Cognition, 33 (1989) 63–109

Neuronal models of cognitive functions

JEAN-PIERRE CHANGEUX

URA CNRS 0210 "Neurobiologie Moléculaire", Département des Biotechnologies, Institut Pasteur, Paris

STANISLAS DEHAENE

INSERM et CNRS, Paris

Abstract

Changeux, J.-P., and Dehaene, S. 1989. Neuronal models of cognitive functions. Cognition, 33:63–109.

Understanding the neural bases of cognition has become a scientifically tractable problem, and neurally plausible models are proposed to establish a causal link between biological structure and cognitive function. To this end, levels of organization have to be defined within the functional architecture of neuronal systems. Transitions from any one of these interacting levels to the next are viewed in an evolutionary perspective. They are assumed to involve: (1) the production of multiple transient variations and (2) the selection of some of them by higher levels via the interaction with the outside world. The time-scale of these "evolutions" is expected to differ from one level to the other. In the course of development and in the adult this internal evolution is epigenetic and does not require alteration of the structure of the genome. A selective stabilization (and elimination) of synaptic connections by spontaneous and/or evoked activity in developing neuronal networks is postulated to contribute to the shaping of the adult connectivity within an envelope of genetically encoded forms. At a higher level, models of mental representations, as states of activity of defined populations of neurons, are discussed in terms of statistical physics, and their storage is viewed as a process of selection among variable and transient pre-representations. Theoretical models illustrate that cognitive functions such as short-term memory and handling of temporal sequences may be constrained by "microscopic" physical parameters. Finally, speculations are offered about plausible neuronal models and selectionist implementations of intentions.

0010-0277/89/$14.60

Introduction

"To think is to make selections"
W. James (1909)

In the course of the past decade, the development of the cognitive sciences has resulted in many significant contributions, and, for many, the science of mental life constitutes a "special science" (Fodor, 1975). But still one must ask how the physical world supports mental processes. The radical proposal was made (Johnson-Laird, 1983) that "the physical nature [of the brain] places no constraints on the pattern of thought ... any future themes of the mind [being] completely expressible within computational terms". The computer thus became the last metaphor, which "needs never be supplanted" (Johnson-Laird, 1983).

In parallel, the sciences of the nervous system have made considerable progress. Studies of single nerve cells and of their molecular components unambiguously rooted the elementary processes of neurons in physical chemistry, thereby introducing a wide spectrum of constraints, in particular upon their dynamics. Moreover, comparative anatomical investigations of higher vertebrate (see Goldman-Rakic, 1987; Rakic, 1988) and human (Geschwind & Galaburda, 1987; Luria, 1973) brain connectivity, biochemistry (e.g., Hökfelt et al., 1986) and physiology (e.g., Edelman, Gall, & Cowan, 1984) have yielded novel views about the complexity of adult brain organization and about its morphogenesis (Changeux, 1983a, 1983b; Edelman, 1987; Geschwind & Galaburda, 1987). On such bases, the alternative radical program was proffered that, ultimately, the science of mental life will be reduced to neural sciences and the tenets of psychology will be eliminated (Churchland, 1986).

The aim of this paper is *not* to deal, once more, with this philosophical debate (Churchland, 1986; Fodor, 1975; see also Mehler, Morton, & Jusczyk, 1984; Stent, 1987), but rather to report recent theoretical and experimental work that actually lies at the frontier between cognitive science and neuroscience and that might ultimately bridge these two disciplines. It is hoped that such neurocognitive approaches will serve to create positive contacts between the two disciplines rather than to stress their differences. The real issue becomes the specification of the relationships between a given cognitive function and a given physical organization of the human brain. From an experimental point of view, our working hypothesis (rather than philosophical commitment) is that levels of organization exist within the brain at which a type-to-type physical identity (Fodor, 1975) might be demonstrated with cognitive processes. Yet, it should be stressed that we address this problem from the neurobiological point of view in an attempt to *reconstruct* (rather than reduce) a function from (rather than to) its neural components. This is, of course, not to deny the bridges that some psychologists have already thrown to the neural sciences in areas like language, neonate cognition, memory or neuropsychology. Our aim is to review recent contributions from the neurosciences to psychological concepts like innateness, learning, internal representations, and intentionality, and to offer relevant models that, in many cases, still remain speculative.

It is beyond the scope of this article to present an exhaustive review of data and models that fall within this approach. The discussion is limited to a few selected issues. First, the notion of level of organization and the relationship between structure and function in biological systems is introduced, together with evidence for the existence of multiple levels of functional organization within the human brain. A generalized Darwinian-like theory is proposed to describe the relationships between levels of organization and, in particular, the transition from one level to the next higher one. This empirically testable theory assumes that selection at a given level takes place in real time as a result of pressure from the immediately higher level, yielding a two-way dependency. Such Darwinian views are applied to the epigenesis of developing neuronal networks. The last two parts deal with the more conjectural neural bases of mental representations and "mental Darwinism", where the above-mentioned two-way dependency is assumed to take place at a higher level, between representations and the goals they subserve. In both instances, Darwinian views are extended to evolutive systems that are *not* based on variations of their genetic material itself, but of their phenotypic expression in developmental and psychological time-scales.

We have deliberately omitted from the discussion complex cognitive functions such as problem solving or language processing, especially those aspects of language interpretation that require access to very large stores of knowledge. Our feeling is that, at this stage, the neurobiological basis is too complex and the relevant data still too scarce for pertinent modelling in neuronal terms.

I. The notion of level of organization in biological systems

The debate over the relative contributions of cognitive psychology, connectionist modelling and neurobiology to the understanding of the higher functions of the human brain hinges primarily upon a basic confusion about the levels of organization at which models of cognitive function and their neural implementations have been designed. The specification of such levels should

precede any theoretical approach, and might even constitute the substance of a full theory. The cleavage of the world into pertinent units – from atoms, molecules and cells, up to, for example, syntactic structures and social classes – is neither trivial nor unequivocal. In highly evolved organisms, the relevant levels are expected to be multiple and intricate. Before dealing with the cognitive level(s), three basic questions must be answered:

(1) What characterizes a given level of organization within a living organism?
(2) To what extent are the levels dependent on each other?
(3) Does a general mechanism account for the transition from one level to the next?

The definition of a given level relies upon the anatomical organization of its elementary components and upon characteristic properties or functions that are unique to this level. The programme of the life sciences is, of course, to specify functions but, most of all, to relate a given function to an appropriate anatomical organization of its components in a causal and reciprocal manner. A quick look at the history of biology shows that unravelling the actual "physical" implementation of a given function has always been at the origin of considerable progress in the understanding of biological functions (see Clarke & O'Malley, 1968, in the case of neurobiology). As a metaphor to illustrate this point, let us consider the well-known example of the catalytic activity of enzymes. Enzymes are proteins made up of linear chains of amino acids that spontaneously fold into three-dimensional edifices of rather large size – macromolecules. At the turn of the century, such a macromolecular organization of enzymes was not known, but their function as catalysts was already well recognized (see Debru, 1983, for a review). The laws of their action on substrates were described in great detail, and in still valid computational terms, by the Henri and Michaelis–Menten *algorithms*, which, to a first approximation, adequately fit the measured kinetics. The rapid progress of molecular biology, in particular the unravelling of the three-dimensional topology of folded polypeptide chains by X-ray diffraction methods at the atomic level, led to an explanation of the way their active sites function. It showed, for instance, that the fixed orientation in space of a few of the protein amino acid side chains suffices for matching to the shape of the substrate and, through chemical bonding, for activating the transformation of the enzyme–substrate complex, a feature that disappears upon unfolding of the protein. The function of the enzyme (which Jacques Monod (1970) already referred to as "cognitive") is thus directly bound to the macromolecular level of organization and may even be viewed as characteristic of this level. The implementation of the function in terms of chemical bonds, including their nature, strength and topology, would never have been predicted from the algorithmic description of the active site function. Yet, this new understanding did not lead to an abandonment of the Michaelis–Menten algorithm, but rather raised many novel functional questions at a more microscopic level.

Another example is that of the material bases of inheritance. Mendel's observations on the heredity of flower colour in peas, and its description in simple algorithmic terms which constitute the Mendelian laws, may, at a glance, appear sufficient from a functional point of view. Yet, the identification of chromosomes and (subsequently) DNA as the physical basis of heredity opened many new avenues of research without, at any moment, contradicting the Mendelian algorithm.

II. The multiple levels of functional organization within the nervous system

The notion of organization and the relationship between physical structure (architecture) and function can legitimately be extended to the nervous system (Changeux, 1983b; Chomsky, 1979; Jackson, 1932). However, to what extent can the entire nervous system be viewed as one such level, above which would stand both a functional algorithmic level and a formal level of abstract grammatical rules (Marr 1982)? Should we follow the classical view (Fodor & Pylyshyn, 1988), which distinguishes among physical, syntactic and knowledge (semantic) "levels", and further assumes that the "functional organization of the nervous system crosscuts its neurological organization" (Fodor, 1975)? Or even, to be more schematic, may we say that the nervous system is the mechanism, the hardware of the machine, above which would stand autonomously the program with its semantic and syntactic levels? We deliberately reject such a view, for within the nervous system, several distinct levels of organization can be defined. Distinct levels of cognitive function relate to these different levels of physical organization, thus shattering the simple-minded (and actually reductionistic) behaviour/algorithm/hardware metaphor for the relationship between psychology and neural science.

The first level, the architecture of which can be related to functional characteristics of the nervous system, is the "*cellular level*", because of the unique ability of nerve cells to make topologically defined networks through their axonal and dendritic branches and their synaptic contacts. This physical architecture is described by the topology of the cell arborization and its synaptic connections. Its function includes both the patterns of electrical and chemical signals produced by the cell (Prince & Huguenard, 1988) and the actions of these signals on effector cells, which lead to either overt behaviour or a

covert process that might contribute to further operations within the system. At the single-cell level, the ability of the neuron to generate electrical impulses (or modulate its spontaneous firing) as a function of the inputs it receives (and several other functions it displays) can be expressed in a global algorithmic form, and implemented in terms of the molecular properties of its synaptic and cell membrane components. Yet, in this instance already, the algorithmic level cannot be viewed simply as autonomous or distinct from the physical one. It would, of course, be absurd to infer from such cellular and molecular data the nature of the architecture of cognitive functions. Yet, nerve cells are the building blocks of cognitive architectures, and as such they exert severe constraints on the coding of mental representations, on the computations accessible to these representations, and on the modalities of storage and retrieval. After all, the rates of propagation of impulses in nerves and across synapses impose inescapable limits on the speed of higher computations; storage of memory traces has to be considered in terms of changes of molecular properties of nerve cells and synapses; and the production of internal thought processes independently of outside world stimulation relies upon the existence of a spontaneous self-sustained and organized firing of nerve cells (see section VI).

Another level of organization, referred to as the "*circuit level*", is reached with the nervous system of invertebrates (sea-slugs like *Aplysia* or *Hermissenda*, insects like *Drosophila* and worms like *Hirudo*). Circuits are made up of thousands (or even millions) of nerve cells organized in well-defined ganglia, with each (or nearly each) cell possessing a well-defined individual connectivity and function within the organism. Satisfactory attempts have been made to account for simple behaviours strictly on the basis of the anatomy of small circuits and the electrical and chemical activity displayed by the circuit (Alkon, Disterhoft, & Coulter, 1987; Grillner, 1975; Grillner et al., 1987; Kandel et al., 1983; Kleinfeld & Sompolinsky, 1987; Stent et al., 1978). The architecture of the most advanced connectionist models, which aim at imitating human-like brains, are no more complex than these rather primitive nervous systems, and often remain inferior in complexity even in the most elaborate attempts.

The mutual relations of individual neural circuits define another level of organization. Thus, for example, classical physiology has emphasized the roles of the spinal cord and brainstem in the mediation of reflexes and various other sensorimotor operations. A vast amount of anatomical, physiological and neuropsychological data, moreover, has led to a closer understanding of the participation of various cortical and other brain domains in sensory perception, motor control, language comprehension and production, memory, etc. (Geschwind & Galaburda, 1987; Kolb & Wishaw, 1980; Luria, 1973). The number of these specialized domains appears much larger than initially

suspected (Rakic & Singer, 1988): in the primate brain, for instance, there are about a dozen representations of the visual world and half a dozen each of auditory and somatic representations. Many of these are interdigitated, and yet, each fragment displays a distinct characteristic, thus "constructing categories in an unlabeled world" (Zeki, 1988).

The cognitive level lies within reach of this "meta-circuit" level, since the parsing of psychological tasks into elementary operations correlates well with a decomposition of the brain into separate areas (Posner, Petersen, Fox, & Raichle, 1988). However, even these ultimate cognitive functions might themselves be cleaved into distinct faculties. Seventeenth- and eighteenth-century philosophers, Kant in particular, distinguished "*reason*" as the "faculty which contains the principles by which we know things absolutely *a priori*" from "*intendment*" which, from the perceived elements, produces concepts and evaluates them. In a simplified manner, the intendment would make the synthesis of sensible elements into concepts, while pure reason would make computations upon the concepts produced by the intendment. Such distinction of different levels of abstraction bears a relation to recent concepts from experimental psychology (see Kihlstrom, 1987). Accordingly, intendment processes would be mostly modular, automatic and unconscious, although attention may regulate their inputs and their access to memory stores, while in pure reason the processing would essentially (though not exclusively) be conscious, non-modular and require attention. What then are the brain structures, if any, for intendment and reasoning?

In attempting to build bridges across disciplines, it may be useful to relate the intendment/reason distinction with a distinction reached through a different approach in the field of artificial intelligence. Newell (1982) raised questions of levels in computers, of their autonomy and reducibility to lower levels, questions that are indeed familiar to neuroscientists. Thus, he proposed that immediately above the symbol (program) level of standard computers stands a knowledge level, where the agent processes its own knowledge to determine the actions to take. The law governing behaviour is the principle of rationality: actions are selected to attain the agent's goal. "If an agent has the knowledge that one of its actions will lead to one of its goals it will select that action" (4.2, p. 17). According to Newell, the "knowledge level is exactly the level that abstracts away from symbolic processes" (4.3, p. 23), a conclusion not far from Kant's definition of reason. Moreover, and this is a crucial point, the actions may add knowledge to the already existing body of knowledge. As mentioned by Simon (1969) in the case of economic agents, "the adjustment to its outer environment (its substantive rationality) is conditioned by its ability to *discover* an appropriate adaptive behavior (its procedural rationality)". In other words, in selecting actions to achieve a goal, the agent not

only relies upon judgments but, in situations of uncertainty, it computes expectations and possible scenarios of complex interactions.

A crucial issue then becomes the specification of the neural architectures that characterize the knowledge level thus defined. Yet, the feasibility of this attempt is under debate. For instance, Pylyshyn (1985) has argued on the basis of Newell's (1982) views, that one should distinguish knowledge-based from mechanism-based explanations. We quote him: "the mechanism is part of the process that itself is not knowledge dependent (it is cognitively impenetrable) hence it does not change according to rational principles based upon decisions, inferences, and the like, but on other sorts of principles, ones that depend on intrinsic properties, which are presumably governed by biological laws" (p. 408). "Ignoring the physics and biology may even be necessary because the categories over which the system's behavior is regular may include such things as the meaning of certain signals and because the entire equivalence class of signals having the same meaning need not have a description that is finitely statable in a physical vocabulary" (p. 405). Our view is that even if the distinction between cognitively penetrable and impenetrable processes may, to some extent and within a limited time-scale, be justified, the separation between the cognitively penetrable and the "biological or biochemical properties" of organisms is not valid. In fact, as will become clear in the following paragraphs, there is hardly any structure in the brain which does not incorporate exterior knowledge during its epigenesis and functioning. We realize that there is more to Pylyshyn's cognitive penetrability than the simple notion of knowledge dependence. For instance, Pylyshyn's mechanistic level, as far as we understand it, comprises the processes that do not vary following a change in conscious beliefs and intentions of an agent. However, the usefulness of this notion is questionable on the grounds that: (1) this particular mechanistic level is not identical to the level of biological laws, since there may be encapsulated psychological processes that cannot be accessed and modified intentionally; (2) its definition depends on the time-scale chosen, since conscious, attention-demanding processes may become automatic and impenetrable in the course of learning (see Baddeley, 1976, 1986); and (3) its definition puts exceptional emphasis on the notions of consciousness and belief, the scientific status of which is now questioned on philosophical as well as neuropsychological grounds.

It is thus part of a concrete scientific programme to describe which (and to what extent) biological structures are penetrable by knowledge at a given level. This programme is necessary if we are to have a biology of goal, knowledge and rational decisions, as well as to know the neural architectures underlying the faculty of reason.

If such a research programme looks plausible, it is far from being achieved (or even undertaken) in neuroscience for many reasons. First of all, in humans, processes as elementary as visual perception, which take place at the intendment level, are deeply impregnated by knowledge from the earliest stages of development (see sections IV and V). Similarly, since storage seems to rely more on semantic than on perceptual cues (Baddeley, 1976, 1986, 1988), it seems almost impossible to analyse the function of long-term memory stores without considering what information they encode. Moreover, from both the phylogenetic and embryological points of view, the faculties for reasoning and for concept formation, among others, cannot be assigned to unique brain domains or areas and, of course, the operations taking place at any level are expected to mobilize important populations of nerve cells with a distributed topology which will *a priori* be difficult to map.

Nevertheless, one may theorize, for example, that the frontal areas of the cortex contribute to the neural architectures of reason (for reviews see Struss & Benson, 1986, and Goldman-Rakic, 1987). Among the many observations supporting this view is the fact that the differential expansion of prefrontal cortical areas from the lowest mammals up to man parallels the development of cognition. Neuropsychological observations of Lhermitte (1983) and Shallice (1982), among others, also relate frontal cortex to high-level cognitive processes. Patients with frontal lobe lesions, for example, display an interesting "utilization" behaviour. They grasp and utilize any object presented to them as if they had become dependent upon sensory stimulation. They also fail in tests such as the "Tower of London", which require planning strategies and control of the execution. Patients no longer employ a general mode of regulation that lets them plan their interactions with the environment, makes them aware of novel situations or errors in the executions of their own strategies, and allows them to generate new hypotheses. Although few studies have investigated the neuronal architecture underlying these functions, brain lesions indicate a dissociation of two levels *within* cognition, which closely approximate the levels of reason and intendment discussed previously.

III. The transition between levels of organization: generalized Variation – Selection (Darwinian) scheme

A given function (including a cognitive one) may be assigned to a given level of organization and, in our view, can in no way be considered to be *autonomous*. Functions obviously obey the laws of the underlying level but also display, importantly, clear-cut dependence on higher levels. Coming back to our favourite metaphor, the function of the enzyme active site is determined by the amino acid sequence of the protein; yet the amino acid sequence is

itself the product of a long genetic evolution, which ultimately rested upon survival (stabilization) rules that constrained the structure of the macromolecule. The macromolecular state, which determines the enzyme catalytic function, is rooted, by its structure, in the underlying levels of physics and chemistry, but also contributes, by its function at a higher level, to the metabolism of the cell and thus to its own existence.

The same interdependence holds for the various levels of neurofunctional organization. If, for instance, the function of the reflex arc can be viewed as strictly determined by a well-defined spinal cord neuronal circuit, its detailed organization is such that higher brain centres may control its actualization within a coordinated motor act. At any level of the nervous system, multiple feedback loops are known to create re-entrant (Edelman & Mountcastle, 1978) mechanisms and to make possible higher-order regulations between the levels.

Our view is that the dual dependence between any two levels must be framed in an evolutionary perspective (see Changeux, 1983b; Delbrück, 1986; Edelman, 1987) and is governed by a generalized variation-selection (Darwinian) scheme. Accordingly, a minimum of two distinct components is required: a generator and a test. "The task of the generator is to produce variety, new forms that have not existed previously, whereas the task of the test is to cut out the newly generated forms so that only those that are well fitted to the environment will survive" (Simon, 1969; p. 52). The initial diversity may arise from combinatorial mechanisms that produce organizations that are novel (or rare) for the considered level and become transitional forms, bridging it to the next higher level. The rules of stabilization (survival) are governed by the function associated with the novel form, thus creating feedback stabilization loops of function upon structure.

Such a scheme is classical in the case of the evolution of species and is evidenced in the development of the immune response, where diversity arises from genomic reorganization and gene expression, and the test arises from the survival of the fittest (including matching to the antigen). The scheme may also account for the transition from cellular to multicellular organisms and for the general morphogenesis of the brain. Our view, in addition, (Changeux, Courrège, & Danchin, 1973; Changeux & Danchin, 1974, 1976; Changeux, Heidmann, & Patte, 1984; Edelman, 1978, 1985, 1987; Jerne, 1967; Young, 1973; see also Ramon y Cajal, 1909; Taine, 1870; references in Heidmann, Heidmann, & Changeux, 1984), is that the interaction between the nervous system and the outside world during the course of postnatal development through adulthood, during which the acquisition of some of its highest abilities, including language and reasoning, occur (see section V), also follows an analogous Darwinian scheme. Yet, such evolution is expected

to take place *within* the brain without any necessary change in the genetic material (at variance with the view of Piaget, 1979, or Wilson, 1975), and inside of short time-scales: days or months in the case of embryonic and early postnatal development and tens of seconds in the case of the processing and reorganization of mental representations. At each level, the generator of variety and the test must be specified, and the time-scale of the evolution must be defined. Moreover, such time-scales are short enough (compared to those spanning the evolution of species) to render the theory experimentally testable.

The justification of such views will constitute the matter of the subsequent sections. In essence the brain will be considered constantly and internally to generate varieties of hypotheses and to test them upon the outside world, instead of having the environment impose (instruct) solutions directly upon the internal structure of the brain (see Changeux, 1983b). The view of the brain as a hardware construct, knowledge-independent, that would be programmed by a computationally autonomous, cognitively penetrable mind has, thus, to be reconsidered for the following reasons:

(1) There exists, within the brain, multiple levels of functional organization associated with distinct neural architectures (see section II).
(2) Several of these neurofunctional levels are cognitively penetrable at some stage of development, and these multiple levels are heavily interconnected via feedback loops, or re-entrant mechanisms, that make possible high-order regulations *between levels* (further examples of such interactions appear below).
(3) The information-processing/input–output scheme has to be abandoned in favour of an internal "generator of variations" continuously anticipating and testing the outside world with reference to its own invariant representation of the world.

IV. The ontogenesis of neural form

A basic principle of cognition is the recognition, storage and internal production of forms in space (patterns) or time (melodies). The Gestalt psychologists have emphasized that this faculty relies on the existence of physiological forms within the organism that display some physical relationship with the psychological ones. Without entering into a debate on the exact nature of this relationship – whether it is isomorphic or not – it is evident that the brain must be viewed as a highly organized system of intertwined architectures whose forms result from pre- and postnatal development. Moreover, as dis-

cussed in the preceding section, several distinct levels of functional organization exist within the brain and develop according to defined biological constraints during both phylo- and ontogenesis. Understanding the formation of neural forms and their hierarchical organizations, for example those involved in the architectures of reason, concept formation or pattern recognition, becomes a fundamental step in the understanding of cognition.

1. Preformation and epigenesis. The problem of the origin of animal form has been at the centre of a long controversy from classical times to the present. According to the extreme *preformationist* view, ontogenesis proceeded strictly as the enlargement of forms that were thought to be already present in the egg (Swammerdam) or in the spermatozoon (Van Leeuwenhoek). Related views are still found in the theory of morphological archetypes advocated by D'Arcy-Thompson (1917), and they have recently been revived by Thom (1980). Accordingly, such archetypes, mathematically formalized in a set of abstract rules, would impose global morphological constraints on, or even direct the ontogenesis of, the adult form. Also, contemporary molecular biologists frequently refer to a DNA-encoded genetic program according to which development would proceed in a strictly autonomous manner (see Stent, 1981, for a criticism of the concept of genetic program). Finally, some contemporary linguists and psychologists posit the innateness of knowledge (or at least of a certain body of information) (Chomsky, 1965) or of mental faculties or structures (Fodor, 1983), although without referring to the actual neural bases of their development. According to such extreme Cartesian views, the internal innate structures are rich and diverse, and their interaction with the environment, although capable of "setting parameters", does not create new order.

The alternative attitude, illustrated by epigenesis, contrasts with radical preformationism by postulating a progressive increase of embryonic complexity (rather than simple enlargement by post-generation and after-production). Such epigenetic conceptualization of the development of animal forms shows analogies in psychology to the associationist attitude (Helmholtz), or mental atomism, which assumes, for instance, that perception relies upon the analysis of external objects into elements and upon their synthesis through association by continuity in time and repetition. According to such views, in their most extreme formulation, mental forms would build up strictly from experience, starting from an initial empty state or *tabula rasa*.

2. The developmental genetics of brain forms. Both the extreme preformationist attitude and the strict epigenetic views are incompatible with current knowledge about development. The contribution of strictly innate, DNA-encoded mechanisms to the development of animal forms is supported by a large body of experimental evidence. Yet, the genes involved are not

expressed all at once, whether in the egg or in the early embryo, as postulated by the extreme preformationist view. Rather, they are activated (or blocked) throughout embryogenesis and postnatal development in a sequential and intricate manner and according to well-defined patterns.

The straightforward comparison of the genetic endowment of the organism with the complexity of the adult brain produces, however, two *apparent* paradoxes (Changeux, 1983a, b). The total amount of DNA present in the fertilized egg is limited to a maximum of 2 million average-sized genes which, because a large number of them consist of non-coding sequences, might in fact be in the range of 100,000 genes. There is thus striking parsimony of genetic information available to code for brain complexity. The second paradox is raised by the evolution of the global amount of DNA in mammals, which appears rather stable from primitive species such as rodents up to humans, whereas the functional organization of the brain becomes increasingly complex. There thus exists remarkable non-linearity between the evolution of total DNA content and that of brain anatomy.

The analysis of the early steps of embryonic development by the methods of molecular genetics primarily in *Drosophila* (see Akam, 1987; Doe, Hiromi, Gehring, & Goodman, 1988; Gehring, 1985; Nüsslein-Volhard, Frohnhöffer, & Lehman, 1987) and in the mouse (see Johnson, McConnell, & Van Blerkom, 1984) begins to resolve these paradoxes. For instance, in *Drosophila*, a variety of gene mutations have been identified that affect early embryonic development. The genes involved have been grouped into three main categories. The first controls the Cartesian coordinates of the embryo; the second controls the segmentation of the body; and the third, called homeotic, specifies the identity of the body segments in well-defined territories of the egg and embryo during oogenesis and embryonic development, respectively, both in a hierarchical and parallel manner, and with cross-regulatory interactions. In the course of expression, the symmetry properties of the developing oocyte or embryo change: symmetry breakings (Turing, 1952) take place. In *Drosophila* the expression of the genes that specify the Cartesian coordinates of the embryo occurs in the mother's ovary through the asymmetric diffusion of morphogens from specialized ovary cells into the egg during the latter's maturation. In the mouse (and most likely in humans), the fertilized egg and early embryo (until the 8-cell stage) appear entirely isotropic. All symmetry breakings take place within the embryo after fertilization. Turing (1952) and followers (Meinhardt & Gierer, 1974) have demonstrated mathematically that such symmetry breakings may be created by random fluctuations that generate defined and reproducible patterns within an initially homogeneous system of morphogens reacting together and diffusing throughout the organism. A spatio-temporal network of gene interac-

tions, with convergence, divergence and re-utilization of regulatory signals, thus governs development of the body plan and, as we shall see, governs the plan of the brain.

Several segmentation and homeotic genes are expressed in the nervous system (Awgulewitsch et al., 1986) and are most likely members of a larger but still unidentified population of genes directly concerned with brain morphogenesis, the parsing of the brain into definite centres and areas and even into asymmetric hemispheres. Their combinatorial expression during development thus offers a first answer to the above-mentioned paradox of gene parsimony. Moreover, a quantitative variation in the expression of a few genes at early stages of development may suffice to account for the increase (or decrease) of surface (or volume) of some brain regions such as the cerebral cortex (or the olfactory bulb) in higher mammals (see Changeux, 1983b), thereby providing an explanation for the paradox of non-linear evolution between the complexity of the genome and that of the organization of the nervous system. For instance, since the total number of neurons in the thickness of cerebral cortex appears uniform throughout vertebrate species (Rockwell, Hiorns, & Powell, 1980), its surface area has been the primary target of evolutive changes (Rakic, 1988). One may then speculate that the extremely fast expansion of the frontal lobe surface that, in part, led to the human brain resulted from the prolonged action of some of these genes in the anterior part of the brain (see Changeux, 1983b). The genomic evolution that underlies this process has been extensively discussed in terms of (classical and non-classical) Darwinian mechanisms (Mayr, 1963).

Later in the course of the cellular organization of the nervous system by the mechanism just mentioned, synaptic connections begin to form. In the higher vertebrates, large ensembles of cells are found to project onto other large ensembles of cells, and maps develop. The best-known maps correspond to the projections of sensory organs, but maps also exist that represent projections from one brain region to another. In the course of such transformations symmetry breaking does not, in general, take place, but regular, geometrical deformations frequently occur. They can be mathematically analysed in terms of the theory of transformations (D'Arcy-Thompson, 1917). The actual mechanisms involved in the establishment of ordered connections of neural maps are still largely unknown. In addition to long-range tropism by chemical substances, the conservation of topological relationships between growing axons and a temporal coding by differential outgrowth of nerve fibres have been invoked. Processes of cell surface recognition might also be seminal in these transformations (Edelman, 1985, 1987).

In conclusion, the main forms of brain architecture develop by principles that can be summarized as follows:

(1) The basic operations of symmetry breaking and transformation that occur during the genesis of the geometric outlines and patterns of central nervous system organization are determined by an ensemble of rules. These rules are genetically coded as defined physico-chemical organizations of the brain. The notion of *tabula rasa* does not hold for the developing brain.

(2) The expression of the genes involved in the determination of brain forms follows well-determined spatio-temporal patterns that lead to the progressive establishment of the adult organization. As a consequence, relationships become established between the progressively laid-down forms, with built-in hierarchies, parallelisms and re-entries.

V. Epigenesis of neuronal networks by active selection of synapses

1. Biological premises

If our view of the development of neural forms accounts for species-specific traits of brain organization and function, it does not suffice for a more detailed description of neural anatomy. Indeed, a significant variability of the phenotype of the nervous system is apparent at several levels of its organization. Examination by electron microscopy of identified neurons from genetically identical (isogenic) individuals reveals minor but significant variability. In the small invertebrate *Daphnia magna*, the number of cells is fixed and the main categories of contacts (between optic sensory cells and optic ganglion neurons) are preserved from one isogenic individual to another. Yet, the exact number of synapses and the precise form of the axonal branches varies between pairs of identical twins (Macagno, Lopresti, & Levinthal, 1973). Similar findings have been reported in the case of the Müller cells of a parthenogenetic fish (Levinthal, Macagno, & Levinthal, 1976) and thus they are not restricted to invertebrates. A fluctuation in the details of the connectivity exists. In mammals, the variation also affects the number of neurons. For instance, in the case of the cerebellum of the mouse, the division and migration of Purkinje cells in consanguineous strains are not subject to as rigorous and precise a determinism as the laying down of neurons in invertebrates (Oster-Granite & Gearhart, 1981; Goldowitz & Mullen, 1982). The variability becomes microscopic and may even affect the chemistry (such as the pattern of transmitters and coexisting messengers synthesized (Hökfelt et al., 1986) of entire populations of neurons. In humans, most of the information available on anatomy derives from individuals taken from genetically heterogeneous populations. Nevertheless, the substantial variability noticed

in the topology of sites specialized for such language functions as naming, syntax or short-term memory, as identified by electrical stimulation mapping (Ojemann, 1983), cannot be accounted for solely by this heterogeneity. In the extreme case where the left hemisphere has been ablated in infants, the right hemisphere systematically takes over the language functions (Geschwind & Galaburda, 1984; Nass et al., 1985, 1989). More subtle is the topological distinction of sites, the stimulation of which causes different errors in the two languages spoken by bilingual subjects (Ojemann, 1983), a situation which, of course, relies upon reorganization by learning. Finally, in monkeys, transection of peripheral sensory nerves or their functional alteration causes striking reorganization of the topology of the relevant sensory cortical maps (Merzenich, 1987; Edelman, 1987). Thus, an important degree of variation of neural phenotype is manifest at several levels of organization of the nervous system, and this variability seems to increase with the complexity of the brain.

It may be useful to introduce the notion of a *genetic envelope* to delimit invariant characteristics from those that show some phenotypic variance (Changeux, 1983a). Thus, as animals evolve from primitive mammals to humans, the genetic envelope opens to more and more variability. This variability in the organization of the adult brain reflects characteristic features of its developmental history, in particular the way neural networks become interconnected. The growth cones of the dendrite and axon tip navigate by trial and error through the developing neural tissue toward their targets and thus contribute to the variability of the adult neuronal network. Another source of variability resides in recognition of, and adhesion to, the target cells. Little evidence exists for a point-to-point pre-addressing between individual neurons by specific chemical cues (the chemo-affinity hypothesis), except perhaps in small invertebrates (Goodman et al., 1985). On the contrary, the emerging view is that growing nerve endings identify ensembles of target cells further grouped into wider categories by a few intercellular adhesion molecules (N-CAM for instance; Edelman, 1985, 1987). Discrimination between cell categories would emerge from the temporal sequence of the expression of such adhesion molecules on the surface of the cells, from their differential topological distribution in regular patterns within populations of cells, and from their graded combination on the surface of cells. In any case, within a given cell category, a limited randomness and even some overlap between growing nerve fibres and individual target neurons from a given category are expected to occur.

At a critical stage (or sensitive period) of development, the axonal and dendritic trees branch and spread exuberantly. There is redundancy but also maximal diversity in the connections of the network. Yet, as mentioned, this transient fluctuation of the growing connections remains bounded within stringent limits of the genetic envelope. These include, in addition to the overall plan of the nervous system, the rules that command the behaviour of growing tips of nerves (the growth cones) regarding recognition among cell categories and the formation and evolution of synapses. Such limited fuzziness of the network makes a final tuning of the connectivity necessary, but also offers the opportunity for an adaptive shaping.

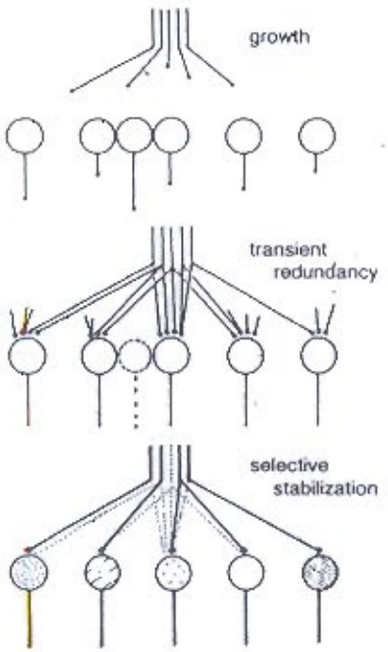
The classical information-processing scheme of the nervous system is based on the notion that its internal states of activity directly result from interactions with the outside world. In fact, from very early on, there is intense spontaneous electrical and chemical activity within the nervous system of the embryo and of the fetus (Hamburger, 1970). Chick embryos move within the egg as early as 2½ days of incubation. These spontaneous movements are blocked by curare and coincide with electrical activity of the same frequency arising in spinal cord neurons. In the human, these movements start during the eighth week of embryonic development and continue and diversify during the following months. Such spontaneous activity develops in a strictly endogenous manner and results from molecular oscillators consisting of slow and fast ionic channels (Berridge & Rapp, 1979). The cost of this activity in terms of structural genes is very small and bears no relation to its enormous potential for interaction and combination, which results from the activity's eventual contribution to the epigenesis of neuronal synaptic networks.

2. *A model for active selection of synapses*

The proposed theory (Changeux & Danchin, 1976; Changeux et al., 1973, 1984; also see Edelman, 1987) deals with the stability of the connections. It postulates that the selection of connections by the activity of the developing network (endogenous, or evoked, or both) contributes to the final establishment of the adult organization. Accordingly, the generator of internal diversity would not be based upon genetically determined recombinations, as in the case for the evolution of species or antibody synthesis, but rather stems from multiple neural configurations with connectivities drawn up during the epigenetic formation of a network at the stage(s) of transient redundancy (Figure 1).

According to our theory, during the sensitive period of maximal connectivity, any given excitatory or inhibitory synapse exists in three states: labile, stable and degenerate. Nerve impulses can be transmitted only in the labile and stable states. A given synapse may undergo transitions from labile to stable states (stabilization), from labile to degenerate states (regression), and from stable to labile states (labilization). The ontogenetic evolution of each

Figure 1. *A diagrammatic representation of the model of epigenesis by selective stabilization. An initial stage of growth precedes the pruning of circuits through elimination and stabilization of synapses.*
(From Changeux, 1983b; reproduced by permission of the publishers, Artheme Fayard, Paris).



synaptic contact is governed by the combined signals received by the cell on which it terminates. The activity of the postsynaptic cell within a given time-window regulates the stability of the synapse in a retrograde manner (Changeux et al., 1973). As a consequence, a given afferent message will cause the long-term stabilization of a matching set of synapses from the maximally connected neuronal network, while the others will regress.

Detailed models of molecular mechanisms possibly involved in this activity-dependent synapse selection have been proposed. Some deal with the topology of the postsynaptic clusters of receptors (Changeux, Courrège, Danchin, & Lasry, 1981), whereas others deal with the selection of multiple afferent nerve endings (Edelman, 1987; Fraser, 1985; Gouzé, Lasry, & Changeux, 1983). In both instances, the competition takes place for a limited component: in the first case, the receptor for a neurotransmitter; and, in the second, a diffusible growth (or stability) factor produced by the postsynaptic cell and taken up by the nerve ending in a differential manner by active and inactive nerve endings. In both cases, the firing of the postsynaptic cell is assumed to limit the component by repressing its biosynthesis. In addition, rules of release (by the postsynaptic cell) and/or uptake (by the presynaptic nerve endings) based upon the timing relationships between pre- and postsynaptic activity have to be specified in order to yield stable morphogenesis (Kerszberg et al., in preparation).

A straightforward consequence of this theory is that the postulated epigenesis contributes to the specification of the network at a low cost in terms of genetic information, which, in addition, can be shared by different systems of neurons. Such a mechanism offers a plausible way for coding organizational complexity from a small set of genes. By the same token, it also accounts for the paradoxical non-linear increase in complexity of the functional organization of the nervous system compared with that of the genome during the course of mammalian evolution.

3. The theorem of variability

In the course of the proposed epigenesis, diversification of neurons belonging to the same category occurs. Each one acquires its individuality or singularity by the precise pattern of connections it establishes (and neurotransmitters it synthesizes) (see Changeux, 1983a, b, 1986). A major consequence of the theory is that the distribution of these singular qualities may also vary significantly from one individual to the next. Moreover, it can be mathematically demonstrated that the same afferent message may stabilize different connective organizations, which nevertheless results in the same input–output relationship (Changeux et al., 1973). The variability referred to in the theory, therefore may account for the phenotypic variance observed between different isogenic individuals. At the same time, however, if offers a neural implementation for the often-mentioned paradox that there exists a non-unique mapping of a given function to the underlying neural organization.

4. Test of the model

Still, only fragmentary experimental data are available as tests of the theory. Elimination of synapses (and sometimes neurons) during development is well documented at the neuromuscular junction (Redfern, 1970; Van Essen, 1982), at the synapses between climbing fibres and Purkinje cells in the cerebellum (Mariani & Changeux, 1981a, b), and at the autonomic ganglia

(Purves & Lichtman, 1980), whereby in each case individual synapses can be easily counted. The phenomenon had already been noticed by Ramon y Cajal (1909) and interpreted as "a kind of competitive struggle" in Darwinian terms. It looks, in fact, to be rather widespread in the central nervous system (Clarke & Innocenti, 1986; Cowan, Fawcett, O'Leary, & Stanfield, 1984; Huttenlocher, De Courten, Garey, & Vander Loos, 1982; Innocenti & Caminiti, 1980; Price & Blakemore, 1985).

Particularly pertinent to the theory is the effect of nerve activity on these regressive phenomena. As noted, the developing nervous system is already spontaneously active at early stages of embryogenesis. This activity persists into maturity, being eventually modulated by the activity evoked by interaction with the outside world. Chronic blocking of the spontaneous undergoing activity prevents or delays the elimination of connections (Benoit & Changeux, 1975; Callaway, Soha, & Von Essen, 1987; Fraser, 1985; Reiter & Stryker, 1988; Ribchester 1988; Schmidt, 1985; Sretavan et al., 1988). In contrast with the classical empiricist views, activity does not create novel connections but, rather, contributes to the elimination of pre-existing ones. Long-term learning results from disconnection but in a growing network. "To learn is to eliminate" (Changeux, 1983b: p. 246, English translation).

At the larger scale comprising neural maps, activity has also been shown to contribute to the shaping of the adult network. Blocking activity by tetrodotoxin in regenerating fish retinotectal connections interferes with the development of a normal map, both by leading to the maintenance of a diffuse topology of connections and by restricting the receptive fields (Edelman, 1987; Fraser, 1985; Schmidt, 1985). The effect of activity on the segregation of ocular dominance columns in the newborn (but also possibly in the fetus) is well documented (Hubel & Wiesel, 1977; also see references in Reiter & Stryker, 1988; Stryker & Harris, 1986). Lastly, the long-term coordination among maps required, for example, for stereoscopic vision or for the unitary perception of visual and auditory spaces, has also been shown to depend on experience, and models for the synaptic mechanisms involved have been proposed (see Bear, Cooper, & Ebner, 1987).

Finally, the theory also accounts for the sensitive phases of learning and imprinting, which may correspond to the transient stage of maximal innervation (or diversity) in which the synaptic contacts are still in a labile state. This stage is well defined in the case of a single category of synapses. In the case of complex systems, such as the cerebral cortex, multiple categories of circuits become successively established, and, accordingly, many outgrowth and regression steps may take place in a succession of sensitive periods. In this sense the whole period of postnatal development becomes critical, but for different sensory inputs and performances! It is worth recalling that in humans this period is exceptionally long. A prolonged epigenesis of the cerebral cortex would not cost many genes, but for the reasons given above would have a considerable impact on the increased complexity and performance of the adult brain. Possible implications of epigenesis by selective stabilization of synapses in left–right hemispheric differentiation (Nass, Koch, Janowski, & Stiles-Davis, 1985, 1989) for all aspects of language learning have already been extensively discussed (see Changeux, 1983b; Gazzaniga, 1987). The developmental loss of the perceptual ability to distinguish certain phonemes in different languages, such as the initial sounds of *ra* and *la* in Japanese, in contrast to Western languages (Eimas, 1975; Miyawaki et al., 1975), the variability in topological distribution of brain areas in different individuals (Ojemann, 1983), and the remarkable segregation of the cerebral territories utilized in the processing of Japanese Kanji and Kana writings (Sasanuma, 1975) may all be interpreted in such terms. But, convincing demonstrations remain to be established on the basis of neurofunctional data.

In summary:

(1) The brain–computer metaphor does not apply to the development of the brain. The brain does not develop via the part-by-part assembly of prewired circuits.
(2) On the contrary, its morphogenesis is progressive, with forms becoming intricated within forms, including possibly, at each step, sensitive phases of limited transient variability and exuberance followed by drastic elimination of connections.
(3) The proposal is made that a selection of synapses takes place at these sensitive steps and is governed by the state of activity of the developing nervous system.
(4) According to this view, the effect of experience is intertwined with innate processes from development through the adult stage. The formation of brain architectures is not independent of cognitive processes but, rather, is deeply impregnated by them, starting from the early stages of postnatal development.

VI. Mental objects and mental Darwinism

1. "Representations" and Hebbian assemblies

J.Z. Young (1964) introduced his insightful book, *Models of the brain*, by defining the brain as the "computer of a homeostat" (p. 14). Being a homeostat the organism can exist in several states, and its adaptation is achieved by selection among possible actions provided by some antecedent process.

Young further speculated that such selection is appropriate in that the homeostat continues its self-maintenance. For it to do so, the organism must adequately represent the situation as a set of physical events (signals) that transmit information. Thus, "The organism is (or contains) a representation of its environment" (p. 20).

The word "representation" has several different meanings. In the cognitive and computer sciences, it refers equally to the structure of internalized information as to its content. Thus, theorists who postulate a "language of thought" (Fodor, 1975) take mental representations to have a combinatorial syntax and semantic content. These representations are often contrasted with the operations performed with them, which are described in algorithmic terms. On the other hand, in neuroscience, the word representation mainly refers to the projections of sensory organs onto defined areas of the brain, or to the mappings of given domains of the brain upon others (Mountcastle, 1978). In his book, *The organization of behavior*, Hebb (1949) introduced the first bridge between the neural and the mental by postulating that "an assembly of association-area cells which can act briefly as a closed system after stimulation has ceased ... constitutes the simplest instance of representative process (image or idea)" (p. 60). For Hebb, the assembly is described by the firing of an anatomically defined population of cells, and "an individual cell or transmission unit may enter in more than one assembly, at different times" (p. 196). Hebb thus posits the assembly as a "three-dimensional lattice", or as a net of neurons, with *coordinated* activity.

Alternative views to Hebb's concept of assembly have been advocated by various authors. For instance, Barlow (1972) has assumed that the coding units of concepts or representations can be identified with the activity of single neurons named "grandmother" or "pontifical" cells. Consistent with such a notion, single cells have been recorded that respond to particular objects, faces or even words (references in Heit, Smith, & Halgren, 1988; Perrett, Mistlin, & Chitty, 1987). In between, there are neuronal groups (Edelman & Mountcastle, 1978) or clusters of cells (Dehaene, Changeux, & Nadal, 1987; Feldman, 1986). Depending on the level at which the coding takes place, pontifical cells or clusters of such cells may be viewed either as autonomous (individual) units or as building blocks for higher order assemblies (Hopfield, 1982).

From both an experimental and theoretical point of view, the actual size of the population of neurons involved in the coding of mental representations remains a debated issue. A wide range of plausible sizes has been suggested, from a single nerve cell to the whole brain. Nevertheless, extensive single-unit recording of populations of cells in awake animals (references in Georgopoulos, Schwartz, & Kettner, 1986; Llinás, 1987; Motter, Steinmetz, Duffy, & Mountcastle, 1987; Steinmetz, Motter, Duffy, & Mountcastle, 1987) as well as large-scale high-resolution imaging (Posner et al., 1988; Roland & Friberg, 1985) give credence to the notion that these mental objects correspond to privileged activity states of widely distributed domains of the brain and have an identifiable, if not yet specified, physical basis.

Nevertheless, Von der Malsburg (1981, 1987) and Von der Malsburg and Bienenstock (1986) have criticized the cell-assembly concept as formalized by Little (1974) and Hopfield (1982), for two main reasons: (1) in a given brain area, a coding assembly is expected to correspond to only a fraction of a larger ensemble of active units; and (2) states which have the same global distribution of features might be confused with each other. Von der Malsburg and Bienenstock thus propose that the discrimination between coding and non-coding units relies on the temporal correlation between active units rather than on the fact that they are active or not. Such correlations become established between action potentials within a time-scale of a few milliseconds, during the overall time-scale of a representation (tenths of second), and are mediated by fast changes of synaptic strength. On this basis, Von der Malsburg and Bienenstock have developed a formalism whereby topologically organized synaptic patterns can be stored and retrieved, and whereby invariant pattern recognition finds a natural solution. The critical aspect of this formalism is to specify the notion of a firing correlation. Two final remarks have to be made about this view. First, the formalism is fully consistent with Hebb's original proposal of synchronization of activity between active cells, and thus only conflicts with the recent reformulations of the cell assembly concept in terms of statistical physics. Second, under physiological conditions, *both* the actual firing of individual nerve cells and the correlation of firing between cells are likely to contribute to the coding of mental representations.

2. Mental Darwinism

According to Hebb (1949), the growth of the cell assembly is determined by the repeated simultaneous excitation of cells consecutive to sensory stimulation in such a manner that each cell assists each other in firing. The time coincidence of firing between two cells increases the efficiency of the synapse linking the cells. In other words, the genesis of mental representations would occur through an instructive or "Lamarckian" mechanism (cf. Rolls, 1987). As noted, the thesis we wish to defend in the following is the opposite; namely, that the production and storage of mental representations, including their chaining into meaningful propositions and the development of reasoning, can also be interpreted, by analogy, in variation–selection (Darwinian)

terms within psychological time-scales (Changeux, 1983b; Changeux et al., 1973, 1984; Dehaene et al., 1987; Edelman & Mountcastle, 1978; Edelman & Finkel, 1984; Finkel & Edelman, 1987; Heidmann et al., 1984).

A basic requirement for such "mental Darwinism" is the existence of a generator of variety (diversity), which would internally produce the so-called Darwinian variations. At psychological time-scales, and at the two levels of intendment and reason, such variations may be viewed as resulting from spontaneous activity of nerve cells. Hebb and many of his followers did not make explicit a possible differential contribution of spontaneous and evoked activity in coding mental representations. But, the occurrence of spontaneous firing by cellular oscillators (see Berridge & Rapp, 1979) and/or oscillatory circuits (see Grillner et al., 1987; Stent et al., 1978) makes possible a strictly internal production of representations.

In a selectionist framework (Changeux, 1983b; Heidmann et al., 1984; Toulouse, Dehaene, & Changeux, 1986), one may thus schematically distinguish: (1) percepts in which the correlation of firing among the component neurons is directly determined by the interaction with the outside world and ceases when the stimulation has stopped; (2) images, concepts and intentions that are actualized objects of memory resulting from the activation of a stabilized trace; and (3) *pre-representations* (analogous to the Darwinian variations), which may be spontaneously developed, multiple, labile, and transient, and which may be selected or eliminated.

Pre-representations would be produced by the neural forms described in the preceding section with both their genetically encoded components (resulting from the evolution of the genome in geological time-scales) and their epigenetic ones (resulting from the consequences of embryonic and postnatal development). Pre-representations would correspond to privileged spontaneous activity states of these wired-in forms, occurring in shorter time-scales (0.1 seconds) and in a transient, dynamic and fleeting manner. At any given time, these pre-representations would be composed of sets of neurons from much larger populations of cells. As a consequence, the correlated firing of diverse combinations of neurons or groups of neurons (or even already stored assemblies) and a wide variety of active neural graphs would be produced successively and transiently. Such pre-representations may arise at the level of intendment and take part in the elaboration and storage of percepts at concepts. They may also arise at the level of reason and contribute to the genesis of higher-order programs and intentions representing the synthesis of concepts (see section VII). The genesis of pre-representations by such a mechanism would thus offer one neural component for the so-called productivity of cognitive processes (Fodor & Pylyshyn, 1988).

At a given stage of the evolution of the organism, some of these spontaneously generated pre-representations may not match any defined feature of the environment (or any item from long-term memory stores) and may thus be transiently meaningless. But, some of them will ultimately be selected in novel situations, thus becoming "meaning full". The achievement of such adequacy (fitness) with the environment (or with a given cognitive state) would then become the basic criterion for selection. The matching between a percept and a pre-representation has been referred to as resonance (Changeux, 1983b; Toulouse et al., 1986) and its unmatching as dissonance. Matching is likely to take place with global, invariant representations of the outside world resulting from the transformation of a set of sensory vectorial components into an invariant functional state (Llinás, 1987; Teuber, 1972).

Finally, mental representations are *not* static states of the brain, but are produced and chained within a system in constant dynamic interaction with the outside world and within itself. They are part of a system in constant evolution within psychological time-scales.

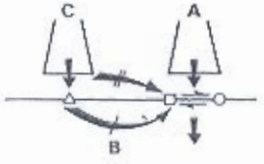
3. *The Hebb synapse and elementary mechanism of resonance*

To account for the growth of the assembly at the first stage of perception, Hebb (1949) postulates that "when an axon of cell A is near enough to excite cell B and repeatedly, or persistently, takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency as one of the cells firing B is increased" (p. 62). This proposal has been much debated from both theoretical and experimental points of view, and many cellular implementations of the Hebb synapse have been suggested since Hebb's original proposal (for instance, see Stent, 1973; Kelso, Ganong, & Brown, 1986; and for a recent review, see Bear et al., 1987).

Heidmann and Changeux (1982) (see also Changeux & Heidmann, 1987; Finkel & Edelman, 1987) have proposed a molecular mechanism whereby synaptic efficacy is governed by the conformational states of the postsynaptic receptor for a given neurotransmitter. The percentage of receptor molecules in activable and inactivable conformations would be a measure of efficacy, and could be modulated by converging electrical and/or chemical postsynaptic signals within a given time-window (Figure 2). This model can account for a variety of modes of regulation of synapse efficacy, as found for instance in *Aplysia* (Abrams & Kandel, 1988; Changeux et al., 1987a; Hawkins & Kandel, 1984) where presynaptic changes appear postsynaptic to a true Hebbian process, or in the cerebellum (Ito, Sakurai, & Tongroach, 1982), and may include so-called non-Hebbian cases.

Most important to our present purpose, such allosteric mechanisms for the Hebb rule and for other rules for synaptic enhancement may serve as basic

Figure 2. *A model of the regulation of synapse efficacy at the postsynaptic level based on the allosteric properties of the acetylcholine receptor (from Dehaene et al., 1987). The conformation of receptor molecules can be affected by intra- or extracellular chemical potential of the postsynaptic cell.*



devices to implement the resonance process defined above. It must be stressed that these mechanisms and rules do not imply any revival of "associationist" ideas (Fodor & Pylyshyn, 1988). Spontaneous firing may also regulate the efficacy of afferent synapses by the same rules. Conditions may be defined to occur for synaptic modifications only during a state of resonance between the activity of afferent synapses and the spontaneous firing of the target neuron (Dehaene et al., 1987). Thus, there is no intrinsic empiricist feature in the way these molecular regulatory devices operate.

4. Modelling mental objects by statistical physics

The test of a biological theory may sometimes be carried out succesfully on qualitative grounds without the help of a mathematical formalism. In most instances, however, the predictions appear difficult to derive intuitively, and the elaboration of a formal model in a coherent and simplified form becomes necessary. This is particularly true for assemblies of neurons, and an abundant literature has recently been published on this matter.

McCulloch and Pitts (1943) first described neurons as simple, all-or-none threshold devices. The further introduction of an analogue of temperature (Little, 1974; see Burnod & Korn, 1989) and of variable but symmetrical synapses (Hopfield, 1982) made possible the application of the ready-made formalism of statistical physics to neuronal networks. Hopfield (1982) showed how a content-addressable memory could be constructed, whereby information is stored as stable attractors of the dynamics via synaptic modifications. Several unrealistic features of the model were later revised, such as symmetrical interactions between neurons (Sompolinsky & Kanter, 1986; Derrida, Gardner, & Zippelius, 1987) or catastrophic deterioration of the memory with overloading (Mézard, Nadal, & Toulouse, 1986; Nadal et al., 1986a).

An important aspect of the attempt to formalize neural networks concerns the actual origin of the firing activity and of the interactions between neurons. In the instructive framework, which is the most commonly adopted (Amit, Gutfreund, & Sompolinsky, 1985a, b; Hopfield, 1982), the interaction with the outside world imposes the internal state of activity of the network. In the initial state, the interactions are vanishingly small, and the energy landscape flat. Among several drawbacks, this hypothesis does not take into account the existence of an already heavily connected network and the occurrence of spontaneous activity within the network (see above). In terms of the spin glass formalism (Toulouse et al., 1986), a selectionist model, in contrast, posits an initially rich energy landscape with pre-existing interactions between neurons and an abundance of hierarchically organized attractors. The interaction with the outside world would not enrich the landscape, but rather would select pre-existing energy minima or pre-representations and enlarge them at the expense of other valleys. As a consequence, the whole energy landscape would evolve, the already stored information influencing the pre-representations available for the next learning event. The system possesses internal rigidity so that not every external stimulus would be equally stored. The crucial issue remains to find a learning rule coherent with such a Darwinian picture. A relatively simple initial model was proposed in Toulouse et al. (1986). It begins with a random distribution of synaptic efficacies, and stores patterns using the same rule as the original Hopfield model. Although this does not allow one to make use of the hierarchical structure of pre-existing attractors, such networks do possess an internal rigidity. Moreover, additional biological constraints, such as excitatory synapses not becoming inhibitory and vice versa, are easily implemented, taking this model a step further toward the selective model. Still, one of the major limitations of this model is that it deals with "spin glass" under static conditions, while selection of mental states in the brain always takes place under constant dynamic conditions (see section VII).

In summary:

(1) Experimental and theoretical evidence from cellular neurophysiology and statistical physics make plausible the hypothesis that mental representations can be defined as states of activity of brain cells.

(2) Models for the Darwinian selection of mental states can be proposed on the basis of:

(a) the spontaneous productions of transient, dynamic and coherent but fleeting activity states referred to as pre-representations, which would be analogues of the Darwinian variations;

(b) the selection and stabilization of some of these pre-representations by matching percepts arising from external and somatic internal stimuli

(e.g., see Damasio, this issue), stimuli with already selected internal states.

VII. Neural architecture and the application of theoretical models to cognitive science

Models inspired from statistical physics deal mostly with networks of fully interconnected neurons, whereby neither the singularity of each individual neuron nor higher-order architectural principles of the networks are specified. To better approach a more realistic neurobiology, some modellists have studied simple systems such as *Limax* (Hopfield & Tank, 1986) or *Tritonia* (Kanter & Sompolinsky, 1987) despite the fact that the cognitive abilities of these creatures are rather rudimentary. Our approach is the opposite. While anatomists, physiologists and neuropsychologists decipher the real functional organization of brain connectivity in mammals (see Goldman-Rakic, 1987; Rakic, 1988) and in humans (Geschwind & Galaburda, 1987), it may be possible starting from simple networks to reconstruct cognitive functions by introducing architectural and functional constraints within, and between, neuronal networks and, in a second step, to look for their presence in real brains. In the following section, we try to establish a parallel between the behaviour of networks and well-known cognitive functions, knowing, *a priori*, that such attempts are a simplistic but necessary preliminary to more complex and plausible reconstruction.

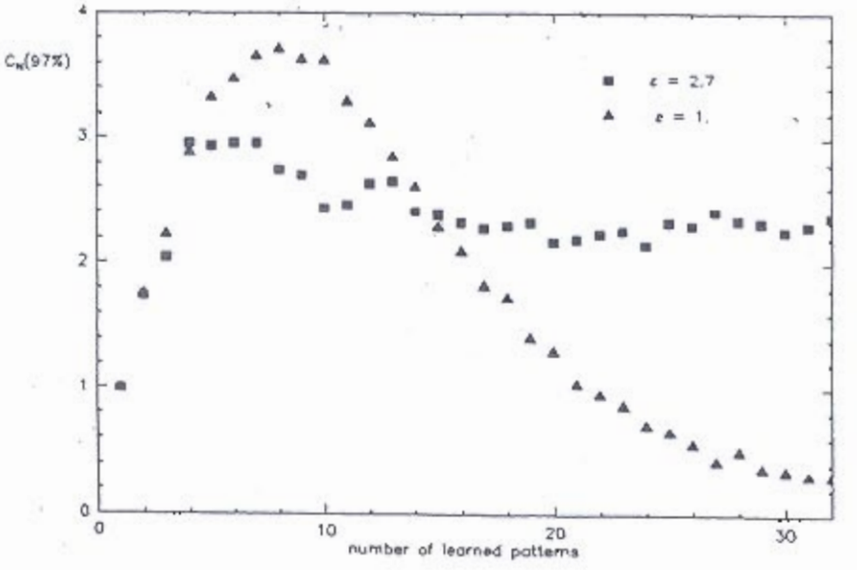
1. Short-term memory of a neuronal network

The simplest feature displayed by formal neuronal networks is the storage of memories within a limited-capacity network. As described above, the Hopfield model can be specified and completed in order to retain only the most recent information it received. A direct relation exists between the steady-state, memory capacity of this network and its connectivity. For instance, a network in which each neuron is connected to 500 other neurons has a memory capacity of seven items (Nadal et al., 1986a, b). Both numbers are within a plausible range. Extension of such evaluations to short-term memory in humans is attractive but still highly hypothetical. Yet, at least for the most perceptually driven memory stores, the application of the Hopfield instructive scheme seems legitimate. A long-standing debate in the field has been the origin of forgetting (see Baddeley, 1976, for discussion). Erasure of the short-term memory traces might result from a spontaneous decay or, alternatively, from an interference of recent memories over older ones. Spontaneous decay

may simply result from the relaxation kinetics of the molecular transitions of proteins engaged in synaptic transmission. The modelling of formal neuronal networks according to the Hopfield scheme, as exemplified by the palimpsest model (Nadal et al., 1986a, b), illustrates, however, the physical plausibility of an interference mechanism. It further accounts for characteristic features of short-term memory stores such as the "primacy effect", according to which the first item stored is more easily recalled than the subsequent ones (Nadal et al., 1986a, b).

An important feature of this model is that it points to the dependence of global functional features (a fixed memory span) upon elementary synaptic parameters such as the average number of synapses per neuron (see Figure 3). As emphasized by Simon (1969), "the most striking limits to subjects' capacities to employ efficient strategies arise from the very small capacity of

Figure 3. *Model of memory palimpsest. The percentage of memorized patterns drops catastrophically as more and more patterns are added into the Hopfield (1982) model (triangles). A minor modification yields a stationary regime (squares) where only recently learned patterns can be retrieved (palimpsest model: from Nadal et al., 1986a; reproduced by permission of the publishers, Editions de Physique, Les Ulis, France).*



the short-term memory structure" (p. 76). These models, at variance with the classical functionalist point of view, thus illustrate how elementary parameters of the brain may place fundamental constraints on the pattern of thought.

2. Organization of long-term memory

The long-term memory stores in humans possess an apparent unlimited capacity, are strongly hierarchical, and are organized along a semantic, rather than perceptual, space (Baddeley, 1976, 1986, 1988; Massaro, 1975). Storage into long-term memory appears as a rare and slow event compared to the short-term storage of information, and may be viewed as a semantically driven, Darwinist selection from representations present in the short-term store.

Several models for storing a hierarchical tree of memories in a neural network have been proposed (Feigelman & Toffe, 1987; Gutfreund, 1988; Parga & Virasoro, 1985). One of them (Gutfreund, 1988) relies on an architecture based upon multiple distinct networks, one for each level of the tree. The retrieval of a particular memory is achieved at the lowest network, and is assisted by the retrieval of its ancestors, and mimics access to long-term stores in humans (see Baddeley, 1986b, 1988). The models inspired from statistical physics thus offer an elementary physical implementation of semantic relationships in long-term memory, which partially covers the notion of "meaning" (Amit, 1988). In this respect it is worth noting that blood-flow studies involving retrieval of specific memories, such as numbers, nine-word jingles or sequence of visual fields reveal something in the activation of topographically distinct though interconnected cortical areas (Roland & Friberg, 1985). Neuropsychological investigations show that even words belonging to different semantic categories may be represented at different loci in the brain (e.g., McCarthy & Warrington, 1988).

Finally, the so-called unlimited capacity of the long-term store is more apparent than real. First of all, reasonable evaluations of its capacity converge at a value of 10^9 bits (see Mitchinson, 1987), which is a small number compared to the 10^{11} neurons of the human brain and its 10^{15} synapses. Second, the transfer from short- to long-term memory looks rather limited considering the large number of representations transiently circulating in the short-term store. Third, only truly new items, distinct from those already present in long-term memory, are stored at any given time. The transfer from short- to long-term stores may thus be viewed as a selection for novelty occurring via the validation of pre-existing hierarchies and the stabilization of small branches, thereby saving considerable space.

3. Plausible molecular mechanism for long-term storage

Mental objects have been defined as transient physical states of neuronal networks with durations in the time-scale of fractions of seconds. Such activity-dependent changes of neuronal and synaptic properties may be extended to longer time-scales by covalent modifications of neuronal and synaptic proteins and, ultimately, by the regulation of protein synthesis.

Yet, at variance with currently accepted views (Goelet, Castellucci, Schacher, & Kandel, 1986; Montarolo et al., 1986), Changeux and Heidmann (1987) have argued that long-term regulations cannot be equated to regulation of protein synthesis, but rather to the perpetuation of an activity-dependent trace beyond protein turnover. The simplest and most plausible general mechanism is that of a self-sustained metabolic steady-state that includes a *positive* feedback loop (or negative ones in *even* numbers (Delbrück, 1949; Thomas, 1981). Such self-reinforcing circuits may be built at the level of neuronal receptors (Changeux & Heidmann, 1987; Crick, 1984; Lisman, 1985), gene receptors (Britten & Davidson, 1969; Monod & Jacob, 1961), and even at the level of the synapse on the basis of a sequence of chemical reactions. In this last instance (Changeux et al., 1987a) a positive feedback loop may be created by the activity-dependent regulation of the production (by the postsynaptic cell) of a growth factor required for the stability of the afferent nerve ending. There is no theoretical time limit to the maintenance of a trace in a system of that sort up to the life span of the organism.

Plausible molecular mechanisms for the extension to long-term changes of synapse efficacies may thus be envisioned at the level of the synapse in which the short-term modification took place. On the other hand, the occurrence of neuropsychological disconnections in the transfer from short- to long-term stores in human patients supports the notion that in the human brain these two compartments might be topographically distinct though highly interconnected.

4. Recognition, production and storage of time sequences

The attempts to model neuronal networks that have been mentioned concern either stable states or relaxation to stable states. However, as emphasized, the nervous system does not process information under static conditions. At the cellular or simple circuits level, it produces coherent patterns of linear or cyclic sequences of activity (Getting, 1981; Grillner, 1975; Grillner et al., 1987; Stent et al., 1978). At higher levels, including the knowledge level, the nervous system possesses the striking faculty to recognize, produce and store time sequences (Lashley, 1951). One basic function of the frontal cortex

mentioned in section II is the control of the temporal evolution of overt behaviour and internal chains of mental representations. An important issue is thus: what are the minimal requirements of neural architecture and function for a network to process temporal sequences?

Complex networks that recognize and produce sequences of higher order have recently been proposed that rely on two sets of synaptic connections: one set which stabilizes the network in its current memory state, while a second set, the action of which is delayed in time, causes the network to make transitions between memories (Kleinfeld, 1986; Sompolinsky & Kanter, 1986; Peretto & Niez, 1986; Amit, 1988). Others (Tank & Hopfield, 1987) are based on delay filters, one for each known sequence. In all these instances, the time delay is built-in as an intrinsic physical parameter of individual neurons, such as postsynaptic potential or axon length.

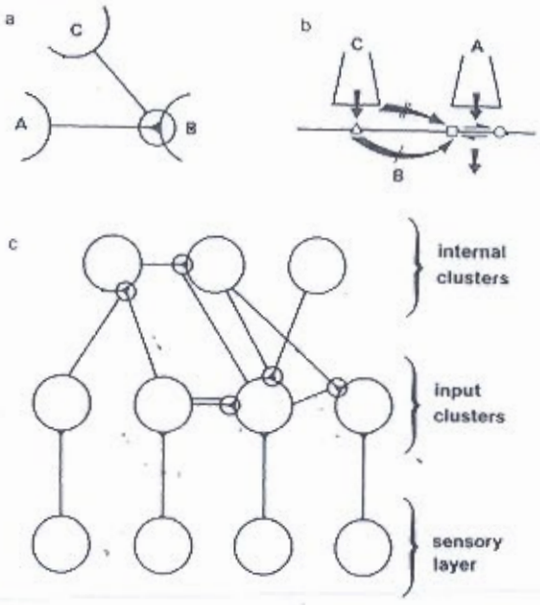
The network model we have proposed (Dehaene et al., 1987) displays capacities for recognition and production of temporal sequences and for their acquisition by selection. It was inspired by the learning of songs in birds such as *Melospiza* by selective attrition of syllables (Konishi, 1985; Marler & Peters, 1982), whereby identified neurons, called song-specific neurons, detect sequences of syllables. It also makes use of the known properties of allosteric receptors (Changeux, 1981; Changeux & Heidmann, 1987; Heidmann & Changeux, 1982), which may potentially serve as regulators of synapse efficacy at the postsynaptic level (including Hebbian mechanisms: see section V).

The model is based on four assumptions about neural architecture (Figure 4): (1) synaptic triads are composed of three neurons A, B, and C. These are connected so that the efficacy of a synapse of A on B is influenced by the activity of a third modulator neuron C, under conditions that make the postsynaptic neuron B behave as a sequence detector for neurons A and C; (2) a Hebbian learning rule increases the maximal efficacy of the A–B synapse toward an absolute maximum if, after its activation, postsynaptic neuron B fires; (3) the network is formed of juxtaposed clusters of synergic neurons densely interconnected by excitatory synapses, thus able to maintain self-sustained stable activity; and (4) the network is subdivided into three superimposed layers: a sensory layer on which percepts are merely imposed, and an internal production network subdivided into two layers for input clusters and internal clusters, respectively.

The network displays several original properties:

(a) Passive recognition and production of temporal patterns. The few architectural designs of the network, the presence of synaptic triads and the existence of clusters with a linear or hierarchical organization, suffice for the recognition of time sequences. As a consequence of the self-excitatory connections within clusters, the activity of previous states remains within the cluster. Individual triads acting as elementary sequence detectors, and the linear and hierarchical arrays of clusters thus behave as detectors of complex sequences that may include repetitions.

Figure 4. *Model of formal neuron networks capable of recognition, production and storage of temporal sequences (from Dehaene et al., 1987) (for explanation see text).*



Furthermore, the same network produces time sequences. A set of triads between clusters transmits activity with a delay. A memory of previous activity is kept through the persistence, or *remanence*, of former activity within the clusters. Remanent activity may thus influence the pathways that the system takes. The temporal span of this remanence will thus determine the complexity of the sequence produced. It is an elementary neural implementation of context dependence which is basic to cognitive psychology and linguistics. Amit (1988) provides a similar demonstration, although in a different framework, that the same stimulus can elicit different network responses depending on the remanent internal state.

(b) Genesis of internal organization by learning. Introduction of the Hebbian learning rule and the subdivision of the network in two distinct layers, one for the input and the other strictly for internal representations, makes possible the differentiation of hierarchically organized sequence detectors from randomly connected clusters. Conditions can be found in which the imposition of a melody to the input clusters leads progressively to elimination of the initial redundancy and to stabilization of hierarchies of sequence detectors. Conversely, sequences of arbitrary complexity may be produced in the network by stabilization of ongoing spontaneous activity during interaction (resonance) with an externally applied melody.

The model thus illustrates how neurons coding for a temporal relationship between activity clusters may differentiate through experience according to a Darwinian mechanism. It also suggests that abstract relations (rules) may be extracted and stored in hierarchically organized neurons or, conversely, imposed to lower-level neurons coding for a variety of more concrete features.

From an experimental point of view the model points to the role of synaptic triads and to the differentiation through experience of neuronal hierarchies. It has global predictions about variability and its reduction in the course of learning, which are strikingly consonant with observations made on children in the course of phonology acquisition (see Ferguson, 1985).

5. Intentions and inventions

Intentionality and meaning appear so basic to cognition that those who do not address these issues are often viewed as missing the whole field! An elementary step toward the physical implementation of meaning has already been made. On the other hand, the problem of intentionality, even in a very limited sense, has not been considered. Any attempt to unravel its neural bases appears unrealistic *a priori* (Searle, 1983). Our aim is not to propose a neural theory of intentionality. Rather, we limit ourselves to a few remarks and arguments illustrating how such a theory might be constructed, with respect to goals and plans.

(a) Intentions and the frontal cortex

The brief discussion on the contribution of the frontal lobe to the architecture of reason (section II) led to the suggestion that this region of the brain contributes to the elaboration of plans, and controls the temporal unfolding of patterned mental operations or behavioural actions according to a goal. Frontal lesions do interfere with planning behaviour and result in a typical unintentional utilization behaviour (Lhermitte, 1983) that one may associate with the lower intendment level. The prefrontal cortex must thus have a *prospective* function in anticipating and planning and a *retrospective* one in maintaining it in a provisional memory until the goal is reached (Fuster, 1980).

Single-unit recordings in the prefrontal cortex during the delay period of a delayed alternation task disclose cells that remain active for seconds before the response and code for the anticipated direction of the motor response (Bruce, 1988; Niki, 1974). Neural activity therefore can be related to the achievement of a goal. From this it is reasonable to conclude that there exists a neural basis for intentions in humans, as there exists a neural basis for goals in monkeys. Intentions may concern both motor acts and thought processes (Searle, 1983). Accordingly, an intention will be viewed as a particular category of mental object characterized by (1) its occurrence at a high level of organization, such as the level of reason in the frontal cortex, and (2) a long-lasting, *predictive*, activity. Self-excitatory clusters of neurons would offer one implementation for the maintenance of intentions, but the involvement of cellular oscillators or, more likely, of reverbatory closed circuits involving positive feedback loops appears equally plausible (see Changeux, 1983b). The intervention of attentional processes (see Posner & Presti, 1987) would contribute to the fixation of such self-sustained states of activity, which would be determined either externally through percepts or internally through the evocation of memory objects.

(b) Intentionality and context dependence

Intentions are viewed as occurring on top of the hierarchy of brain networks and may create a context for underlying motor actions on the environment and the chaining of mental objects by imposing a frame of semantic constraints on these processes. The model proposed for the recognition, storage and retrieval of time sequences (Dehaene et al., 1987) bases context dependence on the joint contribution of the remanence of activity in self-excitatory clusters and of the recognition (or production) of time sequences by synaptic triads. It might serve as a general framework for the elaboration of more complex models for the context created by intentions, for the development of reasonings and in a more general manner for what Fodor and Pylyshyn (1988) refer to as structure-sensitive operations.

(c) The selectionist (Darwinian) test for intentions and inventions

A basic function of the frontal cortex is to capture errors in the unfolding of a motor program. Similarly, intentions might be subjected to internal tests. The validation of a proposition, for example, would then result from a context-dependent compatibility of a chain of mental objects within a given

semantic frame with already-stored mental objects. Such tests for compatibility or adequateness might be viewed, from a neural point of view, as analogues of the matching by resonance (or un-matching by dissonance) of percepts with pre-representations.

The contribution of attention to the processing of sensory information is currently being investigated in great detail by joint psychological and neurological approaches (Posner & Presti, 1987; Posner et al., 1988). It appears plausible that similar, if not identical, attentional systems are involved in maintaining coherent patterns designed to reach a goal (Posner, 1980; Posner & Presti, 1987). "Attention for action" (Posner et al., 1988) is a generic name for the attentional processes involved in the selection of actions, intentions or goals. As in the case of the selection of meaning, but at a different level and on a different time-scale, the selection of actions and intentions may take place via a Darwinian mechanism among internally evoked and context-dependent pre-representations. Of course, such selection will concern higher-order representations including complex chains of objects from the long-term memory stores. Combinatorial processes may produce novel intentions or inventions at this level. The selection will then be carried out by testing their realism from the cognitive as well as affective point of view. The connections existing between the limbic system and the prefrontal cortex offer a material basis for relationships between the emotional and cognitive spheres (Goldman-Rakič, 1987; Nauta, 1971, 1973).

In summary, on the basis of rather simple architectural designs, physical models of neural networks can be proposed that account for some characteristic features of short-term and long-term memory and for the recognition, production and storage of time sequences. Such models make plausible a neural theory of intentions.

VIII. Conclusions

Despite the rather speculative character of some sections of this presentation, we conclude that it is timely to approach cognition in a synthetic manner with the aim to relate a given cognitive function to its corresponding neural organization and activity state. Such a neurocognitive trend (see also Arbib, 1985; Luria, 1973; Struss & Benson, 1986) contrasts with the classical functionalist approach to cognition (Fodor, 1975; Johnson-Laird, 1983), although it is more compatible with a revised, more recent version (Fodor & Pylyshyn, 1988). This approach has several important consequences.

It contends that data from cognitive psychology often overlook many highly intricate levels of functional organization that have to be distinguished within the human brain. Reconstructing architecture on the basis of external observations alone is a complex matter, if not an ill-defined task, with no unique solution. Neuropsychology and neuro-imagery (Posner et al., 1988) usefully complement the psychological approach by offering ways to dissect such global functions into elementary operations that are localized in the brain, and help to show that the cleavage of brain functions into the classical neurological, algorithmic and semantic levels is no longer appropriate and may even be misleading. Ultimately, both theoretical models and experiments must be devised in such a manner that they specify the particular level of neural organization to which a given function is causally related.

It further argues that the classical artificial intelligence approach, which tries to identify the programs run by the hardware of the human brain, loses much of its attractive power. This may be overcome if one conceives brain-style computers based on the actual architectural principles of the human brain and possessing some of its authentic competences rather than simply mimicking some of its surface performances (for references, see Sejnowski, Koch, & Churchland, 1988). Models of highly evolved functions, for example the acquisition of past tense in English (Rumelhart & McClelland, 1987) or the ability to read a text aloud (Sejnowsky & Rosenberg, 1986) among others (Lehkey & Sejnowski, 1988; Zipser & Andersen, 1988) have been implemented in simplistic connectionist machines. But, despite an "appearance of neural plausibility" (Fodor & Pylyshyn, 1988), the architectures involved are far too simple and even naive compared to those that the human brain actually uses for such multimodal performances with deep cultural impregnation (Pinker & Prince, 1988). As stated by Fodor and Pylyshyn (1988), these systems cannot "exhibit intelligent behavior without storing, retrieving or otherwise operating on structured symbolic expressions" (p. 5). A basic requirement of a plausible neurocognitive approach is thus to unravel the physics of meaning or, in other words, the neural bases of mental representations. The attempt to find the implementation of the semantic content of symbolic expressions in neural terms cannot be viewed as secondary to a psychological theory of meaning (Fodor & Pylyshyn, 1988). Theories of the neural implementation of mental representations may raise useful issues such as the capacity of short-term memory, the hierarchical organization of long-term memory, the recognition and storage of time sequences, and context dependence, as illustrated in a still rather primitive form in this paper. An important issue will be the search for the neural representation of rules (in particular, syntax) and their application to restricted classes of mental representations. In this respect, the model of learning temporal sequences by selection (Dehaene et al., 1987) illustrates how neurons coding for relations between mental objects may differentiate as a consequence of experience.

The model recently served as a starting point for an investigation of the role of prefrontal cortex in delayed-response tasks (Dehaene & Changeux, 1989), which are elementary rule-governed behaviours. Such implementations will one day be described in terms of real neural connections and will thus point to critical experimental predictions. Such abstract neurological theories based upon the most advanced progress of brain anatomy and physiology may ultimately unravel novel algorithms and architectures out of the still largely unexplored universe of human brain connectivity. Our view is that there is much more to expect from such an approach than from strict psychological and/or mathematical theorizing.

Another conclusion that we wish to draw from this discussion is that the brain should be viewed as an evolutive system rather than as a static input–output processing machine. The brain is part of an organism belonging to a species that has evolved (and is still evolving) in geological time-scales according to Darwinian mechanisms at the level of the genome. But the complexity of the brain is such that it may itself be considered as a system evolving *within* the organism with, at least, two distinct time-scales: that of embryonic and postnatal development for the process of organizing neuronal somas and connectivity networks, and that of psychological times for the storage, retrieval and chaining of mental objects and for their assembly into higher-order motor programs, behavioural strategies and schemas. The extension of *selectionist* mechanisms to all these levels breaks down the rigidity (sphexishness; Dennett, 1984) of the strictly nativist or Cartesian schemes (Fodor, 1983) by introducing, at each level, a degree of freedom linked with the production of variations. But as long as these variations are constrained by the genetic envelope, it escapes the pitfalls of Lamarckian associationism. The generator of internal diversity produces, at each of these levels, intrinsic richness, and thereby offers possibilities for creating structures within a given level but also between levels, yielding again one plausible component for the productivity requirements of Fodor and Pylyshyn (1988). Moreover, the number of such choices need not be large to cause an important diversity as long as variability exists at all hierarchically organized levels. The criteria for selection of the pre-representations must be defined at each level. At the lower levels, an obvious criterion is the adequateness or fitness to the environment. At a higher one, the internal thought (Gedanken) experiments might refer to the outcome of former experiences stored in the long-term memory and in the genetic endowment of the species which, *in fine*, ensure its survival (see Dennett, 1984, 1987).

Brain would thus be an evolutive system constantly anticipating the evolution of its physical, social and cultural environment by producing expectations and even intentions that create a lasting frame of reference for a selected set of long-term memories. Brain would not only be a semantic engine (Dennett 1984) but an intentional engine. At no level would such a machine be neutra about the nature of cognitive processes. Rather, it would be "knowledge-im pregnated" from the organization of its genome up to the production of it: more labile intentions.

References

Abrams, T., & Kandel, E. (1988). Is contiguity detection in classical conditioning a system or a cellular property? Learning in *Aplysia* suggests a possible molecular site. *Trends in Neuroscience, 11*, 128–135.

Akam, M. (1987). The molecular basis for metameric pattern genes of *Drosophila*. *Development, 101*, 1–22.

Alkon, D., Disterhoft, J., & Coulter, D. (1987). Conditioning-specific modifications of postsynaptic membrane currents in mollusc and mammal: In J.P. Changeux & M. Konishi (Eds.), *The neural and molecular bases of learning*. New York: Wiley.

Amit, D.J. (1988). Neural networks counting chimes. *Proceedings of the National Academy of Science USA. 85*, 2141–2145.

Amit, D.J., Gutfreund, H., & Sompolinsky, H. (1985a). Spin glass models of neural networks. *Physica Review, A32*, 1007–1018.

Amit, D.J., Gutfreund, H., & Sompolinsky, H. (1985b). Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters, 55*, 1530–1533.

Arbib, M. (1985). *In search of the person*. Amherst, MA: University of Massachusetts Press.

Awgulewitsch, A., Utset, M.F., Hart, C.P., McGinnis, W., & Ruddle, F. (1986). Spatial restriction in expression of a mouse homeobox locus within the central nervous system. *Nature, 320*, 328–335.

Baddeley, A.D. (1976). *The psychology of memory*. New York: Harper & Row.

Baddeley, A.D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A.D. (1988). Cognitive psychology and human memory. *Trends in Neuroscience, 11*, 176–181.

Barlow, H.B. (1972). Single units and sensations: a neuron doctrine for perceptual physiology? *Perception, 1*, 371–394.

Bear, M.F., Cooper, L.N., & Ebner, F.F. (1987). A physiological basis for a theory of synapse modification. *Science, 237*, 42–48.

Benoit, P., & Changeux, J.P. (1975). Consequences of tenotomy on the evolution of multi-innervation in developing rat soleus muscle. *Brain Research, 99*, 354–358.

Berridge, M., & Rapp, P. (1979). A comparative survey of the function, mechanism and control of cellular oscillations. *Journal of Experimental Biology, 81*, 217–280.

Britten, R.J., & Davidson, E.H. (1969). Gene regulation for higher cells: a theory. *Science, 165*, 349–357.

Bruce, C. (1988). What does single unit analysis in the prefrontal areas tell us about cortical processing? In P. Rakic & W. Singer (Eds.), *Neurobiology of neocortex*, Dahlem Konferenzen. Chichester: Wiley.

Burnod, Y., & Korn, H. (1989). Consequence of stochastic release of neurotransmitters for network computation in the central nervous system. *Proceedings of the National Academy of Science USA, 86*, 352–356.

Callaway, E., Soha, J., & Von Essen, D. (1987). Competition favoring inactive over active motor neurons during synapse elimination. *Nature, 328*, 422–426.

Cavalli-Sforza, L., & Fedelman, M. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton, NJ: Princeton University Press.

Changeux, J.P. (1981). The acetylcholine receptor: An "allosteric" membrane protein. *Harvey Lectures, 75*, 85–254.

Changeux, J.P. (1983a). Concluding remarks on the "singularity" of nerve cells and its ontogenesis. *Progress in Brain Research, 58*, 465–478.

Changeux, J.P. (1983b). *L'homme neuronal*, Paris: Fayard. English translation by L. Garey (1985). *Neuronal Man*. New York: Pantheon Books.

Changeux, J.P. (1984). Le regard du collectionneur. *Catalogue de la donation Othon Kaufmann et François Schlageter au Departement des peintures, Musée du Louvre*. Paris: Edition de la Réunion des Musées Nationaux.

Changeux, J.P. (1986). Coexistence of neuronal messengers and molecular selection. *Progress in Brain Research, 68*, 373–403.

Changeux, J.P., Courrège, P., & Danchin, A. (1973). A theory of the epigenesis of neural networks by selective stabilization of synapses. *Proceedings of the National Academy of Science USA, 70*, 2974–2978.

Changeux, J.P., Courrège, P., Danchin, A., & Lasry, J.M. (1981). Un mécanisme biochimique pour l'épigénèse de la jonction neuromusculaire. *C.R. Acad. Sci. Paris, 292*, 449–453.

Changeux, J.P., & Danchin, A. (1974). Apprendre par stabilisation sélective de synapses en cours de développement. In E. Morin & M. Piattelli (Eds.), *L'unité de l'homme*. Paris: Le Seuil.

Changeux, J.P., & Danchin, A. (1976). Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. *Nature, 264*, 705–712.

Changeux, J.P., Devillers-Thiéry, A., Giraudat, J., Dennis, M., Heidmann, T., Revah, F., Mulle, C., Heidmann, O., Klarsfeld, A., Fontaine, B., Laufer, R., Nghiêm, H.O., Kordeli, E., & Cartaud, J. (1987b). The acetylcholine receptor: Functional organization and evolution during synapse formation. In O. Hayaishi (Ed.), *Strategy and prospects in neuroscience*. Utrecht: VNU Science Press.

Changeux, J.P., & Heidmann, T. (1987). Allosteric receptors and molecular models of learning. In G. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Synaptic function*. New York: Wiley.

Changeux, J.P., Heidmann, T., & Patte, P. (1984). Learning by selection. In P. Marler & H. Terrace (Eds.), *The biology of learning*. Berlin: Springer-Verlag.

Changeux, J.P., Klarsfeld, A., & Heidmann, T. (1987a). The acetylcholine receptor and molecular models for short and long term learning. In J.P. Changeux & M. Konishi (Eds.), *The cellular and molecular bases of learning*. Chichester: Wiley.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1979). Le débat entre Jean Piaget et Noam Chomsky. In M. Piattelli-Palmarini (Ed.), *Théories du langage – théories de l'apprentissage*. Paris: Le Seuil.

Churchland, P. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Clarke, E., & O'Malley, C. (1968). *The human brain and spinal cord. A historical study illustrated by writings from antiquity to the twentieth century*. Berkeley, CA: University of California Press.

Clarke, S., & Innocenti, G. (1986). Organization of immature intrahemispheric connection. *Journal of Comparative Neurology, 251*, 1–22.

Cowan, M.W., Fawcett, J.W., O'Leary, D., & Stanfield, B.B. (1984). Regressive phenomena in the development of the vertebrate nervous system. *Science, 225*, 1258–1265.

Crick, F. (1984). Memory and molecular turnover. *Nature, 312*, 101.

D'Arcy-Thompson, W. (1917). *On growth and form*. Cambridge: Cambridge University Press.

Debru, C. (1983). *L'esprit des protéines*. Paris: Hermann.

Dehaene, S., & Changeux, J.P. (1989). A single model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience* (in press).

Dehaene, S., Changeux, J.P., & Nadal, J.P. (1987). Neural networks that learn temporal sequences by selection. *Proceedings of the National Academy of Sciences USA, 84*, 2727–2731.

Delbrück, M. (1949). *In Unités biologiques douées de continuité génétique (Publication CNRS)*, 33–35.

Delbrück, M. (1986). *Mind from matter*. Palo Alto, CA: Blackwell.

Dennett, D. (1984). *Elbow room*. Cambridge, MA: MIT Press.

Dennett, D. (1987). *The intentional stance*. New York: Basic Books.

Derrida, B., Gardner, E., & Zippelius, A. (1987). An exactly solvable asymmetric neural network model. *Europhysics Letters, 4*, 167–170.

Derrida, B., & Nadal, J.P. (1987). Learning and forgetting on a symmetric diluted neural network. *Journal of Statistical Physics, 49*, 993–1009.

Doe, C.Q., Hiromi, Y., Gehring, W.J., & Goodman, C.S. (1988). Expression and function of the representation gene *fushi tarazu* during *Drosophila* neurogenesis. *Science, 239*, 170–175.

Edelman, G.M. (1978). Group selection and phasic reentrant signaling: a theory of higher brain function. In G.M. Edelman and V.B. Mounteastle (Eds.), *The mindful brain: Cortical organization and the group-selective theory of higher brain function* (pp. 51–100). Cambridge, MA: MIT Press.

Edelman, G.M., (1985). Molecular regulation of neural morphogenesis. In G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Molecular bases of neural development*. New York: Wiley.

Edelman, G.M. (1987). *Neural Darwinism*. New York: Basic Books.

Edelman, G.M., & Finkel, L. (1984). Neuronal group selection in the cerebral cortex. In G. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Dynamic aspects of neocortical function*. New York: Wiley.

Edelman, G.M., Gall, W.E., & Cowan, W.M. (Eds.) (1984). *Dynamic aspects of neocortical function*. New York: Wiley.

Edelman, G.M., & Mountcastle, V. (Eds.) (1978). *The mindful brain: Cortical organization and the group-selective theory of higher brain function*. Cambridge, MA: MIT Press.

Eimas, P.D. (1975). In L.B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition* (Vol. 2). New York: Academic Press.

Feigelman, M.V., & Toffe, L.B. (1987). The augmented model of associative memory asymmetric interaction and hierarchy of pattern. *International Journal of Modern Physics, B, 1*, 51–68.

Feldman, J.A. (1986). Neural representation of conceptual knowledge. *Technical report, Department of Computer Science, University of Rochester*, TR189, June 1986.

Ferguson, C.A. (1985). Discovering sound units and constructing sound systems: It's child's play. In J.S. Perkell & D.H. Klatt (Eds.), *Invariance and variability in speech processes*. Hillsdale, NJ: Erlbaum.

Finkel, L.H., & Edelman, G.M. (1987). Population rules for synapses in networks. In G.M. Edelman, W.E. Gall, & M.W. Cowan (Eds.), *Synaptic function*. New York: Wiley.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3–71.

Fraser, S.E. (1985). Cell interactions involved in neuronal patterning: An experimental and theoretical approach. In G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Molecular bases of neural development*. New York: Wiley.

Fuster, J.M. (1980). *The prefrontal cortex*. New York: Raven.

Fuster, J.M. (1984). Electrophysiology of the prefrontal cortex. *Trends in Neuroscience, 1*, 408–414.

Gazzaniga, M.S. (1987). The dynamics of cerebral specialization and modular interactions. In L. Weiskrantz (Ed.), *Thought without language*. Oxford: Clarendon Press.

Georgopoulos, A.P., Schwartz, A.B., & Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science, 233*, 1357–1460.

Geschwind, N., & Galaburda, A. (Eds.) (1984). *Cerebral dominance: The biological foundations*. Cambridge, MA: Harvard University Press.

Geschwind, N., & Galaburda, A.M. (1987). *Cerebral lateralization*. Cambridge MA: MIT Press.

Getting, P.A. (1981). Mechanism of pattern generation underlying swimming in *Tritonia*. I. Neuronal network formed by monosynaptic connections. *Journal of Neurophysiology, 46*, 65–79.

Ghering, W. (1985). Homeotic genes, the homeobox and the genetic control of development. *Cold Spring Harbor Symposium for Quantitative Biology, 50*, 243–251.

Goelet, P., Castellucci, V., Schacher, S., & Kandel, E. (1986). The long and the short of long-term memory: A molecular framework. *Nature, 322*, 419–422.

Goldman-Rakic, P. (1987). Circuitry of the primate prefrontal cortex and the regulation of behavior by

representational knowledge. In V. Mountcastle & K.F. Plum (Eds.), *The nervous system: Higher functions of the brain, Vol. 5, Handbook of Physiology*. Washington, DC: American Physiological Society.

Goldowitz, D., & Mullen, R. (1982). Granule cell as a site of gene action in the weaver mouse cerebellum. Evidence from heterozygous mutant chimerae. *Journal of Neuroscience, 2*, 1474–1485.

Gombrich, E.H. (1960). *Art and illusion*. Oxford: Phaidon Press.

Gombrich, E.H. (1983). *L'écologie des images*. Paris: Flammarion.

Goodman, C.S., Bastiani, M.J., Raper, J.A., & Thomas, J.B. (1985). Cell recognition during neuronal development in grasshopper and *Drosophila*. In G.M. Edelman, W.E. Gall, & W.M. Cowan (Eds.), *Molecular bases of neural development*. New York: Wiley.

Gouzé, J.L., Lasry, J.M., & Changeux, J.P. (1983). Selective stabilization of muscle innervation during development: A mathematical model. *Biological Cybernetics, 46*, 207–215.

Grillner, S. (1975). Locomotion in vertebrates. Central mechanisms and reflex interaction. *Physiological Review, 55*, 247–304.

Grillner, S., Wallén, P., Dale, N., Brodin, L., Buchanan, J., & Hill, R. (1987). Transmitters, membrane properties and network circuitry in the control of locomotion in lamprey. *Trend in Neuroscience, 10*, 34–41.

Gutfreund, H. (1988). Neural networks with hierarchically correlated patterns. *Physical Review, 91*, 375–391.

Hamburger, V. (1970). Embryonic mobility in vertebrates. In F.O. Schmitt (Ed.), *The Neurosciences: Second study program*. New York: Rockefeller University Press.

Hawkins, R.D., & Kandel, E. (1984). Is there a cell-biological alphabet for simple forms of learning? *Psychological Review, 91*, 375–391.

Hebb, D.O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Heidmann, A., Heidmann, T., & Changeux, J.P. (1984). Stabilisation selective de représentations neuronales par résonance entre "pré-représentations" spontanées du réseau cérébral et "percepts" évoqués par interaction avec le monde extérieur. *C.R. Acad. Sci. Paris* (série 3), *299*, 839–844.

Heidmann, T., & Changeux, J.P. (1982). Un modèle moléculaire de régulation d'efficacité d'un synapse chimique au niveau postsynaptique. *C.R. Acad. Sci. Paris* (série 3), *295*, 665–670.

Heit, G., Smith, M.E., & Halgren, E. (1988). Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature, 333*, 773–775.

Hökfelt, T., Holets, V.R., Staines, W., Meister, B., Melander, T., Schalling, M., Schultzberg, M., Freedman, J., Björklund, H., Olson, L., Lindk, B., Elfvin, L.G., Lundberg, J., Lindgren, J.A., Samuelsson, B., Terenius, L., Post, C., Everitt, B., & Goldstein, M. (1986). Coexistence of neuronal messengers: An overview. *Progress in Brain Research, 68*, 33–70.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA, 79*, 2554–2558.

Hopfield, J., & Tank, D.W. (1986). Computing with neural circuits: A model. *Science, 233*, 625–635.

Hubel, P., & Wiesel, T. (1977). Functional architecture of macaque monkey visual cortex. Ferrier Lecture. *Proceedings of the Royal Society (London) B, 198*, 1–59.

Huttenlocher, P.R., De Courten, C., Garey, L.J., & Vander Loos, H. (1982). Synaptogenesis in human visual cortex. Evidence for synapse elimination during normal development. *Neuroscience Letters, 33*, 247–252.

Innocenti, G.M., & Caminiti, R. (1980). Postnatal shaping of callosal connections from sensory areas. *Experimental Brain Research, 38*, 381–394.

Ito, M., Sakurai, M., & Tongroach, P. (1982). Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cells. *Journal of Physiology (London), 324*, 113–134.

Jackson, H. (1932). In J. Taylor (Ed.), *Selected Papers* (Vol. 2). London: Hodder & Stoughton.

Jerne, N. (1967). Antibodies and learning: Selection versus instruction. In G. Quarton, T. Melnechuck, & F.O. Schmitt (Eds.), *The Neurosciences*. New York: Rockefeller University Press.

Johnson, M.H., McConnell, J., & Van Blerkom, J. (1984). Programmed development in the mouse embry *Journal of Embryology and Experimental Morphology, 83* (Suppl.), 197–231.

Johnson-Laird, P.N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Kandel, E.R., Abrams, T., Bernier, L., Carew, T.J., Hawkins, R.D., & Schwartz, J.H. (1983). Classic conditioning and sensitization share aspects of the same molecular cascade in *Aplysia*. *Cold Spri Harbor Symposium on Quantitative Biology, 48*, 821–830.

Kanter, I., & Sompolinsky, H. (1987). Associative recall of memory without errors. *Physical Review A, 3* 380–392.

Kelos, S.R., Ganong, A.H., & Brown, T.H. (1986). Hebbian synapses in hippocampus. *Proceedings of t National Academy of Science USA, 83*, 5326–5330.

Kihlstrom, J. (1987). The cognitive unconscious. *Science, 237*, 1445–1452.

Kleinfeld, D. (1986). Sequential state generation by model neural networks. *Proceedings of the Nation Academy of Sciences USA, 83*, 9469–9473.

Kleinfeld, D., & Sompolinsky, H. (1987). Associative neural network model for the generation of tempor patterns: Theory and application to central pattern generators. Unpublished paper.

Kolb, B., & Whishaw, I. (1980). *Fundamentals of human neuropsychology*. San Francisco: Freeman.

Konishi, M. (1985). Bird songs: From behavior to neuron. *Annual Review of Neurophysiology, 8*, 125–170

Lashley, K.S. (1951). *Central mechanisms in behavior*. New York: Wiley.

Lehkey, S.R., & Sejnowski, T.J. (1988). Network model of shape-from-shading: Neural function arises fro both receptive and projective fields. *Nature, 333*, 452–454.

Levinthal, F., Macagno, E., & Levinthal, C. (1976). Anatomy and development of identified cells in isogen organisms. *Cold Spring Harbor Symposium on Quantitative Biology, 40*, 321–332.

Lhermitte, F. (1983). "Utilization behavior" and its relation to lesions of the frontal lobe. *Brain, 106*, 237–235.

Lisman, J.E. (1985). A mechanism for memory storage insensitive to molecular turnover: A bistable autopho phorylating kinase. *Proceedings of the National Academy of Science USA, 82*, 3055–3057.

Little, W.A. (1974). Existence of persistent states in the brain. *Mathematical Bioscience, 9*, 101–120.

Llinás, R.R. (1987). "Mindness" as a functional state of the brain. In C. Blakemore & S. Greenfield (Eds. *Mindwaves*. London: Basil Blackwell.

Lumsden, C., & Wilson, E.O. (1981). *Genes, mind and culture: The coevolutionary process*. Cambridge, MA Harvard University Press.

Luria, A.R. (1973). *The working brain: An introduction to neuropsychology*. New York: Basic Books.

Macagno, F., Lopresti, U., & Levinthal, C. (1973). Structural development of neuronal connections in isogeni organisms: Variations and similarities in the optic tectum of *Daphnia magna*. *Proceedings of the Na tional Academy of Science USA, 70*, 57–61.

McCarthy, R.A., & Warrington, E.K. (1988). Evidence for modality-specific meaning systems in the brain *Nature, 334*, 428–430.

McCulloch, W.S., & Pitts, W.A. (1943). Logical calculus of the ideas immanent in nervous activity. *Bulleti of Mathematical Biophysics, 5*, 115.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Mariani, J., & Changeux, J.P. (1981a). Ontogenesis of olivocerebellar relationships: I – Studies by intracellula recordings of the multiple innervation of Purkinje cells by climbing fibers in the developing rat cerebel lum. *Journal of Neuroscience, 1*, 696–702.

Mariani, J., & Changeux, J.P. (1981b). Ontogenesis of olivocerebellar relationships: II – Spontaneous activit of inferior olivary neurons and climbing fiber-mediated activity of cerebellar Purkinje cells in develop ing rats and in adult cerebellar mutant mice. *Journal of Neuroscience, 1*, 703–709.

Marler, P., & Peters, S. (1982). Development overproduction and selective attrition: New process in the epigenesis of bird song. *Developmental Psychobiology, 15*, 369–378.

Marshall, J.C. (1988). The lifeblood of language. *Nature, 331*, 560–561.

Massaro, D. (1975). *Experimental psychology and information processing*. Chicago: Rand McNally.
Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Harvard University Press.
Mehler, J., Morton, J., & Jusczyk, P.W. (1984). On reducing language to biology. *Cognitive Neuropsychology, 1*, 83–116.
Meinhardt, H., & Gierer, A. (1974). Application of a theory of biological pattern formation based on lateral inhibition. *Journal of Cell Science, 15*, 321–346.
Merzenich, M.M. (1987). Dynamic neocortical processes and the origins of higher brain functions. In J.P. Changeux & M. Konishi (Eds.), *The neural and molecular bases of learning*, New York: Wiley.
Mézard, M., Nadal, J.P., & Toulouse, G. (1986). Solvable models of working memories. *Journal de Physique (Paris), 47*, 1457–1462.
Mitchinson, G. (1987). The organization of sequential memory: Sparse representations and the targeting problem. *Proceedings of the Bad Homburg meeting on Brain Theory*, 16–19 September, 1986.
Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1975). An effect of linguistic experience: the discrimination of |r| and |l| by native speakers of Japanese and English. *Perception and Psychophysics, 18*, 331–340.
Monod, J., & Jacob, F. (1961). General conclusions: Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harbor Symposium for Quantitative Biology, 26*, 389–401.
Monod, J. (1970). *Le hasard et la nécessité*. Paris: Le Seuil.
Montarolo, P.G., Goelet, P., Castellucci, V.F., Morgan, T., Kandel, E., & Schacher, S. (1986). A critical period for macromolecular synthesis in long-term heterosynaptic facilitation in *Aplysia*. *Science, 234*, 1249–1254.
Motter, B.C. Steinmetz, M.A., Duffy, C.J., & Mountcastle, V.B. (1987). Functional properties of parietal visual neurons: Mechanisms of directionality along a single axis. *Journal of Neuroscience, 7*, 154–175.
Mountcastle, V. (1978). An organizing principle for cerebral function: The unit module and the distributed system. In G.M. Edelman & V. Mountcastle (Eds.), *The mindful brain: Cortical organization and the group-selective theory of higher brain function*. Cambridge, MA: MIT Press.
Nadal, J.P., Toulouse, G., Changeux, J.P., & Dehaene, S. (1986a). Networks of formal neurons and memory palimpsests. *Europhysics Letters, 1*, 535–542.
Nadal, J.P., Toulouse, G., Mézard, M., Changeux, J.P., & Dehaene, S. (1986b). Neural networks: Learning and forgetting. In R.J. Cotterll (Ed.), *Computer simulations and brain science*. Cambridge: Cambridge University Press.
Nass, R.D., Koch, D.A., Janowsky, J., & Stile-Davis, J. (1985). Differential effects on intelligence of early left versus right brain injury. *Annals of Neurology, 18*, 393.
Nass, R.D., Koch, D.A., Janowsky, J., & Stile-Davis, J. (1989). Differential effects of congenital left and right brain injury on intelligence (in press).
Nauta, W.J.H. (1971). The problem of the frontal lobe: A reinterpretation. *Journal of Psychiatric Research, 8*, 167–187.
Nauta, W.J.H. (1973). Connections of the frontal lobe with the limbic system. In L.V. Laitiven & K.E. Livingston (Eds.), *Surgical approaches in psychiatry*. Baltimore, MD: University Pack Press.
Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18*, 87–127.
Niki, H. (1974). Prefrontal unit activity during delayed alternation in the monkey. I. Relation to direction of response. II. Relation to absolute versus relative direction of response. *Brain Research, 68*, 185–196.
Nüsslein-Volhard, C., Frohnhöffer, H.G., & Lehmann, R. (1987). Determination of anteroposterior polarity in Drosophila. *Science, 238*, 1675–1681.
Ojemann, G. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *Behavioral and Brain Science, 6*, 189–230.
Oster-Granite, M., & Gearhart, J. (1981). Cell lineage analysis of cerebellar Purkinje cells in mouse chimaeras. *Development Biology, 85*, 199–208.
Parga, N., & Virasoro, M.A. (1985). Ultrametric organization of memories in neural network. *Journal de Physique Lettres, 47*, 1857.

Peretto, P., & Niez, J.J. (1986). Collective properties of neural networks. In E. Bienenstock, F. Fogelman & G. Weisbuch (Eds.), *Disordered systems and biological organization*. Berlin: Springer-Verlag.
Perrett, D.I., Mistlin, A.J., & Chitty, A.J. (1987). Visual neurons responsive to faces. *Trends in Neuroscience, 10*, 358–364.
Personnaz, L., Guyon, I., & Dreyfus, G. (1985). Information storage and retrieval in spin-glass like neur. networks. *Journal de Physique Lettres, 46*, L359.
Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., & Raichle, M.E. (1988). Positron emission tomograph studies of the cortical anatomy of single-word processing. *Nature, 331*, 585–589.
Piaget, J. (1979). *Behavior and evolution*. London: Routledge & Kegan Paul.
Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel model of languag acquisition. *Cognition, 28*, 73–913.
Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.
Posner, M.I., Petersen, S.E., Fox, P.T., & Raichle, M.E. (1988). Localization of cognitive operations in th human brain. *Science, 240*, 1627–1631.
Posner, M., & Presti, D.F. (1987). Selective attention and cognitive control. *Trends in Neuroscience, 10*, 13–1
Price, D.J., & Blakemore, C. (1985). Regressive events in the postnatal development of association projectior in the visual cortex. *Nature, 316*, 721–723.
Prince, A., & Pinker, S. (1988). Rules and connections in human language. *Trends in Neuroscience, 1*, 195–202.
Prince, D.A., & Huguenard, J.R. (1988). Functional properties of neocortical neurons. In P. Rakic & W Singer (Eds.), *Neurobiology of the neocortex*. Chichester: Wiley.
Purves, D., & Lichtman, J.W. (1980). Elimination of synapses in the developing nervous system. *Scienc 210*, 153–157.
Pylyshyn, Z. (1985). Plasticity and invariance in cognitive development. In J. Mehler & R. Fox (Eds. *Neonate cognition*. Hillsdale, NJ: Erlbaum.
Rakic, P. (1988). Intrinsic and extrinsic determinants of neocortical parcellation: a radial unit model. In I Rakic & W. Singer (Eds.), *Neurobiology of neocortex*. Chichester: Wiley.
Rakic, P., & Singer, W. (Eds.) (1988). *Neurobiology of the neocortex*. Chichester: Wiley.
Ramon y Cajal, S. (1909). *Histologie du système nerveux de l'homme et des vertébrés* (2 vols.). Paris: Maloine.
Redfern, P.A. (1970). Neuromuscular transmission in newborn rats. *Journal of Physiology (London), 20* 701–709.
Reiter, H.O., & Stryker, M.P. (1988). Neural plasticity without postsynaptic action potentials: Less-activ inputs become dominant when kitten visual cortical cells are pharmacologically inhibited. *Proceedings of the National Academy of Science USA, 85*, 3623–3627.
Ribchester, R.R. (1988). Activity-dependent and -independent synaptic interactions during reinnervation partially denervated rat muscle. *Journal of Physiology, 401*, 53–75.
Rockwell, A., Hiorns, R., & Powell, T. (1980). The basic uniformity in structure of the neocortex. *Brain 103*, 221–224.
Roland, P.E., & Friberg (1985). Localization of cortical areas activated by thinking. *Journal of Neurophysiology, 53*, 1219–1243.
Rolls, E. (1987). Information representation, processing and storage in the brain: Analysis at the single neuro level. In J.P. Changeux & M. Konishi (Eds.), *The neural and molecular bases of learning*. Chichester Wiley.
Rugg, M. (1988). Stimulus selectivity of single neurons in the temporal lobe. *Nature, 333*, 700.
Rumelhart, D.E., & McClelland, J.L. (1987). Learning the past tenses of English verbs: Implicit rules parallel distributed processing? In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
Sasanuma, S. (1975). Kana and Kanji processing in Japanese aphasics. *Brain and Language, 2*, 369–383.
Schmidt, J. (1985). Factors involved in retinotopic map formation: Complementary roles for membrane reco

nition and activity-dependent synaptic stabilization. In G.M. Edelman, W.E. Gall, & W.N. Cowan (Eds.), *Molecular bases of neural development*. New York: Wiley.
Searle, J.R. (1983). *Intentionality: An essay in the philosophy of mind*. New York: Cambridge University Press.
Sejnowsky, T.J., Koch, C., & Churchland, P.S. (1988). Computational neuroscience. *Science, 241*, 1299–1306.
Sejnowsky, T.J., & Rosenberg, C.R. (1986). *NET-talk: A parallel network that learns to read aloud*. Technical report JHU/EECS-86/01. Department of Electrical Engineering and Computer Science, John Hopkins University.
Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B, 298*, 199–209.
Simon, H.A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.
Sompolinsky, H., & Kanter, I. (1986). Temporal association in asymmetric neural networks. *Physical Review Letters, 57*, 2861–2864.
Sperber, D. (1984). Anthropology and psychology: Towards an epidemiology of representations. *Man* (N.S.), *20*, 73–89.
Sretavan, D.W., Shatz, C.J., & Stryker, M.P. (1988). Modification of retinal ganglion cell axon morphology by frontal infusion of tetrodotoxin. *Nature, 336*.
Steinmetz, M.A., Motter, B.C., Duffy, C.J., & Mountcastle, V. (1987). Functional properties of parietal visual neurons: Radial organization of directionalities within the visual field. *Journal of Neuroscience, 7*, 177–191.
Stent, G. (1973). A physiological mechanism for Hebb's postulate of learning. *Proceedings of the National Academy of Science USA, 70*, 997–1001.
Stent, G. (1981). Strength and weakness of the genetic approach to the development of the nervous system. *Annual Review of Neuroscience, 4*, 163–194.
Stent, G. (1987). The mind–body problem. *Science, 236*, 990–992.
Stent, G.S., Kristan, W.B., Friesen, W.O., Ort, C.A., Poon, M., & Calabrese, R.L. (1978). Neuronal generation of the leech swimming movement. *Science, 200*, 1348–1356.
Stryker, M.P., & Harris, W.A. (1986). Binocular impulse blockage prevents the formation of ocular dominance columns in cat visual cortex. *Journal of Neuroscience, 6*, 2117–2133.
Stuss, D., & Benson, F. (1986). *The frontal lobes*. New York: Raven.
Taine, H. (1870). *De l'intelligence*. Paris: Hachette.
Tank, D.W., & Hopfield, J.J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Science USA, 84*, 1896–1900.
Teuber, H.L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiology Experiments (Warsz.), 32*, 615–656.
Thom, R. (1980). *Modèles mathématiques de la morphogènèse*. Paris: Bourgeois.
Thomas, R. (1981). On the relation between the logical structure of systems and their ability to generate multiple steady-states or sustained oscillations. *Springer Series in Synergetics, 9*, 180–193.
Toulouse, G., Dehaene, S., & Changeux, J.P. (1986). Spin glass model of learning by selection. *Proceedings of the National Academy of Science USA, 83*, 1695–1698.
Turing, A.M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society (London), 237*, 37–72.
Van Essen, D.G. (1982). Neuromuscular synapse elimination: Structural, functional and mechanistic aspects. In N.C. Spitzer (Ed.), *Neuronal development*. New York: Plenum.
Von der Malsburg, C. (1981). *The correlation theory of brain function*. Internal report 8–12 July 1981, Department of Neurobiology, Max Plank Institute for Biophysical Chemistry, Göttingen.
Von der Malsburg, C. (1987). Synaptic plasticity as basis of brain organization. In J.P. Changeux & M. Konishi (Eds.), *The neural and molecular bases of learning*. Chichester: Wiley.
Von der Malsburg, C., & Bienenstock, E. (1986). Statistical coding and short-term plasticity: A scheme for knowledge representation. In E. Bienenstock, F. Fogelman, & G. Weisbuch (Eds.), *Disordered systems and biological organization*. Berlin: Springer-Verlag.

Wilson, E.O. (1975). *Sociobiology*. Cambridge, MA: Harvard University Press.
Young, J.Z. (1964). *A model of the brain*. Oxford: Clarendon Press.
Young, J.Z. (1973). Memory as a selective process. *Australian Academy of Science Reports: Symposium on the Biology of Memory*, 25–45.
Zeki, S. (1988). Anatomical guides to the functional organization of the visual cortex. In P. Rakic & W. Sing (Eds.), *Neurobiology of neocortex*. Chichester: Wiley.
Zipser, D., & Andersen, R.A. (1988). A back-propagation programmed network that stimulates respon properties of a subset of posterior parietal neurons. *Nature, 331*, 679–684.

Résumé

Comprendre les bases neurales de la cognition est devenu un problème abordable scientifiquement, et d modèles sont proposés dans le but d'établir un lien causal entre organisation neurale et fonction cognitiv Dans ces conditions, il devient nécessaire de définir des niveaux d'organisation dans l'architecture fonctionel des systèmes de neurones. Les transitions d'un niveau d'organisation à l'autre sont envisagées dans u perspective évolutive: elles ont lieu avec différentes échelles de temps, et reposent sur la production transitoi de nombreuses variations et la sélection de certaines d'entre elles au cours de l'interaction avec le mon extérieur. Au cours du développement et chez l'adulte, cette évolution interne est de nature épigénétiqu elle ne requiert pas d'altération du génôme. L'activité (spontanée ou évoquée) d'un réseau de neurones cours du développement stabilise de manière sélective certaines synapses et en élimine d'autres, contribuar de ce fait, à la mise en place de la connectivité adulte à l'intérieur d'une enveloppe de potentialités défini génétiquement. A un niveau supérieur, la modélisation de représentations mentales par des états d'activi de populations restreintes de neurones est réalisée par les méthodes de la physique statistique: la mémorisati de ces représentations est envisagée comme un processus de sélection parmi des "pré-représentations" va ables et instables. Des modèles théoriques montrent que des fonctions cognitives comme la mémoire à cou terme ou la manipulation de séquences temporelles peuvent dépendre de paramètres physiques élémentaire Une implémentation neuronale et sélectionniste des intentions est envisagée.