united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR/1308/8

Third Antonio Borsellino College on Neurophysics:

Evolution of Intelligent Behaviour

(23 April - 4 May 2001)

On the representation of information in life, brains and consciousness

J. Roederer
(University of Alaska)

strada costiera, 11 - 34014 trieste italy - tel. +39 040 2240111 fax +39 040 224163 - sci_info@ictp.trieste.it - www.ictp.trieste.it

ON THE REPRESENTATION OF INFORMATION IN LIFE, BRAINS AND CONSCIOUSNESS [(i)]

Juan G. Roederer
Geophysical Institute
University of Alaska-Fairbanks
JGR@gi.alaska.edu

1. INFORMATION AND LIFE

As we look around us we realize that much of the human environment appears to be made up of discrete clumps of matter with clearly defined, sharp boundaries. When we explore the environment with scientific instruments, we realize that also at the microscopic level the cells of living matter, their organelles, viruses, and the molecules, atoms and elementary particles can all be considered "clumps of matter". In the macroscopic domain, even fluids appear organized in discrete ways: the ocean, the atmosphere and the ionized plasmas that make up stars, stellar winds and magnetospheres all appear arranged into regions separated by thin boundary layers, discontinuities or surface sheets. The late Nobel Laureate Hannes Alfven called it "The Cellular Universe".

Our senses of vision and touch are geared toward the perception of shape, texture and other features of the boundary surfaces of the every-day clumps of matter or *objects*, and our brain is endowed with the computational machinery to identify their position and orientation, appraise their relative motion and, ultimately, determine their identity and purpose. In addition to these "objects in space" there are also "objects in time": the discrete trains of acoustical waves to which our sense of hearing responds. Although we do hear continuous sounds (just as our eyes perceive continuous things like the blue sky), relevant acoustical information is extracted from *temporal changes* in the wave field. Of course, in vision the information is also carried by waves, but the key aspect for perception is their spatial (angular) distribution and related changes.

It is our experience from daily life (and from precise observations in the laboratory) that the presence of one object may alter the state of other objects in some well-defined ways. We call this process an "*interaction*". Quite generally, we may define "science" as the systematic and quantitative study of the interactions between objects or ensembles of objects. Each scientific discipline studies the interactions affecting objects belonging to one common class (e.g., elementary particles, stars, molecules, cells, people, plants, etc.). The interactions observed in the natural environment can be divided into two broad classes: one comprises the *physical interactions* between inanimate objects, the other represents the *biological interactions* in which at least one component is a living organism.

In a physical interaction between two bodies, we observe that the presence of one modifies the properties of the other (its motion, structure) in a definite way that depends on the relative configuration of both components of the system (e.g., their relative positions). Primary physical interactions are always two-way (true *inter*-actions) and strictly reversible, but in the macroscopic domain they usually appear connected in irreversible cause-and-effect chains. *Physics*, in particular, studies interactions between objects in the most general terms, regardless of their chemical composition, shape and color, and their actual meaning to humans. In a physical interaction between two bodies, the concept of *force* is introduced as a measurable agent responsible for the change that occurs during the interaction. For the fundamental interactions such as gravity and electromagnetism, the concept of *force field* is introduced as a property of space surrounding the interacting bodies. The energy necessary for change in a physical interaction is drawn from the interaction mechanism (the force field) itself. At the microscopic, subatomic level, everything can be reduced to four basic interactions between elementary particles [1]. The environment in which humans and all living beings exist, however, pertains to the *macroscopic domain* in which objects consist of the order of 10^{23} or more molecules, and in which all physical quantities of relevance are averages over enormously large ensembles of particles. Still, in that domain, all physical interactions can be reduced to two elementary ones: gravitation and electromagnetism (for instance, friction, elasticity and chemical reactions are all, ultimately, electromagnetic in nature).

Biological interactions are of a fundamentally different character: they are *information-driven*. Consider the two examples sketched in Fig. 1a. On the left we show several satellites orbiting the Earth, on the right we show insects "in orbit" around a light bulb. Both have in common some well-defined paths that are followed at a regular, well-defined pace. The motion of a satellite is governed by the force of gravitational interaction f, which is a function of position and masses—and nothing else. The motion of an insect is governed by the force of propulsion imparted by the wings (which also balances gravity and air friction), and controlled by a sensory system with a complicated mechanism of light detection and pattern recognition—a process that involves *information acquisition and processing*. In other words, we have the chain: light emission → detection → pattern analysis → muscle activation. No such set of algorithms appears in the case depicted on the left side: the concept of information is totally alien to gravitational interactions—and to *all* purely physical interactions between inanimate objects. The latter just "happen": they don't require intermediate operations of information detection and processing. A second pair of examples is shown in Fig. 1b. At left we have a schematic view of the "electric charge separator" in a particle detector, on the right, a "sex separator".

We call the interactions between inanimate objects "*force-field driven interactions*", and those between a living object and other objects (living or not) "*information-driven interactions*". The latter, of course, ultimately are coupled to physical interactions; the key aspect, however, is their *control* by information processing operations.

Let us proceed systematically with our description of information-driven interactions and related definitions, and point out both analogies and differences with the case of physical interactions between inanimate bodies. First of all, we note that information-driven interactions can occur only between two bodies the complexity of which exceeds a certain minimum degree; they cannot occur between elementary particles, classical "mass points" or periodically structured solids such as

[1] In the subatomic domain, interactions between elementary particles are described by the interchange of "field quanta", which themselves are particles (for instance, photons in the case of the electromagnetic interaction between two electrically charged particles). The interaction between the two original particles thus really becomes a *sequence* of at least two more fundamental interactions "particle 1→field quantum→particle 2" and vice versa. When very large distances or very short time intervals are considered, the fact that energy cannot propagate faster than light (Theory of Relativity) also leads to a *chain* of more fundamental interactions (e.g., when we say "Sun-Earth interaction" we really mean the chain Sun→radiation field→Earth).

crystals. We say that a (complex) body *A* is in information-driven interaction with (complex) body *B* if the spatial or temporal structure of body *A* causes specific changes in the structure of body *B* that are invariant with respect to geometric circumstances such as the position of body *A* with respect to body *B*. In other words, in an information-driven interaction a *correspondence* is established between a spatial or temporal pattern in body *A* and specific changes triggered in *B*; this correspondence is independent of any other circumstances. At this stage we are assuming that both *A* and *B* are "natural bodies", i.e., *not deliberately manufactured* by an intelligent being (thus keeping man-made technological systems out of the picture). We further note that primary information-driven interactions are unidirectional (i.e., they are really actions, *not* interactions), going from one of the bodies (the "source") to the other (the "receiver"). Finally, while in physical interactions the energy involved is tapped from the interactive force field, in information-driven interactions the energy is provided by a reservoir *external* to the interaction mechanism.

We now introduce the concept of *information* as a measurable agent responsible for the specific changes in the "receiver" (body *B*): it is what links the particular features of the structure of the "emitting" body *A* with the specific changes caused in the structure of *B*. Thus defined, information is not an "absolute" concept; it relates to a whole process, and we say that "*B* has received and processed information from *A*". While in a certain sense this concept seems equivalent to that of force or force field in a physical interaction ("that which links the changes of one body to the presence of the other"), there are fundamental differences: (i) information is unidirectional (going from *A* to *B*), (ii) it is independent of the particular constellation of the pair *A*, *B* (e.g., their mutual distance), and (iii) there is no energy transfer involved (although locally supplied energy may be necessary to emit, transmit and process the information).

Information can be expressed in many equivalent ways; what must remain invariant is the *code* linked to the particular arrangement or order of relevant features in the emitter, and the particular changes caused in the receiver [2]. It should be clear that information-driven interactions in *natural* systems must *evolve*, they cannot appear "by chance" out of purely physical interactions because to operate, a *common* code must always exist at both ends of the interactive link. We contend that only living systems can entertain information-driven interactions, and now offer the following definition [(ii)]: *a life system is a natural (not-man-made) system exhibiting interactions that are controlled by information processing operations* (the proviso in parentheses is there to emphasize the exclusion of computers and robots!) Information thus becomes the defining concept that separates the living world from the inanimate world. Note that in the two examples of Fig. 1 involving living organisms, the electromagnetic waves (light) themselves do not drive the interactions—it is the *information content* in the patterns of the wave trains, not their energy, that plays the controlling role. Quite generally, an information-driven interaction behaves like a *trigger* of physical/chemical processes, but does not participate in them.

[2] Science in general defines information as "a statement that describes the outcome of expected alternatives". This is the definition of what is usually called "semantic information". The statement itself that represents the information in question may be given in numerical form, digital when the number of alternatives is finite (e.g., the faces of a die) or real numbers when there is an infinite number of alternatives (e.g., the coordinates of a mass point). We note that this definition is only a special case of the more basic concept given above. Communications theory introduces a mathematical expression for the *amount* of information: for n equally probable alternatives, knowledge of the outcome represents $\log_2 n$ "bits" (binary digits) of information (the outcome of a coin toss thus represents 1 bit of information).

The perhaps most important biological information-processing operation is that of *pattern recognition*. This process, in its most general form, represents the generation of identical responses to different complex input configurations which bear a certain common spatial or temporal feature (the pattern that is being recognized), independently of the particular circumstances (place, size, orientation, or order) in which the feature appears among other components of the input. Examples are a particular spatial arrangement of nucleotides in the DNA molecule exerting a pattern-specific effect on its chemical environment; the geometric features of the contour of an object whose optical image is projected on the retina, triggering pattern-specific cognitive operations in the brain; and the set of harmonics of a musical tone that elicits distinct resonance regions on the basilar membrane in the inner ear, giving rise to a single pitch and timbre sensation.

In the abiotic world there is information, but it is not being used. A lunar rock that has been lying on the lunar surface for billions of years only becomes a source of "information" when it is looked at or analyzed by a human being or sniffed by a dog; information played no role in its formation and erosion [3]. The absence of information in a life-less world may come as a surprise to many. After all, both in thermodynamics and quantum mechanics one always talks about "information"! In thermodynamics, for instance, information indeed seems to be playing a controlling role in how a system behaves. A loss of information is always accompanied by a gain in entropy and vice versa, and it takes a minimum of energy to store a given amount of information [4]. Does this mean that our definition of "life" is no longer valid because the role of information has lost its exclusiveness? This is not so, because in our definition of life, information controls the *interactions per se*. In a thermodynamic system (and also in quantum mechanics), "information", whenever it appears, relates to the observer, experimenter or thinker, *not* to the system per se. It does not influence the system, whose molecules do what they are supposed to do according to basic physical laws in which information plays no role. What does influence the behavior of a system is *what we humans do to it*: it is *us humans* who prepare a system, set up its boundaries and initial conditions—even if only in thought experiments. And it is those human-induced actions that condition the response of a physical, inanimate thermodynamic system, whether in reality or only in thought: in Maxwell's Demon paradox, the demon (or its robot surrogate) is placed there *by a human being*, it has no chance of appearing naturally! In thermodynamics, the concept of information appears because we humans imagine or fabricate *unnatural* situations. In life systems, instead, the concept of information is an active participant in their *natural* evolution and behavior.

2. INFORMATION AND BRAIN

The formation of large, complex organic molecules was the primary condition for the emergence of life on Earth; the early ocean-atmosphere system provided the appropriate medium. Chemical reactions are governed by physical interactions in the quantum domain, in which, according to our discussion above, "information as such" should play no role—these interactions just "happen". In the process of molecular synthesis, however, large and complex polymer-like macromolecules

---

[3] This absence of information may be viewed in equivalence to the force field surrounding a *single* particle (for instance, the electric field around an electron): to reveal its existence, a *second* electric charge must be placed in the field—with key differences, however: the second charge will exert a reaction on the first one, and energy will be involved in the process!

[4] Entropy ΔS = k ln2 per bit lost; minimum energy $E_{\min}$ = _ kT per bit stored. k: Boltzmann constant (= 1.38×10^{-23} Joule °K^{-1}).

emerged as *code-carrying templates* whose effect on other molecules in their environment is to bind them through a catalysis-like process into conglomerates according to patterns represented in the code. Some of these polymers might also have served as templates for the production of others like themselves—those more efficient in this process would multiply. Replication and natural selection may indeed have begun already in a pre-biotic chemical environment.

Concerning these macromolecules, we can say that beyond a certain degree of complexity, *information as such* (on the existence of certain patterns, like the code in the sequence of nucleotides in RNA) begins to play the decisive role in organizing the chemical environment. It is because of the supremacy of information as the controlling agent that these molecules can catalyze chemical reactions that would be highly improbable to occur naturally under the same environmental conditions and energy sources. In summary, the perhaps first "grand moment" in the evolution of life occurred when molecules of sufficiently high complexity appeared, for which the information expressed in their structural patterns began to take a highly selective control over their interaction with the surrounding medium [5].

From the very beginning of the evolutionary process, very slow changes of the environment (with time-scales orders of magnitude longer than the life-span of one generation of the species) were incorporated in the memory code of the DNA through the combined action of mutation and survival of the fittest. Special mechanisms developed inside the cells to accelerate the process of random genetic change to facilitate this process. With the appearance of *locomotion* the number of relevant environmental variables affecting the organism increased drastically, with time-scales of change down to a fraction of a second. Sophisticated sensory organs developed, and with them, nervous systems. What started out as a simple environmental signal conversion and transmission apparatus evolved into the central nervous system of the higher vertebrates, with sophisticated input information analysis and behavioral response-setting capabilities. The *brain* emerged as the "central processor" to carry out the fundamental operations of sensory information processing, cognitive operations of object recognition and environmental representation, and the planning of motor response based on the momentary state of the environment, on the state of the organism, on innate instructions (instincts) and on learned information.

In its evolution, the brain developed in a way quite different from the development of other organs of the vertebrate body. Separate layers appeared, with distinct functions, *overgrowing* the older structures but not replacing them, thus preserving "older functions"—the "hard-wired" memories or *instincts* that a species has acquired during evolution. The outermost layer, the cerebral *cortex*, executes all higher order cognitive operations. These functions are, listed here in oversimplified form: (i) to analyze the information received from the senses and the detectors that monitor the posture of the body in the environment; (ii) in cooperation with subcortical structures (see below), to sort out that which is of relevance for the organism's well-being and intentions; (iii) to construct new or improved neural representations or "mental maps" of the surrounding space and events therein; (iv) to determine and store in memory relevant cause-and-effect relationships; and (v) to

[5] No doubt that we are dealing here with a "threshold" period in the evolutionary schedule in which a transition takes place to information-driven interactions that later will culminate in the development of full-fledged information-processing operations of the human brain. It is interesting to note that in the molecular domain this transition is conceptually similar to the boundary that separates the domains of quantum and classical behavior of matter, the so-called region of wave function de-coherence.

activate the musculature to elicit the most appropriate behavioral response. The human brain, of course, has some very distinct additional capabilities. In this Section we will discuss some common aspects of animal and human brain function [(iii)].

We first turn briefly to perception. The most refined senses are vision and audition and we will refer mostly to these in what follows. The neural circuitry in the periphery and afferent pathways up to and including the so-called primary sensory receiving area of the cortex (Fig. 2) carries out some basic preprocessing operations mostly related to *feature detection* (e.g., detection of edges and motion in vision, spectral pitch and transients in hearing). The next stage is *feature integration* or *binding*, needed to sort out from an incredibly complex input those features that belong to one and the same spatial or temporal object (i.e., binding those edges together that define the boundary of the object; sorting out those resonance regions on the basilar membrane that belong to one musical tone). At this stage the brain "knows" that it is dealing with an object, but it does not yet know *what* the object is. This requires a complex process of comparison with existing, previously acquired information. The recognition process can be "automatic" (associative recall) or require a further analysis of the full sensory input in the frontal lobes. As one moves up the stages of Fig. 2, the information processing becomes less automatic and more centrally controlled; more *motivation-controlled* actions and decisions are necessary, and increasingly the *previously stored* (learned) information will influence the outcome.

The motivational control deserves special attention. One of the lower, phylogenetically much older parts of the vertebrate brain is the so-called *limbic system*. We shall use this term as a short-hand for a system that comprises several deep structures including the amygdala, the ventral tegumental area and the hippocampus. In conjunction with the cingulate cortex and the hypothalamus (the region of the brain that receives and integrates signals from the autonomic nervous system and regulates the neuro-chemical information system), the limbic system constructs a map of the state of the organism, "polices" sensory input, selectively directs memory storage according to the relevance of the information, and mobilizes motor output (Fig. 3). In short, the aim of this system is to ensure a behavioral response that is most beneficial to the organism according to genetically acquired information—the so-called *instincts* and *drives*. Emotion and motivation are integral manifestations of the limbic system's guiding principle to assure that all cortical processes are carried out to maximum benefit of the organism and the propagation of the species [6]. The limbic system constantly challenges the brain to find solutions to alternatives, to probe the environment, and to perform certain actions even if not needed at that moment (e.g., animal play for the purpose of training in skilled movement). In all its tasks, the limbic system communicates interactively with the cortex, particularly the prefrontal regions, relating everything the brain perceives and plans to the organism and vice versa. This interplay gives rise to animal consciousness (Damasio's "core consciousness" [(iv)]).

To carry out its functions the limbic system works in a curious way by dispensing sensations of reward or punishment; pleasure or pain; love or anger; happiness or sadness or fear. Of course, only we humans can report to each other on these feelings, but on the basis of behavioral and neurophysiological studies we have every reason to believe that higher vertebrates also experience them. What kind of evolutionary advantage was there to this mode of operation? Why does pain

[6] The functions of the limbic system are sometimes referred to in English as "the four F's": **F**eeding, **F**ighting, **F**leeing and **F**... reproducing!

hurt? Why do we feel pleasure scratching a mosquito bite or eating chocolate? How would we program similar reactions into a robot? [7] Obviously this has to do with evoking the *anticipation* of pain or pleasure whenever certain constellations of environmental events are expected to lead to something detrimental or favorable to the body, respectively. Since such anticipation comes *before* any actual harm or benefit could arise, it helps guide the organism's response in a direction of maximum chance of survival. In short, the limbic system directs a brain to *want* to survive and to find out the best way of doing so given current, genetically unprogrammable circumstances. Plants cannot respond quickly and plants do not exhibit emotions; their defenses (spines, poisons) or insect-attracting charms (colors, scents) develop only through the slow process of evolution.

How is "information" actually represented in the brain? At the individual neuron level, it is encoded in the temporal succession of pulses of electric trans-membrane potential when the information is being transmitted or processed, and in the form of changes in the synaptic connections between neurons, when it is stored in long-term memory [8]. In this discussion, however, we are more interested in the neural representation of higher-level cognitive information or "maps" that involve many processing stations in the brain and millions of neurons. The new non-invasive techniques of neural activity imaging (functional nuclear magnetic resonance or fNMR, positron emission tomography or PET) are providing a wealth of new information. For instance, observations clearly confirm the processing stages sketched in Fig. 2, and, moreover, demonstrate that they do unfold exactly in *reverse* order during the process of mental imaging (imagining things). These techniques do not, however, provide details of the neural activity per se—they only identify the regions that are activated in a given task and the corresponding timing [(v)].

Concerning the encoding of integral sensory and interoceptive information per se, or of *any* other kind of cognitive information, there is now a convincing ensemble of data that show that this encoding appears in the form of a *specific spatio-temporal distribution of neural impulses*. For instance, according to this description the mental representation or image of an object (visual, acoustic, olfactory, tactile) appears in the form of a specific distribution of electrical signals in the neural network of the cerebral cortex *that is in one-to-one correspondence with the specific features sensed* during the perception of this object. By "one-to-one" we do not mean a "geometric" or isomorphic correspondence but, rather, a distribution which, however complex, is always the *same* (within limits) whenever information on that particular object is involved in brain processing. According to this result, "cognition" is nothing else but the occurrence of a specific neural activity in certain areas of the cortex that is in one-to-one correspondence with the object that is being recognized, remembered or imagined. For instance, the distributions of neural activity displayed on the intervening cortical areas by the perception of the following objects—a big red apple, an apple tree, a piece of an apple pie—though widely different, would all bear in common some subset of neural activity, namely the one that appears in correspondence with, and defines the cognition of "apple". Every time a dog hears the voice or sees the face of its master, no matter

[7] We could program a robot to emit a crying sound whenever it loses a part, to reach out toward loose screws and tighten them, or to seek an electrical outlet whenever its batteries are running low, but how do we make it to actually feel "pain" or "pleasure"?

[8] It is important to point out that there is also a *neurochemical* information system (neurotransmitters) which regulates the general state of the brain and organism. This system is "slow" and only *transmits*, but does not process, information.

under what circumstances, some unique distribution of neural activity will occur in its brain that is specific to the dog's recognition of its master.

It takes some time to get used to the meaning and relevance of this "specific spatio-temporal activity distribution" and the information it represents. We could describe such a distribution mathematically with a function $\psi(\boldsymbol{r}, t)$, representing the neural activity (e.g., neural firing rate) at point $\boldsymbol{r}$ in the brain tissue at the time t. Unfortunately, this function would be *ultra-discontinuous*: two neighboring neurons may participate in totally different tasks, and thus fire unrelated electrical impulses. We could use a vector function $\psi = [f_1(t), f_2(t), \ldots\ldots f_m(t), \ldots\ldots f_n(t)]$, where $f_m(t)$ is the firing rate of the m-th neuron in an ensemble of n cells. For a given distribution, an $f_m(t)$ would be zero if the corresponding neuron does not participate in the particular cognitive task represented by ψ. An even more frustrating problem is that in each cognitive task there may be hundreds of millions if not billions of neurons involved. In any case, if we *could* determine ψ, there would be *one specific* vector function for each mental image (and *that* number is, of course, beyond comprehension!) and a typical duration of a triggered imaging event would be only 50-200 ms. Although it seems hopeless to use this description in a mathematical sense, it does help organize one's ideas about the way information is encoded in the brain. In particular, it helps understand the *categorical* and *hierarchical* way in which information is treated in brain processing: the representation (ψ) of more complex concepts can be thought of as the sum of the representations of its simpler parts (think of: apple orchard $\Rightarrow$ apple tree $\Rightarrow$ apple, all having ψ_{apple} in common)! Recent experiments on categorical perception strongly confirm the main aspects of this description [vi]. This is precisely how memory storage and recall works.

The act of remembering, or memory recall, consists of the re-elicitation or "replay" of that particular distribution of neural activity which is specific to the object or concept that is being remembered. This pattern can be triggered by external sensory information: when a dog looking at a group of people suddenly recognizes his master, the corresponding "ψ_{master}" was triggered. A memory recall can also be triggered internally by the perception (or, for humans, imagination) of *correlated* events: a hungry feeling may trigger ψ_{master} in that dog's brain and he may run to his master to beg for food. These examples show that one must recognize the neural activity distribution as the fundamental "physical quantity" that represents "information" in the working brain.

It is important to emphasize that the integral neural representation ψ is not limited to just one brain processing center, but that it involves much of the cortex and many underlying nuclei. What characterizes this specific neural distribution is monolithic *coherence* and *synchronism* [vii], and *consistent specificity* with each cognitive act, feeling or motor output. Borrowing a concept from condensed matter physics, we say that it represents the cooperative action creating unity in an ultracomplex system. While no doubt ψ involves many "subservient" programs or "subroutines" that trigger or control very specific information-processing operations which never reach consciousness, *there is only one "main program"*; this leads to unity of perception and behavior, and represents the *basic conscious state* of the animal or human brain (close but not equal to Damasio's "core consciousness" [iv]). There are operational and evolutionary reasons for having a single state of consciousness. First of all, if several "main programs" were to run at the same time, there could be no coherence between the many subsystems, and simultaneous but conflicting orders would ensue: the brain would fall into a sort of epileptic state. Second, the instantaneous state of brain activity as represented by the specific neural activity distribution ψ must be spread over processing networks that are responsible for associative memory recall in all modalities, in order to be able to activate an appropriate behavioral response. And finally, it must be intimately coupled to the limbic system, to be able to implement the dictates of the latter for the benefit of the organism (see Fig. 3).

3. INFORMATION, HUMAN INTELLIGENCE AND SELF-CONSCIOUSNESS

Another "grand moment"—the last and grandest of all—came over two million years ago when the evolution of the *human* brain began. Interestingly, from the neurophysiological and neuroarchitectonic points of view the human brain is not particularly different from the brain of a primate like the chimpanzee. It has a cortex with more neurons and some of the intercortical fasciculae have more fibers, but this difference is of barely one order of magnitude—except for the number of synapses in the adult brain, which in humans is orders of magnitude larger. Is the difference in information processing capabilities only one of quantity but not one of substance?

Aristotle already recognized that "animals have memory and are able of instruction, but no other animal except man can recall the past at will". More recently, J. Z. Young [viii] stated this in the following terms: "Humans have capacity to rearrange the "facts" that have been learned so as to show their relations and relevance to many aspects of events in the world with which they seem at first to have no connection". More specifically, the most fundamentally distinct operation that the human, and only the human, brain can perform is to recall stored information, images or representations (i.e. ψ's), manipulate them, and re-store modified or amended versions thereof *without any concurrent external sensory input.* The acts of information recall, alteration and re-storage without any external input represent the *human thinking process* or *reasoning* [ii].

It appears that humans have the ability of making a "*representation of representations*", i.e., that there is a second order level of representations in the brain which has cognizance of consciousness and can manipulate independently the primary representation ψ of current brain activity (Fig. 4). The feeling of being able to observe and control one's brain function, the capacity of making independent decisions and overrule (or independently stimulate) the limbic dictates, and the feeling of "being just one", collectively represent what we call *human self-consciousness* (close but not equal to Damasio's "extended consciousness" [iv]). There is no need to assume the existence of a separate neural network or separate entity ("mind") for this second order representation; there is enough informational space in the cortical and subcortical networks that can be shared with the primary representation ψ (except, perhaps, that there may be a greater participation of the prefrontal cortex in the former). The "mental singleness" we humans experience reflects the fact that a thought can establish itself only if none of the participating modalities throws in a veto [vii]; coherence of brain function assures the latter. We should note that consciousness really is not a state of the brain but *a process* extended over a certain window of time. To summarize: in an animal (including human) brain, consciousness creates unity of perception and response; in a human being, *self*-consciousness creates the ability of sensing basic consciousness and changing its course in total independence of real-time sensory input (Fig. 4).

All the aforesaid had vast consequences for human brain evolution. The capability of re-examining, rearranging and altering images led to the discovery of previously overlooked cause-and-effect relationships (creation of *new* information!), to a quantitative concept of elapsed time, and to the awareness of future time. Along came the possibility of long-term prediction and planning [9], the postponement of behavioral goals and the capability to overrule the dictates of the limbic system (the instincts): the body started serving the brain instead of the other way around. Mental images could thus be created that had no relationship with sensory input; abstract thinking and artistic creativity began; this also brought the development of beliefs and values. Concurrently came the ability to encode complex mental images into specific acoustic signals (concurrent with the development of a more sophisticated muscular control of the larynx) and the emergence of *human language.* This was of such decisive importance to the development of human intelligence that certain parts of the auditory and motor cortices began to specialize in verbal image coding (in the left cerebral hemisphere for 97% of all persons), and the human thinking process began to be influenced and even controlled by the language networks. Finally, though much later in human evolution, came the deliberate storage of information in the environment; this externalization of memory led to the documentation of feelings and events through written language, music, visual artistic expression, and science—to human culture as such.

ACKNOWLEDGMENT

I am grateful to Professor Valentin Braitenberg of the Max Planck Institute for Biological Cybernetics for enlightning discussions on the concept of "information".

REFERENCES

[i] Based on J. G. Roederer (2000). Information, life and brains, in *Astrobiology*, J. Chela-Flores et al., eds., 179-194.

[ii] Roederer, J. G. (1978). On the relationship between human brain functions and the foundations of physics, *Found. of Phys.*, *8*, 423-438.

[iii] For a summary and references, see for instance: Roederer, J. G. (1995). *Physics and Psychophysics of Music*, Springer-Verlag, New York. (Translations: *Acústica y Psicoacústica de la Música*, Ricordi Americana, Buenos Aires, 1997; *Acustica e Psicoacustica da Musica*, Editora da Universidade de Sao Paulo, 1998; *Physikalische und Psychoakustische Grundlagen der Musik*, Springer-Verlag, Heidelberg, 1998).

[iv] Damasio, A. (1999). *The Feeling of What Happens*, Harcourt Inc., San Diego, New York, London.

[v] e.g., Thorpe, S. J. and M. Fabre-Thorpe (2001). Seeking categories in the brain, *Science*, *291*, 260-263.

[vi] Freedman, D. J., M. Riesenhuber, T. Poggio and E. K. Miller (2001). Categorical representation of visual stimuli in the primate prefrontal cortex, *Science*, *291*, 312-316.

[vii] von der Marlsburg, Chr. (1997). In *Cognition, Computation and Consciousness*, M. Ito, Y. Miyashita and E. T. Rolls, eds., Oxford University Press, Oxford, New York and Tokyo, pp. 193-204.

[viii] Young, J. Z. (1987). *Philosophy and the Brain*, Oxford Univ. Press, Oxford, New York.

[9] It may well be that higher primates have "bursts" of self-consciousness during which internally recalled images are manipulated and a longer-term future is briefly "illuminated". But there is no clear and convincing evidence that any outcome is stored in memory for later use. In other words, it is conceivable that some higher mammals may exhibit bursts of human-like thinking, but they seem not to be able to do anything long-lasting with the results. There is a contentious debate on this issue between animal psychologists and some brain scientists.

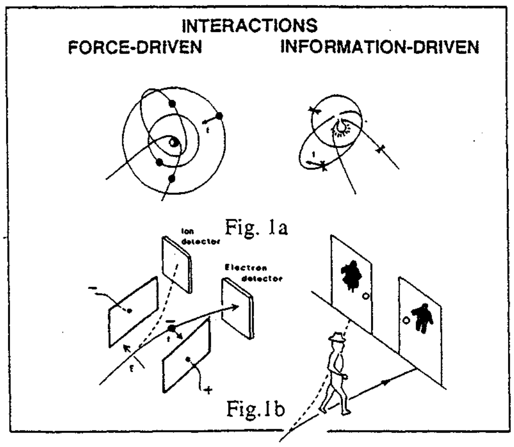


Fig. 1a

Fig.1b

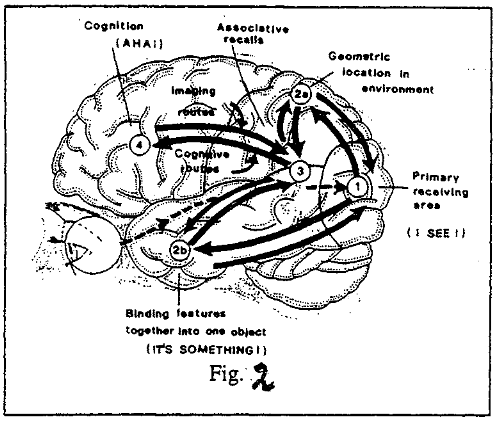


Fig. 2

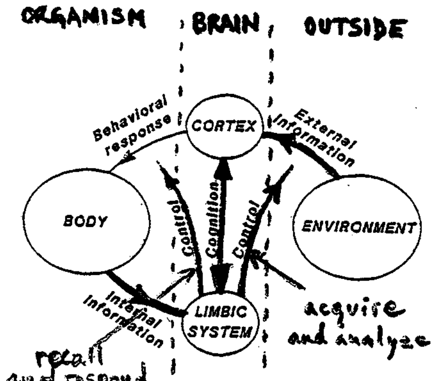
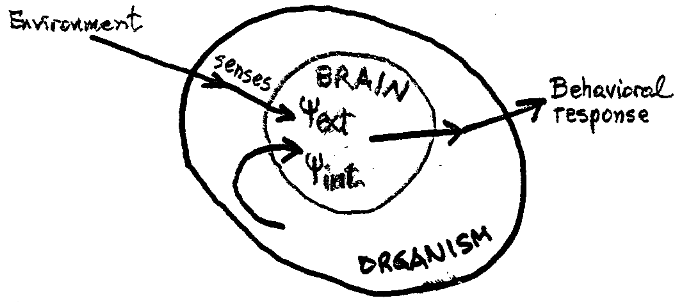
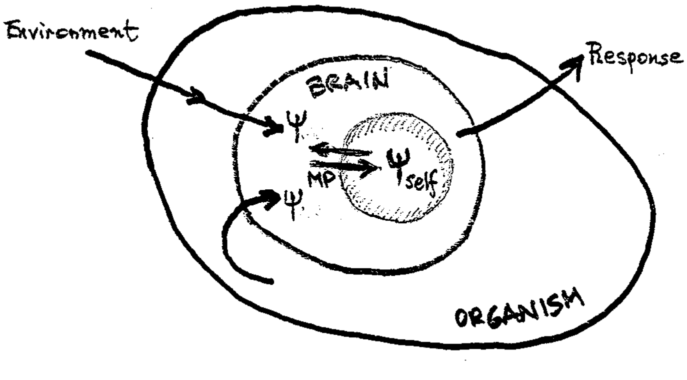


FIGURE 3



$\Psi_{ext} + \Psi_{int}$ + memory ⇒ Main program (MP) (current goal)
Subsidiary programs (coordinated by MP)
Only current info from senses and organism can change MP
→ Core consciousness



ditto, plus consultation or scanning of memory
without external or somatic input
→ Self-consciousness (extended or autobiographical consciousness)

Fig 4