Cognition without Neurones: Adaptation, Learning and Memory in the Immune System

JOHN STEWART
Institut Pasteur

Abstract

This paper proposes a definition of cognition as a system capable of both action and perception, in which the coupling of action and perception is such that the emergent behaviour of the system in its environment satisfies a meaningful constraint. The immune system of vertebrate animals reacts to nonself antigens by producing a specific immune response which triggers destruction of the antigen; it reacts to self antigens by incorporating them into the regulatory dynamics of a self-organizing network. It is argued that this behaviour meets the requirements of the definition, and hence that the immune system is a cognitive entity capable of adaptation, learning and memory. Consequences of this perspective are that the molecules and cells of a multicellular organism are not, as such, "cognitive"; that the full articulation of neurophysiology into cognitive science requires a study of emergent behaviour at the level of meaningful interactions between a system and its environment; and that the objects of cognition, with or without neurones, are brought into existence by the coupled perceptions and actions of the cognitive system itself.

Key Words: *Adaptation, Cognition, Learning, Memory; Computer Simulations; Network; Immune System*

Cognición sin neuronas: adaptación, aprendizaje y memoria en el sistema inmunitario

Resumen

Este artículo propone una definición de cognición como un sistema capaz de acción y percepción, de modo que esta confluencia acción-percepción determina que el comportamiento emergente del sistema en su medio satisfaga una constricción significativa. El sistema inmunitario de los animales vertebrados reacciona ante los antígenos ajenos produciendo una respuesta inmune específica que desencadena la destrucción del antígeno; reacciona ante los propios antígenos incorporándolos en el mecanismo regulador de una red auto-organizada. Se argumenta que esta conducta se ajusta a los requisitos de la definición y, por tanto, que el sistema inmunitario es una entidad cognitiva capaz de adaptación, aprendizaje y memoria. Las consecuencias de esta perspectiva son: las moléculas y células de un organismo multicelular no son, por sí mismas, cognitivas; una plena articulación de la neurofisiología como ciencia c ognitiva requiere un estudio de la conducta emergente en el nivel de las interacciones significativas entre un sistema y su medio; y los objetos de la cognición, ya sea con o sin neuronas, se materializan al confluir las percepciones y acciones del propio sistema cognitivo.

Palabras clave: *Adaptación, c ognición, aprendizaje, memoria, simulación en ordenador, red, sistema inmunitario.*

Acknowledgements: The work on modelling the immune system described in this paper was carried out in collaboration with Francisco Varela and Antonio Coutinho, and would not have been possible without their intellectual stimulation, friendship and support.

Author's Address: Unité d'Immunobiologie, CNRS URA 359, Institut Pasteur, 25 rue du Dr Roux, F-75724 Paris, France.

I. INTRODUCTION

It is widely accepted as self-evident that neurones (or neuromimetic automata in the case of computer simulations) are a necessary and sufficient basis for cognitive phenomena, and hence that neurophysiology is both automatically a constitutive element of the cognitive sciences, and virtually the only biological discipline attaining to that status. In this paper, I shall question these assumptions by providing a possible counter-example. The science of immunology is characterized by the fact that immunologists spontaneously attribute "cognitive" capacities to the immune system: recognition, learning, memory and self/non-self discrimination (Klein, 1982). In order to examine the validity of the suggestion that the immune system may be "cognitive", I shall propose some theoretical criteria for defining what we might mean by "cognition". These criteria can then be used to elaborate certain requirements that the neurosciences should satisfy in order to constitute a fully articulate part of cognitive science.

II. THE IMMUNE SYSTEM

All vertebrate animals possess an immune system composed primarily of a special class of cells, the *lymphocytes,* which circulate freely in the blood and lymph. The interactions of lymphocytes, both amongst themselves and with other cells and molecules, are mediated primarily by a special class of protein molecules, the *immunoglobulins,* which exist in two forms, either bound in the cell-membrane of the lymphocytes, or as free molecules secreted into the body fluids. The membrane-bound immunoglobulins function as receptors: when they are "occupied" by other molecules, the lymphocyte is activated, either to *secretion* of large numbers of immunoglobulin molecules, or to *proliferation*, forming a clone of lymphocytes with a common ancestor. It is possible to distinguish two different regions in each immunoglobulin molecule: the "framework" region which is virtually the same in all immunoglobulins, and which ensures that they all have the same overall shape and structure; and the "variable" region, which is different from clone to clone[1] and which forms the "combining site" responsible for the specific interactions of the immunoglobulin with other molecules. The diversity of the variable regions is generated by a remarkable process of somatic DNA rearrangement which, as far as is known, occurs only in lymphocytes. This diversity is so great that no two clones have identical combining sites; at the same time, in a population of 10^8 lymphocytes[2], there exists at least one clone which will interact specifically, via its immunoglobulin receptors, with any molecular shape whatsoever (of the appropriate size). Immunologists describe this phenomenon by saying that the "repertoire" of the immune system is "complete" (Coutinho et al., 1984).

III. FUNCTIONS OF THE IMMUNE SYSTEM

The immune system possesses at least two distinct modes of operation, each of which gives rise to a characteristic phenomenological domain.

III.1. Classical immunology: the immune response

Classical immunology, ever since its inception at the beginning of the century (Stewart & Coutinho, 1991), has focussed primarily on the activation of the immune system by external "antigens", typically bacteria and viruses. When an antigen is introduced into the system, the specific lymphocyte clones that interact[3] with it will be stimulated, both to proliferate[4], and to secrete large quantities of the corresponding immunoglobulins. Thus, in the absence of "network" interactions between the various lymphocyte clones themselves (see below), the introduction of an antigen leads to a strong "immune response" which triggers the destruction of the antigen. This mode of operation of the immune system is known as the "clonal selection theory" (Burnet, 1959).

III.2. A network model of the immune system

In recent years, immunological theory has been enriched by a new approach to understanding the organization of the system. The initial impetus was given by Jerne (1974), who pointed out that if the immune repertoire is indeed complete, then one would necessarily expect that the immunoglobulin receptors of one clone interact with the immunoglobulins produced by certain other clones. This has led to a concept of the immune system as an autonomous, self-activating network. In order to gain some idea as to how such a network might work, we have built a mathematical model of the immune system (Varela et al., 1988; Stewart & Varela, 1989). Even in its simplest form, to be described below, this model is a highly nonlinear dynamic system. Since exact analytical solutions are inaccessible, we have performed computer simulations to study its behaviour.

The basic constituents of this model are a set of lymphocyte clones, numbered i = 1...n. Throughout the life of an individual animal, new clones are continuously generated (in higher vertebrates, in the bone marrow) and presented as candidates for recruitment into the population of activated lymphocytes. In order to represent the interactions between clones, we have used the "shape-space" concept (Stewart & Varela, 1991) according to which the "shape" of the combining site of each immunoglobulin can be defined by p stereo-chemical parameters. The values of these parameters can be taken as co-ordinates in a p-dimensional "space", so that each molecular shape corresponds to a point in "shape-space". This representation has the virtue that molecular shapes which are similar, and hence have similar profiles of interaction with other shapes, will correspond to points which are close in the shape-space. However, it is also important to represent a relationship of complementarity between any two molecular shapes, since this gives rise to high-affinity interactions between the shapes in question. We therefore postulate that each point in shape-space corresponds not to a single shape, but to a pair of complementary shapes (distinguished as "black" and "white") having high-affinity interactions between them. This makes it possible to calculate the affinity m_{ij} between any two shapes i and j as:

$$m_{ij} = \exp(-d_{ij}^2) \qquad \text{Equation (1)}$$

(where d_{ij} is the distance in shape-space between a black/white point i and a white/black point j).

In the simple form of the model presented here, there is no proliferation; clones are either present at unit concentration, or absent. The net effect of all the other clones in the network on any given clone i is then defined by the "field" h_i, where

$$h_i = \Sigma m_{ij} \qquad \text{Equation (2)}$$

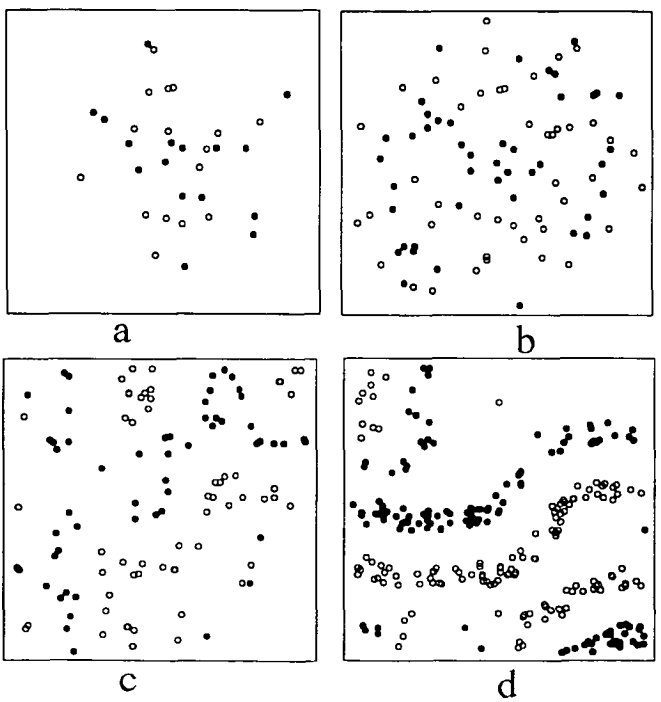
(summation over j=1. .n)

On this basis, the model postulates that the criteria for the recruitment af a new clone and for the maintenance of a pre-existing clone are the same: if the field experienced by a clone is within a "window" between a lower threshold and an upper threshold, the clone is recruited/maintained; outside the window, the clone is not recruited/eliminated from the population.

The simulations to be shown here employ a 2-dimensional shape-space, not because this is a particularly realistic assumption (to the extent that the shape-space concept is valid at all, the dimensionality is certainly higher), but because this lends itself to a graphic visualisation of the results which contributes greatly to their interpretation. Each simulation commences with a single "black" clone at the centre of the shape-space. "Black" and "white" clones are generated at random points in the shape-space and proposed a candidates for recruitment into the network. After each recruitment, the entire population is re-evaluated and clones with a field outside the "window" are eliminated. The result of this process is illustrated in Figure 1; schematically, it comprises two phases.

FIGURE 1

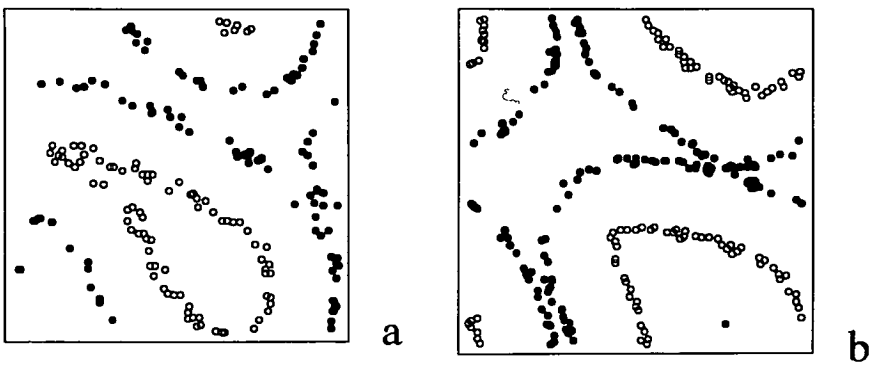
Four successive stages in the self-organizing constitution of an immune network, as modelled by computer simulation in a 2-dimensional shape-space. After an initial chaotic phase (la and lb), chains of parallel black and white clones appear (1c) and stabilize (1d). From Stewart & Varela (1991).



Initially, the population grows until the whole of the available shape-space is occupied by blak and white clones, but in a rather disorganized manner (Figures 1a and 1b). During this phase, a phenomenon of "local collapse" is frequently observed, due to the following mechanism: (i) the presence of a "black" (or white) clone creates a field permitting the recruitment of a neighbouring "white" (or black) clone; (ii) this recruitment often has the effect of increasing the field for the first clone above the upper threshold, so that it is eliminated; and (iii) this elimination abolishes the field for the second clone, so that it too is eliminated. In other words, the process displays a form of "self-organized criticality" characterized by numerous local "avalanches". The initial phase comes to an end when this perpetual self-reorganization leads to the emergence of certain configurations in shape-space which appear to be quasistable (Figures 1c and 1d). The general nature of these configurations is even clearer if the "edge effects" are eliminated by joining opposite edges of the 2-dimensional square, so that the shape-space has the topology of a torus (Figure 2).

FIGURE 2

Two further examples of the emergent configurations in a simulated immune network, when the shapce-space has the topology of a torus. The regularity of the patterns is particularly striking. Prom Stewart & Varela (1991).



The quasi-stable configurations in question are based on the formation of linear "chains" of black and white clones. Each black chain runs parallel to a facing white chain (and vice versa). The region "behind" a black (or white) chain is bounded by a second black chain which runs parallel to its own facing white chain, and so on. The quasi-stability of this configuration can be explained as follows. In the region between a pair of facing black and white chains, the field is above the upper threshold, too high for clones of either colour to be recruited. In the region "behind" a chain, on the other hand, the field is too low for clones of the same colour to be recruited. In other words, the configuration of a pair of parallel facing black and white chains is stable because each chain creates a field (at a distance corresponding to the interval between them) within the window for recruitment/maintenance of the other. Finally, there is one remaining form of recruitment which could potentially disrupt this overall configuration: that of white clones in the back-to-back region bounded by black chains (or vice versa). Sporadically, this can and does occur (cf the isolated black clone in Figure 2b). However, the isolated clone immediately creates a field leading to the recruitment of neighbouring clones of the other colour; because of the asymmetry due to the surrounding chains, the field for the isolated clone will systematically exceed the

upper threshold, leading to its elimination by a process basically similar to that involved in the phenomenon of "local collapse" discussed above [5].

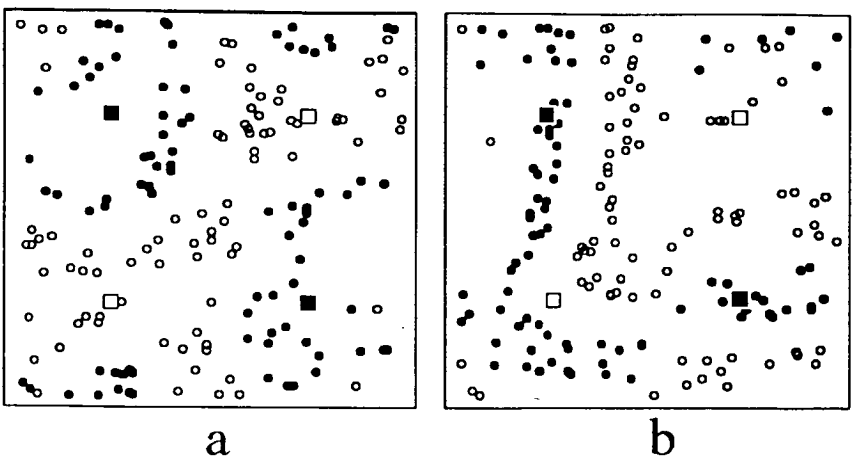
In conclusion, we can say that certain characteristic configurations emerge precisely because of their relative stability. This process is somewhat akin to natural selection, but with two notable differences compared to the "clonal selection" of the classial theory. Firstly, the entities selectecl are not individual clones, but entities at a higher level of organization, i.e. global configurations of organized sets of clones. Secondly, the operative agent of selection is not a fixed, pre-given, external entity; rather, the constraints are endogenously generated by the process itself.

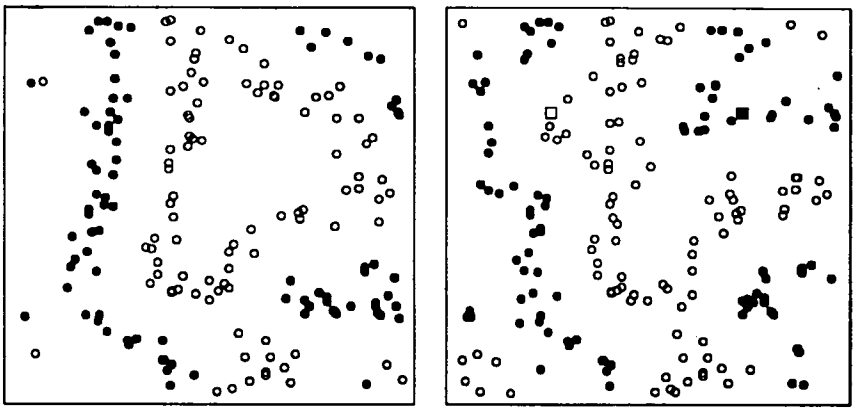
Having thus established some of the major self-organizing characteristics of the network, the next question concerns the interaction of the network with antigens. We have modelled antigens as molecular shapes whose presence is independent of the field they experience. Since antigens can, in principle, have any shape whatsoever, there will be both "black" and "white" antigens. The results of simulations in which the ontogenesis of the network occurs in the presence of such antigens are shown in Figure 3; the antigens are represented by square symbols in order to distinguish them from the circular symbols representing lymphocyte clones.

What we see is that the network retains its self-organizing capacity, giving rise to the same sort of emergent configurations as previously. The novel feature is that the black and white antigens appear to be systematically incorporated into chains of the same colour (Figures 3a and 3b). In other words, the previously unconstrained disposition of the chains is now constrained by the coupling of the network with the antigens. Figure 3c shows that this constrained disposition is maintained even if the constraining antigens are removed[6]. That this is a dynamic feature, rather than the result of static inertia, is shown in Figure 3d: the introduction of discordant antigens (black antigen in a white chain, white in a black) results in a realignment of the chains such that they incorporate antigens of like colour.

FIGURE 3

The effect of coupling an emergent immune network, as in Figures 1 and 2, with fixed antigens (black and white squares). 3a and 3b: two examples showing the incorporation of the antigens into the black and white "chains". 3c: the configurations remain quasi-stable even when the antigens are removed, showing "memory". 3d: the rcadjustmcnt of the configurations when discordant antigens are introduced. From Stewart & Varela (1991).





c d

IV. DISCUSSION

IV. 1. A definition of "cognition"

We are now in a position to turn to the central question of this paper: is there any sense in which it might be meaningful to say that the immune system is "cognitive"? In order to discuss this question with any clarity, we require some sort of definition of the term "cognitive". In this paper, I propose to examine the fruitfulness of the following (definition: an entity is "cognitive" if it is capable of both perception and action; moreover, its actions should be guided by its perceptions in such a way that a meaningful constraint is satisfied. This is very close to the concept of "adaptive systems" proposed by Meyer and Guillot (1990); it is of course a minimal requirement, a sort of "degree zero" of cognition. It is not, however, trivial; to clarify this point, I shall elaborate on what I mean by a "meaningful" constraint.

The sort of constraint I have in mind is what may be termed a "proscriptive" constraint rather than a prescriptive one. A prescriptive constraint specifies precisely, in every detail, exactly what the system should do in every possible situation. In "Good Old-fashioned Artificial Intelligence" (Boden, 1987), systems are typically subjet to prescriptive constraints. In order to cover all possible situations, the specification of such constraints tends to be long, complicated and tedious. Moreover, since a classical system of this sort functions by simply internalizing these external constraints, it is "heteronomous" in the sense of being strictly determined by the outside. In a way, such a system is actually rather stupid, since it does just what it is told to do, neither more nor less.

In recent years, mainstream cognitive science has ben considerably enriched by the advent of neo-connexionism. In particular, the technique of adjusting connexion weights by back-propagation in a network with hidden layers overcomes the awkward bottleneck by making it possible to associate sensory inputs with symbolic outputs on the basis of a set of examples rather than by complete prescriptive specification. However, to the extent that an appropriate set of "training examples" must be furnished from outside the system itself, such neo-connexionist networks remain essentially heteronomous.

A proscriptive constraint, on the other hand, is often conveniently expressed in

negative terms. Thus, we might require of a robot on a table that it should not stay still, that it should not bump into obstacles, and that it should not fall off the edge of the table. However, although this may sound somewhat negative, the capacity to satisfy a proscriptive constraint is actually a positive achievement. Since the strategy to be employed is not dictated from the outside, a proscriptive constraint not only allows but generally requires a degree of autonomy on the part of the system. This creative aspect is particularly clear in what we may take as the prototype of a proscriptive constraint: the capacity of living organisms to remain within their domain of viability (Aubin, 1991). The "negative" constraint (not departing from viability) is, of course, a remarkable, positive achievement: staying alive, or, in the terms of Maturana and Varela (1980) maintaining a state of autopoïesis. At the same time, it is clearly not a prescriptive constraint, with a single optimal solution; on the contrary, an astounding plurality of qualitatively different strategies are possible, as witnessed by the millions of species and life-forms on the planet Earth.

The definition of cognition that I propose here is, as I have said, an elementary, minimal one; it is already satisfied by primitive living organisms (but not by viruses in isolation). Even bacteria, for example, are capable of guiding their actions (their cilia either flow in coherent waves, producing locomotion in a straight line, or wave in chaotic fashion, poducing a "tumbling" motion and randon reorientation) as a function of their perceptions (their sensors discriminate between a situation in which the environmental sugar concentration is increasing and one in which it is decreasing) in such a way that they tend to move towards a nutritious source of sugar. This performance is minimally "cognitive" in the sense of the proposed definition.In order to avoid misunderstanding, I should state explicitly that I do not seek for a moment to suggest that this is *all* cognition ever is; in the subsequent course of evolution, a whole series of new phenomenological domains emerge successively. To quote just a few: the appearance of multicellular organisms having an ontogeny; semiotic communication in social animals; the appearance of language during hominisation; the invention of writing and the entry into human history; reflexive self-consciousness and the invention of philosophy; modernity and the birth of science. Any one of these could plausibly be taken as defining "cognition". However, I maintain that all subsequent forms of cognition are rooted in this most elementary form. Moreover, the radical idea that "Life=Cognition" (Stewart, 1992, 1993a) has some strong and counter-intuitive consequences. In particular, as we shall see, it means that brains - and computers - are not, in themselves, "cognitive".

In what follows, I shall therefore accept this minimal definition of cognition as constrained action-perception, and examine its fruitfulness by applying it to the immune system. I should emphasize that this definition is not simply tailor-made to suit my present purposes in this article. It is consonant with the general approach of the vigorous new field of research known as "Artificial Life" (Langton, 1989, ECAL, 91; ECAL, 93). Moreover, this approach has definite positive implications for the study of higher-level cognition, notably the phenomena of communication, representation and language. I shall return to this question in my conclusions.

IV.2. Is the immune system "cognitive"?

Defining cognition as constrained action-perception, the first question is this: does the immune system "perceive" its environment? By "perceive", I mean that the interactions between the system and its environment reliably lead to characteristic

modifications in the internal state of the system. The perspective here is not one of transferring "information" from the environment to the system; rather, the interaction triggers a switch from one internal state to another, but what these states are is determined by the self-organizing properties of the system itself.

To answer the question properly, therefore, we need to take into account the fact that the immune system has two distinct modes of operation: the "classical" mode of clonal selection, and the "network" mode modelled in the simulations described above. In the first mode, the system perceives an antigen by mounting an immune response, i.e. by producing high concentrations of immunoglobulins with strong affinity for the antigen. In the second mode, coupling with an antigen leads to realignment of the chains of "black" and "white" clones such that they incorporate the antigen in question. In both cases, we can conclude that the immune system does indeed "perceive" its environment in the sense defined here.

Secondly, is the immune system capable of "actions" guided by its perceptions? Again, the answer is yes, on condition that we accept to describe its operation in terms that are inhabitual in classical immunology. By mechanisms that it is not necessary to describe in detail here [7], an immune response triggers the destruction of the antigen that provoked it. In the case of the network mode, a positive answer depends on a speculative hypothesis: it has been suggested that the incorporation of antigens into immunoglobulin "chains" may create functional connections between the set of antigens thus incorporated (Varela et al., 1992). Briefly, the idea is that if two antigens, at different points in shapespace, are linked by such chains, then a variation in concentration of one antigen will cause variation in the field experienced by the other antigen. If the chains of lymphocyte clones are already in place - an ontogenetic process occurring on a time-scale of weeks to months - then this form of "action at a distance in shape-space" can occur in seconds or minutes by a molecular process of association/dissociation of immunoglobulin-immunoglobulin complexes. When the antigens in question are molecules serving as chemical signals - hormones, neurotransmitters, their receptors etc. - such connections could have an integrative physiological function, accounting perhaps for psychosomatic phenomena.

Thirdly, does the way in which the actions are guided by the perceptions conform to some non-trivial, meaningful constraint? With this question, we are at the heart of the matter. The classical mode of operation is appropriate in the case of external antigens such as bacteria and viruses: their destruction, triggered by an immune response, defends the organism against infectious disease. The network mode of operation is appropriate in the case of "self-antigens", i.e. molecules which are a part of the organism itself. Conversely, a network response to bacteria and viruses would fail to protect the organism from infection; an immune response to self-antigens would cause auto-immune disease. Thus, the immune system may be regarded as "cognitive" precisely to the extent that it systematically enters into the appropriate mode of operation. As a matter of empirical fact, this does seem to be the case; but immunology will only fully qualify as a "cognitive science" if the mechanisms by which this comes about can be elucidated. This is the issue conventionally known as "self/non-self discrimination" (Bernard et al., 1990); in view of its central importance, it is worth examining the matter more closely.

To begin with, it is to be noted that the two modes of operation are mutually exclusive: if the system responds in the network mode to a given antigen, that effectively precludes an immune response to the antigen in question. This can be clearly seen by considering the fate of a lynphocyte freshly emerged from the bone-marrow. If the lymphocyte is recruited into the network, it will be activated and become part

of what has been called the "Central Immune System" or CIS (Varela & Coutinho, 1991). If it is not recruited, it will remain as a small resting cell for two or three days before dying. The rate of bone-marrow production is extremely high (it is sufficient to reconstitute the entire immune system in less than a week), so that at any single point in time these inactivated, moribund cells makc up fully 80% of the total lymphocyte population; they constitute what has been called the "Peripheral Immune System" or PIS. It is to be noted that although normally moribund, the cells in the PIS will be activated if they encounter an antigen having high affinity with their immunoglobulin receptors; in this case, since they do not possess a network organization, they will give rise to a classical immune response.

It is important to appreciate the relationship between the CIS and the PIS. It is the CIS which is positively defined: it is composed of cells integrated into a self-activating network. The PIS is negatively defined: it is simply composed of the residue of cells which are not recruited into the network. Since the "repertoire" of lymphocytes produced by the bone-marrow is "complete", it follows that the repertoire of the PIS is "complete minus the repertoire of CIS".

We are now in a position to specify more precisely the conditions under which the distribution of roles between the network mode of operation (instanciated by the CIS) and the classical mode of operation (instanciated by the PIS) will be "appropriate" in the sense defined above. The essential requirements are that the repertoire of the CIS should reliably include all the self-antigens of the body; but that it should not extend too far beyond this, otherwise the repertoire of the PIS will be correspondingly reduced and may not cover the totality of pathogenic micro-organisms whose function it is to recognize and destroy. The first requirement is probably met by the timing of developmental events in ontogeny: the CIS is the first to develop, in the pre-natal embryo, at a time when only self-antigens are present. The second requirement is more problematic, and is in fact a major focus of our current research. The problem is that in the very simple model described above, the "network" expands to occupy the whole of the available shape-space; this would reduce the repertoire of the PIS to nothing. This problem is, however, clearly posed, and it should be possible to solve it by an appropiate extension of the model[8].

IV.3. Self-identity and learning

Summing up, then, it seems that it is both plausible an fruitful to envisage the hypothesis that the immune system is indeed "cognitive" in the sense defined here. This signifies that the employment of cognitive metaphors is not just loose heuristic talk, but may correspond lo a scientific reality. As we have seen, in immunology a central issue is the positive constitution of a self-identity; the capacity for self/non-self discrimination is a secondary consequence. It is also worth noting that because (in an outbreeding population) each individual is genetically unique, being in particular different from both its parents, such "self/non-self" discrimination cannot be encoded in the germ-line; it is the result of a process of epigenetic "learning". Turning the cognitive metaphor around, our analysis of the immune system suggests that learning is an epigenetic phenomenon, fundamentally inseparable from the ontogenetic organization of the system as a whole. This perspective is in harmony with the suggestion that "memory" may basically be nothing other than a reflection of the fact that the current state of a cognitive system arises from the total history of its ontogenetic constitution (Rosenfield, 1988).

V. CONCLUSIONS

V. 1. Generalizing from the immune system

What are the potentially generalizable lessons to be gleaned from this investigation of whether it is scientifically reasonable to attribute cognitive faculties to the immune system? One point that emerges very clearly is that the question cannot be answered by analysis of the individual components of the system. Molecules and differentiated cells (such as neurones and lymphocytes) may or may not be parts of a cognitive system, but in and of themselves there is nothing cognitive about them. Nor is it sufficient to analyse the low-level interactions between these components. Thus, the stereo-chemically specific interactions of immunoglobulins, both with each other and with other molecules, are basic to the dynamic functioning of the system; but again, as such, they are not cognitive. It is important to emphasize that *at this level of organization* there is absolutely nothing which distinguishes a "self-antigen" from an "external antigen": the physico-chemical forces involved are of exactly the same nature.

If molecular interactions, as such, are not "cognitive", neither is the activation of a cell by the interaction between its membranebound receptor molecules and particular ligands. It only becomes legitimate to speak of such interactions as instances of "recognition" if they are considered in the context of a global system which is indeed cognitive. Even the entire immune system is not, in itself, cognitive; the decisive questions turn on the quality of the interactions between the system and its environment. Thus, the minimal level of hierarchical organization at which it is possible to speak of cognition involves perceptually-guided actions in an ecological niche; the key question is whether the coupling between action and perception is such that the emergent behaviour of the system satisfies some meaningful constraint.

If it is possible to give an affirmative answer to this last question –and I have suggested that this may be the case for the immune system– then (on the definition used here) it is legitimate to speak of cognition. Now, however, a fascinating question arises: what is the "object" of this cognition? In the case of the immune system, we have seen that the answer is: the incorporation of selfantigens into the dynamic configurations of lymphocyte clones which are constitutive of the immunological self-identity; and discrimination between these and external non-self antigens which evoke destructive immune responses. What is remarkable is that this "object" can only be specified by reference to categories constituted by the functioning of the immune system itself. Thus, the immune system is capable of perception and learning; it is undoubtedly adaptive, in the sense that it makes a major contribution to the survival of the organism of which it is part; but it does not perceive, learn or adapt to anything pre-existing or definable in the realm of purely objective reality. It perceives, learns, and adapts (under strong constraints) to aspects of reality that have only been brought into existence through the participation of its own cognitive activity. To sum up in a phrase: the object and subject of immune cognition are inseparable.

V.2. The case of neuronal cognition

Are these considerations applicable to neurophysiology? Before deciding, let us see what would be the implications.

To begin with, this perspective would suggest that neurotransmitters and voltage-dependent ionic channels in neuronal membranes are not, in themselves, cognitive. Nor are neurones and synapses; nor even is the brain, considered as such in isolation. In order to be cognitive, the nervous system must be considered in the functional context of an animal accomplishing perceptually-guided actions in an environment. In this respect, it is surely significant that neurones first appear in evolution in the role of linking sensory inputs to locomotory actions. An interesting feature here is that such actions will, via the animal's movement in its environment, give rise to correlatable variations in its sensory perceptions. In a classic study, Held and Hein (1958) showed that kittens are only able to learn to see objects if their perceptions are related to their own actions. The significance of a central nervous system with a connexionist architecture may be that it makes it possible for the animal to identify such action-perception correlations, and thus to develop conceptual categories based on its active involvement in the world. This would indicate that to interpret the function of the nervous system in terms of "feature detectors" or "grandmother neurones" is to go badly astray.

This brief evocation may be sufficient to show that the perspective put forward here, although at first sight rather negative concerning much current research in neurophysiology, can actually serve to integrate the neurosciences into the mainstream of cognitive science by focussing on what is cognitively significant in the underlying molecular and cellular mechanisms.

V.3. Implications for higher-level cognition

As already indicated, the approach proposed in this paper has definite implications for the study of higher-level cognition. In the usual computational theories of mind, symbolic representations are defined as theoretical primitives, with a syntactical structure analogous to that of formalist mathematics and logic (Lakoff, 1987). By contrast, the elementary definition of cognition suggested here invokes neither symbols nor representations. This opens up the possibility of studying the emergence of higher-level cognition as natural events occurring in the course of biological evolution and human history. Three such emergent phenomena are of particular importance: semiotic communication, representation, and language. Finally, I will comment on some of the perspectives for technological applications opened up by this approach.

a) *Communication.* Given a population of interacting cognitive agents, the basic schema of constrained action-perception can be extended so that the repertoire of actions includes the emission of signals, and the repertoire of perceptions includes the interpretation of these signals. For a tightly restricted set of conditions on the emission and interpretation of the signals, the result will be a co-ordination of the actions of the agents contributing decisively to their individual and collective viability. A good example is the "swarm intelligence" of social insects (Deneubourg et al 1991). This approach to the understanding of communication as an emergent phenomenon is also amenable to computer simulation (Booth & Stewart, 1993, Victorri & Cazoulat, 1993). It is worth noting that the signals involved (for example, ant pheromones) are arbitrary (Saussure, 1985); their semiotic significance is not due to any "information" encoded in their structure, but to their rôle in mediating a co-ordination of actions. It is also to be noted that, in contrast with computational theories of mind, social organisation is fundamental as both the condition and the consequence of semiotic communication.

b) *Representation.* In computational theories of mind, symbolic representations acquire their semantic content by virtue of correspondance relations with aspects of a metaphysically postulated objective reality. By contrast, in the approach presented here, the "objects" of representations are not separable from the cognitive subject. Given an elementary cognitive agent successfully engaged in constrained action-perceptions, the agent's actions will have definite consequences for its own subsequent perceptions; on this view, "representations" are representations of such consequences, and not of external reality "in itself". An agent endowed with this sort of representational capacity will be able to anticipate, and actually to elaborate an action sequence intended to achieve a particular perceptual state specified in advance as a goal. Thus, "representations" are a sort of internal "doubling-up" of elementary cognition; in practice, this probably requires a brain in the case of living organisms or, in the case of robots, a connexionist network autonomously generating its own "training set" of anticipated and actualised perceptual states (Mel, 1990). It follows that the emergence of representations, and their corollary of intentional action, must have been a quite definite and identifiable event, probably situated in vertebrate evolution somewhere between lampreys (the most primitive known vertebrates, with practically no head) and the higher apes who are clearly capable of harbouring intentions (Lestel, 1993). Dennett (1987) has proposed the virtual elimination of intentionality as a scientific concept, reducing it to an epistemic projection: to say that an entity is "intentional" would mean nothing more than that we can spontaneously describe its behaviour in terms of intentions. On Dennett's own admission, this would imply that thermostats, for example, are "intentional". On the view that I am proposing, this is unnecessarily pessimistic; neither thermostats nor bacteria have "representations" (in the precise sense defined here), and there are thus sound scientific grounds for saying that their actions are not intentional. Conversely, in cases where it can be shown that representations are effectively used to elaborate action-sequences designed to achieve a pre-specified goal, intentionality can be reinstated as a meaningful concept.

c) *Language* . What is the difference between the semiotic communication of social animals, as discussed in (a) above, and fully-fledged human language ? I suggest that a crucial difference lies in the much greater variability in the meaning of the signals used in linguistic communication. In animal communication, both the conditions of emission and the interpretation of the signals are highly stereotyped. Thus, for example, vervet monkeys have three distinct types of alarm calls: one for eagles, one for snakes, and one for big cats. The actions that are triggered in response are likewise stereotyped: respectively squatting on the ground under cover, clambering into trees while looking at the ground, and hiding while peering anxiously into the distance. The actions thus coordinated clearly contribute to the animals' viability by avoiding each of the three specific predators, and this therefore counts as communication on the definition that I have proposed. The stereotyped nature of this communication is probably neither innate nor acquired, but results from the particular organisation of the ontogenetic process which ensures regularity of outcome for all normal members of the species for anatomical and behavioural characters alike (Oyama, 1985; Stewart, 1993b). In human linguistic communication, on the other hand, the meaning of words is so context-dependent that, in the limit, we hardly use the same word twice with exactly the same meaning. This is related to the basically metaphorical nature of language (Lakoff, 1987). Given this great variability, mutual comprehension is probably only possible because, in normal human conversation, we are constantly communicating about our communication (Reddy,

1979) - "Do you mean that....?", "No...", "Yes, I see", etc., to say nothing of non-verbal communication - raised eyebrows, frowns, nods, etc. Thus, linguistic communication is distinct from animal communication in that it normally includes meta-communication.

These characteristics of human language, which are integrated quite naturally in the emergent approach outlined here,. serve to substantiate a point forcibly made by Raccah (1993), to wit that natural languages (English, French, German, Spanish etc.) are entities quite different in kind from the formal languages used in mathematical logic and computer programming. This point is important because computational theories of mind assume that natural languages can be treated as though they were formal languages (Montague, 1970), with the corollary that syntax takes pride of place in the study of language (Chomsky, 1957; for an opposing view, see Lakoff 1987). Indeed, these theories further assume that all cognition takes place in a "language of thought" (Fodor, 1975) amenable to symbolic computation. It is salutory to recall that in fact formal languages are an extremely recent human invention; they achieved their canonical form only in the first part of the 20th century, in the very special context of formalist mathematics and logic (Lakoff, 1987). Their redeployment as a basis for cognitive science was principally due to Turing and the notion of "universal computation". It should be clear that, from the point of view presented in this paper, taking a computational theory of mind as the basis for the whole of cognitive science amounts to putting the cart before the horse with a vengeance.

d) *Technological perspectives.* Computational theories of mind have immediate and well-known technological applications, such as expert systems and Artifical Intelligence in general. The alternative approach suggested here has its own characteristic applications, notably in the field of autonomous mobile robots which are less well-known and may be worth mentioning. A noted pioneer in this field, Brooks (1987), has stated quite explicitly that in designing such robots it is actually advantageous to completely discard conventional symbolic representations of the external world. By superimposing layers of perception-action couplings, Brooks has been able to produce legged robots that learn to walk in unknown, irregular terrain. The perception-action coupling in Brooks' robots has to be "wired-in" by human design; recent work by Cliff et al. (1992) suggests that it may be possible to overcome this bottleneck by the evolution of a carefully simulated population of robots using genetic algorithm techniques. Another recent development is experimentation with populations of mobile robots which "communicate" according to the definition given in section IV.3a above (Deneubourg, personal communication). Finally, a highly speculative long-term goal may be worth mentioning because of its theoretical implications. To the extent that autonomous robots of the sort mentioned here "interpret" signals in terms that are directly meaningful for their own "cognition", it is not inconceivable that they may one day "understand" language in a manner qualitatively different from current AI machines which no more "understand" language than the rote operator in Searle's famous "Chinese room" thought-experiment.

V.4. General conclusions.

This overview of some of the implications of the elementary definition of cognition proposed here for the study of higher-level cognition is partial and schematic in

the extreme. It may nevertheless suffice to show that my aim in proposing this definition is in no way to reduce all cognitive phenomena to the level of behaviour exhibited by bacteria. In fact, my aim is far more ambitious; it is to renew the whole of cognitive science by proposing a thorough-going alternative to current computational theories. This being so, I would like to conclude by reiterating certain implications suggested by the detailed analysis of the immune system which forms the central theme of this paper. The following general conclusions stem directly from the elementary definition of cognition as looped perception-action subject to a proscriptive constraint. Cognitive perception, learning and adaptation are anything but arbitrary: they are strongly constrained by a "reality principle" which, in the case of living oranisms, relates to their viability (Aubin, 1991) and their capacity to maintain their autopoietic organization (Varela, 1989). It is a general condition for the actual realization of cognition that, in each case, it instanciates a particular, specific strategy for meeting a proscriptive constraint. Within the framework provided by any one such strategy, the collstraint takes on the form of severe limitations on the permissible range of all the parameters which affect the emergence of the required global behaviour of the system; in fact, the parameters will often appear to be subject to optimization around a single ideal solution. However, such optimization is always relative to the particular strategy adopted; and as a matter of general principle, there is always a plurality of potential strategies. It follows that the objects of cognition) with or without neurones) are neither pre-existent, nor definable in terms of a totally independent and purely objective reality; they are brought into existence by the coupled perceptions and actions of the cognitive system itself.

References

AUBIN, J.P. (1991). Viability Theory. Birkhauser.

BERNARD, J., BESSIS, M. & DEBRU, C. (Eds) (1990). *Soi et non-soi*. Editions du Seuil, Paris.

BODEN, M.A. (1987). *Artificial Intelligence and Batural Man*. 2nd Edition, MIT Press, London.

BOOTH . & STEWART, J. (1993). Un modèle de l'émergence de la communication. Actes des Premières Journées Francophones "Intelligence Artificielle Distribuée et Systèmes Multi-Agents", Toulouse 7-8 avril 1993, 9-18.

BROOKS, R.A. (1987). Intelligence without representation. MIT Artificial Intelligence Report.

BURNET, F.M. (1959). *The Clonal Selection Theory of Acquired Immunity*. Cambridge University Press, Cambridge.

CLIFF, D., HARVEY, I. & HUSBANDS, P. (1992). Incremental evolution of neural network architectures for adaptive behaviour. University of Sussex School of Cognitive and Computing Sciences Technical Report CSRP256.

CHOMSKY, N. (1957). *Syntactic Structures*. Mouton, The Hague.

COUTINHO, A. FORNI, L., HOLMBERG, D., IVARS, F. & VAZ, N. (1984). From an antigen-centered, clonal perspective of immune responses to an organism-centered, network perspective of autonomous activity in a self-referential immune system. *Immunol. Revs.* 79, 151169.

DENEUBOURG, J.L., THERAULAZ, G. & BECKERS, R. (1992). Swarm-made architectures. In: P. Bourgine & F. Varela (Eds), *Towards a Practice of ,Autonomous Systems*. Proceedings of the First European Conference on Artificial Life. MIT Press/Bradford Books .

DENNETT, D. (1987). The Intentional Stance. MIT Press, Cambridge.

ECAL 91. *Towards a Practice of Autonomous Systems*. P. Bourgine & F. Varela (Eds), MIT Press/Bradford Books.

ECAL 93. *Self organisation and life: from simple rules to global complexity*. Mmay 24-2, 1993, Brussels, Belgium.

FODOR, J. (1975). *The Language of Thought*. Crowell, New York.

HELD, R. & HEIN, A. (1958). Adaptation of disarranged hand-eye coordination ontingent upon re-afferent stimulation. *Perceptual-Motor Skills* 8, 87-90.

JERNE, N.K. (1974). Towards a network theory of the immune system. *Ann. Immunol. (Paris)* 125C, 373-389.

KLEIN, J. (1982). *Immunology: the science of self-nonself discrimination*. Wiley, New York.

LAKOFF, G. (1987). *Women, Fire and Dangerous Things*. Chicago University Press, Chicago.

LANGTON, C,. Ed. (189). *Artificil Life.* Addison Wesley.
LESTEL D. (1993). Une multimodalité problématique: communications symboliques des primates non humains, tromperie tactique et socialité postulée. 4e Ecole d'Eté de l'ARC, "Communication et multimodalité dans les systèmes naturels et artificiels", Bonas, 4-17 July 1993.
MATURANA, H.R. & VARELA, F.J. (1980). Autopoiesis and Cognition: the realization of the living. Reidel, Dordrecht.
MEL, B. (1990). Connexionist robot motion planning: a neurally-inspired approach to visually-guided reachng. Academic Press, New York.
MEYER, J.A. & GUILLOT A. (1990). Simulation of adaptive behavior in animats: review and propects. In Meyer & Wilson (Eds), *From animals to animats.* Proceedings of the First International Conference on the Simulation of Adaptive Behavior. MIT Press/Bradford Books.
MONTAGUE, R. (1970). Universal Grammar. *Theoria* 36, 373.
OYAMA, S. (1985). *The Ontogeny of Information: Developmental systems and evolution.* Cambridge University Press, Cambridge.
RACCAH, P.-Y. (1993). Significado e Inferencia Argumentativa. In: Bustos, E.,García-Bermejo, J.C., Pérez Sedeno, E., Rivadulla & Zofio, J.L. Eds., *Panorama de lógica, methodología y filosofía de la ciencia.* Siglo XXI, Madrid.
REDDY, M.J. (1979). The conduit metaphor - a case of frame conflict in our language about language. In: Ortony A. Ed., *Metaphor and Thought,* Cambridge University Press, Cambridge,. 284-324.
ROSENFIELD, I. (1988). *The Invention of Memory: a New View of the Brain.* Knopf, New York.
STEWART, J. & VARELA, F.J. (1989). Exploring the meaning of connectivity in the immune network. *Immunol. Revs.* 110, 37-6 1 .
STEWART, J. & COUTINHO, A. (1991), A hundred years of immunology: paradigms, paradoxes and perspectives. In: P.A.,Cazenave & P.,Ialwar Eds., *Immunology: Pasteur's Inheritance.* Wiley, New York.
STEWART, J. & VARELA, F.J. (1991), Morphogenesis in Shape-space. *J. Theoret. Biol.* 153, 477-498.
STEWART, J. (1992). Life=Cognition: the epistemological and ontological significance of Artificial Life. In: P. Bourgine & F. Varela (Eds), *Towards a Practice of Autonomous Systems.* Proceedings of the First European Conference on Artificial Life. MIT Press/Bradford Books.
STEWART, J. Ed. (1993a). *Biologie et Cognition. Intellectica* 16.
STEWART, J. (1993b). Au-dela de l'inné et de l'acquis. *Intellectica* 16, 151-174.
VARELA, F. J. (1989). *Autonomie et Connaissance: Essai sur le Vivant.* Editions du Seuil, Paris.
VARELA, F., Coutinho, A., DUPIRE, B. & VAZ, N.M. (1988). Cognitive networks: immune, neural and otherwise. In: *Theoretical Immunology.* A.S. Perelson Ed., Addison-Wesley, New York, Vol II 359-374.
VARELA, F.J . & COUTINHO, A. (1991). Second generation immune networks. *Immunology Today* 12, 159-166.
VARELA, F., COUTINHO, A.& STEWART, J. (1992). What is the immune network for? In: W. Stein & F. Varela (Eds), *Thinking about biology: an invitation to current theoretial biology.* Addison Wesley (SFI Series on Complexity), New Jersey.
VICTORRI, B. & CAZOULAT, R. (1993). Auto-organisation et émergence des symboles. Actes des Journées "Formation des symboles dans les modèles de la cognition", Grenoble l0-11 juin 1993.

Notes

[1] The variable region is identical in all the cells within a given clone.

[2] This is the number of lymphocites in a mouse.

[3] Inmmunologists speak of "recognition"; I will examine in the discussion whether the use of this cognitive metaphor can be justified.

[4] Hence the term "clonal selection", in analogy with the natural selection of neo-darwinian evolution theory.

[5] Conoisseurs of the Japanese game of Go will recognize striking –and aesthetically pleasing– analogies, both in the actual forms of the emergent configurations and in the dynamics of the strategic reasons underlying them.

[6] I shall examine whether this can be considered as a example of "memory" in the discussion.

[7] Briefly: inflammation, macrophage ingestion and the complement cascade (Klein 1982).

[8] In technical terms, one promising possibility involves including interactions between the immunoglobulin-producing B-lymphocytes (the only type considered here) and the T-lymphocytes which recognize digested protein fragments "presented" on cell surfaces by the molecules of the "Major Histocompatibility Complex".