Neural Systems and Artificial Life Group,
Institute of Psychology,
National Research Council, Rome

Evolutionary Connectionism and Mind/Brain Modularity

Raffaele Calabretta and Domenico Parisi

Technical Report NSAL 01-01

February 16, 2001
(revised June 4, 2001)

To appear in: *Modularity. Understanding the development and evolution of complex natural systems*. The MIT Press, Cambridge, MA.

Department of Neural Systems and Artificial Life
Institute of Psychology, Italian National Research Council
V.le Marx, 15 00137 Rome - Italy
Phone: +39-06-86090233, Fax: +39-06-824737
E-mail: rcalabretta@ip.rm.cnr.it, parisi@ip.rm.cnr.it
http://gral.ip.rm.cnr.it

Evolutionary Connectionism and Mind/Brain Modularity

Raffaele Calabretta

(rcalabretta@ip.rm.cnr.it)

Domenico Parisi

(parisi@ip.rm.cnr.it)

Institute of Psychology, National Research Council, Rome, Italy
http://gral.ip.rm.cnr.it

Abstract

Brain/mind modularity is a contentious issue in cognitive science. Cognitivists tend to conceive of the mind as a set of distinct specialized modules and they believe that this rich modularity is basically innate. Cognitivist modules are theoretical entities which are postulated in "boxes-and-arrows" models used to explain behavioral data. On the other hand, connectionists tend to think that the mind is a more homogeneous system that basically genetically inherits only a general capacity to learn from experience and that if there are modules they are the result of development and learning rather than being innate. In this chapter we argue for a form of connectionism which is not anti-modularist or anti-innatist. Connectionist modules are anatomically separated and/or functionally specialized parts of a neural network and they may be the result of a process of evolution in a population of neural networks. The new approach, Evolutionary Connectionism, does not only allow us to simulate how genetically inherited information can spontaneously emerge in populations of neural networks, instead of being arbitrarily hardwired in the neural networks by the researcher, but it makes it possible to explore all sorts of interactions between evolution at the population level and learning at the level of the individual that determine the actual phenotype. Evolutionary Connectionism shares the main goal of Evolutionary Psychology, that is, to develop a psychology informed by the importance of evolutionary process in shaping the inherited architecture of human mind, but differs from Evolutionary Psychology for three main reasons: (1) it uses neural networks rather than cognitive models for interpreting human behavior; (2) it adopts computer simulations for testing evolutionary scenarios; (3) it has a less pan-adaptivistic view of evolution and it is more interested in the rich interplay between genetically inherited and experiential information. We present two examples of evolutionary connectionist simulations that show how modular architectures can emerge in evolving populations of neural networks.

1 Connectionism is not necessarily anti-modularist or anti-innatist

In a very general and abstract sense modular systems can be defined as systems made up of structurally and/or functionally distinct parts. While non-modular systems are internally homogeneous, modular systems are segmented into modules, i.e., portions of a system having a structure and/or function different from the structure or function of other portions of the system. Modular systems can be found at many different levels in the organization of organisms, for example at the genetic, neural, and behavioral/cognitive level, and an important research question is how modules at one level are related to modules at another level.

In cognitive science, the interdisciplinary research field that studies the human mind, modularity is a very contentious issue. There exist two kinds of cognitive science, *computational* cognitive science and *neural* cognitive science. Computational cognitive science is the more ancient theoretical paradigm. It is based on an analogy between the mind and computer software and it views mind as symbol manipulation taking place in a computational system (Newell & Simon, 1976). More recently a different kind of cognitive science, connectionism, has arisen which rejects the mind/computer analogy and interprets behavior and cognitive capacities using theoretical models which are directly inspired by the physical structure and way of functioning of the nervous system. These models are called neural networks, large sets of neuron-like units interacting locally through connections resembling synapses between neurons. For connectionism mind is not symbol manipulation and is not a computational system but is the global result of the many interactions taking place in a network of neurons modeled with an artificial neural network and consists entirely of quantitative processes in which physico-chemical causes produce physico-chemical effects. This

new type of cognitive science can be called neural cognitive science (Rumelhart & McClelland, 1986).

Computational cognitive science tends to be strongly modularistic. The computational mind is made up of distinct modules which specialize in processing distinct types of information, have specialized functions, and are closed to interference from other types of information and functions (Chomsky, 1980; Fodor, 1983). Computational cognitive models are schematized as "boxes-and-arrows" diagrams (for an example see Figure 1**)**. Each box is a module with a specific function and the arrows connecting boxes indicate that information processed by some particular module is then passed on to another module for further processing. In contrast, connectionism tends to be antimodularistic. In neural networks information is represented by distributed patterns of activation in potentially large sets of units and neural networks function by transforming activation patterns into other activation patterns through the connection weights linking the network's units. Most neural network models are not divided up into any kind of modules except for the distinction between input units, output units, and one or more layers of intermediate (hidden or internal) units **(**for an example see Figure 4, left**)**.

One cannot really understand the contrast between modularism and antimodularism in cognitive science, however, if one does not consider another contrast which opposes computational cognitive science (cognitivism) to neural cognitive science (connectionism). This is the contrast between innatism and anti-innatism. Cognitivists tend to be innatist. Modules are assumed to be specified in the inherited genetic endowment of the species and of each individual. For evolutionary

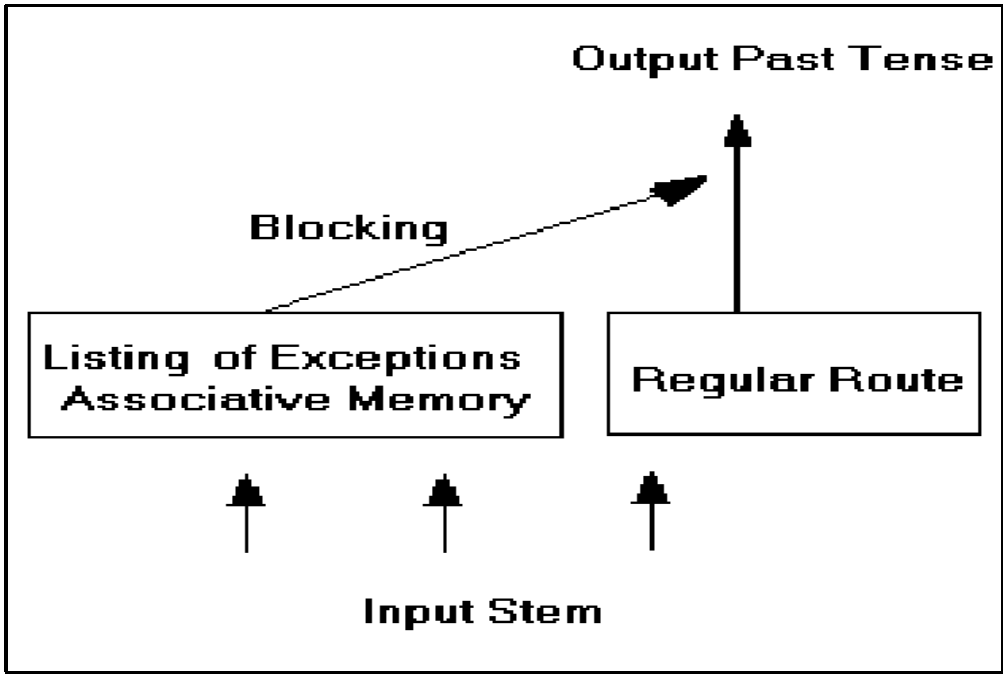

Figure 1. An example of "boxes-and-arrows" model: the dual-route model for the English past tense (Pinker & Prince 1988). "The model involves a symbolic regular route that is insensitive to the phonological form of the stem and a route for exceptions that is capable of blocking the output from the regular route" (Plunkett, 1996).

psychologists, who tend to be cognitivists, the modular structure of the mind is the result of evolutionary pressures and evolutionary psychologists are convinced that it is possible to identify the particular evolutionary pressures behind each module. Hence, evolutionary psychologists (Cosmides & Tooby, 1994) embrace a strong form of adaptivism. They not only think that modules are already there in the genetic material but they think that modules are in the genes because in the evolutionary past individuals with a particular module in their genes have generated more offspring than individuals without that genetically specified module. This pan-adaptivism is not shared by all cognitivists, however. For example, the linguist Noam Chomsky believes that the mind is computational and that there is a specific mental module specialized for language (or for syntax) but he does not believe that language in humans has emerged under some specific evolutionary pressure (cf. Fodor, 2000). As some evolutionary biologists, in particular Gould (1997), have repeatedly

stressed, what is genetically inherited is not necessarily the result of specific evolutionary pressures and is not necessarily adaptive but it can also be the result of chance, it can be the adaptively neutral accompaniment of some other adaptive trait, or an exaptation, i.e., the use for some new function of a trait which has evolved for another function (Gould & Vrba, 1982). More recently, the contrast between Steven Pinker and Jerry Fodor, who are both well-known cognivists and innatists, has shown how the adaptive nature of inherited traits can divide computational cognitive scientists. Pinker (Pinker, 1999) has argued for a strong form of adaptive modularism while Fodor is in favor of a strong form of non-adaptive modularism (Fodor, 1998).

In contrast to cognitivists, connectionists tend to be anti-innatist. Connectionism is generally associated with an empiricist position that considers all of mind as the result of learning and experience during life. What is genetically inherited, in humans, is only a general ability to learn. This general ability to learn, when it is applied to various areas of experience, produces the diverse set of capacities which are exhibited by humans.

The matter is further complicated if one considers development. Development is the mapping of the genetic information into the adult phenotype. This mapping is not instantaneous but is a process that takes time to complete, and in fact development consists of a temporal succession of phenotypical forms. When one recognizes that the genotype/phenotype mapping is a temporal process, the door is open for an influence of learning and experience on the phenotype. Therefore, cognitivists tend to be not only innatists but also antidevelopmentalists. Cognitivist developmental psychologists (e.g., Spelke *et al.*, 1992; Wynn, 1992) tend to think that modules are already there in the phenotype since the first stages of development and that there is not much of real importance that actually changes during life. Furthermore, as innatists, they think that even if something changes during development it is due not to learning and experience but to some temporal scheduling encoded in the genetically inherited information, like sexual maturity which is not present at birth but it is genetically

scheduled to emerge at some later time during life. On the contrary, developmental psychologists who are closer to connectionism (Karmiloff-Smith, 2000) tend to think that modules are not present in the phenotype from birth, i.e., in newborns or in infants, but develop later in life and, furthermore, they believe that modules are only very partially encoded in the genotype but are the result of complex interactions between genetically encoded information and learning and experience.

In the present paper we want to argue for a form of connectionism which is not anti-modularist or anti-innatist. Connectionism is not necessarily anti-innatist. Even if many neural network simulations use some form of learning algorithm to find the connection weights that make it possible for a neural network to accomplish some particular task, connectionism is perfectly compatible with the recognition that some aspects of a neural network are not the result of learning but they are genetically inherited. For example, since most simulations start from a fixed neural network architecture one could argue that this network architecture is genetically given and the role of learning is restricted to finding the appropriate weights for the architecture. In fact, Elman *et al.* (1996) have argued that connectionist networks allow the researcher to go beyond cognitivism, which simply affirms that this or that is innate, to explore in a detailed way what can be innate and what can be learned by showing how phenotypical capacities can result from an interaction between what is innate and what is learned. These authors distinguish among different things that can be innate in a neural network: the connection weights (and therefore the neural representations as patterns of activation across sets of network units), architectural constraints (at various levels: at the unit, local, and global level), and chronotopic constraints (which determine when things happen during development). One could also add that the connections weights may be learned during life but there may genetically inherited constraints on them, for example their maximum value or their "sign" (for excitatory or inhibitory connections) may be genetically specified or the genotype may encode the value of learning parameters such as the learning rate and momentum (Belew *et al.*,

1992). As we will show later in this chapter, modularity can emerge in neural networks as a function of genetically inherited architectural constraints and chronotopic constraints.

However, to argue that something is innate in a neural network it is not sufficient that some of the properties of the neural network are hardwired by the researcher in the neural network but it is necessary to actually simulate the evolutionary process that results in these genetically inherited properties or constraints. Artificial Life simulations differ from the usual connectionist simulations in that Artificial Life uses genetic algorithms (Holland, 1992) to simulate the evolutionary process and to evolve the genetically inherited properties of neural networks (Parisi *et al.*, 1990; Calabretta *et al.*, 1996). Unlike traditional connectionist simulations Artificial Life simulations simulate not an individual network that learns, based on its individual experience, some particular capacity, but they simulate an entire population of neural networks made up of a succession of generations of individuals each of which is born with a genotype inherited from its parents. Using a genetic algorithm, the simulation shows how the information encoded in the inherited genotypes changes across the successive generations because reproduction is selective and new variants of genotypes are constantly added to the genetic pool of the population through genetic mutations and sexual recombination. At the end of the simulation the inherited genotypes can be shown to encode the desired neural network properties that represent innate constraints on development and behavior. We call this type of connectionism Evolutionary Connectionism.

We can summarize the three options that are currently available to study the behavior of organisms with the Table 1.

Evolutionary connectionist simulations do not only allow us to study how genetically inherited information can spontaneously emerge in populations of neural networks, instead of being arbitrarily hardwired in the neural networks by the researcher, but they make it possible to explore

COMPUTATIONAL COGNITIVE SCIENCE or COGNITIVISM	Mind as symbol manipulation taking place in a computer-like system	INNATIST	MODULARIST
NEURAL COGNITIVE SCIENCE or CONNECTIONISM	Mind as the global result of the many physico-chemical interactions taking place in a network of neurons	ANTI- INNATIST	ANTI-MODULARIST
EVOLUTIONARY CONNECTIONISM	Mind as the global result of the many physico-chemical interactions taking place in a network of neurons	INTERACTION BETWEEN EVOLUTION AND LEARNING	MODULARIST

Table 1. Three options for studying behavior and mind

all sorts of interactions between evolution at the population level and learning at the level of the individual that determine the actual phenotype.

In this chapter we describe two evolutionary connectionist simulations that show how modular architectures can emerge in evolving populations of neural networks. In the first simulation every network property is genetically inherited (i.e., both the network architecture and the connection weights are inherited) and modular architectures result from genetically inherited chronotopic constraints and growing instructions for units' axons. In the second simulation the network architecture is genetically inherited but the connection weights are learned during life. Therefore, adaptation is the result of an interaction between what is innate and what is learned.

2 Cognitive vs. neural modules

Neural networks are theoretical models explicitly inspired by the physical structure and way of functioning of the nervous system. Therefore, given the highly modular structure of the nervous system it is surprising that so many neural network architectures that are used in connectionist

simulations have internally homogeneous architectures and do not contain separate modules. Brains are not internally homogeneous systems but they are made up of anatomically distinct parts and distinct portions of the brain are clearly more involved in some functions than in others. Since it is very plausible that human brains are able to exhibit so many complex capacities not only because they are made up of 100 billion neurons but also because these 100 billion neurons are organized as a richly modular system, future connectionist research should be aimed at reproducing in neural networks the rich modular organization of the brain.

However, even if, as we will shown by the two simulations described in this chapter, connectionist simulations can address the problem of the evolution of modular network architectures, it is important to keep in mind that the notion of a module is very different for cognitivists and for connectionists. Cognitivistic modularism is different from neural modularism.

For cognitivists modules tend to be components of theories in terms of which empirical phenomena are interpreted and accounted for. A theory or model of some particular phenomenon hypothesizes the existence of separate modules with different structure and/or function which by working together explain the phenomenon of interest. Therefore, cognitivist modules are *postulated* rather than *observed* entities. For example, in formal linguistics of the Chomskian variety syntax is considered as an autonomous module of linguistic competence in that empirical linguistic data (the linguistic judgements of the native speaker) are interpreted as requiring this assumption. Or, in psycholinguistics, the observed linguistic behavior of adults and children is interpreted as requiring two distinct modules, one supporting the ability to produce the past tense of regular English verbs (e.g., *worked*) and the second one underlying the ability to produce the past tense of irregular verbs (e.g., *brought*) (Pinker & Prince, 1988; see Figure 1). This purely theoretical notion of a module is explicitly defended and precisely defined in Fodor's book *The Modularity of mind* (Fodor, 1983), one of the foundational books of computational cognitive science.

The same is true for evolutionary psychology which, as we have said, has a cognitivist orientation. Evolutionary psychology's conception of the mind as a "Swiss knife", that is, as a collection of specialized and genetically inherited adaptive modules, is based on a notion of module according to which modules are theoretical entities whose existence is suggested by the observed human behavior.

Neuroscientists also have a modular conception of the brain. For example, Mountcastle (cited in Restak, 1995, p. 34) maintains that "the large areas of the brain are themselves composed of replicated local neural circuits, modules, which vary in cell number, intrinsic connections, and processing mode from one brain area to another but are basically similar within any area." However, the neuroscientists' conception of the brain is based on empirical observations of the anatomy and physiology of the brain rather than on theory (see Figure 2). The brain obviously is divided up into a variety of 'modules' such as distinct cortical areas, different subcortical structures, interconnected sub-systems such as the retina-geniculate-visual cortex for vision or the basal ganglia-frontal cortex subsystem for attention. This rich modularity of the brain, both structural (anatomical and cytoarchitectonic) and functional (physiological), is evidenced by direct (instrumental) observation, by data on localization of lesions in various behavioral/mental pathologies and on neuropsychological dissociations, and more recently and increasingly, by neuroimaging data.

One can look for correspondences between the two types of modules, the theoretical modules of computational cognitive science and the observed 'modules' of the brain. This is what cognitive neuropsychologists are supposed to do. They interpret the behavioral deficits of patients using the "box-and arrows" theoretical models of cognitive psychology (see Figure 1) – where boxes are modules and arrows indicate the relationship between modules – and then they try to match this

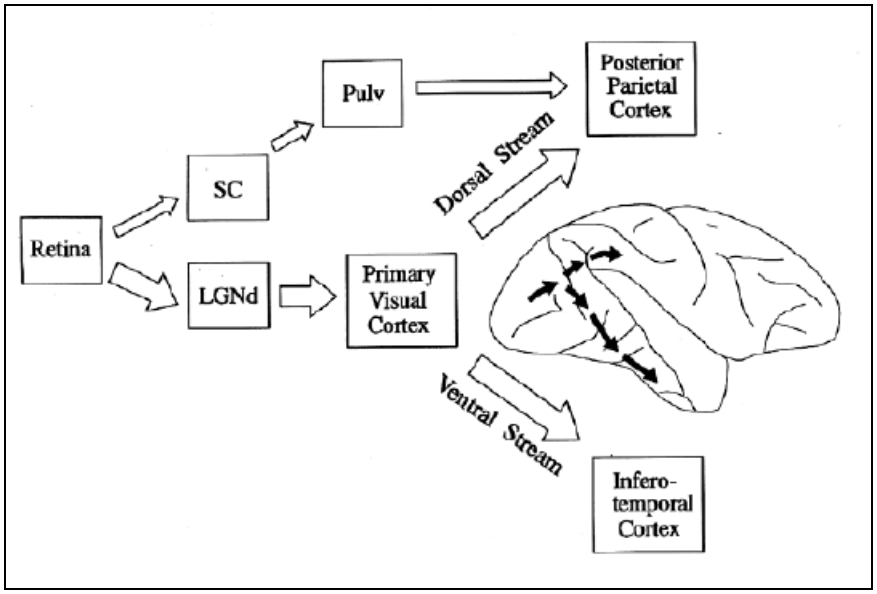

Figure 2. The major routes of visual input into the dorsal and ventral streams. The diagram of the macaque brain shows the approximate routes of the cortico-cortical projections from the primary visual cortex to the posterior parietal and the inferotemporal cortex, respectively (Milner & Goodale, 1998). According to Ungerleider and Mishkin (1982), the ventral stream plays a critical role in the identification and recognition of objects (i.e., the 'what' task), while the dorsal stream mediates the localization of those same objects (i.e., the 'where' task).

modular analysis with observations and measurements on localization of lesions and other physical data on patients' brain. However, one cannot assume that the modular theoretical models of computational cognitive science necessarily correspond to the observed modular structure and functioning of the brain. Cognitive modules may not match the physical (neural) structural or functional 'modules' of the brain, and the brain can be organized into distinct 'modules' which do not translate into distinct components of the theoretical models in terms of which psychologists and cognitive scientists interpret and explain behavioral data.

This is particularly important to keep in mind when one turns to an alternative type of theoretical models which can be used to interpret and explain behavioral data and cognitive capacities, i.e.,

neural networks. Neural networks are theoretical models which, unlike the theoretical models of cognitivist psychology and computational cognitive science, are directly inspired by the physical structure and way of functioning of the brain. Hence, neural networks are at the same time models of the brain and models of the mind. The neural networks used in most simulations so far have been nonmodular. They are a homogeneous network of units with minimal structure constituted by an input module (i.e., set of units), an output module, and (almost always) a single internal module in between. However, this should be a considered as a limitation of current neural network models, not as an intrinsic property of these models. If neural networks claim to be inspired by the structure and functioning of brain, they must be modular since the brain is modular. Notice, however, that the modules of neural networks will be more similar to the modules of the brain than to the theoretical "boxes" of the "boxes-and-arrows" models of computational cognitive science. A module in a modular neural network is a (simulated) physical module, not a postulated theoretical construct. A neural module can be a sub-set of network units with more internal connections linking the units of the module among themselves than external connections linking the units of the module with units outside the module. Or, more functionally, a neural network module can be an observed correlated activity of a sub-set of the network's units, even without 'anatomical' isolation of that sub-set of units. If the modular structure of a neural network is hardwired by the researcher, the researcher should be inspired by the actual modular structure of the brain rather than by theoretical considerations based on cognitive models. If, more in the spirit of neural cognitive science, the network architecture is not hardwired by the researcher but is a result of evolution, development, and/or learning, the researcher should be interested in ascertaining if the emerging modular structure matches the actual modularity of the brain.

As we have said, connectionist research tends to be considered as anti-modularist, in contrast to the strongly modular cognitive models. This is factually correct because most neural network architectures actually used in connectionist simulations are nonmodular and because connectionism

tends to underscore the role of general learning mechanisms rather than that of genetically inherited specific modules in shaping the behavior of organisms. However, as we have also said, neural network research need not be anti-modularist and need not downplay the role of genetically inherited information. The real contrast between neural network models and cognitive models does not concern modularity in itself but rather the nature of modules and the question of what theoretical models are appropriate to explain behavior and cognition.

Consider the cognitivist hypothesis that English speakers produce the past tense of verbs using two distinct modules, one for regular verbs and the other for irregular verbs (Pinker, 1999; see Figure 1). There appears to be some empirical evidence that these two modules might reside in physically separate parts of the brain. Patients with lesions in the anterior portion of the brain tend to fail to produce regular past tense forms while their ability to produce irregular past tense forms appears to be preserved. In contrast, patients with lesions in the posterior portion of the brain tend to show the opposite pattern. They find it difficult to produce irregular past tenses whereas they are able to produce regular ones. This may indicate that two distinct neural modules actually underlie past tense production. This is completely acceptable for a connectionist (at least for the variety of connectionism represented by the authors of this paper), who will try to simulate the behavior of producing the past tense of verbs using a modular network with two distinct modules, one for regular verbs and another for irregular verbs. (These two modules could be either structural or functional, in the sense defined above.)

What distinguishes the cognitive and the neural approach to the treatment of past tense is the nature of the modules. Cognitivists claim that the regular past tense module is a rule-based module. When producing the past of the verb *to work*, the brain is applying the rule: “Add the suffix *-ed* to the verb root”. In contrast, the irregular past tense module is an association-based module containing a finite list of verb roots each associated with its irregular past tense form. The brain just consults this list of

associations, finds the appropriate verb root (for example, *bring*), and produces the corresponding past tense form (*brought*).

This theoretical interpretation of past tense behavior is rejected by a connectionist simply because his or her theoretical tools, i.e., neural network models, do not allow for this interpretation. Neural network models are inspired by the brain, and brains are physical systems made up of physical entities and processes in which all that can ever happen is the production of physico-chemical effects by physico-chemical causes. Hence, in principle a neural network cannot appeal to a rule as an explanation of any type of behavior and cognitive ability. A connectionist can accept that separate and distinct portions of the brain, and of the neural network that simulates the brain, may be responsible for the production of regular past tense forms and of irregular past tense forms. However, both neural modules cannot but function in the same basic way: units are activated by excitations and inhibitions arriving from other connected units. This does not rule out the possibility that one can discover differences in the organization and functioning of the two different neural modules for regular and irregular English verbs and of course this requires an explanation of why the brain has found it useful to have two separate modules for controlling verb past tense behavior instead than only one. This poses the question of the origin of modules to which we turn in the next Section.

3 Evolutionary connectionist simulations: an evolutionary and developmental approach to the study of neural modularity

In this Section we describe two evolutionary connectionist simulations in which modular network architectures evolve spontaneously in populations of biologically reproducing neural networks. The two simulations address only some of the many different problems and phenomena that may arise as a result of the complex interactions between the adaptive process at the population level

(evolution) and the adaptive process at the individual level (learning) and that may be studied using evolutionary connectionist simulations. In the first simulation a modular architecture emerges as part of a process of development taking place during the life of the individual which is shaped by evolution but does not takes experiential and environmental factors during development into consideration. Furthermore, the connections weights for this network architecture are also genetically inherited. In the second simulation evolution actually interacts with learning because the network architecture evolves and is genetically inherited while the connection weights for this architecture are learned during life. (For other simulations on the evolution of modular network architectures, cf. Murre, 1992.)

3.1 Evolution and maturation of modules

Cecconi & Parisi (1993) have described some simulations of organisms which live in an environment containing food and water and which to survive have to ingest food when they are hungry and water when they are thirsty. The behavior of these organisms is influenced not only by the external environment (the current location of food and water elements) but also by the motivational state of the organism (hunger or thirst) which is currently driving its behavior. The body is hungry until a given number of food elements have been eaten and then it becomes thirsty and, similarly, thirst becomes hunger after a given number of water elements have been drank. At any given time the motivational state of the organism is encoded in a special set of "motivational" units representing an internal input (coming from inside the body) which, together with the external input encoding sensory information about the location of food and water, sends activation to the network's hidden units and therefore determines the network's output. The network's output encodes the displacements of the organism in the environment to reach food or water.

In Cecconi and Parisi's simulations the network architecture is fixed, hardwired by the researcher, and nonmodular. By using a genetic algorithm for evolving the connection weights, the authors demonstrate that the organisms evolve the appropriate weights for the connections linking the motivational units to the hidden units in such a way that the current motivational state appropriately controls the organisms' behavior. When the organisms are hungry they look for food and ignore water. When they are thirsty they look for water and ignore water.

But what happens if, instead of hardwiring it, we try to evolve the architecture by means of a genetic algorithm? Is the evolved architecture modular or nonmodular?

To answer this question Cangelosi *et al.* (1994) added a model of neural development to the simulation of Cecconi and Parisi (1993). In the new model the network architecture, instead of being hardwired by the researcher, is the eventual result of a process of cell division and migration and of axonal growth and branching which takes place during the life of the individual organism. Unlike most simulations using genetic algorithms to evolve the architecture of neural networks (Yao, 1999), in Cangelosi *et al.*'s model the genotype does not directly encode the connectivity pattern of the network. What is specified in the genotype is the initial spatial location (in bidimensional space) of a set of simulated neurons (network units), the rules that control the migration of each neuron within the bidimensional space, and the growth parameters of each neuron's axon after the neuron has reached its final location. When a new individual is born a process of neural development takes place in the individual. First, each of the individual's neurons is placed in the bidimensional space of the nervous system according to the x and y coordinates specified in the genotype for that neuron. Second, each neuron displaces itself in neural space according to other genetically specified information until it reaches its final location. Third, after reaching its final location the neuron grows its axon according to growth instructions (orientation and length of axonal branches) also specified in the genotype. When the axonal branch of a neuron

reaches another neuron a connection between the two neurons is established and the connection is given a connection weight which is also specified in the genotype.

A genetic algorithm controls the evolution of the population of organisms. Starting from an initial population with randomly generated genotypes, the best individuals, i.e., those that are best able to eat when hungry and drink when thirsty, are selected for reproduction and the offspring's genotypes are slightly modified by some random genetic mutations. The result is that after a certain number of generations the organisms are able to reach for food when they are hungry while ignoring water and to reach for water when they are thirsty while ignoring food.

Notice that in the genotype neurons are not specified as being input neurons, output neurons, or hidden neurons. The total bidimensional space of the brain is divided up into three areas, a lower area that will contain input units (both external sensory units and internal motivational units), a intermediate area that will contain hidden units, and a higher area that will contain motor output units. If during development a neuron ends up in one of these three areas it takes the function (input, hidden, or output) specified by the area. Furthermore, if a neuron ends up in the input area it can either be a sensory neuron encoding environmental information on location of food and water or a motivational neuron encoding internal (bodily) information on whether the organism needs food (is hungry) or water (is thirsty). Individual organisms can be born with a variety of defective neural networks (no input units for food or water or for hunger/thirst, no motor output units, no appropriate connectivity pattern) but these individuals do not have offspring and their defective genotypes are eliminated from the population's genetic pool.

What network architectures emerge evolutionarily? Are they modular?

Evolved network architectures contain two distinct neural pathways or modules: one for food and the other for water. When the motivational state is "hunger" only some of the hidden units have activation states that vary with variations in input information about food location, while variations in input information about water location do not affect these hidden units (Cf. Figure 3, left). This is the food module. Conversely, when the motivational state is "thirst", water input information controls the activation level of the remaining hidden units which are insensitive to sensory information about food. This is the water module. All successful architectures contain motivational units that send their connections to both the food module and the water module and, on the basis of their activation (hunger or thirst), alternatively give control of the organism's behavior to either food or water.

This shows that, unlike the network architecture hardwired by Cecconi and Parisi (1993) which was nonmodular, if we allow evolution to select the best adapted network architectures, the evolved architectures are modular. The neural network prefers to elaborate information about food and information about water in dedicated sub-networks that we can call modules.

However, as real brain modules as contrasted with cognitive "boxes-and-arrows" models and even hardwired modular architectures, evolved neural modules are not completely isolated or insulated modules. In the evolved architectures of Cangelosi *et al.* the water pathway includes some units which are specialized for processing information about water but also some units which are also used to process information about food. In other words, while information about water is blocked by the network's connection weights when the organism is hungry and it is trying to approach food, information about food has some role even when the organism is thirsty and it is trying to approach

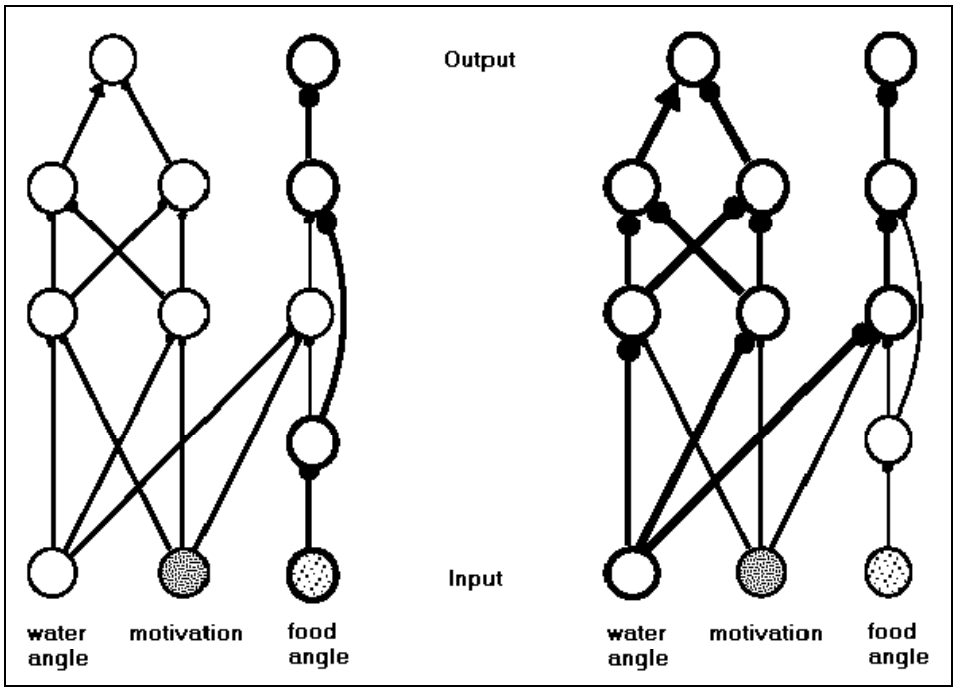

Figure 3. Food pathway and water pathway are shown in bold on the left and right side, respectively (Cangelosi *et al.*, 1994).

water (Cf. Figure 3, right). Interestingly, the asymmetry between the two neural pathways or modules appears to be related to the history of the evolution of the two abilities of finding food and finding water since the ability to find food begins to emerge evolutionarily in this population earlier than the ability to find water.

The fact that the water module includes some units that are also part of the food module, together with the historically contingent fact that the food module emerges earlier than the water module, demonstrate the role of historical contingency in evolved systems. Since for chance reasons the water pathway emerges evolutionarily after the food pathway (i.e., some generations later), the evolutionary process cannot but take what has already evolved into consideration. As a consequence, some of the hidden units dedicated to processing food-related sensory information will end up among the hidden units dedicated to processing water-related information. The lesson

that can be derived from this result is that it can in some cases be erroneous to explain the morphological or functional characteristics of organisms in exclusively adaptivistic terms (as evolutionary psychologists tend to do; cf. Barkow, Cosmides, and Tooby, 1992; Buss, 1999). As suggested by Gould and others (cf. Gould & Lewontin, 1979), evolutionary reality is more complex and some evolved characteristics can be just the by-product of other, directly selected, characteristics or be the result of chance. Artificial Life simulations can help us demonstrate these different mechanisms and processes that result in the evolutionary emergence of organismic characteristics.

The results obtained with these very simple simulations demonstrate how evolving the network architectures, instead of hardwiring (i.e., postulating) them, might have important consequences for the study of neural modularity in organisms that must accomplish different tasks to survive (finding food and finding water).

In Cangelosi *et al.*'s simulation both the network architecture and the network's connection weights are genetically inherited and they evolve at the population level. The particular experience of the individual in its environment has no role in determining the individual's phenotype. It is true that the individual develops, in that the adult neural network is the result of a succession of developmental stages (the displacements of the network's units in bidimensional space and the process of axonal growth), but those changes should be called maturation rather than development since the environment and the individual's experience has no role in determining them. Rather, it is evolution that selects, at the population level, the most appropriate maturational sequence.

In the next section we describe another simulation in which evolution and individual learning during life both contribute in shaping the individual's phenotype. More specifically, evolution

creates modular architectures as the most appropriate ones for the particular task the individual faces during life and learning identifies the connection weights for these architectures.

3.2 Evolution and learning in the emergence of modular architectures

In 1982 Ungerleider and Mishkin (1982) proposed the existence in primates of two visual cortical pathways, the occipito-temporal ventral pathway and the occipito-parietal dorsal pathway, which were respectively involved in the recognition of the identity ("What") and location ("Where") of objects (see Figure 2). (More recently, what was interpreted as the representation of the location of an object has been reinterpreted as representing what the organism has to do with respect to the object (“How”). Cf. Milner & Goodale, 1995.)

This work has been very influential both in neuroscience and cognitive science and in 1989 Rueckl, Cave and Kosslyn (1989) used a neural network model for exploring the computational properties of this "two-systems" design. In their model, neural networks with different fixed architectures were trained in the What and Where task by using the back-propagation procedure (Rumelhart and McClelland, 1986) and their performances were compared. The results of the simulations show that modular architectures perform better than nonmodular ones and they construct a better internal representation of the task.

One way of explaining the better results obtained with the What and Where task with modular networks than with nonmodular ones is to point out that in nonmodular architectures one and the same connection weight may be involved in two or more tasks. But in these circumstances one task may require that the connection weight’s value be increased whereas the other task may require that it be decreased (see Figure 4, left). This conflict may affect the neural network’s performance by giving rise to a sort of neural interference. On the contrary, in modular architectures modules are

sets of "proprietary" connections that are only used to accomplish a single task and therefore the problem of neural interference does not arise (see Figure 4, right). Rueckl *et al.* (1989) hypothesize that this might be one of the reasons for the evolutionary emergence of the two distinct neural pathways in real organisms.

To test this hypothesis Di Ferdinando, Calabretta, and Parisi (2001) repeated the experiment of Rueckl *et al.* (1989) by allowing the evolution of the network architecture. In Rueckl *et al.*'s simulations the network architectures are hardwired by the researcher and the authors are able to find the best possible architecture (which is a modular architecture with more hidden units assigned to the more difficult What task and fewer hidden units assigned to the easier Where task) by trying many different hardwired architectures and testing them. Jacobs and Jordan (1992) have used a developmental model in which the network architecture emerges as a result of a process of development in the individual. The individual network starts as a set of units each placed in a particular location of a bidimensional physical space and then pairs of units may establish connections based on a principle of "short connections" according to which two units are more likely to become connected the more close they are in space. This is an interesting proposal based on a principle that favors short connections which is likely to play a role in neural development. However, the resulting network architecture is not really self-organizing because it is the researcher who decide the location of units in physical space and therefore in a sense hardwire the network architecture. (In the simulations described in Section 3.1 there is also development of the connectivity pattern as in Jacobs and Jordan's simulations but both the location of the network's units in space and the rules controlling the growth of connections are genetically inherited and they are the result of a self-organizing evolutionary process.) In the simulations described in this Section, although there is no development, the network architectures are the spontaneous outcome of a process of evolution which is independent from the researcher.

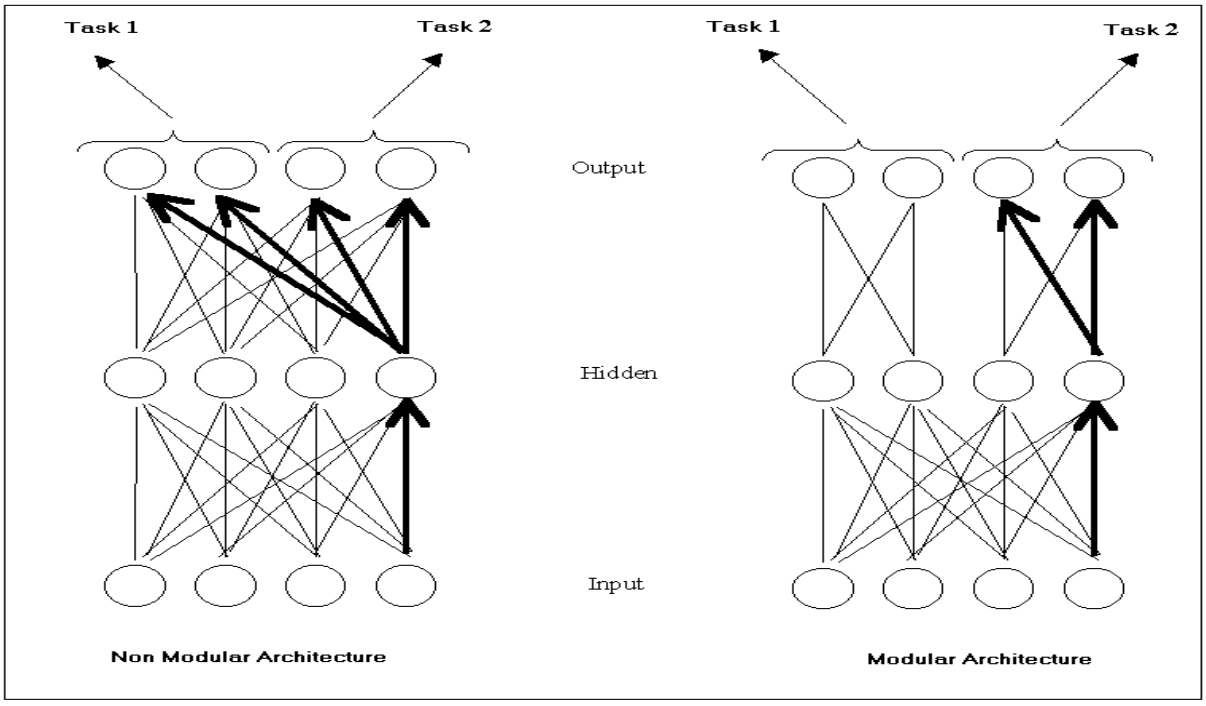

Figure 4. Neural interference in neural networks that have to learn to perform two tasks. Consider in the non-modular architecture (left) the highlighted connection that connects one of the input units with one of the hidden units. Since the hidden unit sends its connections (highlighted) to both the output units involved in task 1 and those involved in task 2, a modification of the connection's weight value would affect both tasks. In this kind of architecture neural interference may arise because optimization of one task may require that this connection's weight value be increased whereas optimization of the other task may require that it be decreased. In the modular architecture (right) the problem of neural interference does not arise because the connection goes to a hidden unit which sends its connections (highlighted) only to the output units involved in task 2 and therefore its value can be changed to satisfy the requirements of task 2 only.

In a first set of simulations Di Ferdinando *et al.* (2001) used a genetic algorithm for evolving both the architecture and the connection weights of the neural networks. The results showed that the genetic algorithm was unable to evolve both the architecture and the weights. Furthermore, the network architecture that tended to evolve was different from the best architecture of Rueckl *et al.* in that it assigned more resources (hidden units) to the easier Where task than to the mode difficult What task. In other words, the evolutionary algorithm was not able to allocate the appropriate resources to the two tasks.

The failure of the genetic algorithm to find the best architecture for the What and Where task and therefore to reach appropriate levels of performance when both the network architecture and the connection weight are genetically inherited appears to be due not only to the fact that a mutation affecting the architecture can suddenly make a set of weights evolved for the preceding architecture inappropriate for the new architecture but also to a phenomenon analogous to genetic linkage. In simulations in which the architecture is fixed and is the best modular architecture (more units allocated to the What task compared to the Where task), the genetic algorithm appears to be unable to evolve the appropriate connection weights because a favorable mutation falling on the weights of one module can be accompanied by an unfavorable mutation in the weights of the other module. This interference at the genetic level appears to be unexpected according to models of population genetics (Wagner, personal communication).

Further analyses of the simulation results reveal other interesting phenomena that are due to the co-evolution of architecture and weights, for example freezing of the architecture at low mutation rates and oscillation of the evolved architecture at high rates.

The best results - i.e., the appropriate modular architecture and high levels of performance - are obtained in simulations in which evolution cooperates with learning. More specifically, the best solution, as suggested by Elman *et al.* (1996), is to have evolution take care of the architecture and learning of the connection weights. With this solution evolution is free to zero in on the best network architectures without fear that inherited weights that were appropriate for previous architectures may turn out to be inappropriate for mutated architectures (genetic linkage) and learning during life is free to find out the best connection weights for each inherited architecture. These simulation results clearly show that evolution and learning are not dichotomous as empiricists and nativists sometimes seem to be believe but that their cooperation is necessary if organisms must be able to acquire complex capacities.

As a final observation we note that, as in the simulations described in the previous Section, the evolved neural modules are not completely isolated and the modular architecture is not as clean as a "boxes-and-arrows" model. While most connections are proprietary of the two modules, the What module and the Where module, there are some connections that are shared by the two modules.

6 Conclusions

In this chapter we have described a new approach to studying brain/mind modularity which takes into consideration the phylogenetic history of an organism's brain modules. This approach, Artificial Life, allows us to simulate in the same model an organism at the genetic, neural, and behavioral level and may helps us in revealing how modules at one level may be related to modules at another level.

Brain/mind modularity is a contentious issue in current cognitive science. Cognitivists tend to conceive the mind as a set of distinct specialized modules and they believe that this rich modularity is basically innate, with evolutionary psychologists even thinking that each module is adaptive in that it has been biologically selected as a result of specific evolutionary pressures. (But other cognitivists, such as Chomsky and Fodor, believe that modules are innate but not necessarily adaptive (Fodor, 2000).) On the other hand, connectionists tend to think that the mind is a more homogeneous system that basically genetically inherits only a general capacity to learn from experience and that if there are modules they are the result of development and learning rather than being innate.

We have maintained that connectionism is not necessarily anti-modularist and anti-innatist. On the contrary, since neural network models are said to be inspired by the brain they cannot but be modular (if even most network architectures used in connectionist simulations are nonmodular) because the brain is a rich structure of specialized modules. Viewing neural networks in the perspective of Artificial Life allows us to develop an appropriately modular and innatist connectionism, Evolutionary Connectionism. Artificial Life simulations simulate evolving populations of organisms that inherit a genotype from their parents which together with experience and learning determines the individual phenotype. The way is open then for simulations that explore whether modular or nonmodular network architectures emerge for particular tasks and how evolution and learning can cooperate to shape the individual phenotype.

In any case, even if connectionism can be modularistic, this does not imply that when connectionists talk about modules they mean the same thing as cognitivists. Cognitive modules are theoretical entities which are postulated in "boxes-and-arrows" models used to explain behavioral data. Connectionist modules are anatomically separated and/or functionally specialized parts of the brain. There may be only partial co-extensiveness between the two types of modules and in any case research on neural modules is very differently orientated than research on cognitive modules and it considers different types of empirical evidence.

Evolutionary Connectionism shares the main goal of Evolutionary Psychology, that is, to develop a "psychology informed by the fact that the inherited architecture of human mind is the product of the evolutionary process" (Barkow *et al.*, 1992), but it differs from Evolutionary Psychology for three main reasons: (1) it uses neural networks rather than cognitive models for interpreting human behavior; (2) it adopts computer simulations for testing evolutionary scenarios; (3) it has a less pan-adaptivistic view of evolution and it is more interested in the rich interplay between genetically inherited and experiential information. The simulation of evolutionary scenarios allows us to take

chance and other nonadaptive evolutionary factors into consideration and therefore prevents us from explaining all the morphological or functional characteristics of organisms in exclusively adaptivistic terms.

We have presented two Artificial Life simulations in which the genetic algorithm actually selects for modular architectures for neural networks. In one simulation both the network architecture and the network weights are genetically inherited and they evolve but evolution selects for appropriate maturational sequences and in the other simulation evolution and learning cooperate in that evolution selects for the network architecture and learning finds the weights appropriate for the inherited architecture. These simulations weaken Marcus' criticism when he says that " none of [… connectionist] models learn to divide themselves into new modules" (Marcus, 1998, p. 163).

The first of the two simulations described in this chapter also shows that modules can be inherited (innate) but their exact structure is not necessarily the result of specific evolutionary pressures and adaptive but it can be the result of other evolutionary forces such as chance and pre-adaptation.

More Artificial Life simulations are of course needed to explore how modular architectures evolve or develop during life and how selective pressures at the population level or experience during life may shape the existing modules. But the Artificial Life perspective allows us to explore other research directions that involve the interactions of modules at the genetic, neural, and behavioral level (Calabretta *et al.*, 1998). For example, using neural networks in an Artificial Life perspective one can explore if genetic duplication would facilitate the evolution of specialized modules. This would represent an important confirmation of the general hypothesis that gene duplication facilitates the evolution of functional specialization which was originally proposed by Ohno (1970) and modified by Hughes (1994). We have already shown with Artificial Life simulations that this might actually be the case (Calabretta *et al.*, 2000; Wagner *et al.*, this volume). Another research

direction would be testing whether sexual reproduction might decrease the kind of genetic interference (linkage) postulated by Di Ferdinando *et al.* (2001) and this would strengthen one of the several hypotheses formulated about the role of sexual reproduction in evolution (Michod & Levin, 1988).

References

Barkow, J., Cosmides, L., & Toby, J. (1992). *The Adapted Mind: Evolutionary psychology and the generation of culture.* Oxford University Press, New York, NY.

Belew, R. K. McInerney, J. & Schraudolph, N. N. (1992). Evolving networks: using the genetic algorithm with connectionist learning. In Langton, C. G., Taylor, C., Farmer, J. D. & Rasmussen, S. (Eds.), *Artificial Life II*, pp. 511-547. Addison-Wesley, New York.

Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind*. Allyn & Bacon, Boston, MA.

Calabretta, R., Galbiati, R., Nolfi, S. & Parisi, D. (1996). Two is better than one: a diploid genotype for neural networks. *Neural Processing Letters* 4, 149-155.

Calabretta, R., Nolfi, S., Parisi, D. & Wagner, G. P. (1998). A case study of the evolution of modularity: towards a bridge between evolutionary biology, artificial life, neuro- and cognitive science. In Adami, C. Belew, R. Kitano, H. & Taylor, C. (Eds.), *Proceedings of the Sixth International Conference on Artificial Life*, pp. 275-284. The MIT Press, Cambridge, MA.

Calabretta, R., Nolfi, S., Parisi, D. & Wagner, G. P. (2000). Duplication of modules facilitates the evolution of functional specialization. *Artificial Life* 6, 69-84.

Cangelosi A., Parisi D., & Nolfi S. (1994). Cell division and migration in a 'genotype' for neural networks. *Network: Computation in Neural Systems* 5, 497-515.

Cecconi, F. & Parisi, D. (1993). Neural networks with motivational units. In Meyer, J.-A., Roitblat, H. L., and Wilson, S. W. (Eds.), *From animals to animats 2. Proceedings of the 2nd International Conference on Simulation of Adaptive Behavior*, pp. 346-355. The MIT Press, Cambridge, MA.

Chomsky, N. (1980). *Rules and representations*. Columbia University Press, New York.

Cosmides, L., Tooby, J. & Barkow, J. (1992). Introduction: evolutionary psychology and conceptual integration. In Barkow, J., Cosmides, L., & Tooby, J. (1992), pp. 3-18.

Cosmides, L. & Tooby, J. (1994). The evolution of domain specificity: the evolution of functional organization. In Hirschfeld, L. A., and Gelman, S. A. (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*. The MIT Press, Cambridge, MA.

Di Ferdinando, A., Calabretta, R. & Parisi, D. (2001). Evolving modular architectures for neural networks. To appear in French, R. & Sougné, J. (Eds.), *Proceedings of the Sixth Neural Computation and Psychology Workshop Evolution, Learning, and Development*. Springer Verlag, London.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking innateness. A connectionist perspective on development*. The MIT Press, Cambridge, MA.

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

Fodor, J. (1998). The trouble with psychological darwinism. *London Review of Books* Vol. 20(2) (cover date 15 January 1998).

Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. The MIT Press, Cambridge, MA

Gould, S. J. (1997). Evolution: the pleasures of pluralism. New York Review of Books, June, 26, 1997.

Gould, S. J. & Lewontin, R. (1979). The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London, Series B* 205, 581-598.

Gould, S. J. & Vrba, E. S. (1982). Exaptation - a missing term in the science of form. *Palebiology* 8(1), 4-15.

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. The MIT Press, Cambridge, MA.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication.

Proceedings of Royal Society. Series B 256, 119-124.

Jacobs, R. A. & Jordan, M. I. (1992). Computational consequences of a bias toward short connections. *Journal of Cognitive Neuroscience* 4, 323-335.

Karmiloff-Smith, A. (2000). Why babies' brains are not Swiss army knives. In H. Rose and S. Rose (Eds.), *Alas, poor Darwin*, pp. 144-156, Jonathan Cape, London.

Marcus, G. F. (1998). Can connectionism save constructivism? *Cognition* 66, 153-182.

Michod, R. E. & Levin, B. R. (1988). *Evolution of Sex: An Examination of Current Ideas*. Sinauer Associates.

Milner, A. D. & Goodale, M. A. (1998). The visual brain in action. *PSYCHE*, 4(12) (http://psyche.cs.monash.edu.au/v4/psyche-4-12-milner.html).

Murre, J. M. J. (1992). *Learning and categorization in modular neural networks*. Harvester, New York.

Newell, A. & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery* 19, 113-126.

Ohno, S. (1970). *Evolution by Gene Duplication*. Springer Verlag, New York.

Parisi, D., Cecconi, F. & Nolfi, S. (1990). Econets: Neural networks that learn in an environment. *Network* 1, 149-168.

Pinker, S. (1999). *Words and Rules. The Ingredients of Language*. Weidenfeld and Nicolson, New York.

Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28, 73-193.

Plunkett, K. (1996). Development in a connectionist framework: rethinking the nature-nurture debate. *CRL Newsletter*, University of California, San Diego. http://crl.ucsd.edu/newsletter/10-4/

Restak, R. M. (1995). *The modular brain*. Simon & Schuster, New York, NY.

Rueckl, J. G., Cave, K. R. & Kosslyn, S. M. (1989). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience* 1, 171-186.

Rumelhart, D. & McClelland, J. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. The MIT Press, Cambridge, MA.

Spelke, E. S., Breinlinger, K., Macombe, J. & Jacobson, K. (1992). Origins of knowledge. *Psychological Review* 99, 605-632.

Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W. (Eds.), *The Analysis of Visual Behavior*. The MIT Press, Cambridge, MA.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, August 27, 749-750.

Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE* 87, 1423-1447.