of Memory. New York, Guilford Press (1984b) 424-442.

21. Albus, J. S., A Theory of Cerebellar Function. *Math Biosci.*, 10 (1971) 25-61.

Eccles, J. C., "An Instruction-selection Theory of Learning in the Cerebellar Cortex," *Brain Research*, 127 (1977) 327-352.

Ito, M., "The control Mechanism of Cerebellar Motor Systems," E. O. Schmitt (Ed.), *Neuroscience: Third Study Program*. Cambridge, MA, MIT Press (1974).

Marr, D., "A Theory of Cerebellar Cortex," *Journal of Physiology (London)*, 202 (1969) 437-470.

Grossberg, S., "On Learning of Spatiotemporal Patterns by Networks with Ordered Sensory and Motor Components, I. Excitatory Components of the Cerebellum," *Studies in Applied Mathematics*, 48 (1969) 105-132.

Brindley, G. S., "The Use Made by the Cerebellum of the Information that it Receives from Sense Organs," *International Brain Research Organization Bulletin*, 3 (1964) 80.

22. Thompson, R. F., McCormick, D. A., & Lavond, D. G., "Localization of the Essential Memory Trae System for a Basic Form of Associative Learning in the Mammalian Brain," S. Hulse, (Ed.), *G. Stanley Hall Centennial Volume*, Baltimore, MD, Johns Hopkins University Press (in press).

23. Steinmetz, J. E., McCormick, D. A., Baier, C. A., & Thompson, R. F., "Involvement of the Inferior Olive in Classical Conditioning of the Rabbit Eyelid," *Neuroscience Abstracts*, 10:122 (No. 36.4).

24. Mauk, M. D. & Thompson, R. F., "Classical Conditioning Using Stimulation of the Inferior Olive as the Unconditioned Stimulus," *Neuroscience Abstracts*, 10:122 (No. 36.5).

25. Ito, M., *The Cerebellum and Neural Control*. New York: Raven Press (1984). Llinas, R., Walton, K., Hillman, D. E., & Sotelo, C, "Inferior Olive: Its Role in Motor Learning," *Science*, 190 (1984) 1230-1231.

26. Steinmetz, J. E., Lavond, D. G., & Thompson, R. F. "Classical Conditioning of the Rabbit Eyelid Response with Mossy Fiber Stimulation as the Conditioned Stimulus," *Bulletin of the Psychonomic Society*, in press.

27. Eccles, J. C. (1977). "An Instruction-selection Theory of Learning in the Cerebellar Cortex," *Brain Research*, 127 (1977) 327-352.

28. Hawkins, R. D. & Kandel, E. R. (1984). Is There a Cell-biological Alphabet for Simple Forms of Learning?" *Psychological Review*, 91 (1984) 375-391.

29. Damasio, A. R. & Geschwind, N. (1984). "The Neural Basis of Language," W. M. Cowan, E. M. Shooter, C. F. Stevens and R. F. Thompson, (Eds.), *Annual Review of Neuroscience*, 7 (1984) 127-147.

30. Gluck, M. A. & Thompson, R. F., "Models of Basic Associative Learning in Invertebrate and Vertebrate Preparation," *Workshop on Theory in Neurobiology*, La Jolla, November. (1984).

PROFILE IN SCIENCE

Professor Richard F. Thompson is the Bing Professor and Chair of the Program in Human Biology and Professor of Psychology at Stanford University. His area of research and scholarly interest is the broad field of Psychobiology with a focus on the neurobiological substrates of learning and memory. This work has provided the basis for his current work as a Principal Investigator in the Office of Naval Research's Learning and Memory Special Focus Program.

Previous positions include Professor of Psychobiology in the School of Biological Sciences at the University of California, Irvine, Professor of Psychology (Karl Lashley's Chair) at Harvard University and Professor of Medical Psychology and Psychiatry at the University of Oregon Medical School. Professor Thompson wrote the *Foundations of Physiological Psychology*, a highly regarded text which analyzes behavioral phenomena in neuronal terms.

Professor Thompson has written several other texts, edited several books and published over 200 research papers. In addition to being elected to the National Academy of Science in 1977, he has received the Distinguished Scientific Contribution Award of the American Psychological Association, has been elected President of Division 6 (Physiological Psychology) of the American Psychological Association, and appointed to the Presidential Task Panel on Research in Mental Health. Professor Thompson is currently Chief Editor of the journal *Behavioral Neuroscience*, and Regional Editor for *Psychology & Behavior*, *Behavioral Brain Research*, and *Experimental Animal Behavior*, and is on the editorial board of a number of other scientific journals.

Dr. Thompson's laboratory has an outstanding group of postdoctoral fellows, visiting professors, graduate, and undergraduate students working with him on basic neuronal mechanisms of learning and memory.

NEURAL MECHANISMS OF LEARNING AND MEMORY: CELLS, SYSTEMS AND COMPUTATIONS

James L. McGaugh, Norman M. Weinberger,
Gary Lynch and Richard H. Granger
University of California, Irvine

Introduction

The function of the brain is to process information which provides a basis for behavior. During the course of each day we have countless experiences. Our experiences create memories which constitute our representations of the world as well as our repertoires for behaving. Our knowledge is preserved in a variety of forms of memory. We learn and remember the unique events of the day, general facts, perceptual skills, and motor skills, as well as cognitive skills such as those required for problem solving, mathematics, and language. And, our ability to acquire new information rests heavily on our existing knowledge. Reading, understanding, and remembering a sentence such as this, for example, involves the use of a highly complex set of well-learned skills. Clearly, the capacity to learn and remember is essential for information processing.

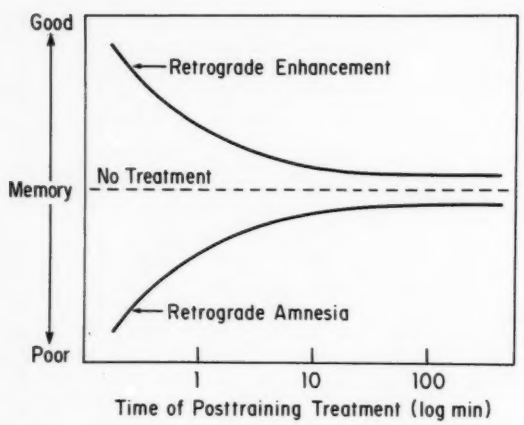
Our research focuses on the neurobiological processes underlying learning and memory. Experiences have many effects on the brain. Some are fleeting and some are lasting. Most likely, only a portion of the effects will be involved in memory storage. Some of the effects no doubt serve to regulate or modulate memory storage. Understanding the mechanisms by which the brain stores information will require knowledge of physiological systems which are activated by experiences and which are causally related to cellular changes that, we assume, underlie memory. The changes observed in nerve cells must also be causally linked to memory as it is expressed in behavior. Thus, the study of memory requires coordinated investigation at several levels of analysis—from cellular chemistry and physiology to behavior. The study of memory also requires the development of a theoretical framework which will integrate the experimental findings. This paper summarizes our recent efforts to understand the neurobiological processes underlying the storage of experiences and to develop a conceptual model of learning which is faithful to the experimental findings.

Modulation of Memory Storage

It is well-established that in humans as well as animals permanent memories are not established instantaneously at the time of an experience. The well-known clinical observation that head injuries can produce retrograde amnesia indicates that memory storage processes remain susceptible to disruptive influences for some time following learning.[1] Experiments with laboratory animals have shown that memory can be impaired by a wide variety of treatments affecting brain functioning, including electrical stimulation of the brain as well as drugs, if the treatments are administered shortly after training.[2] Typical retrograde gradients are shown in Figure 1. From one perspective, the susceptibility to retrograde amnesia would seem to suggest that the brain is badly designed. However, the evidence that memory can also be enhanced by posttraining treatments[3] suggests that the susceptibility of the brain to posttraining modulating influences may have a highly adaptive function: Following an experience, memory storage processes may be modulated by endogenous physiological systems.[4] Our recent findings provide strong support for this view. Our evidence indicates that memory storage is modulated by hormones released by experiences and that the hormones act through influences involving brain systems that regulate memory storage.

Figure 1:

Gradients of retrograde amnesia and retrograde enhancement of memory.



Mildly stimulating conditions, such as those typically used in training animals, elicit the release of several stress-related hormones, including ACTH, corticosteriods, vasopressin, oxytocin, epinephrine, and norepinephrine, into the brain and blood. Several lines of evidence indicate that these hormones influence memory storage.[5] First, numerous studies have shown that retention is influenced by posttraining administration of these hormones. For example, in rats and mice, retention of recently learned information can be enhanced by peripheral injections of low doses of epinephrine. High doses impair retention. Since such effects are seen only if the hormone is administered shortly after learning, it seems likely that the hormone affects retention by modulating memory storage processes. Many experiments investigating hormone influences on memory have used a mild footshock as punishment in training the animals. For example, in a one-trial inhibitory avoidance task, animals are simply given a punishing footshock as they walk from one region of an apparatus to another. Retention is evidenced by the animals' avoidance of the shocked area (measured as response latency) on a retention test. Posttraining epinephrine enhances the animals' avoidance of the shocked area. Typical findings are shown in Figure 2. Comparable effects are obtained with other types of learning tasks using punishment, as well as tasks using a reward such as water for motivation. Thus, the effects of epinephrine on memory are not restricted to aversively motivated tasks. In other recent work, we have found that epinephrine enables animals to learn under conditions where learning would otherwise not occur. Rats under general anesthesia were stimulated with legshock paired with a sound. Retention was tested later, following recovery from the anesthesia, by measuring suppression of drinking at the onset of the sound. Animals given epinephrine prior to training showed clear evidence of learning. The epinephrine did not alter the depth of anesthesia. Obviously, it will be of considerable interest to understand how and where epinephrine acts in enabling learning to occur in an anesthetized brain.[6]

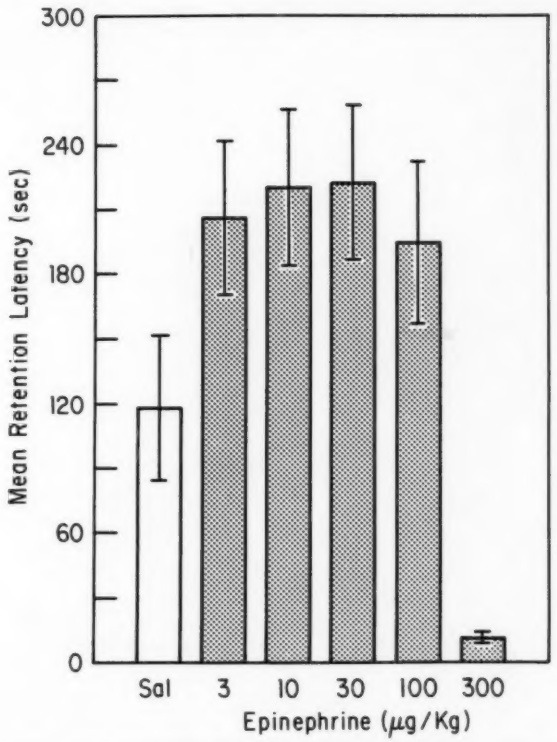
It is clear that posttraining hormones can affect memory. But, do endogenously released hormones serve to modulate memory? If they do, then learning and memory should be impaired by blocking the release of the hormones. Epinephrine is released from the adrenal medulla. We, as well as other investigators, have found that denervation of the adrenal medulla or adrenal demedullation impairs learning of inhibitory and active avoidance tasks. Further, retention in such surgically treated rats is normal if the animals are given posttraining injections of epinephrine.[7] The finding that hormone replacement attenuates the retention impairment supports our interpretation that endogenous hormones are involved in modulating memory storage.

There is extensive evidence that brain regions, located in the medial temporal region, including the amygdala and hippocampus, play an important role in memory storage. It is well-known that in monkeys as well as humans lesions of these brain regions produce a form of anterograde amnesia—impairment in learning new information.[8,9] In rats, electrical stimulation of these brain regions produces retrograde amnesia.[10] In recent work we have found that adrenal denervation or demedullation alters the effects of amygdala stimulation on memory. Posttraining amygdala stimulation which produces amnesia in normal rats produces enhancement in adrenal-denervated or demedullated rats. Further, the effect is normalized, that is, the stimulation produces

retrograde amnesia, if the animals are given epinephrine shortly before the amygdala stimulation.[7] Such findings have suggested the possibility that endogenously released hormones may influence memory storage through effects involving the amygdala.

Figure 2:

Effects of posttraining administration of epinephrine of memory: dose response effect.

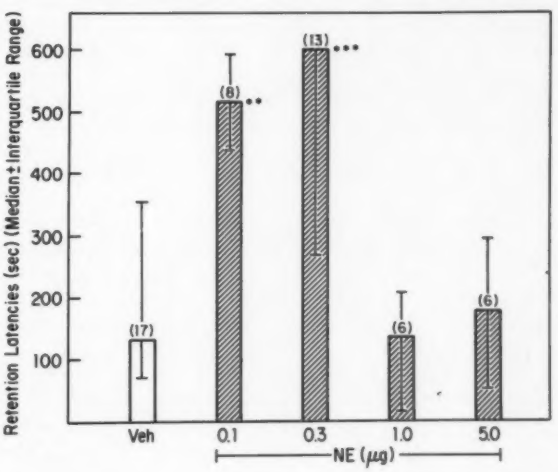


Other recent findings provide additional support for this view. For example, we have found that lesions of the stria terminalis, one of the major pathways connecting the amygdala to other regions of the brain, block the effects of peripheral epinephrine on memory.[11] Since we have found that stria terminalis lesions also block the memory-modulating effects of the opioid antagonist naloxone, it may be that the effects of several hormonal systems are mediated by influences involving the amygdala. Other investigators have reported that, in rats, memory is modulated by intra-amygdala administration of opioid and adrenergic agonists and antagonists. In recent work, we have found that retention is enhanced by injections of norepinephrine administered directly into the amygdala immediately after training (Figure 3).[12] The effect is blocked by concurrent administration of propranolol, a drug which blocks β-adrenergic receptors. Such findings suggest that the effects of peripheral epinephrine on memory may involve adrenergic activation of the amygdala. In support of this view, we have found that posttraining intra-amygdala injections of propranolol block the effects of peripherally administered epinephrine on memory. Further, we have found that the memory-impairing effects of adrenal demedullation are attenuated by posttraining intra-amygdala injections of norepinephrine.

Our findings are beginning to reveal the locus of neural systems through which hormones act in influencing memory storage. It seems highly likely that subsequent research will indicate that several brain systems are involved in modulating memory. Further, the involvement of hormones in memory is undoubtedly more complicated than is suggested by this brief summary of our findings. It is known, for example, that the effects of epinephrine on memory depend upon levels of other hormones, including corticosterone and vasopressin[13,14] as well as the opioid hormone β-endorphin (Introini and McGaugh, unpublished findings). Such information should provide highly useful in efforts to understand the way or ways in which hormones modulate cellular activity at the sites in the brain where the structural changes underlying memory occur.

Figure 3:

Effects of posttraining intra-amygdala administration of norepinephrine on memory: dose response effect.



Memory in Sensory Systems

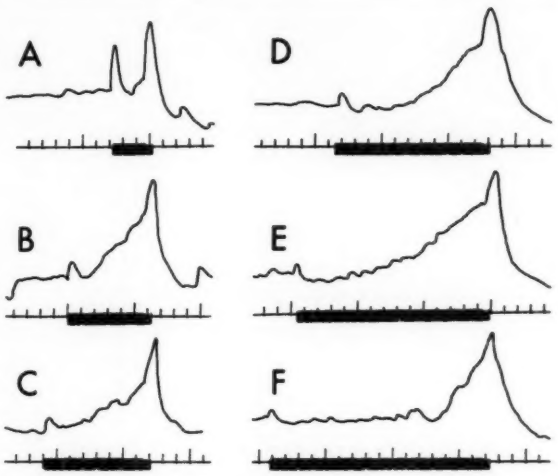
Two major informationally-based physiological disciplines have emerged during the third quarter of this century: sensory physiology and the neurobiology of learning and memory. Unfortunately, they have trod different paths, with few intersections and a marked degree of mutual indifference. Happily, there are clear signs of a developing symbiosis between these fundamental disciplines, some of which have emerged from our research program. It will be helpful to begin with a consideration of environmental events, stimuli that are the inputs to the informational nervous system.

Stimuli are described well by their physical parameters. In the case of acoustic stimuli, which we use, these include frequency (in KHz), intensity (in db referenced to a standard), location in space (referenced in azimuth and elevation to the intersection of the midline and interaural planes), and others which can be measured with equal precision. However, physical parameters alone are insufficient to fully characterize stimuli that are the inputs to nervous systems. Stimuli that impinge upon sensory receptors also have another set of parameters, their meaning or significance. These psychological parameters are measured by the behavioral output of nervous systems. Units of measure comparable to those used for physical parameters have not yet been developed for the psychological parameters of stimuli. Nonetheless, the standardization of many behavioral techniques has led to highly replicable findings among laboratories world-wide, and to an impressive amount of knowledge about the role of stimulus significance in neural information processing. For example, when an innocuous stimulus is given prior to a biologically important "reinforcing" stimulus, the first stimulus ("signal") usually acquires the ability to elicit responses that previously were evoked by the second stimulus ("reinforcer"). This standard paradigm, termed Pavlovian conditioning in honor of its discoverer, has proven to be a powerful way of studying information processing during learning.

In our studies of the effects of learning on the processing of sensory signals in the auditory system, we have made extensive use of such conditioning. It has proven highly advantageous to study the conditioned pupillary dilation response that develops when a sound is paired with cutaneous stimulation. This response is an extremely sensitive and accurate index that the signal-reinforcer relationship has been learned (Figure 4).[15,16,17,18,19] A major advantage of pupillary conditioning is that the learned response develops within a very small number of pairings; even a single pairing is often sufficient to produce a learned association between the two stimuli (Figure 5)[20] Pupillary conditioned responses belong to a class of acquired behaviors that all develop significantly faster than do specific muscular responses, such as those of eye blink and leg flexion.[21,22,23] In contrast to the latter class of specific responses, only one of which is acquired in a given learning situation, pupillary (and also cardiovascular, respiratory and non-directed muscular) conditioned responses are acquired regardless of the particular stimulus that reinforces a signal. The brain processes which produce such rapidly acquired responses may well be necessary for the learning of eye blinks, leg flexions and the like.[21,24]

Figure 4:

Pupillary dilation responses for a single cat as the interval between the beginning of an acoustic signal and the cutaneous reinforcer is increased from 2.5 sec (A) to 16 sec (F). A rise in pupillometer output indicates increased dilation. Upward deflections indicate one second marker, downward vertical lines indicate auditory pips and cutaneous stimulation. (The pupil response may appear to precede the stimulus marker because of the curvilinear nature of the polygraph paper.) Note that the latency of maximal dilation increases as the interval increases, indicating that the animal has learned the relationship between the stimuli and so can predict the occurrence of the cutaneous stimulus.

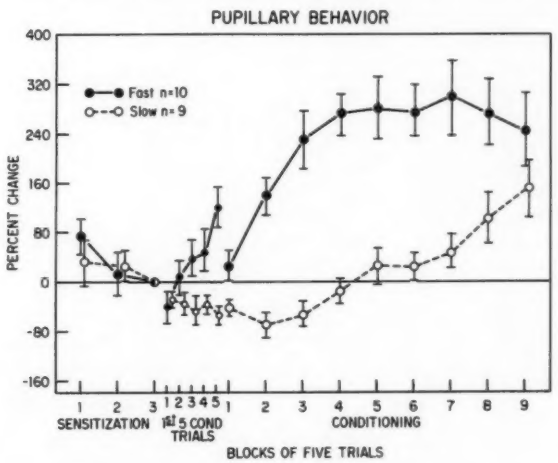


Because of the rapidity and reliability of this learning, it has proven possible to record the electrical discharges of brain cells within the auditory system at the same time that an animal is actually acquiring a pupillary dilation conditioned response. As in the case of other sensory systems, the auditory system is comprised of neural stations that extend from its receptors at the periphery up through major subdivisions of the brain to the cerebral cortex. Our initial studies revealed that the responses of cortical neurons to acoustic signals changed systematically during learning.[25] Furthermore, a very high percentage of individual neurons exhibit this response plasticity (70-95%).[20,26] The development of this neuronal plasticity was as rapid as that of the behavioral pupillary conditioned response. These findings indicated that the auditory cortex was responsive to the psychological significance of acoustic stimuli, not merely to their physical parameters.

The finding of learning effects on the processing of acoustic information within the auditory system presented a paradox. It is universally agreed that sensory systems must provide an accurate and detailed representation of the environment. How, then, can a sensory system perform this function if its cellular responses to stimuli are governed not only by the physical characteristics of environmental stimuli, but also by their acquired importance of signal significance? How, for example, is an animal able to distinguish between a loud unimportant sound and a weaker significant sound if its auditory neurons respond in the same way to both events? A resolution of this paradox was suggested by studies of the part of the auditory system which controls input to the auditory cortex. This is the medial geniculate nucleus (MGN) which comprises the thalamic auditory system.

Figure 5:

Pupillary learning curves. Data are divided into two groups of animals based on the rate of acquisition of conditioned responses. The slow learners were exposed to many acoustic stimuli prior to training; their acquisition of the pupillary dilation response is retarded because this prior exposure predicted the absence of a reinforcer. The rapid learners reveal the fundamental learning curve in the absence of such prior inhibitory experience. Note that these subjects display conditioned responses within the first 10 trials of the conditioning phase and reach asymptote by the 20th trial. The rate of acquisition is illustrated in finer detail for the first five trials of conditioning which shows that a conditioned response had actually developed on trial #2 relative to trial #1, hence one-trial learning. (Sensitization refers to a control period in which the acoustic signal and the reinforcer were presented randomly).



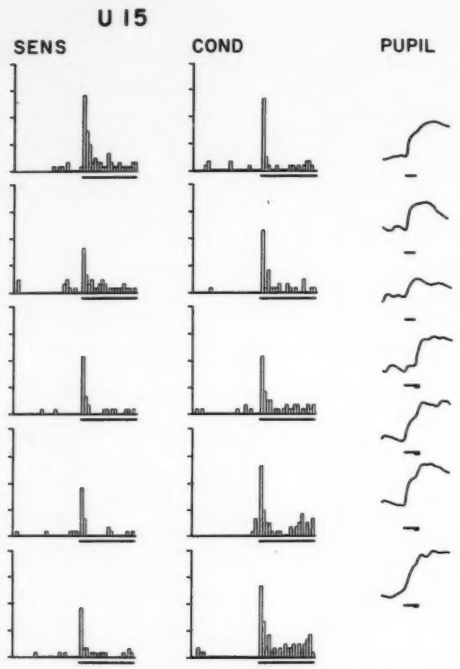
Ascending auditory information is obliged to synapse in the medial geniculate nucleus on its way to the auditory cortex. Prior anatomical and physiological studies have revealed that this nucleus is actually comprised of subdivisions[27] that have different properties. Cells in the ventral part of the MGN are each narrowly tuned to specific acoustic frequencies.[28] Indeed, the frequency organization of the cochlea is mapped topographically in this region, as it is at all other levels of the auditory system, including the primary auditory cortex. In contrast, although cells in the medial part of the MGN also are responsive to auditory stimulation, they are not narrowly tuned. We recorded cellular responses to acoustic stimuli simultaneously in these subdivisions of the medial geniculate. Only neurons in the medial division developed discharge plasticity; they did so rapidly, as quickly as the acquisition of the pupillary conditioned response. Cells in the ventral division were not plastic (Figures 6, 7).[29,30,23,31] Similar findings have been obtained in other laboratories using different species and different training tasks.[32,33] So the paradox described above may be resolved by realization that the auditory system has anatomically and functionally distinct components which separately handle the physical and psychological aspects of stimuli.

Figure 6:

The medial geniculate nucleus showing its major subdivisions. Note that the ventral region (VL, VO) has small cells arranged in lamina. These neurons are each tuned to a narrow frequency band; they do not change responses during learning. The medial region (M) has large cells, each of which responds to most frequencies. These cells are highly sensitive to the significance of sounds as indicated by rapid changes in response during learning.

Figure 7:

Histograms and pupillary records for a cell in the medial part of the MGN. Each histogram is the sum of neuronal discharges for two consecutive trials; each bar represents 50 msec. Data are for the ten trials of sensitization and the first ten trials of conditioning. Note the decrement during sensitization and the extremely rapid increment during conditioning. The pupil response, as always, shows a decrease during the former and an increase during the latter. Calibration: 6 spikes per division; horizontal bar, 1 sec.



These findings raise yet another problem. How are the physical and psychological aspects of auditory information processing brought together, so that ultimately meaning can be attributed to a particular stimulus, the details of that stimulus can be perceived, and both aspects of the stimulus can be stored? No definitive answer can yet be given, but it does appear that the two aspects of a stimulus are kept separate through the level of the thalamus. They appear to converge first at the auditory cortex. Thus, the auditory cortex is not strictly auditory, nor is it strictly cognitive. It seems to be both, and might better be termed "auditory perceptive cortex," for perception is well-known to be influenced by both the physical and the psychological parameters of stimuli.

Having established auditory cortex as a site of particular interest, it became possible to combine the techniques of learning with those of sensory physiology to begin to determine exactly which aspects of auditory information processing were altered by learning. Learning might simply change the excitability of auditory cortical neurons. On the other hand, learning could alter information processing in a highly specific manner, perhaps by changing neuronal responses to particular acoustic frequencies only. Therefore, we initiated studies in which both techniques were employed and in which we asked whether learning causes specific changes in the frequency tuning of cells, for frequencies that we employed as the conditioned signal during learning.

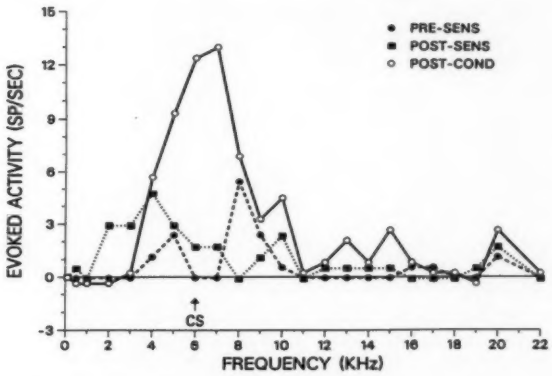
The techniques of sensory physiology and learning are complementary. For the former, the physical parameters of a stimulus are systematically changed, while the meaning of a stimulus is kept constant. In the case of learning studies, the physical parameters are kept constant while the meaning is changed by arranging a relationship between the stimulus and a reinforcer. In our studies, we employ the sensory physiology approach before and after various phases of conditioning studies. First, for example, a neuron's frequency response function is determined by presenting a series of tone bursts (at equal intensity at the tympanic membrane) from approximately 0.2 KHz to 25.0 KHz. (We have found that such frequency functions are very stable for cortical neurons in the absence of any behavioral training.) Next, we select one frequency to be used as the acoustic signal; this frequency is not the best frequency of the neuron, but one to which the neuron could in principle increase or decrease its discharges. We present this signal randomly with the cutaneous reinforcer. This training phase, referred to as sensitization, yields an estimate of the non-associative effects of stimulus presentation. Following this test, we condition the animal by pairing the signal frequency with the reinforcer. Pupillary conditioned responses develop at this time. Immediately thereafter, we again retest the cell with the full frequency range, as is done routinely in sensory physiology studies. We have found that conditioning can induce a frequency-specific change in the tuning function of a cell.[34] This change is generally highly specific to the frequency that was employed as the acoustic signal (Figure 8). If we carry the training further, by presenting the acoustic signal alone and deleting the reinforcer, the conditioned pupillary response disappears ("experimental extinction"). So also does the frequency-specific change in a cell's tuning function. Thus, learning can involve a highly specific change in the processing of information about acoustic frequency. This change can be rapidly induced and rapidly reversed.

These findings indicate that the effects of learning upon auditory cortex are not merely indicative of general changes in excitability, but actually engage mechanisms involved in the cell's tuning. The data suggest that the receptive fields of cortical neurons may be more reflective of stimulus importance than of the physical parameters of the acoustic environment. Further, they reveal that a combined approach of learning and sensory physiology can be a powerful tool in the study of information processing in the brain. Studies of the mechanism of the specific effects of learning on information processing in the cortex need to be done. It is equally important to develop a comprehensive characteriza-

tion of the learning effects, both in auditory cortex and in other sensory systems. Only then will we begin to develop adequate conceptualizations of the actual role of sensory systems in behavioral adaptation and the interplay between the physical and the physiological worlds.

Figure 8:

Learning produces a frequency-specific change in the "frequency tuning" function of single cells in the auditory cortex. For this neuron, the response to various frequencies (0.2-22.0 KHz) was determined at three times: before any training ("pre-sens"), following random presentation of twenty presentations each of a 6.0 KHz tone and cutaneous stimulation ("post-sens"), and immediately after twenty paired presentations of these stimuli in the 6.0 KHz tone became a conditioned signal for the cutaneous stimulation ("post-cond"). (During pairing, the animal developed a pupillary dilation conditioned response.) This particular cell initially responded only weakly to some frequencies, and there was no systematic change following sensitization. In contrast, after learning that the 6 KHz tone was a signal, the cell greatly increased its response to this and adjacent frequencies. This indicates that the cell's frequency response reflects the meaning of specific tonal frequencies.



Cellular Mechanisms of Memory

Although sites of learning-induced response plasticity continue to be identified, we do not know what chemistries the brain uses to store memory. However, we can make some reasonable inferences about the site and general properties of the mechanism. Most theorists conclude that the bulk of human memory is encoded as structural changes in discrete subpopulations of the thousands of billions of synapses in cerebral cortex (see reference 35, for a discussion). Structural changes are indicated by the duration of memory since the molecular constituents of the neurons and glia are broken down on a daily or weekly basis. Any purely biochemical change will therefore be replaced in a time frame far too short to account for memory. A synaptic locus for memory is required to account for the enormous capacity of memory. That is, if the whole neuron, or all of its axonal or dendritic processes were modified by single encoding events, the gradual accretion of memory over years should begin to deplete the pool of unchanged cells available for modification and hence storage. The same is true of synapses, but since these are present in 10,000 to 100,000 greater numbers than neurons, the limit would be approached much more slowly. Finally, cortex is chosen as the region of storage because it alone combines vast numbers of synapses, dense interconnectivity, and intimate access to all brain processing systems; the second and third of these features are required to deal with the ability of the memory system to associate disparate elements. Putting all this together, we find that the memory chemistry should be capable of irreversibly modifying the structure of cortical synapses and that it must be able to do this to a small group of contacts on a single cell without influencing the remaining inputs. This is a demanding set of constraints and one that can only be satisfied by a fairly exotic (to brain scientists at least) type of cell chemistry.

Recently, members of our group have identified a specific enzymatic process that appears capable of producing the effects described above, and have advanced the hypothesis that this process is responsible for certain types of memory. Before describing the hypothesis, it is necessary for us to first discuss a problem, in fact *the* problem that has dogged virtually every study of memory substrates in behaving animals. Consider this: It follows from the vast capacity of memory that any single item is encoded by changes in an infinitesimally small number of storage elements, but to detect biochemical changes we obviously need a very sizable sample of changed elements. How then can we hope to observe the chemistry of memory in action? One approach that is enjoying increasingly widespread usage begins with the study of physiologically induced synaptic plasticity: The idea here is that if memory involves long-lasting synaptic modification elicited by neuronal impulses, then we should be able to produce the same effects (but on a massive scale) by stimulating brain circuitries with the appropriate patterns of activity. This strategy has been in the field for many years, but it was not until 1973 that investigators[36,37] obtained positive results. In those studies, it was shown that a few seconds of high frequency stimulation of the major input to the hippocampus, the simplest of cortical structures, produced an increase in the strength of the connections formed by that input and that this effect could persist for weeks and months. This was followed by an intense search for the types of synaptic changes associated with LTP, resulting in the discovery by one of the laboratories in our group, that the stimulation caused changes in both the shape of synapses and, most remarkably, the formation of new contacts.[38,39] Other experiments have confirmed and extended these observations.[40,41] Alterations of these types can be expected to change the chemistry of synapses and, indeed, we now know that LTP is also correlated with an increase in the number of receptors for glutamate, one of the chemicals used as a transmitter in hippocampus.[42]

How could intense synaptic activity produce effects of this type? This is a subject now under investigation in many laboratories, but one clue comes from the recent observation that injection in many laboratories, but one clue comes from the recent observation that injection into neurons of a chemical that buffers calcium to very low levels blocks the induction of LTP.[43] This finding strongly suggests that the lasting potentiation is triggered by a sudden rise in intracellular calcium in post-synaptic processes. Summing up, it does appear that cortical synapses possess some type of chemical mechanism that is activated by physiological activity and that produces significant structural alterations. Moreover, we have reason to suspect that this mechanism is triggered by calcium and, among its other effects, produces an increase in the number of glutamate receptors.

Membrane-Associated Calcium-Activated Proteases

Most cell biological processes exist in a world of checks and balances and have effects that are fully reversible. The proteases are a conspicuous exception to this rule; these enzymes break up their substrate proteins, an effect that can only be counteracted by replacing the protein. This irreversible effect of proteases makes them interesting candidates for long-lasting phenomena such as memory. Moreover, work over the past ten years has shown that many cell types, including neurons,[44,45,46] contain a group of calcium-activated proteases ("calpain") that degrade proteins involved in forming the cytoskeleton,[47] an action that we can assume leads to structural change. Calpain activation thus offers a means through which the types of alterations found in the LTP experiments might be realized, and there is now a considerable body of experimental work that supports this idea.

First, the number of glutamate binding sites in synaptic membranes purified from hippocampus and cortex of rats is more than doubled when the membranes are exposed to low levels of calcium.[48,49] This effect is irreversible in that the added receptors remain after the added calcium is removed.[50,48] Thus, calcium, the suspected trigger of LTP, reproduces one of the correlates of LTP; moreover, calcium's effect on receptors is selectively and dramatically blocked by inhibitors of calpain.[51,49] Biochemical experiments have demonstrated that calpain is present in high concentrations in synaptic membrane fractions,[52] where it acts to degrade microtubule associated proteins (MAP's) and fodrin (or brain spectrin).[53] Fodrin, like all members of the spectrin family, forms a dense meshwork immediately beneath the membrane and serves to link the cytoskeleton with transmembrane proteins. There is considerable evidence that this meshwork restricts the mobility of membrane elements such as receptors. Proteolytic action in the fodrin network would be expected to produce changes in surface chemistry and thus could explain the effects of calcium on glutamate receptors. To directly test this idea, Siman et al.[54] used antibodies against fodrin to protect it from calpain and found that under these conditions calcium no longer induced glutamate binding sites.

Calpain, by locally disrupting the cytoskeleton, might also produce structural reorganization of synapses as well as prepare membranes for the development of new contacts. It is difficult to test this idea in brain but recent experiments strongly suggest that shape change and structural reorganization in red blood cells[55] and platelets are accompanied by calpain-mediated breakdown of spectrin and other cytoskeletal-linking proteins.

These observations suggest the following hypothesis[56] about the cell biology responsible for producing lasting alterations in cortical synapses:

(1) high frequency stimulation transiently increases intracellular calcium, thereby activating calpain;
(2) calpain degrades the links between membrane and cytoskeleton;
(3) these effects result in the appearance of new receptors and anatomical changes in synaptic contacts;
(4) calcium homeostasis is rapidly restored but the effects of calpain activation remain.

Calpain and Memory in Cortical Networks

There is much in the proposed calpain hypothesis as described immediately above that requires futher testing. We need to know much more about the precise localization of the enzyme and its substrates—it will also be necessary to determine if it is activated by physiological events and produces lasting changes in synaptic physiology. But the experimental results to date indicate that calpain provides an excellent means for selectively and irreversibly changing synapses and thereby offers a very specific candidate for a memory mechanism. It is therefore only appropriate to begin investigating the possible role of the enzyme in memory. Unfortunately (for the experimenter) all behavioral paradigms may not be equally appropriate in this regard. Many investigators now believe that humans have at least two fundamentally different memory systems, one which is involved with specific data (e.g., names, faces, telephone numbers, locations, etc.) and a second which is concerned with the procedures or rules needed to solve a problem (e.g., operating a car, solving puzzles of the same type, etc.). Note also that it is possible that still other categories exist and that some of the most typically used paradigms may sample memories that are neither "data" nor "procedural" in type. This is particularly relevant here because it is very unlikely that calpain is involved in the production of all variants of information storage: We cannot detect the enzyme in fish brains[57], and fish certainly learn. Even within mammals, calpain varies greatly across brain regions and cell types, as shown by immunocytochemical experiments,[58] and it does not appear to expose glutamate receptors in certain brain stem regions.[51] Thus, any forms of memory stored in the lower parts of the brain may also involve processes other than calpain. Accordingly, we decided to begin our experiments using memory problems that clearly involved circuitries running through mammalian forebrain and in particular cortex, areas in which we are confident that the calpain mechanism is present. This led us

to the olfactory system, since its projections have been well-studied and are restricted to forebrain regions. In our first studies, we injected an inhibitor of calpain, leupeptin, into the cerebral ventricles using continuously active osmotic minipumps implanted under the skin of rats. The drug caused no observable effects on feeding, drinking, body temperature, or exploratory behavior, but it produced a substantial impairment of the rat's ability to remember which of two odors led to a water reward.[59,60] To test the generality of this result we measured spatial memory using an 8-arm radial maze; while little is known of the circuitries used by rats to solve this problem, it is well-established that the hippocampus is critical.[61] Leupeptin produced a sizable impairment in the rat's performance in this test.[59] In striking contrast to these results, the calpain inhibitor had virtually no effect on shock avoidance learning. This dissociation raised again the possibility that different forms of memory with different chemical substrates co-exist in the rat brain. In support of this, we have found that drugs that *do* block shock-avoidance memory have no effect on olfactory memory[62]. This provided the first double dissociation of memory using pharmacological agents and confirms the selective nature of the effects produced by leupeptin.

Calpain pharmacology is very limited at this time although new drugs are gradually being developed. Future studies of the role of the enzyme in memory will be greatly facilitated when these become available for testing. But beyond this, we need to begin testing for the effects of drugs on the physiology of those circuitries involved in the encoding and storage processes. This will allow us to determine if calpain inhibitors actually do interfere selectively with behaviorally induced synaptic plasticity; moreover, if the projections and connections underlying the memory process can be identified, it may be possible to look for structural and chemical events, including calpain activation, associated with storage. As discussed in the next section, the olfactory system is uniquely suited for analyses of this type and has a number of other advantages as well.

Olfaction and Memory in Cortical Networks

The projections of the olfactory system are relatively simple. Fibers from the nasal receptors terminate in the olfactory bulbs which in turn distribute their axons to the amygdala and an extended sheet of cells that constitute the outermost layer of the piriform/entorhinal (olfactory) cortex. This latter region then innervates the hippocampus and the dorso-medial nucleus of the thalamus. The smell system is unique among the senses in having virtually direct connections with three structures (DMN, amygdala, and hippocampus) involved in the processing of certain kinds of memory in mammals, including humans. This may explain why odors are so potent in evoking memories and may also explain the recent findings that olfactory problems are common in human amnestics.[63,64] It is also interesting that smell memory in rodents displays a number of striking similarities to everyday memory in humans. Recently we demonstrated that rats with lesions separating olfactory inputs from the hippocampus exhibit a rapid forgetting syndrome that parallels the anterograde amnesia reported for humans with damage to the temporal lobes and hippocampus.[60] These animals were able to acquire smell discriminations but within an hour gave no evidence of remembering the correct odor. Work from other laboratories strongly suggests that this type of lesion does not interfere with the animals' ability to learn, remember, and use procedures needed to solve olfactory problems. Thus, damage to the hippocampal system is selective to one kind of memory in rats as it is in humans (see reference 65 for a review of the human literature).

The reason that olfaction is not widely used in learning experiments is that odors cannot be used as punctate, well-defined (temporally or spatially) cues; this vastly complicates the job of measuring physiological responses in the brain. However, preliminary studies suggest that it may be possible to substitute electrical stimulation of the olfactory bulb for odors as cues in rats trained on a series of olfactory discriminations (unpublished observations). If this proves correct, then it will be possible to exploit the extreme simplicity of the olfactory system to follow the events involved in processing and storing information in a cortical network and to test where and how drugs influence these events. Together with new pharmacological tools, those developments should permit much more rigorous correlational and manipulative tests of the calpain hypothesis.

Computational Psychobiology

The field of the neurobiology of learning and memory has two sides: one is the question of neurobiological mechanisms underlying animals' information-processing abilities (i.e., the ability to acquire, store and retrieve memories); the other, reciprocal question is: precisely what are these information-processing abilities themselves? In a simple sense, animals can store and retrieve information based on experience. But there are notable constraints on those abilities; not all experiences are learned with equal facility, and not all things that are learned are recalled accurately. Work in our laboratory on computational psychobiology has attempted to identify formal characterizations of precisely what is learned, and under what circumstances, and to relate those computational findings to neurobiological mechanisms that may underlie them.

For example, a rat in a discrimination-learning environment learns over trials to pick out the key sensory feature cues that accurately predict a shock, or a food reward, etc., from among a wide array of possible candidate cues. At the neurobiological level, we might want to ask how it is that this discrimination ability arises from brain function. But at the computational level, we also must ask which of the many candidate sensory cues are learned, and why? There are many reasonable algorithms for a human or animal to use in selecting some subset of a set of candidate cues, and each different algorithm makes different predictions about what the animal will actually learn from particular experiences. Before we can attempt to find a circuit that performs

this learning behavior, it would help to know precisely what it is that the circuit is supposed to compute, i.e., to be able to specify formally what the learning behavior actually is.

There are a number of model systems of human associative learning behavior, ranging from sea slugs, to rats, to artificial-intelligence computer models. For any such model to be a useful substrate for the study of human cognition, it must contain similarities to human processing; this is indeed the motivating assumption in all model work. It is hoped that sea slugs have neural circuitry that works according to the same mechanisms and principles as human neural circuitry; the same applies to pigeons, rats, and rabbits. There is, of course, no such hope for computers, yet it is possible to study human cognition on at least three levels via computer models: (1) At the biochemical level, computer models can aid in the analysis of low-level mechanisms of synaptic change; (2) At the circuit level, models of connectivity and information-flow can aid in the testing of theories of neural networks; (3) At the behavioral level, accurate characterization of the constraints on what is learned can aid in the testing of theories of associative conditioning. Each level, in isolation, has shortcomings; accurate behavioral characterization can make only a limited contribution toward the development of a theory of the mechanisms of learning, and a biophysical mechanism of synaptic change will answer only a part of the larger question of how experiences are encoded, indexed and retrieved.

What is Computed In Contingency–Based Associative Learning?

Our work has focused primarily on the level of behavioral characterization of learning. Our recent work has attempted to formalize results on contingency and latency in classical and instrumental conditioning. At the computational level, a good deal of work on animal learning and memory[66,67,68,69] has identified necessary constraints on behavior, i.e., input-output specifications that must be accounted for by any algorithms intended as theories of contingent learning and memory. The constraints focused on in this section are those of the salience assignment task, i.e., identifying which of several features are predictive of a particular result.

Imagine a rat attempting to decide, over trials, which of several environmental features or events should be learned to be a predictor of a recurring shock event. We can view the animal's task as having to hypothesize potential prediction relationships between the shock and the various features, individually and in various combinations, and to weigh these hypothesized relations against each other.

We might initially assume that a particular feature or event would be inferred to be the predictor of the shock depending on the number of times that feature actually occurred immediately before the shock. Each time the feature is paired with the shock, the association between the feature and the shock might be strengthened (see reference 70). Extensive experimental evidence in the psychological literature shows this to be false. This condition is insufficient to warrant the animals' inference that the tone is more likely to predict the occurrence of shock. Rather, this association depends upon the fact that, of all stimuli present, the tone is the best predictor of the shock. Hence, the naive idea that the number of pairings alone determines the predictiveness (or salience) of candidate predictive features, or that strengthening alone could be the mechanism for learning associations, is false.

Contingency vs. Number of Pairings

This result requires a somewhat counterintuitive computation on the part of the animal: not only must the animal be computing, somehow, the conditional probabilities of the shock occurring, given the light and given the tone, but it must also be able to compute the relative conditional probabilities (RCP), i.e., the comparison or ratio of the likelihood of various stimuli preceding the shock.

Saying that this "must be computed" speaks only of a behavioral constraint, but does not yet address the question of the algorithm that the animal might be using to arrive eventually at behavior that conforms to the constraint. Marr's three levels of explanation for theories of visual processing[71] can be extended to the domain of learning and memory: At the computational level, we may speak simply of the computations that must somehow be arrived at (such as the contingency or salience relationship between cues and results), without reference to how those computations may be carried out. At this level, we simply specify accurate empirical generalizations that capture the data. The algorithm level constitutes the level of mathematical function that performs the necessary computations. Finally, these algorithms are instantiated in hardware (neural or computer) at the implementation level.

These three levels are not wholly independent. In particular, the algorithm level must conform to the constraints provided by each of the other two levels: it must compute all and only those things that have been identified as actually occurring in humans and animals (at the computational level), and the proposed algorithms must be able to be carried out somehow, given the tools and limitations provided in the substrate (at the implementation level). Rescorla's[72,73,74] formulation of the necessary computation underlying contingency is that the probability of the shock outcome (the US), given the conditional stimulus feature (the CS, e.g., the tone), must be greater than the probability of the outcome occurring without that feature having occurred, or $p(US|CS) > p(US|\overline{CS})$.

It is this measurement of relative conditional probabilities that accurately captures the empirical results of studies on contingency-based learning. Animals, and humans in analogous circumstances, exhibit contingency-driven learning in the sense that they somehow maintain incrementally-updated knowledge of the relative predictiveness or salience of features. Through experience the animal must pick out the relevant (salient) features from the background of uncorrelated features and use only these salient features to predict future events.

Thus, although the term 'contingency' is often used simply to denote the selection (by any algorithm) of salient features from a group of candidate features during associative learning, true contingency-based learning can be more formally stated as the assignment of associative salience to a feature (or feature-cluster) E1 relative to an outcome E2 always and only when the conditional probability of E2 occurring given the occurrence of E1 is greater than the conditional probability of E2 given the absence of E1, or as it was stated above, $p(E2|E1) > p(E2|\overline{E1})$.

Differential Predictions

Because we have stated the constraint of true contingency in a formal way at the computational level, we can differentiate between algorithms that conform to this formal constraint (i.e., conform to this accurate characterization of the data) and those that do not. Some particular algorithm might loosely account for associative learning phenomena by providing a mechanism for selecting salient from irrelevant cues, but that algorithm might not accurately conform to the actual data on contingency, as captured in the statement of relative conditional probabilities.

In particular, algorithms that operate by strengthening associations by some amount each time a specific cue is paired with some outcome, and weakening them at non-pairings, cannot in principle compute true contingency. This is primarily because there is no differential weighting performed for different types of non-pairings, i.e., extra spurious CSs versus extra spurious USs over trials; in general, linear equations cannot correctly compute contingency, which is a nonlinear equation.

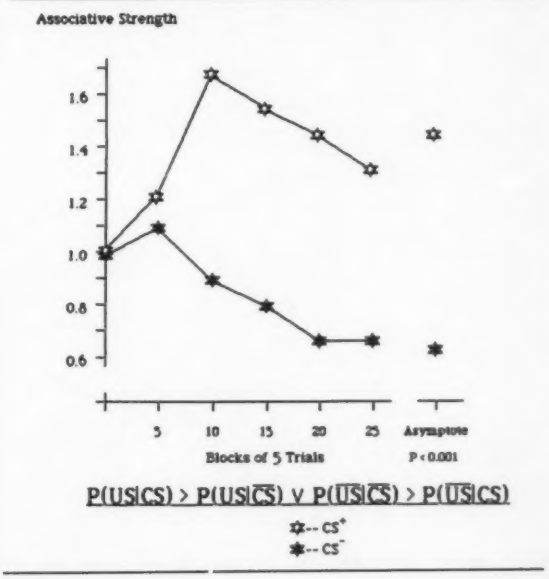
Granger, Schlimmer and Young[75] give a detailed account of our own formulation of the problem and algorithm for the correct computation of contingency using only "local" calculations, that is, not making any assumptions about any extraordinary ability of the animal to hold lots of data in its head. To summarize the results, there are four categories of possible relationship among stimuli that are calculated: for any given US, a CS may be either a positive cue (i.e., predicts the US), a negative cue (predicts the absence of the US, i.e., a 'safety signal'), a context cue (behaves in an ambiguous way with respect to prediction), or an uncorrelated cue (i.e., is random with respect to the US).

We have constructed a working computer model of this algorithm, which we call CAP-CEL (Contingent Associative Processing in the CEL framework).[76] Figures 9 and 10 illustrate the different predictions that would be made by (a) a true contingency algorithm (in this case our CAP-CEL algorithm) versus (b) a strengthening (linear) algorithm. The white and black items represent two different candidate cues that are sometimes paired with a shock; the animal is trying to decide which of them is the predictive cue. Animals would choose the white cue as the predictive one over the black cue, since the white cue is the one that is paired contingently with the shock. The stars on the graph in Figure 9 illustrate that a computer simulation of CAP-CEL's algorithm based on relative conditional probabilities correctly orders the two cues, white over black. The circles in Figure 10, however, denote a computer simulation of a strengthening algorithm (one that strengthens or weakens the associative strength of cues based on the number of times the cues have been paired with the shock via a linear equation). This algorithm chooses the black cue over the white cue for this example set of trials. Hence, this associative learning algorithm does not correctly model the animal data; i.e., it does not compute the right thing. Our computational-level approach to learning and memory therefore provides a set of analytical tests against which proposed mechanisms of learning can be measured. We are currently attempting to analyze a number of proposed animal models of learning, including Hawkins and Kandel's *Aplysia* circuits[77], Alkon's *Hermissenda* circuits[78], and Lynch's mammalian hippocampal slice circuits[56,42], as well as a number of proposed computer and analytical models, including Gluck and Thompson[79], Sutton and Barto[80], and Gibbon et al.[81], to identify precisely what these models compute.

Thus, algorithmic interpretations of empirical results in neurobiology must conform both to "top-down" constraints of behavioral data (i.e., they must compute the right thing) and to "bottom-up" constraints of the substrate (e.g., known constraints of various brain structures and circuits).

Figure 9:

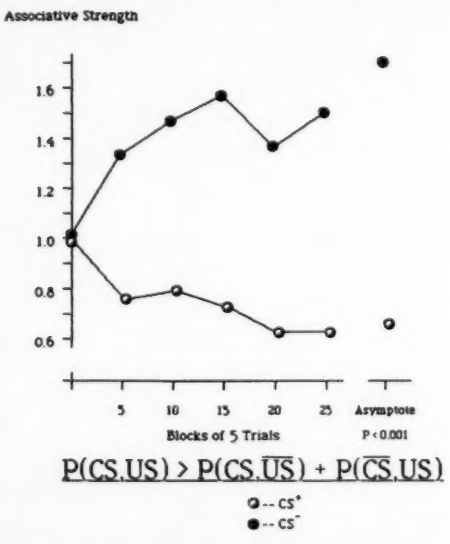
Discrimination learning curve using contingency. The open symbols denote acquisition of association to the predictive cue; closed symbols, to the unpredictive cue. Associative strengths greater than one indicate predictiveness. Note that the contingency mechanism correctly learns the predictive cue.



All experimental research in neurobiology is (at least implicitly) guided by some top-down hypotheses. At their simplest, these hypotheses can simply be statements such as: There are long-lasting changes in behavior during learning (at the computational level), so there must be some mechanism for causing long-lasting changes in the brain (algorithm level) which is grounded in some biophysical process (implementation level). However, any algorithm or neurobiological (implementation) mechanism proposed for long-lasting changes must conform to more stringent computational requirements than that the change simply lasts a long time once it has occurred. Computational-level data on the duration of the trace, its onset, and its extinction are crucial (and measurable) constraints on any such candidate mechanism; further crucial constraints are that the trace must be part of a system that enables the trace to be "retrievable" exactly when appropriate at some later time, the ability to store "sequences" of long-lasting traces, and a mechanism for "generalizations" of traces. These computational constraints on a memory mechanism are more difficult to collect data on, and yet any proposed mechanism that does not conform to them is clearly not the mechanism of long-term memory. It is crucially important for the computational work formally specifying these constraints to move forward, so that candidate mechanisms for long-term memory may be evaluated by these formal specifications.

Figure 10:

Discrimination learning curve using strengthening. The open symbols denote acquisition of association to the predictive cue, and closed symbols denote acquisition of association to the unpredictive cue as in Figure 9. However, note that the strengthening mechanism incorrectly learns the unpredictive cue.



Concluding Comments

Rapid progress in the subfields of memory research is beginning to afford opportunities for the development of synthetic theories in which various levels of analysis (and phenomenological observation) inform and interact with each other. Thus, we now have good evidence that certain cortical regions are involved in storage and that the hippocampus and amygdala promote the encoding process. How do these latter two structures influence physiological events in cortex? In what way is the transmission of information passing through olfactory and auditory cortices modified by hormonal states acting upon the amygdala? It should be possible to begin addressing questions about the interactions at the level of neuronal systems. Similarly, tools are available for analyzing the similarities and differences between memory types found at the behavioral level. We have found drugs that dissociate some categories of memory tasks. This can be carried much further, particularly if we add hormonal treatments to the list of manipulations. Studies at a still more conceptual level can also be envisioned. As we more clearly define the various components of the neural systems that cooperate in producing memory, it will become possible to match these against the subcomponents of models of memory processing. It is also the case that the neural networks involved are in some cases so precisely defined that they can be used to construct computer models of memory machines.

Biographies

James L. McGaugh is the Director of the Center for the Neurobiology of Learning and Memory at the University of California, Irvine, where he was also the founding chair of the Department of Psychobiology. Dr. McGaugh received the Distinguished Scientific Contribution Award of the American Psychological Association for his extensive and pioneering studies of the drug and hormonal modulation of memory.

Norman M. Weinberger served as chair of the Department of Psychobiology at the University of California, Irvine from 1975-1978. He is a founding member of the Center for the Neurobiology of Learning and Memory. Dr. Weinberger's research concerns the neurophysiological bases of learning in the auditory system.

Gary Lynch, a member of the Center for the Neurobiology of Learning and Memory, is a professor in the Department of Psychobiology at the University of California, Irvine. Dr. Lynch is a prolific and highly regarded researcher whose work on the cellular basis of memory is receiving considerable attention.

Richard H. Granger is a member of the Center for the Neurobiology of Learning and Memory and an associate professor of computer science and of cognitive science at the University of California, Irvine. His interdisciplinary work focuses on computer models of psychobiological phenomena.

References

1. Russell, W. R., and Nathan, P. W., "Traumatic Amnesia," *Brain,* 69 (1946) 280-300.

2. McGaugh, J. L., and Herz, M. J. *Memory Consolidation.* San Francisco, Albion Publishing Company, 1972.

3. McGaugh, J. L., "Drug Facilitation of Learning And Memory," *Annual Review of Pharmacology,* 13 (1973) 229-241.

4. Gold, P. E., and McGaugh, J. L., "Endogenous Processes in Memory Consolidation," H. Weingartner and E. Parker (Editors), *Memory Consolidation,* Hillsdale, N. J., Lawrence Erlbaum Associates (1984) 65-83.

5. McGaugh, J. L., "Hormonal Influences on Memory," *Annual Review of Psychology,* 34 (1983) 297-323.

6. Weinberger, N. M., Gold, P. E., and Sternberg, D. B., "Epinephrine Enables Pavlovian Fear Conditioning Under Anesthesia," *Science,* 223 (1984) 605-607.

7. Liang, K. C., Bennett, M. C., and McGaugh, J. L., "Peripheral Epinephrine Modulates the Effects of Posttraining Amygdala Stimulation on Memory," *Behavioural Brain Research,* 15 (1985) 93-100.

8. Mishkin, M., Malamut, B., and Bachevalier, J., "Memories and Habits: Two Neural Systems," G. Lynch, J. L. McGaugh and N. M. Weinberger (Editors), *Neurobiology of Learning and Memory,* New York, Guilford Press (1984) 65-77.

9. Squire, L. R., and Cohen N. J., "Human Memory and Amnesia," G. Lynch, J. L. McGaugh, and N. M. Weinberger (Editors), *Neurobiology of Learning and Memory,* New York, Guilford Press (1984) 3-64.

10. McGaugh, J. L., and Gold, P. E., "Modulation of Memory by Electrical Stimulation of the Brain," M. R. Rosenzweig and E. L. Bennett (Editors), *Neural Mechanisms of Learning and Memory,* Cambridge, MIT Press (1976) 499-525.

11. Liang, K. C., and McGaugh, J. L., "Lesions of the Stria Terminalis Attenuate The Enhancing Effect of Post-training Epinephrine on Retention of an Inhibitory Avoidance Response," *Behavioural Brain Research,* 9 (1983) 49-58.

12. Liang, K. C., Juler, R., and McGaugh, J. L., "Modulating Effects of Posttraining Epinephrine on Memory Involvement of the Amygdala Noradrenergic System," *Brain Research,* in press.

13. Borrell, J., de Kloet, E. R., and Bohus, B., "Corticosterone Decreases the Efficacy of Adrenaline to Affect Passive Avoidance Retention of Adrenalectomized Rats," *Life Sciences,* 34 (1984) 241-258.

14. Borrell, J., de Kloet, E. R., Versteeg, D. H. G., and Bohus, B., "The Role of Adrenomedullary Catecholamines in the Modulation of Memory by Vasopressin," E. Endroczi (Editor), *Integrative Neurohumoral Mechanisms,* Amsterdam, Elsevier/North Holland (1983) 85-90.

15. Ashe, J. H., Cassady, J. M., and Weinberger, N. M., "The Relationship of the Cochlear Microphonic Potential to the Acquisition of a Classically Conditioned Pupillary Dilation Response," *Behavioral Biology,* 16 (1976) 45-62.

16. Ashe, J. H., Cooper, C. L., and Weinberger, N. M., "Mesencephalic Multiple-unit Activity During Acquestion of Conditioned Pupillary Dilation," *Brain Research Bulletin,* 3 (1978) 143-154.

17. Oleson, T. D., Vododnick, D. S., and Weinberger, N. M., "Pupillary Inhibition of Delay During Pavlovian Conditioning in the Paralyzed Cat," *Behavioral Biology,* 8 (1973) 337-346.

18. Oleson, T. D., Westenberg, I.S., and Weinberger, N. M., "Characteristics of the Pupillary Dilation Response during Pavlovian Conditioning in Paralyzed Cats," *Behavioral Biology,* 7 (1972) 829-840.

19. Weinberger, N. M., Oleson, T. D., and Haste, D., "Inhibitory Control of Conditional Pupillary Dilation Response in the Paralyzed Cat," *Behavioral Biology,* 9 (1973) 307-316.

20. Diamond, D. M., and Weinberger, N. M., "Physiological Plasticity of Single Neurons in Auditory Cortex of Cat during Acquisition of the Pupillary Conditioned Response: II. Secondary Field (AII)," *Behavioral Neuroscience,* 98 (1984) 189-210.

21. Konorski, J., *Integrative Activity of the Brain: An Interdisciplinary Approach.* Chicago, The University of Chicago Press, 1967.

22. Schneiderman, N., "Response System Divergencies in Aversive Classical Conditioning," A. H. Black and W. F. Prokasy (Editors), *Classical Conditioning II: Current Research and Theory,* New York Appleton-Century-Crofts (1972) 341-376.

23. Weinberger, N. M., "Effects of Conditioned Arousal on the Auditory System," A. L. Beckman (Editor), *The Neural Basis of Behavior,* New York, Spectrum Publications (1982) 63-91.

24. Mackintosh, N. J., "Varieties of Conditioning," N. M. Weinberger, J. L. McGaugh and G. Lynch (Editors), *Memory Systems of the Brain: Animal and Human Cognitive Processes,* New York, Guilford Press, 1985.

25. Oleson, T. D., Ashe, J. H., and Weinberger, N. M., "Modification of Auditory and Somatosensory System Activity during Pupillary Conditioning in the Paralyzed Cat," *Journal of Neurophysiology,* 38 (1975) 1114-1139.

26. Weinberger, N. M., Hopkins, W., and Diamond, D. M., "Physiological Plasticity of Single Neurons in Auditory Cortex of Cat During Acquisition of the Pupillary Conditioned Response: I. Primary Field (AI)," *Behavioral Neuroscience,* 98 (1984) 171-188.

27. Morest, D. K., "The Neuronal Architecture of the Medial Geniculate Body of the Cat," *Journal of Anatomy,* 98 (1964) 611-630.

28. Calford, M. B., and Webster, W. R., "Auditory Representation Within Principal Division of Cat Medial Geniculate Body: An Electrophysiological Study," *Journal of Neurophysiology*, 45 (1981) 1013-1028.

29. Ryugo, D. K., and Weinberger, N. M., "Differential Plasticity of Morphologically Distinct Neuron Populations in the Medial Geniculate Body of the Cat During Classical Conditioning," *Society for Neuroscience Abstracts*, 2 (1976) 435.

30. Ryugo, D. K., and Weinberger, N. M., "Differential Plasticity of Morphologically Distinct Neuron Populations in the Medial Geniculate Body of the Cat during Classical Conditioning," *Behavioral Biology*, 22 (1978) 275-301.

31. Weinberger, N. M., "Sensory Plasticity and Learning: the Magnocellular Medial Geniculate Nucleus of the Auditory System," C. D. Woody (Editors), *Conditioning: Representation of Involved Neural Function*, New York, Plenum (1982) 697-710.

32. Birt, D., and Olds, M., "Associate Response Changes in Lateral Midbrain Tegmentum and Medial Geniculate during Differential Appetitive Conditioning," *Journal of Neurophysiology*, 46 (1981) 1039-1055.

33. Gabriel, M., Miller, J. D., and Saltwick, S. E., "Multiple Unit Activity of the Rabbit Medial Geniculate Nucleus in Conditioning, Extinction and Reversal," *Physiological Psychology*, 4 (1976) 124-134.

34. Weinberger, N. M., Diamond, D. M., and McKenna, T. M., "Initial Events in Conditioning: Plasticity in the Pupillomotor and Auditory Systems," G. Lynch, J. L. McGaugh and N. M. Weinberger (Editors), *Neurobiology of Learning and Memory*, New York, Guilford Press (1984) 197-227.

35. Lynch, G., "Synapses, Circuits, and The Beginnings of Memory," Cambridge, Mass., MIT Press, in press.

36. Bliss, T. V. P., and Gardner-Medwin, A., "Long-Lasting Potentiation of Synaptic Transmission in the Dentate Area of the Unanesthetized Rabbit Following Stimulation of the Perforant Path," *Journal of Physiology London*, 232 (1973) 357-374.

37. Bliss, T. V. P., and Lomo, T., "Long-lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anesthetized Rabbit Following Stimulation of the Perforant Path," *Journal of Physiology London*, 232 (1973) 331-356.

38. Lee, K., Oliver, M., Schottler, F., and Lynch, G., "Electron Microscopic Studies of Brain Slices: The Effects of High Frequency Stimulation on Dendritic Ultrastructure," G. Kerkut, and H. V. Wheal, (Editors), *Electrical Activity in Isolated Mammalian C.N.S. Preparations*, New York Academic Press, 1981 189-212.

39. Lee, K., Schottler, F., Oliver, M., and Lynch, G., "Brief Bursts of High Frequency Stimulation Produce Two Types of Structural Changes in Rat Hippocampus," *Journal of Neurophysiology*, 44 (1980) 247-258.

40. Chang, F.-L. F., and Greenough, W. T. "Transient and Enduring Morphological Correlates of Synaptic Activity and Efficacy Change in the Rat Hippocampal Slice," *Brain Research*, 309 (1984) 35-46.

41. Wenzel, J., and Matthies, H., "Morphological Changes in the Hippocampal Formation Accompanying Memory Formation and Long-term Potentiation," N. M. Weinberger, J. L. McGaugh, and G. Lynch (Editors), *Memory Systems of the Brain: Animal and Human Cognitive Processes*, New York, Guilford Press 1985.

42. Lynch, G., Halpain, S., and Baudry, M., "Effects of High Frequency Synaptic Stimulation on Glutamate Binding Studied with a Modified *in vitro* Hippocampal Slice Preparation," *Brain Research*, 244 (1982) 101-111.

43. Lynch, G., Larson, J., Kelso, S., Barrionuevo, G. and Schottler, F., "Intracellular Injections of EGTA Block the Induction of Hippocampal Long-term Potentiation," *Nature*, 305 (1983) 719-721.

44. Guroff, G., "A Neutral, Calcium-activated Proteinase from the Soluble Fraction of a Rat Brain," *Journal of Biological Chemistry*, 239 (1964) 149-155.

45. Murachi, T., Hatanaka, M., Yasumoto, Y., Nakayama, N., and Tanaka, T., "Quantitative Distribution Study on Calpain and Calpastatin in Rat Tissue and Cells," *Biochemistry International*, 2 (1981) 651-656.

46. Simonson, L., Baudry, M., Siman, R., and Lynch, G., "Regional Distribution of Soluble Calcium Activated Proteinase Activity in Neonatal and Adult Rat Brain," *Brain Research*, 327 (1985) 153-159.

47. Schlaepfer, W. W., and Micko, S., "Chemical and Structural Changes of Neurofilaments in Transected Rat Sciatic Nerve," *Journal of Cell Biology*, 78 (1978) 369-378.

48. Baudry, M., and Lynch, G., "Regulation of Glutamate Receptors by Cations," *Nature*, 282 (1979) 748-750.

49. Baudry, M., and Lynch, G., "Regulation of Hippocampal Glutamate Receptors: Evidence for the Involvement of a Calcium-activated Protease," *Proceedings of the National Academy of Sciences*, (USA) 77 (1980) 2298-2302.

50. Baudry, M., Kramer, K., and Lynch, G., "Irreversibility and Time Course of Calcium Stimulated [^{3}H] Glutamate Binding to Rat Hippocampal Membranes," *Brain Research*, 270 (1983) 142 145.

51. Baudry, M., Bundman, M., Smith, E., and Lynch, G., "Micromolar Calcium Stimulates Proteolysis and Glutamate Binding in Rat Brain Synaptic Membranes," *Science*, 212 (1981) 937-938.

52. Siman, R., Baudry, M., and Lynch, G., "Purification from Synaptosomal Plasma Membranes of Calpain I, a Thiol Protease Activated by Micromolar Calcium Concentrations," *Journal of Neurochemistry*, 41 (1983) 950-956.

53. Siman, R., Baudry, M., and Lynch, G., "Brain Fodrin: Substrate for Calpain I, an Endogenous Calcium-activated Protease," *Proceedings of the National Academy of Sciences* (USA), 81 (1984) 3572-3576.

54. Siman, R., Baudry, M., and Lynch, G., "Regulation of Glutamate Receptor Binding by the Cytoskeletal Protein Fodrin," *Nature*, 313 (1985) 225-228.

55. Siman, R., Baudry, M., and Lynch, G., "Calcium Activated Proteases as Possible Mediators of Synaptic Plasticity," G. M. Edelman, W. M. Cowan, and W. E. Gall (Editors), *New Insights Into Synaptic Function*, New York, Wiley, in press.

56. Lynch, G., and Baudry, M., "The Biochemistry of Memory: A New and Specific Hypothesis," *Science*, 224 (1984) 1057-1063.

57. Baudry, M., Simonson, L., DuBrin, R., and Lynch, G., "A Comparative Study of Soluble Calcium-Dependent Proteolytic Activity in Brain," *Journal of Neurobiology*, in press.

58. Siman, R., Gall, C., Perlmutter, L., Christian, C., Baudry, M., and Lynch, G., "Distribution of Calpain I, an Enzyme Associated with Degenerative Activity in Rat Brain," *Brain Research*, in press.

59. Staubli, U., Baudry, M., and Lynch, G., "Leupeptin, a Thiol Proteinase Inhibitor, Causes a Selective Impairment of Spatial Maze Performance in Rats," *Behavioral and Neural Biology*, 40 (1984) 58-69.

60. Staubli, U., Ivy, G., and Lynch, G., "Hippocampal Denervation Causes Rapid Forgetting of Olfactory Information in Rats," *Proceedings of the National Academy of Sciences* (USA), 81 (1984) 5885-5887.

61. Olton, D. S., Becker, J. T., and Handelmann, G. E., "Hippocampus, Space, and Memory," *The Behavioral and Brain Sciences*, 2 (1979) 313-365.

62. Staubli, U., Faraday, R., and Lynch, G., "Pharmacological Dissociation of Memory: Anisomycin, A Protein Synthesis Inhibitor, and Leupeptin, a Protease Inhibitor, Block Different Learning Tasks," *Behavioral and Neural Biology*, 43 (1985) 287-297.

63. Eichenbaum, H., Morton, T. H., Potter, H., and Corkin, S., "Selective Olfactory Deficits in Case H.M," *Brain*, 106 (1983) 459-472.

64. Potter, H., and Butters, N. M., "An Assessment of Olfactory Deficits in Patients with Damage to Prefrontal Cortex," *Neuropsychologia*, 18 (1980) 621-628.

65. Squire, L. R., "The Neuropsychology of Human memory," *Annual Review of Neuroscience*. 5 (1982) 241-273.

66. Kamin, L. J., "Predictability, Surprise, Attention, and Conditioning," M. R. Jones (Editor), *Miami Symposium on the Prediction of Behavior, Adversive Stimulation*, Coral Gables, University of Miami Press, 1968.

67. Kamin, L. J., "'Attention-like' Processes in Classical Conditioning," B. A. Campbell, A. Baron and R. M. Church (Editors), *Punishment and Adversive Behavior*, New York, Appleton-Century-Crofts, 1969.

68. Mackintosh, N. J., *Conditioning and Associative Learning*. New York, Oxford University Press, 1983.

69. Rescorla, R. A., and Wagner, A. R., "A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Non-reinforcement," A. H. Black and W. F. Prokasy (Editors), *Classical Conditioning II: Current Research and Theory*, New York, Appleton-Century-Crofts, 1972.

70. Anderson, J. R., *Architecture of Cognition*. Cambridge, Mass., Harvard University Press, 1983.

71. Marr, D. *Vision*. San Francisco, W. H. Freeman and Co., 1982.

72. Rescorla, R., "Predictability and Number of Pairings in Pavlovian Fear Conditioning," *Psychonomic Science*, 4 (1966) 383-384.

73. Rescorla, R. A., "Pavlovian Conditioning and Its Proper Control Procedures," *Psychological Review*, 74 (1967) 71-80.

74. Rescorla, R. A., "Probability of Shock in the Presence and Absence of CS in Fear Conditioning," *Journal of Comparative and Physiological Psychology*, 66 (1968) 1-5.

75. Granger, R. H., Schlimmer, J. C., and Young, M. T., "Contingency and Latency in Associative Learning: Computational, Algorithmic and Implementation Analyses," J. Davis, E. Wegman, and R. Newburg (Editors), *Brain Structure, Learning and Memory*, in press.

76. Granger, R. H., "Components of Episodic Learning: The CEL Model of Early Learning and Memory," *Journal of Cognition and Brain Theory*, 6 (1983) 5-38.

77. Hawkins, R. D. and Kandel, E. R., "Is There a Cell-biological Alphabet for Simple Forms of Learning?" *Psychological Review*, 91 (1984) 376-391.

78. Alkon. D. L., "Membrane Depolarization Accumulates During Acquisition of an Associative Behavioral Change," *Science*, 210 (1980) 1375-1376.

79. Gluck, M. A., and Thompson, R. F., "A Computational Approach to Modeling the Neural Substrates of Associative Learning and Memory," *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, University of California, Irvine (1985) 36–41.

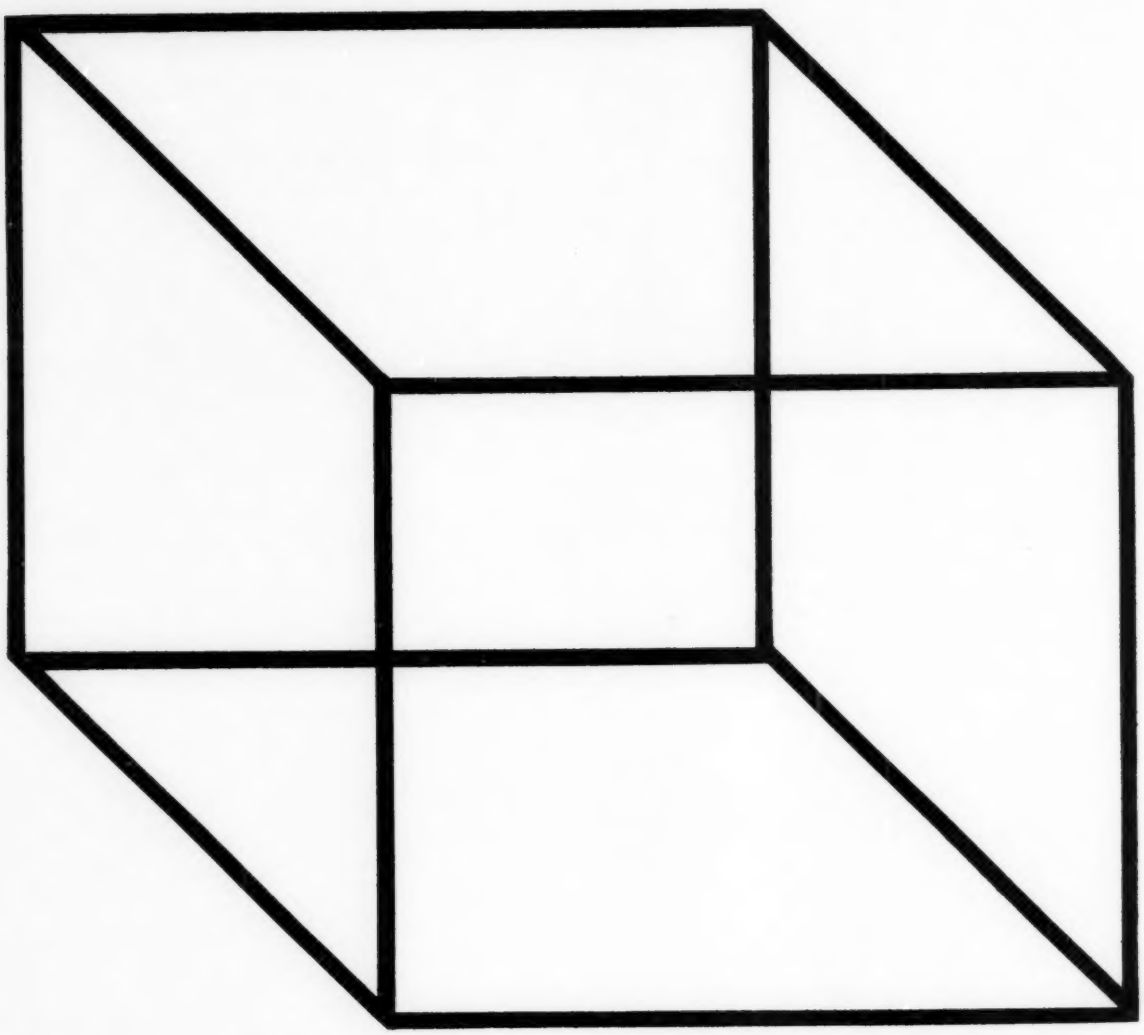
80. Sutton, R. S., and Barto, A. G., "Toward a Modern Theory of Adaptive Networks: Expectation and Prediction," *Psychological Review*, 88 (1981) 135-170.

81. Gibbon, J., Berryman, R., and Thompson, R. L., "Contingency Spaces and Measures in Classical and Instrumental Conditioning," *Journal of the Experimental Analysis of Behavior*, 21 (1974) 585-605.

Computational Neuropsychology:

A New Perspective on Mental Imagery

Stephen M. Kosslyn, Harvard University



Historically, in America neuropsychology and cognitive psychology have been distinct enterprises, with neuropsychology being largely clinical in focus and cognitive psychology being experimental in focus. However, cognitive methodologies have been profitably used in conjunction with clinical populations in Europe for decades.[1,2] The value of combining the two approaches is now becoming apparent to a wide audience, as witnessed by the founding last year of the journal *Cognitive Neuropsychology*. However, the power of cognitive/neuropsychological methodologies is greatly amplified when combined with computational theorizing. Computational theories specify the components of information processing in a way that can be programmed into a computer, and hence are precise and generate explicit predictions. There has been dramatic recent progress in studying cognitive processing due to this new approach. In this paper, we will illustrate this "computational neuropsychological" approach by considering one cognitive activity, mental imagery.

To look at them, computers have almost nothing in common with brains, and it is not clear how neuropsychology or cognitive methodologies can be augmented by considering ways of programming computers. To see the salubriousness of a combined approach, we must first appreciate the relevance of computers to understanding the brain. It may be most useful to begin by considering the stomach. Would you be satisfied that you had truly understood the nature of the stomach if all you had was a detailed description of its physical composition? Say you knew its precise shape, which cells excreted which enzymes in what amounts, and so on; would that be enough? Probably not. You probably would want to know what all of this was *for*. Understanding why the stomach is constructed as it is requires an understanding of its role in digestion. You need to know what kinds of material go into the stomach and what are the requirements of organs that receive the output from the stomach. To truly understand the stomach, you need to understand the "problems" that it must "solve" (breaking down specific compounds into specific constituents) in the context of the digestive system as a whole.

Similarly, a complete understanding of the brain will require more than a description of its physical composition (cells and their connections, various chemical and electrical interactions); if someone provided you with only a complete physical description of the brain, including a detailed blueprint of all the connections and pathways, this probably would not be entirely satisfying. We also need to know what all of this is for. The function of brains (or at least cortices, which "mental" functions carry out) is to process information. That is, the brain is the organ that allows us to recognize friends and foes, to store information, to communicate, and so on. The way it carries out these tasks involves the processing of incoming sensory information and the use of stored information. Thus, a complete understanding of the brain requires an understanding of how it processes information.

The notion of function is, of course, hardly foreign to neuroscientists. But the language of computation provides the only precise and rigorous vocabulary for understanding information processing. By studying what seems to be required to build a computer model, we learn something about the nature of organic intelligences. Constructing computer programs that process information the way brains do helps us to discover the problems that must be solved for an entity to see, learn, and so on, which are the analogues to the stomach's job of breaking down foods into usable constituents.

Thus, although they may not look alike, at one level of analysis computers work like brains. But a computer is a general purpose machine; it operates only in accordance with precise instructions. A computer will mimic a brain's information processing only if one programs (instructs) it appropriately. How does one develop a theory that can guide construction of a computer simulation model? A promising answer to this question was provided by the late David Marr.[3] In this paper, we will consider how Marr's approach can be applied to cognitive processes, using mental imagery as a concrete example. Thus, let us now consider the benefits of a computational neuropsychology of mental imagery.

Mental Imagery

Visual mental imagery is a kind of "seeing" in the absence of the appropriate immediate sensory input; it is "re-seeing" information previously seen and stored in memory. Imagery is an important component of many kinds of thinking. For example, one can anticipate the trajectory of a moving object by imaginally projecting its path, or one can discover an efficient way of packing irregularly-shaped objects into a container by imagining them arranged in different ways. It is important to understand imagery not only for practical reasons (e.g., training people to be better navigators), but because it is a fundamental cognitive ability that is used in numerous tasks.

As the vocabularly used to characterize imagery suggests, there are some deep similarities between imagery and like-modality perception. Consider the purposes of visual perception: first, we use vision to recognize objects and, secondly, we use vision for navigation and tracking. The purposes of imagery neatly parallel these purposes of vision: First, "recognition" is used in imagery as a way of recalling visual information. That is, imagery is used when one tries to answer questions from memory about subtle visual properties of objects, resulting in one's imagining an object and "recognizing" previously-unconsidered characteristics. For example, imagery tends to be used when one tries to answer questions like: "what shape are a beagle's ears?" "which is larger, a goat or a hog?" or "which is higher off the ground, a race horse's knees or the tip of its tail?" One is "recycling" stored visual memories, internally "looking" at the object again and recognizing previously-unconsidered properties.

Second, mechanisms used in navigation and tracking also may be used in the service of visual thinking. That is, one can mentally "move" imaged objects, or can "move" the self in relation to imaged objects, to "see" what would happen in the corresponding actual situation. For example, one might want to anticipate whether a vehicle will fit in a space between two trees, and so might project an image of the vehicle forward to "see" if it will fit; or one might want to know whether a tree would bridge a stream if it were chopped down and made to fall in a certain direction, and so might image the tree falling and "see" whether it would cross the stream.

There is much empirical data indicating that imagery and like-modality perception share common mechanisms. For example, if one is holding a visual image (e.g., of a flower), this will impair visual perception more than it impairs auditory perception, but vice versa if one is holding an auditory image (e.g., the sound of a telephone ringing[4]). Perhaps most interesting, manipulating objects in images reveals time-courses like those observed in the real world. For example, Figure 1 illustrates pairs of stimuli used in a classic study by Shepard and Metzler.[5] They asked subjects to decide if the objects were the same or different shape, irrespective of their orientation. Figure 2 presents the results, indicating a highly linear increase in decision time as more mental rotation was required to bring the forms into congruence. This result is impressive because images are not actual, rigid objects and hence are not constrained by physics to have to pass through intermediate positions when the orientation of an imaged object is changed. Similar results are obtained with image scanning. Kosslyn, Ball and Reiser asked subjects to close their eyes and imagine the map illustrated in Figure 3.[6] This map had seven locations, which were positioned such that there were 21 distinct inter-location distances between all possible pairs. The subjects began by "focusing" on a given location on the imagined map (e.g., the tree), and then decided whether a second named location was or was not on the map (e.g., the hut versus a bench); they were asked to respond in the affirmative only after they had the second object clearly in focus. As is evident in Figure 4, increasingly more time was required to scan between pairs of locations that were increasingly farther apart on the map, indicating that an imaged map can "stand in" for the actual one. Kosslyn[7] reports another finding that is especially suggestive: If an object is imagined at a small size, more time is required to "see" its parts than if the object is imagined at a larger size.[8,9] This result is intriguing because it suggests that objects in images are subject to spatial summation, a well-known property of neural mechanisms used in vision.

Figure 1:

Stimuli used in the Shepard and Metzler experiment on mental rotation.[5] The stimuli in pairs 1 and 2 are congruent when rotated appropriately; the stimuli in pair 3 cannot be rotated into congruence.

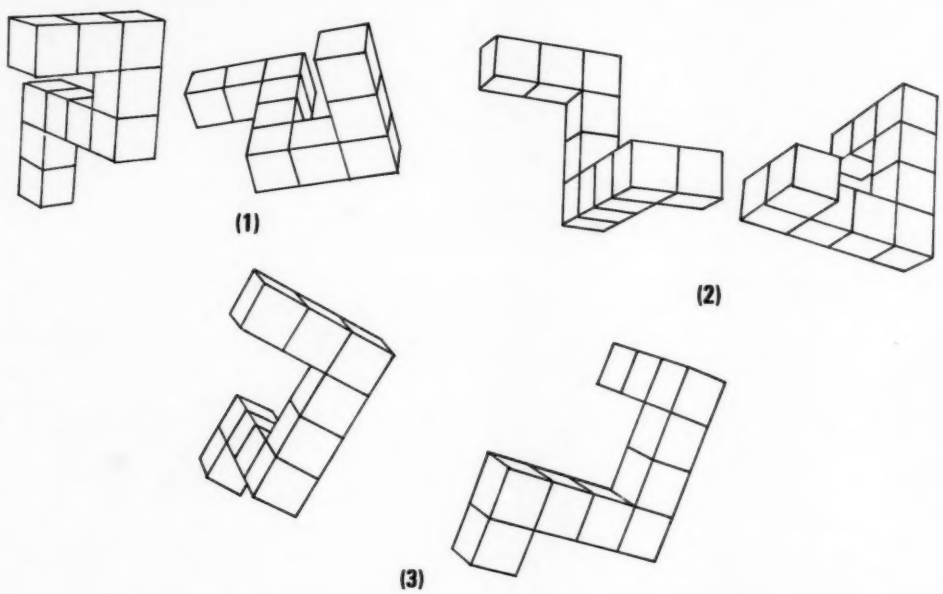


Figure 2:

Results from the Shepard and Metzler experiment on mental rotation.[5] Time to compare the shapes of forms increases linearly with the angular disparity between the forms. Rotation can be carried out both in the picture plane and in depth.

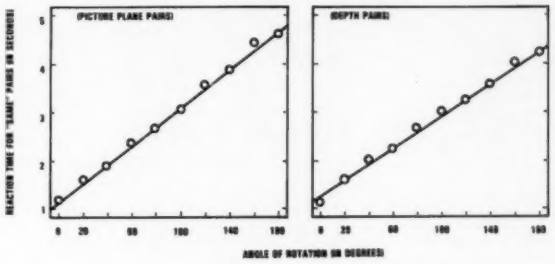
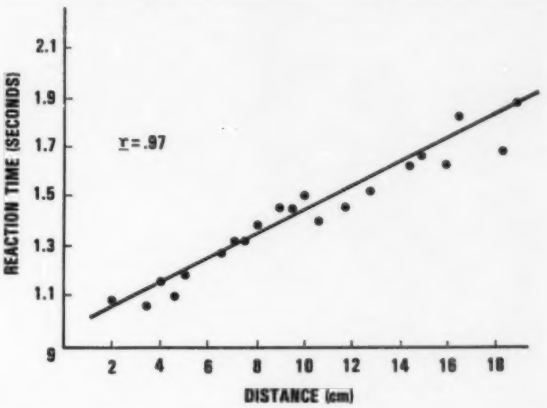


Figure 3:

The fictional map used in the Kosslyn, Ball and Reiser experiment on image scanning.[6] The map was constructed so that there were 21 distinct distances between all possible pairs of seven locations.

Figure 4:

Results from the Kosslyn, Ball and Reiser experiment on image scanning.[6] Time to scan across the imaged map increased linearly with the distance scanned.



Three Problems in Vision

Although the behavioral phenomena are very interesting, they do not serve to specify a precise theory of how the brain is forming and using mental images. However, such data become very useful when we consider them in combination with neuropsychological findings and computational theorizing. It is at this point that one of Marr's ideas becomes relevant. We can now consider an example of what Marr calls a "theory of the computation."[3] A computation can be regarded as a "black box" that transforms input in a systematic way. A theory of a computation specifies *what* must be computed and why. Such a theory justifies positing a given computation by an analysis of what problems must be solved and the requirements on the solution to the problems. That is, the goal of a computation is specified, as well as the nature of the input and constraints on the solution. This sort of theory is to be distinguished from a "theory of the algorithm," which specifies the specific steps actually used to carry out a computation. The theory of the computation is a first step, outlining the basic "processing components" that should be included in the theory; the theory of the algorithm fleshes out the details of how the computations are performed. In this paper, we will concentrate on the first level, focusing on a theory of the computations used in one aspect of imagery.

We can use the inference that imagery shares mechanisms with perception to discover a remarkable amount about the structure of the information processing system underlying imagery. We do so by first considering some fundamental problems that must be solved by a visual system. Our brains have apparently solved these problems in specific ways, and the outlines of these solutions are now apparent in the literature on the neurophysiology and neuroanatomy of visual perception; these solutions have direct implications for a theory of visual processing.

Thus, in this section we will begin to develop theories of some of the computations that are performed by the visual system. We will do so by considering three problems (which must be solved by any visual system) and the apparent solutions to these problems adopted by primate brains. In the following section, we will explore the implications of these inferences for a theory of imagery, assuming that visual imagery makes use of visual processing mechanisms.

1. The Problem of Position Variability

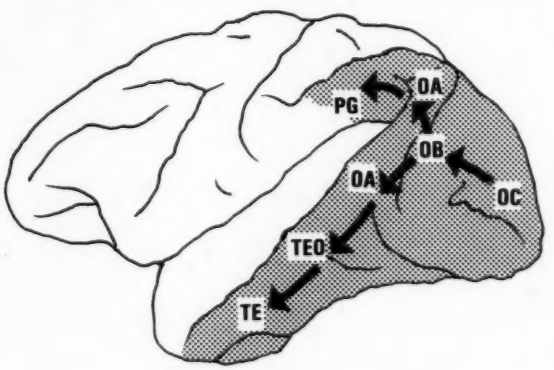
The same object is likely to occur at various positions in the visual field. Nevertheless, once we have seen an object, we can recognize it just as easily when it subsequently is in a different position in the field. Logically, there are two ways we could perform this feat: On the one hand, when an object is encoded initially, the visual system could associate a separate representation with each of the possible positions of the object. This seems to be the choice made by McClelland and Rumelhart in their theory of word perception.[10] They associate a representation of each letter with each position in the field. (This was done so that the same letter could be detected in more than one position in a word.) On the other hand, when an object is encoded initially, it could be stored using representations that are associated with a set of positions in the field. In the limit, only one representation would be used for all positions. This is the solution Marr offered for the position variability problem; Marr suggested that the appearance of objects is stored in "object-centered" representations.[3] In such representations, the locations of parts of objects are specified relative to other parts, not to positions in space.

The solution adopted by primate visual systems to the problem of position variability is now evident in the neurophysiological literature: It has been found in primates that visual cells in area TE (near the anterior end of the inferior temporal lobe) have very large receptive fields, and respond when patterns are present over a wide range of positions (the receptive field sizes are usually larger than 20 × 20 degrees of visual angle). This area of the brain has been shown to be critically involved in recognition per se.[11] Thus, the primate's solution to the position variability problem relies on not representing the position of a pattern in the high-level shape representation system. (Incidentally, this is a good example of how facts about the neurological underpinnings of behavior can have direct bearing on theories of cognition; this finding is a significant challenge to the McClelland and Rumelhart model.)[10] One implication of this solution is that only one shape can be recognized at a time (although we could rapdily switch back and forth between stimuli, only one would be processed at any given instant); if multiple stimuli were being processed simultaneously, the large receptive fields would often result in the system's not being able to tell if there is one stimulus or two of the same stimuli being presented in different locations (e.g., the letter A in a word). Hence, figure/ground segregation is necessary to isolate individual patterns before they can be processed further. If this is done, then duplicate patterns can be isolated and processed separately, preventing confusions about how many of a pattern are present in the field.

However, we *do* know where an object is when we see it. Thus, there must be a separate representation of an object's location, which implies two separate mechanisms—one to represent a shape independently of its position and one to represent its position. And in fact, Ungerleider and Mishkin summarize evidence for "two cortical visual systems."[12] Their claim is that the ventral system, running from OC (primary visual cortex) through TEO down to TE, is concerned with analyzing *what* an object is, whereas the dorsal system, running almost directly from circumstriate area OB to OA and then to PG (in the parietal lobe) is concerned with analyzing *where* an object is. Figure 5 illustrates the relevant areas of the brain.

Figure 5:

Visual areas of a rhesus monkey; these areas directly correspond to analogous areas in other primate brains. Ungerleider and Mishkin described two cortical visual systems; the "dorsal system" runs through area OA to area PG in the parietal lobe, whereas the "ventral system" runs through area OA to area TE in the inferior temporal lobe.[12,13]



Two sorts of data are relevant to Ungerleider and Mishkin's claim. First, the neuroanatomy and physiology support this distinction. There are well-known neural connections running along both pathways, and the visual properties of these areas have been well documented.[13] The visual areas of the parietal lobe appear to have different properties than those of the ventral visual system (e.g., they include the fovea in their receptive fields less often). Second, behavioral evidence suggests that animals are severely impaired in their ability to learn to discriminate between patterns if the inferior temporal lobes are removed. However this lesion does not disrupt their ability to learn locations. On the other hand, if the parietal lobes are removed, animals are severely impaired in their ability to discriminate on the basis of location, although they retain the ability to discriminate between patterns (see Ungerleider and Miskin, 1982a,b).

This interesting design of the processing mechanisms leads to difficulties that must be overcome by the system, as is evident when we consider another problem of visual perception.

2. The Problem of Figure/Ground Segregation.

Before we can recognize an object, "figure" must be segregated from "ground"; one must somehow pick out regions that are likely to correspond to distinct objects. The magnitude of difficulty of this problem becomes evident if you look at a digitized representation of a picture, with numbers representing the intensity of light at each point; the objects are overwhelmed by differences in lighting, texture, and so on, and it is very difficult to pick them out. A figure must be selected on the basis of physical properties of the input, such as regions of homogeneous color or texture, or contiguous zero-crossings in the second derivative of the function relating intensity to position (which occur at the edges of objects).[3] That is, because one has not yet identified the object (segregating its form from the background is a logical prerequisite to recognition), one can only use physical parameters to parse figure from ground. There are numerous proposals in the computer vision literature for ways of organizing input into regions likely to correspond to figures.[14]

An interesting problem arises here because the visual system processes input at different spatial frequency bandwidths.[15] Higher spatial frequencies correspond to more light/dark alternations per degree of visual angle; thus, higher resolution is required to detect higher spatial frequencies. The system can be described as having a number of different "channels," each differing in resolution. At average viewing distances, the lowest spatial frequency channel produces an output that will often correspond to the general shape envelope of an object. But consider what will happen at higher spatial frequency channels: the same factors that result in the parse of the object from the background will result in parts of a single object (e.g., the arms, legs, and head of a person) being parsed from one another. That is, the system cannot "know" what is an object and what is a part of an object; it just organizes regions in the input on the basis of physical parameters of the input array. And here lies a difficulty: once parsed, the shape representation system "ignores" the location in the visual field of the stimulus. Thus, the representation of the shapes of the parts will not preserve their positions. But the arrangement of parts is important for many recognition tasks. The relations must be represented somehow.

The most straightforward solution to this problem requires a minor revision to the Ungerleider and Mishkin theory. It seems clear that "what" and "where" are not so distinct conceptually: sometimes the spatial relations among the parts are critical for identifying the form; for example, the difference between a *Q* with a long tail and an *O* with a diagonal slash through it is a matter of where the diagonal is positioned. Rather than "where," the dorsal system seems specialized for representing spatial relations, including those among parts of a single object. The relations among high-resolution representations of parts presumably are represented the same way as are the spatial relations among separate objects in a scene. (However, note that the parts and their relations are also implicit in the low-resolution, low spatial frequency representation; for example a handle of a mug will be a bulge on the blob-like representation of the mug.) In the relevant experiments, animals have never been required to discriminate among patterns that differ only in the relations among parts; usually stimuli differ in terms of numerous features, and the relationships among them are not important (e.g., as is true for the square and plus sign used by Ungerleider and Mishkin, 1982b, which can be discriminated between simply by looking in the center of the figure and seeing if there is a line). In short, it would appear that once figure is segregated from ground, regardless of whether "figure" is an object or a part of an object, the location of that figure is represented in the parietal lobes.

3. The Problem of Non-rigid Transformations.

We can gain some insight into the way spatial relations are represented by considering another problem that must be contended with by a visual system: namely, the problem that many objects are subject to a near-infinite number of transformations, and so may not look the same from instance to instance. For example, a human form can be configured in a huge number of different ways, crouching, arms raised, standing on one toe with the arms held out to the side, and so on. Similarly, letters of the alphabet can occur in numerous fonts, which are not simply linear transformations of each other. We cannot store a separate representation of all the possible configurations of such objects, with the aim of being able to match input to a specific stored representation—there are simply too many possible configurations, and one often may encounter configurations not previously seen. Thus, it is useful to have a representation that will be general across a wide range of transformations. Two kinds of attributes remain constant under such transformations: First, the individual parts remain the same; although some may be hidden depending on the configuration, no parts are

actually added to or deleted from the object. Second, the topological-like relations among parts remain constant under all of these transformations. Topological relations are more abstract than the precise relative position of two parts as they appear in any given case (i.e., the topographic relations); they indicate which parts are connected to each other and which are contained within each other. For example, the topological relation between the arm and shoulder remains constant under all of the different positions the arm can take. However, literally topological relations are too weak; the relation of ears to the side of the head, or the thumb to a hand, are important and will remain constant under transformations. Thus, some general categories of relations must be used, not the actual topographic appearances.

This problem places requirements on what the dorsal system must do: this system must be able to derive a description of relations which will remain true no matter how the object is configured. These descriptions themselves cannot be images; images are concrete, being representations of specific instances. Instead, the dorsal system must be able to make use of more abstract, "categorical" representations. Such representations capture general properties of a relationship without specifying the details (e.g., "left of" without specifying how much or exactly what angle).

Finally, it would seem necessary that at some point the representations of perceptual units and their relationships must come together. A possible locus of that nexus is the association cortex near Wernicke's area (in the posterior, superior temporal lobe), which appears to be involved in semantic processing. This arrangement is somewhat awkward, in that the relations must be delivered in synchrony with the related units; if the inputs fall out of phase, one may make "illusory conjunctions." That is, one may conjoin units using the wrong relations. Interestingly enough, *Treisman and Gelade* report that humans will see just such illusory conjunctions when they are pushed to perform well in a difficult task.[16]

Computations Used in Imagery

Given the sorts of empirical findings discussed in the introduction to this paper, it seems safe to assume that visual mental images are formed on the basis of representations initially encoded during perception. If so, then we are in a position to exploit our analysis of object representation to formulate a theory of the computations used in imagery. Before continuing, it may be useful to summarize where we have arrived up to here. The visual system appears to have two classes of mechanisms. The "ventral system" forms representations of an individual part or the overall shape envelope. Because the receptive fields are so large, this system does not register the locations of parts; hence, only one part can be processed at a time (otherwise it would not be able to tell when two of the same part—such as the same letter repeated in a word—are present). The representation of shape used in the ventral system should be concrete, capturing the precise shape and surface details of the part or object. This system does not processs relations among parts, except insofar as they are implicit in a low resolution representation of the entire object (a kind of blurry blob-like form). In contrast, we are led to assume that the "dorsal system" must be able to derive abstract representations of the spatial relations among parts or objects. The use of categorical representations of spatial relations (e.g., "connected to") is especially appropriate for classes of objects whose members are subject to non-rigid transformations. In such cases, the parts can be arranged in a large number of topographical configurations.

Image Generation

A fundamental question about imagery is, how are mental images formed? That is, when you try to decide what shape a beagle's ears are, you will probably visualize the dog's head and "look" at its ears. Until you tried to provide the information, there was no image. How was the image created? In this section we will focus on one aspect of an overall theory of imagery, namely the theory of image generation. Our discussion of the three problems in vision has direct implications for a theory of the computations used in image generation. In this section, we will draw out these implications, and then will turn to an empirical test of these ideas in the following section.

Presumably, the stored visual representations used in recognition can also be used in imagery, being activated to form an image. These representations are "concrete", containing enough information to allow one to reconstruct the actual appearance. Image generation requires a computation that activates representations stored in memory; we can call this the PICTURE computation. The purpose of this computation is to make explicit the topographical layout of parts of an object, which allows one to note previously unnoticed spatial relations (when imagery is being used in memory retrieval) and to "see" what would happen if one were moving through space or the object were manipulated (when imagery is being used for visual thinking). Thus, the PICTURE computation presumably produces a pattern of activation in a "visual buffer"; this pattern of activation is an "image representation". The "visual buffer" is a functionally-defined storage medium that probably corresponds to the joint operation of numerous topographically organized areas of cortex.[17] A pattern of activation in this structure makes explicit the local geometric relations among parts of an object. We assume that these visual parts of cortex also can be activated from stored information, resulting in a mental image. This buffer is equivalent to the buffer that supports Marr's "2 1/2 D sketch" in vision.[3]

The PICTURE computation activates an image of a single stored unit, which corresponds either to a part or a low-resolution image of the entire object (provided it was encoded). Because the spatial relations among parts are represented as descriptions, and presumably stored in that form, other computations must be used if a multipart or detailed object is to be imaged. For one, we need a computation that can access stored descriptions of relations and use

them to juxtapose separate parts in the correct relative positions in an image. Let us call this the PUT computation. Part of what this computation must do is look up and interpret a description of how parts are to be arranged. For example, in generating a detailed image of a car, it might look up "front wheel" and discover the location description "under front wheelwell."

According to our analysis of the problem of position variability, units are encoded sequentially. This analysis, in conjunction with our analysis of requirements on representing spatial relations among parts of non-rigid objects, suggests that local part-by-part relations are likely to be stored for non-rigid objects (such as part 1 is "left of," "above," or "connected to" part 2). Local part-by-part spatial relations are easiest to encode if parts are encoded one after the other, and are likely to remain constant under transformations of the object. If so, then the stored descriptions of inter-part relations will specify how one part is positioned relative to another (previously encoded) part. Thus, in generating multi-part images one must know where a reference point (e.g., the "wheelwell," for a car's wheel) is before being able to position another part correctly in an image. In order to locate a reference point, a third computation, which we can call the FIND computation, must be used. The FIND computation takes as input a pattern in the visual buffer and produces a label for it (obviously, this computation is in fact a collection of sub-computations). It is used to locate the "foundation part" on an already-imaged part (i.e., where the new part belongs, "front wheelwell" in the example). In order to position a new part correctly, the PUT computation must use the output from the FIND computation (e.g., the location on the body where the front wheelwell belongs) plus the description of the relation ("under") to compute parameter values for the PICTURE computation. These values are provided to the PICTURE computation, allowing the new part to be imaged in the correct relation to the foundation part.

Brain Damage and Image Generation

The theory of image generation developed here is a theory that posits a set of distinct computations carried out by the brain. One source of support for such a theory would be the finding of "functional dissociations" among the putative components. That is, some type of brain damage may leave some of the components intact while disrupting the others. As a specific example, let us review recent evidence supporting the existence of a distinct PUT computation.

Kosslyn, Holtzman, Farah, and Gazzaniga investigated a possible dissociation between the FIND and PICTURE computations on the one hand, and the PUT computation on the other hand.[18] These studies examined the isolated cerebral hemispheres of "split-brain" patients, who have had their corpus callosa—the major connection between the hemispheres—surgically severed (for medical reasons). Because the PUT computation putatively involves operations also used in language (accessing and interpreting descriptions), we expected that it might typically be more effective in the left cerebral hemisphere, which is specialized for language processing in most people (for a more precise theory underlying this expectation, see Kosslyn reference 19). In contrast, because both hemispheres can recognize and remember shapes, we expected the PICTURE and FIND computations to be equally effective in both hemispheres. That is, the PICTURE computation simply activates stored visual representations, which exist in both hemispheres, and the FIND computation identifies patterns, which also is done in both hemispheres. Thus, we hoped to show that the isolated left hemisphere could perform image generation tasks that involve all three computations, whereas the right hemisphere could only perform image generation tasks that require the PICTURE and FIND computations but not the PUT computation. This prediction is especially interesting because the common wisdom has it that the right cerebral hemisphere is the seat of mental imagery.[20,21] Thus, if we can show that the left hemisphere is selectively able to perform imagery tasks better than the right hemisphere, counter to the common sense notions about imagery and the right hemisphere, this demonstration will be good evidence of the usefulness of the computational approach. In addition, in the course of implicating a specific processing deficit in the right hemisphere, we will see the usefulness of cognitive methods applied to the investigation of neuropsychological phenomena.

Before beginning the investigation proper, preliminary experiments were conducted with normal subjects to show that people actually use imagery to perform the tasks. The task of most interest simply required subjects to decide from memory if upper case letters of the alphabet are composed only of straight lines or include some curved lines. The upshot of the preliminary work was that imagery is indeed used in this task, and that images of upper case letters are usually generated a part at a time—which putatively requires descriptive relations among the parts. The weak link here, of course, is the assumption that because letters are imaged a part at a time, they must be imaged on the basis of a stored description. This claim is reasonable in light of the fact that letter fonts differ widely, and hence letters are objects subject to non-rigid transformations. If so, then one typically will not store representations of particular letters (e.g., a B seen on page 4 of yesterday's *Times*), but will store an abstract description of the relations among the parts, and this description later be used to form an image.

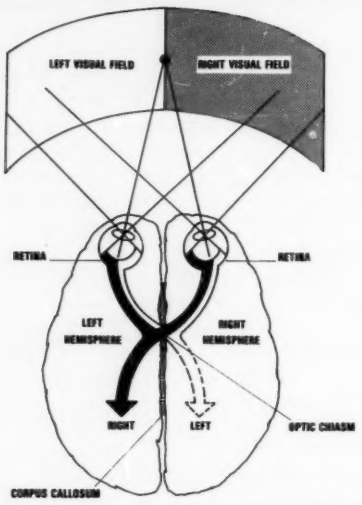
An Image Generation Deficit

The visual system is constructed conveniently for the scientist interested in presenting a stimulus to only a single hemisphere of the brain. As is illustrated in Figure 6, the right half of each retina projects only to the right cerebral hemisphere, and the left half of each retina projects only the left cerebral hemisphere. Thus, if a subject is asked to stare at a point directly in front of him or her, a stimulus flashed to the right side will strike only the left half of each retina—resulting in its being projected to only the left hemisphere;

similarly, a stimulus flashed to the left side will project only to the right hemisphere. If the stimulus is presented for under about 200 msec, the subject will be unable to move his or her eyes before it is gone (an eye movement requires about 200 msec to perform), ensuring that the stimulus falls only one half of each retina.

Figure 6:

A schematic top view of the visual system illustrating how the left half of each retina projects only to the left cerebral hemisphere and the right half of each retina projects only to the right cerebral hemisphere.



To get a sense of how computational theories can be tested using cognitive neuropsychological methodologies, we will focus on a series of experiments done with split-brain patient, J.W. This patient had his corpus callosum sectioned about three years prior to testing, as a treatment for severe intractable epilepsy; he has been extensively tested and it has been found that his right hemisphere is capable of comprehending involved verbal instructions and of making simple deductions and classifications.[22]

In the first experiment, J.W. was asked to look straight ahead, and read the lower case letters flashed to either side. These letters were exposed for only 150 msec, ensuring that only one hemisphere saw them. J.W. was to decide whether or not the corresponding upper case version of each letter had any curved lines. He pressed one button if he thought the upper case letter had curves, and another if he thought it only had straight lines. The left arm was used for all responding (due to ipsilateral efferents, both hemispheres have some control over the major arm movements).

The results of the first sets of trials were straightforward: J.W. performed almost perfectly in his left hemisphere, and showed severely degraded performance in his right hemisphere.

A task analysis suggests that this deficit could be due to problems in carrying out any one or combination of seven processing stages. That is, to peform this task successfully one must: ENCODE the cue letter, ACCESS the upper case representation, GENERATE the image, RETAIN the image long enough to use it, INSPECT the image, EVALUATE the shape, and RESPOND. Furthermore, the deficit could be due to the right hemisphere's not understanding the instructions and/or to its not being able to combine stages. Thus, we had to show that the right hemisphere's deficit could be ascribed to difficulty in generating the image per se, and then we had to try to isolate which image generation computation or computations was deficient.

A number of control experiments were conducted to implicate a deficit in image generation per se in the right hemisphere. For example, in one control experiment the upper case letters themselves were presented, and J.W. was asked to classify what he saw; both hemispheres performed virtually perfectly. In another experiment, 3-letter words composed of upper case letters (e.g., MUG) were presented to one hemisphere or the other, and J.W. was cued as to which letter (first, second, or third) to classify as being straight or curved. The cue was given either two seconds before the word was presented, making it a perceptual classification task, or two seconds after the word was presented, requiring him to retain and classify an image. There was no difference in how well each hemisphere performed in the imagery and perceptual conditions. Thus, the problem was not that the right hemisphere a) could not ENCODE the letters, b) could not RETAIN an image, c) could not INSPECT an image, d) could not make the straight/curved EVALUATION, e) could not RESPOND, of f) could not understand the instructions. In addition the word-task required combining these stages together. In another task, it was found that the right hemisphere could use a lower case letter cue to ACCESS the other-case representation: when shown the lower case letter, both hemispheres could pick out the corresponding upper case version. Numerous other control experiments were conducted to provide convergent evidence for the inference that all stages but image generation were intact in the right hemisphere of this patient.[19]

Implicating a Computation

The results described so far only serve to demonstrate an image generation deficit in J.W.'s right hemisphere. They do not implicate a deficit in one computation or another. In order to implicate the PUT computation it must be shown that both hemispheres can use the PICTURE and FIND computations. These are the only computations required to generate images of a single part or overall shape; when multiple parts are composed, the PUT computation is putatively required. If so, then J.W.'s right hemisphere should be able to form an image of the general shape envelope of an object, which is presumed to be encoded as a single perceptual unit at a low spatial frequency channel (as a degraded blob-like form, similar to the appearance of a blurred slide). And indeed, it was demonstrated that both of

J.W.'s hemispheres could perform tasks requiring the imaging of general shapes: In one experiment, the names of animals were presented to one hemisphere or the other, and J.W. was asked whether that animal was larger than a goat (pressing one button to indicate yes and another to indicate no). The named animals were of very similar sizes to a goat (e.g., hog, wolf), and this task previously had been shown to require imagery in normal subjects (see chapter 9 of reference 8). Thus, it is of interest that both of J.W.'s hemispheres performed this task virtually perfectly. In another experiment, it was found that J.W. also could decide equally well in both hemispheres whether a named object (e.g., a cup) was higher than it is wide. These tasks require the PICTURE computation to generate images of the general shapes of the objects, and the FIND computation to inspect the objects in the image. Because no parts need to be added to the general shape, the PUT computation is not necessary.

To make the case that there is a deficit in one particular computation in J.W.'s right hemisphere, we must consider one more result. So far the results demonstrate that J.W.'s right hemisphere cannot generate images of upper case letters, but can image and compare object shapes. Perhaps the problem is in generating images of letters per se. To eliminate this possibility, the animal names used in the size judgment experiment were again presented separately to each hemisphere. Now, however, J.W. was to decide whether each animal's ears protrude above the top of the skull or not (e.g., horse versus sheep). This task requires imaging the ears in the correct relative location to the head. In this task, J.W.'s right hemisphere performed at chance—in contrast to the virtually perfect performance exhibited by the left hemisphere. This was true even though the right hemisphere judged the same physical stimuli (animal names) virtually perfectly in the size comparison task. Thus, it appears that the right hemisphere has selective difficulty when parts must be arranged correctly in an image.

These experiments have also been conducted on another split-brain patient, V.P., which also produced evidence for a right-hemisphere deficit in image generation. Unlike J.W., however, V.P. eventually learned to generate images of letters in her right hemisphere—but only for the specific letters she had previously practiced imaging. It is of interest that unlike J.W., whose corpus callosum is cleanly severed, parts of the splenium were inadvertently spared in V.P. (as was recently discovered through NMR scans). The splenium is part of the callosum known to transfer visual information between the hemispheres. In addition, similar results have been obtained with normal subjects when we examine the time to rsspond depending on which hemisphere sees the cue first.

The results of these experiments, then, documented a highly selective deficit in the right hemisphere for one computation involved in image generation. The computational theory proved its worth by leading us to investigate a counter-intuitive claim that probably would not have arisen using traditional approaches.

Conclusions

Perhaps the greatest indicator of success in science is the discovery of underlying structure where none was originally apparent. By this measure, the present approach has promise. A computational analysis of imagery and vision led to a novel analysis of the structure of one imagery ability, the ability to form visual mental images. In addition, cognitive methods applied to neuropsychological phenomena allowed us to test and verify a non-intuitive implication of the theory, namely that the left cerebral hemisphere would be better than the right at specific imagery tasks.

This approach not only has promise of helping us to discover the underlying mechanisms of mental imagery, but may help us to understand the functional organization of "mental" processes in the brain more generally. The notion that the left and right hemispheres have distinct functions is not new, and in fact has penetrated deeply into the popular consciousness. Indeed, many people attempt to characterize themselves as predominently "left-" or "right-brained," assuming that the hemispheres differ along some simple dichotomy (e.g., "verbal" versus "imaginal"). If the present approach is correct, these dichotomies are hopelessly oversimplified and have little utility (e.g., both hemispheres are imaginal, but in different ways). Ã more refined theory will specify the components of processing that are better developed in one hemisphere relative to the other, and will specify how these components are brought to bear in helping one to perform specific tasks.

Detailed theories of the computations used in processing will be very complicated and, without question, computer simulation will be required to develop them and explore their implications. That is, we will have to embody the putative computations in a computer model, and observe its behavior in various tasks. This conclusion is particularly evident when we compare the results we have obtained for different subjects in the experiments described above, which vary slightly from person to person. Presumably, different computations are more or less effective for different people, and any detailed simulation will have to account for this by specific parameter variations. Fortunately, the technology is now available for just such future developments.

Biography

Dr. Kosslyn is Professor of Psychology at Harvard University. His publications have been mainly in the field of cognitive science, specializing in visual information. Included among his books are *Image and Mind* and *Ghosts in the Mind's Machine*. In 1983, he received the Inititives in Research Award from the National Academy of Sciences.

References

1. Warrington, E. K., and James, M., "Disorders of Visual Perception in Patients with Unilateral Cerebral Lesions," *Neuropsychologia* 5 (1967) 253-266.

2. Ellis, A. W., *Normality and Pathology in Cognitive Functions*. New York, Academic Press (1982).

3. Marr, D., *Vision*. San Francisco, CA, W. H. Freeman (1982).

4. Segal, S. J., Processing of the Stimulus in Imagery and Perception," S. J. Segal (Ed.), *Imagery: Current Cognitive Approaches*, New York, Academic Press, (1971).

5. Shepard, R. N., and Metzler, J., "Mental Rotation of Three-dimensional Objects," *Science*, 171 (1971) 701-703.

6. Kosslyn, S. M., Ball, T. M., and Reiser, B. J., "Visual Images Preserve Metric Spatial Information: Evidence From Studies of Image Scanning," *Journal of Experimental Psychology: Human Perception and Performance* 4 (1978) 47-60.

7. Kosslyn, S. M., (1975) "Information Representation in Visual Images," *Cognitive Psychology* 7 (1975) 341-370.

8. Kosslyn, S. M., *Image and Mind*. Cambridge, MA, Harvard University Press, (1980).

9. Kosslyn, S. M., *Ghosts in the Mind's Machine*, New York, W. W. Norton (1983).

10. McClelland, J. L., and Rumelhart, D. E., An Interactive Activation Model of Context Effects in Letter Perception: Part I. An Account of Basic Findings. *Psychological Review*, 88 (1981) 375-407.

11. Mishkin, M., "A Memory System in the Monkey," *Philosophical Transactions of the Royal Society of London B*, 298 (1982) 85-95.

12. Ungerleider, L. G., and Mishkin, M., "Two Cortical Visual Systems," D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior*, Cambridge, MA, MIT Press, (1982a).

13. Ungerleider, L. G., and Mishkin, M., Contribution of Striate Inputs to the Visuospatial Functions of Parieto-preoccipital Cortex in Monkeys," *Behavioural Brain Research*, 6 (1982b) 57-77.

14. Ballard, D. H., and Brown C. M., *Computer Vision*, New York, Prentice Hall (1982).

15. Shapley, R., and Lennie, P., Spatial Frequency Analysis in the Visual System," W. M. Cowan (Ed.), *Annual Review of Neuroscience, Vol 8*. Palo Alto, CA, Annual Reviews (1985).

16. Treisman, A., and Gelade, G., "A Feature Integration Theory of Attention," *Cognitive Psychology*, 12 (1980) 97-136.

17. Van Essen, D. C., and Maunsell, J. H. R., "Hierarchical Organization and Functional Streams in Visual Cortex," *Trends in NeuroScience*, September (1983) 370-375.

18. Kosslyn, S. M., Holtzman, J. D., Farah, M. J., & Gazzaniga, M. S., A Computational Analysis of Mental Image Generation: Evidence from Functional Dissociations in Split-Brain Patients, *Journal of Experimental Psychology: General* (in press).

19. Kosslyn, S. M., "Visual Hemispheric Specialization: A Computational Theory," ONR Technical Report #7, Department of Psychology and Social Relations,—Harvard University.

20. Erlichman, H., and Barrett, J., "Right Hemisphere Specialization for Mental Imagery: A Review of the Evidence," *Brain and Cognition*, 2 (1983) 55-76.

21. Springer, S. P., and Deutsch, G., *Left Brain, Right Brain*, San Francisco, W. H. Freeman (1981).

22. Sidtis, J. J., Volpe, B. T., Wilson, D. H., Rayport, M., and Gazzaniga, M. S., Variablility in Right Hemisphere Language Function: Evidence for a Continuum of Generative Capacity," *Journal of Neuroscience*, 1 (1981) 323-331.

Research Notes

Establishment of a New Center for Brain Research

New York University Medical Center and Biomagnetic Technologies, Inc. have announced a collaborative research program in magnetoencephalography (MEG). The basic research underlying this venture has been supported by the Office of Naval Research principally at NYU and the University of California, Los Angeles. (The original research used a single channel second order gradiometer). A 100 channel device is now being built for delivery in about two years. Neuromagnetometry offers a non-invasive way to monitor human brain activity more precisely than previously possible. The early recognition by ONR of the potential of the MEG technique and the continued support over the past ten years is directly responsible for this latest development.

(D. P. Woodward, ONR)

Installation of a Multiple Channel Neutromagnetometer at UCLA

The world's second five channel DC neuromagnetometer was delivered to the Human Neurophysiology Laboratory at the University of California, Los Angeles. This device will greatly speed the neuromagnetic investigations of cognitive functions by Jackson Beatty and his coworkers. This project has been sponsored by the Physiology Program of the Office of Naval Reserach.

The instrument itself is somewhat more advanced than the original multichannel system, both of which were manufactured by Biomagnetic Technologies, Inc., of San Diego. The UCLA system has an improved signal to noise ratio, made possible in large part by recent advances in shielding technologies. Further, the UCLA laboratory is a magnetically quiet site. The combination of low instrument and enviromental noise levels permits operation of the neuromagnetormeter at unusually high grain levels. The result is a major enhancement in total measurement sensitivity.

The five recording channels are configured as second derivative gradiometers with extended baselines for improved sensitivity to deep cranial signals. Coil diameter is 1.5 cm with 2 cm inter-coil spacing. An additional four magnetometer channels provide for adaptive noise cancellation.

The instrument will be used in a program of research designed to locate and measure intracranial currents mediating early vision, linguistic processing, and attention. Neuromagnetic imaging provides a unique view of cortical function, permitting both three dimensional spatial resolution and millisecond temporal resolution. The technique has already proven to be a powerful new tool for brain research; the multichannel system represents an improvement of measurement technology that will speed experimental research by at least one order of magnitude.

(D. Woodward, ONR)

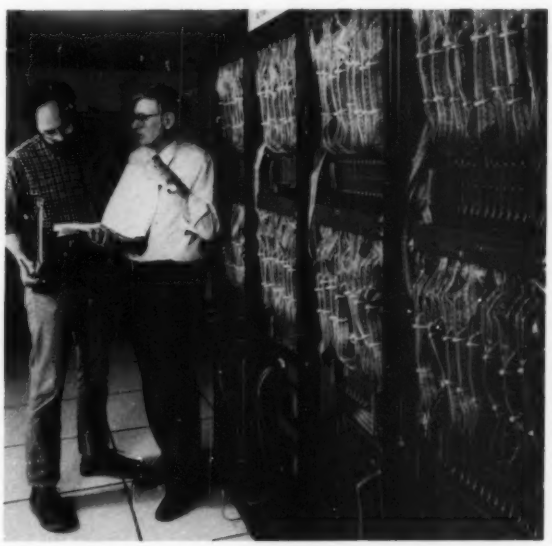
The Butterfly Computer

BBN Laboratories' Butterfly™ parallel processing system achieves nearly linear speedup with 128 processors, for a total of 60 million instructions per second (MIPS). The 128-processor Butterfly system is configured in 4 racks. Each processor node consists of an 8 MHz MC68000 processor, a co-processor, memory management, switch interface and I/O interface.

Parallel computers, such as the Butterfly, are playing an increasingly important role in Computer Science research and practice. It has been established that powerful parallel computers can be built and programmed to be extremely cost-effective for certain problems. The challenge is to extend the range of tractable problems and to provide programming languages and environments that support parallel computer, and several research groups are finding fruitful interactions among studies of natural and artificial computation.

One of these interdisciplinary efforts is at the University of Rochester, which will install the second 128-processor Butterfly computer later this year. In a project jointly sponsored by Office of Naval Reserach, National Science Foundation, and Defense Advanced Research Projects Agency, the Rochester group will be studying both conventional programming and "massively parallel" computational models for the Butterfly. The massively parallel approach attempts to decompose problems into computations done by millions or billions of simple units. This approach has been effective for a variety of tasks and the obvious connection to neurons is the basis for interdisciplinary research, some of which is described in this issue.

(Joel Davis, ONR)



Butterfly Computer of BBN Laboratories, Cambridge, Massachusetts.

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
ARLINGTON, VA. 22217

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

THIRD-CLASS MAIL
POSTAGE & FEES PAID
USN
PERMIT No. GS-9