ON THE CONCEPT OF INFORMATION AND ITS ROLE IN NATURE

Juan G. Roederer

Geophysical Institute, University of Alaska-Fairbanks, USA[*]
and The Abdus Salam Centre for Theoretical Physics, Trieste, Italy

Abstract: In this article we address some fundamental questions concerning information: Can the existing laws of physics adequately deal with the most striking property of information, namely to cause specific changes in the structure and energy flows of a complex system, without the information in itself representing energy in any of its characteristic forms? Or is information irreducible to the laws of physics and chemistry? Are information and complexity related concepts? Does the Universe, in its evolution, constantly generate new information? Or are information and information-processing exclusive attributes of *living* systems, related to the very definition of life? If that were the case, what happens with the physical meanings of entropy in statistical mechanics or wave function in quantum mechanics? How many distinct classes of information and information processing do exist in the biological world? How does information appear in Darwinian evolution? Does the human brain have unique properties or capabilities in terms of information processing? In what ways does information processing bring about human self-consciousness?

We shall introduce the meaning of "information" in a way that is detached from human technological systems and related algorithms and semantics, and that is not based on any mathematical formula. To accomplish this we turn to the concept of *interaction* as the basic departing point, and identify two fundamentally different classes, with information and information-processing appearing as the key discriminator: force-field driven interactions between elementary particles and ensembles of particles in the macroscopic physical domain, and information-based interactions between certain kinds of complex systems that form the biological domain.

We shall show that in an abiotic world, information plays no role; physical interactions just happen and do not require any operations of information processing. Information only enters the non-living physical world when a living thing interacts with it—and when a scientist extracts information through observation and measurement. But for living organisms, information is the very essence of their existence: to maintain a long-term state of unstable thermodynamic equilibrium with its surroundings, consistently increase its organization and reproduce, an organism has to rely on information-based interactions. This latter class comprises *biomolecular* information processes controlling the metabolism, growth, multiplication and differentiation of cells, and *neural* information processes controlling animal behavior and intelligence. The only way new information can appear is through the process of biological evolution and, in the short term, through sensory acquisition and the manipulation of images in the nervous system. Non-living informational systems such as books, computers, AI systems and other artifacts, as well as living organisms that are the result of breeding or cloning, are *planned* by human beings and will not be considered here.

[*] Permanent address: Geophysical Institute, University of Alaska-Fairbanks, Fairbanks AK 99775-7320, USA. jgr@gi.alaska.edu.

CONTENTS

INTRODUCTION

If we want to understand the role of information in Nature, particularly in the biological world, we must find an appropriate definition of this ubiquitous and seemingly trivial concept—a definition that is objective and *independent* of human actions and human-generated devices. Conventional information theory works mainly with computers and technological communications systems; it is not so much interested in an independent formal definition of "information" as it is in the precise mathematical expression for the information content of a given message and its degradation during transmission, processing and storage. In general two classes of information are considered: "statistical" or "semantic" information describing the outcome of expected alternatives for the behavior of a system, and "algorithmic" information, defined as the minimum-bit statement that describes a given thing [1]. The content of information in the former is related to the probability of occurrence of a given alternative; in the latter it is given by the actual number of bits of the minimum-bit statement. In this article we are interested in the existence and meaning of information as a concept that is unrelated to human communication and technological systems, and that is not based on any formula. This is no trivial matter: one cannot tell by examining a complex organic molecule or a slice of brain tissue whether or not it has information (beyond that generated in our senses by its own composition and structure)—as yet there is no "information dynamics" comparable to the dynamics of matter [2]. In the natural world information requires complexity and organization but complexity and organization alone do not represent information.

At the present time, some of the most fundamental questions under discussion are: Can the rigorous methods of physics adequately describe and explain that strange but well-recognized property of information, namely that the mere *form*, *shape* or *pattern* of something—*not* its field, energy or forces—may lead to dramatic and specific change in a system? Or is information irreducible to the laws of physics and chemistry? Are information and complexity related concepts? Does the Universe, in its evolution, constantly generate new information? Or are information and information-processing exclusive attributes of *living* systems, related to the very definition of life? If that were the case, what happens with the physical meanings of entropy in statistical mechanics or wave function in quantum mechanics? How many distinct classes of information and information processing do exist in the biological world? How does information appear in Darwinian evolution? Is information inextricably linked to purpose and intentionality? Does the human brain have unique properties or capabilities in terms of information processing? In what ways does information processing bring about human self-consciousness?

We shall try to accomplish our task by starting with the process of "*interaction*" as the underlying basic, primordial concept. We shall identify two fundamentally different classes of interactions between the bodies that make up the universe as we know it, with the concepts of information and information processing appearing as the key discriminators between the two. As much as possible, we shall follow an objective line of reasoning without appealing to "made things" such as computers, artificial communications systems, Turing machines or robots.

The main objective of this article is to stimulate interest among physicists, especially those who teach, in a strongly multidisciplinary subject to which physics, its methods, instruments and "ways of thinking" have a lot to contribute. It does not address philosophical issues or any social aspects. References are mostly limited to articles in journals or books of ready access to physicists. Unfortunately, many topics are still rather speculative and controversial—but to be fair to all points of view would require writing a whole book! Thus, I must apologize to the researchers working in the different disciplinary areas discussed below for taking sides in issues where there is no consensus yet, or for occasionally sacrificing parochial detail to improve ecumenical understanding.

1. PHYSICAL AND BIOLOGICAL INTERACTIONS

As we look around us we realize that much of the environment appears to be made up of discrete "things": clumps of matter—complex bodies which in general have clearly defined boundaries. When we explore the environment with scientific instruments, we realize that also at the microscopic level, the cells of living matter, molecules, atoms and elementary particles can all be considered "clumps of matter", although for the latter we may not be able to define a "clearly defined boundary". In the macroscopic domain, even fluids appear organized in such a way: the ocean, the atmosphere and the ionized plasmas that make up stars, stellar winds and magnetospheres tend to arrange themselves into discrete regions separated by boundary layers or interfaces with transverse dimensions small compared to the size of the regions in question. The late Nobel Laureate Hannes Alfvén called it "the cellular Universe". As the Universe evolved after the Big Bang, the gradual condensation and stratification of materials led to increasing inhomogeneities of the regions [3] and an increasing complexity of their structure and the interfaces separating them.

Our senses of vision and touch are geared toward the perception of borders, shapes and patterns of the boundary surfaces of everyday clumps of matter or *objects*, and our brain is endowed with the cognitive machinery to restitute the three-dimensional nature of objects from the perceived 2-D boundaries, 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 carried by waves, too, but the most relevant aspect for perception is their two-dimensional angular distribution in space 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*—without attempting any formal definition (a "metaphysical primitive"). We just note that in an interaction a *correspondence* is established between certain properties of one object (its position, speed, form, etc.) and specific changes of the other, and maybe vice versa. Quite generally, we may consider "science" as the systematic and quantitative study of interactions between objects or ensembles of objects. Each scientific discipline studies the interactions affecting objects belonging to some common class (e.g., elementary particles, stars, molecules, cells, people, plants, etc.). The interactions observed in the *natural* environment (i.e., leaving out all human-made artifacts) can be divided into two broad classes.

The first class comprises the *physical interactions* between two bodies, in which we observe that the presence of one modifies the properties of the other (its motion, structure, temperature) in a definite way that depends on the relative configuration of both components of the system (their relative positions, velocities, etc.). Primary physical interactions between elementary particles are two-way (true *inter*-actions) and reversible, but in the macroscopic domain the observable and measurable interactions usually appear as chains of irreversible cause-and-effect relationships. They are irreversible because to unfold *exactly* in reverse order without external intervention, too many conditions concerning all intervening interactions must be fulfilled *exactly*, making the natural occurrence of the reverse process extremely unlikely. In a non-thermal physical interaction between two bodies isolated from the rest, the concept of *force* is introduced as the agent responsible for the change (acceleration or deformation) that occurs during the interaction [4]. For fundamental interactions such as gravity and electromagnetism, the concept of a *force field* is introduced as a property of the space surrounding the interacting bodies. The end effect of a physical interaction will always depend on some "initial conditions", such as the initial configuration (positions, velocities) of the interacting bodies. A most fundamental characteristic is the fact that during the interaction, there is a direct transfer of energy from one body to the other, or to and from the *interaction mechanism* itself (in the case of fundamental interactions, the force field; in more complex cases, some process linking the two bodies). In other words, the changes that occur in the two interacting bodies are coupled energy-wise; the energy is provided by the interaction mechanism itself. For macroscopic chains of interaction processes, the more complex and irreversible the system, the weaker the energy coupling will be between one end and the other of the chain.

At the microscopic, subatomic level, everything can be reduced to four basic interactions between elementary particles. The environment with which humans and other living beings interact, however, pertains to the *macroscopic domain* in which objects consist of the order of 10^{20} or more molecules, and in which all physical quantities of relevance are averages over enormously large ensembles of particles. Indeed, our senses in general respond to signals that are ensemble-averages over myriads of signals from the individual atomic constituents of macroscopic bodies. We should point out that at the macroscopic level all bodies are extraordinarily complex, but it is the high degree of order and organization of this complexity that makes them amenable to cognition and mathematically describable with continuous functions representing ensemble averages of the microscopic variables. For that very reason, in physics we can describe the behavior of interacting complex bodies in terms of simple *models* (mass points, rigid bodies, ideal gas, etc.) and use *linear* mathematical relationships as a first approximation. Finally, it is important to note that in the macroscopic domain, all physical interactions can be reduced to two elementary ones, namely gravitation and electromagnetism (friction, elasticity, thermal interactions, and chemical reactions are all, ultimately, electromagnetic in nature). We shall call the physical interactions between inanimate objects *force-field driven interactions*.

To turn to the other broad class of interactions, consider the two examples shown in *Fig. 1*: several satellites orbiting a central massive body on the left, and some insects "in orbit" around a light bulb, on the right. Both have in common motions along some well-defined paths 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 of 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 complex mechanism of light detection and pattern recognition—a process that involves *information acquisition and processing*. In other words, we have the chain: light emission → pattern detection → pattern analysis → muscle activation; the energy required at the insect's end is not provided by the incoming light waves. No such set of algorithms appears in the case of the satellites: the concept of information is totally alien to gravitational interactions—and to *all* purely physical interactions between inanimate objects. The latter just happen and don't require intermediate operations of information detection and processing. The configuration of satellite orbits may vary greatly depending entirely (and exclusively) on the initial conditions, whereas in the case of the insects the final orbital configuration may depend very little on the actual "injection point" of each insect. And, most importantly, there is strict energy coupling between each satellite and the gravitational potential of the central body, whereas the energy required for the physiological processes in the insects is totally unrelated to that of the bulb's light.

We call this second class of interactions *information-based* [5]. Of course, the intervening processes always involve physical mechanisms; the key aspect, however, is their *control* by information and information-processing operations. There is no energy coupling between the interacting bodies (although energy must be supplied locally for the intervening processes). In the example of *Fig. 1* involving the insects, the electromagnetic waves (light) themselves do not drive the interactions—it is the *information* in the patterns of the wave trains, not their energy, which plays the controlling role; the energy needed for change must be provided locally. Quite generally, in all information-based interaction mechanisms the information is the *trigger* of physical/chemical processes, but has no relationship with the energy or energy flows needed by the latter to unfold. Physical and chemical processes provide a medium for information, but do not represent the information *per se*.

2. PROTOTYPE EXAMPLES

The following question, already formulated in the preceding section, arises again: Can the existing laws of physics adequately deal with the most striking property of information, namely to cause specific changes in the structure and energy flows of a complex system, without the information in itself representing energy in any of its characteristic forms? The answer hinges on finding an objective definition of information. To that effect, let us analyze a few examples to elucidate more clearly the characteristics of force-field driven and information-based interactions, identify different prototypes, and compare them with each other.

1. *An electron interacting electrostatically with a proton at rest*. From the point of view of quantum electrodynamics, the interaction mechanism consists of the emission and absorption of virtual photons by the interacting particles. In a first-order Feynman diagram the electron (or the proton) emits a photon which is then absorbed by the proton (or electron); the energy and momentum balance at each node (emission/absorption process) accounts for a change in the motion of each particle. The end result, strictly reversible, is the sum total of all possible photon exchange processes; the total energy (and momentum) of the pair of particles is conserved. We are tempted to say (and often do so) that in this interaction, a virtual photon "carries information" from one particle to the other; however, this is purely subjective parlance: no information, information-processing and purpose are at work at either end of the emission/absorption process.

2. *A small body orbiting an irregularly shaped massive asteroid* (e.g., asteroid Ida with its tiny satellite Dactyl). This is a macroscopic physical interaction. The satellite's orbit, viewed from a system fixed to the asteroid, is very complicated due to the latter's irregular mass distribution; the resulting motion is governed by the sum total of the gravitational actions (forces) of each element of mass of the asteroid [6], in

each of which information and information-processing do not play any role. There is strict to-and-fro energy coupling (and functional continuity) between the orbital characteristics and the irregular features of the asteroid's gravitational potential. In view of the correspondence between the detailed features of a particular orbit and the irregularities of mass distribution, we are again tempted to say that the orbital characteristics "carry information on the pattern of mass distribution in the asteroid" (indeed, this is how the mass distribution is determined). However, no information is actually at work in this natural system; each element of the pattern at the "source" (the asteroid) contributes with a direct dynamic effect on the "recipient" (the satellite) and there is no *need* to invoke the concepts of information and information processing. A similar case is represented by the *footprint of a rock* that has been lying on the soil for a long time. Here we have a complex pattern on the ground carved by the elastic interaction with the rock's irregular surface—an interaction in which a fixed pattern has led to specific change in another "body". However, the change was produced in point-by-point physical interactions in which there was a direct coupling (force- and energy-wise) between the "source" and the "recipient". Again, no information, information-processing or purpose can be identified in the process of the formation of the imprint.

In both above examples we have a physical interaction between two objects mediated by a "field", with energy coupling and continuity. There is no information, signals, messages, codes, decisions or purpose involved—unless we, humans, put them there by making a mental model of the interaction process in which analogies are drawn from subjective experience. The following are prototypes of information-based interactions.

3. *A dog walking around an obstacle*. The dog responds to the visual perception of the obstacle, a complex process that involves information-processing and decision-making at the "receiver's" end with a definite *purpose*. At the obstacle's (the sender's) side, we have scattering and/or reflection of incident light waves; no information and no purpose are involved—only physical processes are at work here. There is no energy coupling between sender and receiver; like in the example of the insects of *Fig. 1*, what counts is *not* the energy of the electromagnetic waves but the pattern of their spatial distribution. This is the prototype of information-based interactions that involves *information extraction*: it is a fundamental element in the interaction of any organism with the environment. The electromagnetic waves emitted by the obstacle and detected by the dog's sensors are *signals* [7]. Note that the mechanism for this class of interactions must *evolve*; it cannot arise spontaneously. As a matter of fact, Darwinian evolution itself embodies a gradual, species-specific information extraction from the environment (see section 4). Other examples would be that of a geologist's brain processing the visual image of the rock's footprint from example 2 above, and, at the cellular/molecular level, a cell that has sensed the lack of a given nutrient, and has responded by starting up the synthesis of enzymes necessary to utilize other nutrients. The signal (e.g., glucose concentration) has neither purpose nor information, but becomes information when detected by the metabolic machinery in the bacterium.

4. *An ant leaving a scent mark along its path*. This is the reverse process of the preceding example, in which information is *deposited* in the environment for later use. The sender has a specific purpose and information is involved. At the other end of the interaction (the

path) only physical (chemical) processes are at work (but later, another ant can extract information from these physical signals). This is an example of deliberate environmental modification with a purpose; writing an article is another. At the cellular or molecular level we can refer to the methylation process, which changes the cell's ability to transcribe certain portions of its DNA into RNA, without changing the information content of the DNA. This alters the program of the genes that they can express, thereby rendering these cells different from the others. Methylation is heritable, so this lets the change persist across multiple cell divisions.

5. *A person talking to another*. Here clearly we have information processes and purpose at *both* ends of the interactive link. The corresponding acoustical waves carry a *message*; again, what counts in this interaction is not the energy but the acoustical patterns. To function, a common code must exist at both ends (see next section). This interactive process is a prototype of *communication*. In principle, we can think of this prototype as a succession of types 4 and 3, in that order. An example at the molecular/intracellular level is the genome-regulated manufacture of proteins (DNA-guided synthesis of RNA by an enzyme at one end, manufacture of the protein in a ribosome at the other).

3. INFORMATION AND LIFE: DEFINITIONS

Let us now formalize our description of information-based interactions and related definitions, and point out again both analogies and differences with the case of physical interactions between inanimate bodies. First of all, we note that information-based interactions occur only between bodies or, rather, between systems the complexity of which exceeds a certain minimum degree. We say that a (complex) system *A* is in information-based interaction with a (complex) system *B* if the configuration of *A*, or, more precisely, the presence of a certain spatial or temporal feature or pattern in system *A* causes a specific alteration in the structure or the dynamics of system *B*, with a final state that depends *only* on whether that particular pattern was present in *A*. Moreover, it is a condition that (a) both *A* and *B* be *decoupled energy-wise* (meaning that the energy needed to effect the changes in system *B* must come from sources other than energy reservoirs or flows in *A* or the physical mechanism linking *A* with *B*), and (b) *no lasting changes* occur as a result of this interaction in system *A* (which thus plays a catalytic role in the interaction process). In other words, in an information-based interaction a one-to-one *correspondence* is established between a spatial or temporal feature or pattern in system *A* and a specific change triggered in *B*; this correspondence depends *only* on the pattern in question, being independent of other circumstances.

We should emphasize that to keep man-made artifacts and technological systems (and also clones and artificially bred organisms) out of the picture, both *A* and *B* must be "natural" bodies or systems, i.e., *not deliberately manufactured or planned* by an intelligent being. We further note that primary information-based interactions are *unidirectional* (i.e., they are really "actions", despite of which I shall continue calling them *inter*actions), going from one of the bodies (the "source" or "sender") to the other (the "recipient" or "receiver"). There is a true cause-and-effect relationship in which the cause remains unchanged (the pattern in *A*) and the effect is represented by a specific change elsewhere (in *B*). While in basic physical interactions there is an energy flow between the interacting bodies or to and from the interaction mechanism, in information-based interactions any energy required for the participating (but not controlling) physical processes must be provided (or absorbed) by reservoirs *external* to the interaction process. Energy distribution and flow play a defining role in complex systems [3]; however, they have no direct bearing on the information-based interactions between them. In physical interactions the end effect always depends on the initial conditions of the combined system *AB*; in information-based interactions in general it does not. Finally, information-based interactions are usually "discontinuous", in the sense that if the original pattern at the source is modified in a steady, continuous way, the response in the recipient may not vary at all in a continuous way.

Although we have been talking about information all the time, we must now provide a more formal definition: information is *the agent that mediates the above described correspondence*: it is what links the particular features or pattern in the source system *A* with the specific changes caused in the structure of the recipient *B*. In other words, information represents and defines the uniqueness of this correspondence; as such, it is an irreducible entity [8]. We say that "*B* has received information from *A*" in the interaction process. Note that in a natural system we cannot have "information alone", detached from any interaction process past, present or future: information is always there *for a purpose* [9]. Given a complex system, structural order alone does not represent information—information appears only when structural order leads to specific change elsewhere, without involving any specific transfer or interchange of energy. We can only speak of information when it has both a sender and a recipient which exhibits specific changes when the interaction occurs; this indeed is what Küppers has called "pragmatic information" [10]. It is important to note that the pattern at the sender's site (system *A*) in itself does *not* represent information; indeed, the same pattern can elicit quite different responses in different recipients (think of a Rorschach test!).

Concerning the interaction mechanism per se, i.e., the physical and chemical processes that intervene between the sender's pattern and the corresponding change in the recipient, note that it does not enter in the above definition of information. What counts for the latter is the uniqueness of the correspondence—indeed, many different mechanisms could exist that establish the same correspondence. In the transmission from *A* to *B* information has to "ride" on something that is part of the interaction mechanism, but that something is not *the* information. Note that there can be information processing at the sender's end (example 4, section 1), at the recipient's end (example 3 of section 1) or at both ends (example 5). In all cases (excuse for a moment the anthropomorphic language!) an "accord" or common code must exist between sender and recipient so that an interaction can take place (more on this in section 5). Indeed, example 3 requires the evolution of a sensory apparatus able to detect and interpret the presence of an obstacle; example 4 requires the existence of other ants that will respond to the scent marks, and example 5 requires a common language. While the interaction mechanism does not intervene in the definition of information, the accord or common code is integral part of the concept of information—and that of purpose. Information, information-processing and purpose are thus tied together in one package: the process of information-based interaction (*Fig. 2*). We never should talk about information alone without referring to, or at least having present in our mind, the interaction process in which that information mediates a correspondence pattern→change. As an independent, stand-alone concept, information cannot be defined objectively.

Since in the natural world only biological systems can entertain information-based interactions, we can turn the picture around and offer the following definition [11]: *a biological system is a natural (not-human-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). More recently, Küppers stated (ref. [10], p. 66): "…systems of the degree of complexity of organisms, even at the molecular level, can arise and maintain themselves reproductively only by way of information-storing and information-producing mechanisms." By adopting the process of interaction as the primary algorithm, we are really taking one step further and recognize information as the defining concept that *separates* life from any natural (i.e., not made) inanimate complex system.

4. INFORMATION AND PHYSICS

In the abiotic world there is no information (examples 1 and 2 in section 2), unless it appears through the interaction with a living organism (examples 3 and 4). A rock lying on the lunar surface has been interacting physically during billions of years with the Moon's gravity, the soil on which it came to rest, solar radiation, solar wind and cosmic ray particles. But it only becomes a source of "information" when it is looked at or analyzed by a human being (example 3); information played no role in its entire past (example 2). In short, without "observers" or "users" such as life forms (or devices built or designed by intelligent beings), there is no information.

It is rather difficult to imagine a physical, prebiotic world without information. This happens because the mere act of imagination necessarily puts information into the picture. Yet all interactions in the inanimate world are governed by forces and force fields in which information as a controlling agent does not appear. To physicists in particular, the absence of information in a life-less world may seem quite odd. In statistical thermodynamics, for instance, doesn't information play a crucial role in how a system behaves? A loss of information (about the microscopic state of a given system) is accompanied by a gain in entropy and vice versa. And in a double-slit electron diffraction experiment, doesn't the electron going through the open slit have information on whether or not the other slit is closed?

Physics works with information obtained from the environment and sets up a mathematical framework that quantitatively describes the behavior of idealized *models* of "reality". The extraction of information is done through a process called "measurement", which involves a measuring device and two types of interactions: a physical interaction with the system being measured (but it is a willful intervention, so I would put it in the same category as example 4 in section 1, *not* examples 1 or 2) and, at one time or another, an information-based interaction with the perceptional and cognitive apparatus of *a human being* (*prima facie* like example 3, but what the cognitive apparatus perceives is the result of another human action, so it rather should be compared to example 5). Based on the results of measurements, the human brain (see section 7) constructs a simplified model of the system under study and its interactions and formulates physical laws. The spatial and temporal features of these models are in specific *correspondence* with some, but not all, features of the systems being modeled. In view of the definition given in the previous section, it is clear that information as such appears in physics because of the paradigms (correspondences) involved and because measurement and experimentation are processes imposed on nature with a pre-designed purpose—but not because information was present and controlled the physical world "before we looked".

The physical laws reveal the existence of tremendous constraints for the domains of variability of the physical magnitudes that describe the dynamics of bodies in interaction. From the point of view of the traditional concept of information used in communications theory, the more such constraints exist, the less (algorithmic) information is needed to describe the system and predict its evolution. The collapse of any physical law (e.g., a non-conservation, or a breaking of symmetry) will originate new (algorithmic) information. But, again, information *per se* as defined in section 3 plays no role in the control of the dynamical processes involved in any physical, inanimate system.

In a thermodynamic system "information", whenever it appears, relates to the observer, experimenter or thinker, *not* to the system per se [5]. Our senses only respond to macrosystems, and we describe the physical interactions between macroscopic systems using mathematical relationships between state variables. On the other hand, we can infer from observations with special instruments that there also exists a micro description of a thermodynamic system (unattainable in exact, non-statistical terms for practical reasons), in which the interactions between the constituent molecules are mostly reversible, with information *per se* playing no role. It is only when we connect both descriptions that the concept of information creeps in as an active participant. Similar considerations apply to quantum mechanics: it is *us* who interfere with a quantum system by creating "unnatural" situations, such as the application of a measuring device (a process that unavoidably perturbs the state of a quantum system); the setting up of laboratory experiments that have no natural equivalents; the asking of macroscopic-world questions such as "where *exactly* is the electron now?"; or the sending of quantum messages from "Bobs" to "Alices" in situations that would never arise naturally in an abiotic world.

Here we can build a bridge between our focus on pragmatic information and the use of traditional information theory in statistical thermodynamics. Consider *Fig. 2* and view it as a measuring device or "detector" to determine whether or not the given pattern, to which the recipient is "tuned", is present or not in the sender. In other words, take a digital approach to the information-based interaction scheme depicted in the figure. If we then apply this detector to an *ensemble* of complex bodies each one of which may or may not carry the given pattern, it will be possible to assign an *information-content value* to that pattern in the ensemble. In note [13] we use Gibbs' paradox as an illustrative example (see also [5]).

In summary, what does influence the behavior of a physical system under study is *what we humans do to it or with it*: it is us humans who prepare a system, set up its boundaries and often contrive some unlikely initial conditions [12]—even if only in thought experiments. As a matter of fact, what we consider the "initial conditions" in a natural system is only a result of its *previous evolution*, unless those initial conditions are deliberately *set* (or imagined) by a human being (the only cosmologically "true" natural initial condition is the state "right after" the Big Bang [e.g., within a Planck-time interval of 10^{-43} s]). It is those human-induced actions that condition the response of a physical, inanimate system, whether in the laboratory or only in thought. For instance, in Maxwell's demon paradox, the demon (or its robot surrogate) is placed there for a specific purpose *by a human being*—it, and its internal information-processing mechanism have no chance of appearing naturally! In a statement like "the process of cosmic evolution itself generates information" (e.g., p. 131 in ref. [3]) we should realize that this is information *for us* observers *now*—it did not control any of the natural processes involved (unless they pertained to living matter). In a biological system, instead, information is an *active* participant in its own organization, behavior and multiplication, and information exists and operates in total independence of any outside observer [2].

Finally, a few remarks about *products* of life systems which in themselves are inanimate, but which have been produced or created by a living organism. For the discussion let us consider a bird's nest, a person's house, a newspaper and a cruise missile flying to its target. These are all inanimate objects, but they are *products of life systems*; specifically, they are the result of some quite specific information processing, they all carry information, and they all have something in common, not reducible to their components: a *purpose* [5], [13]. However, there are fundamental differences among them. The nest is an instinctive, "automatic" result of a blueprint: information acquired during the long process of evolution in the genes of the bird and hardwired into its brain (see section 7). The house is the result of a blueprint conceived on the spur of the moment by a human brain that had a vision of the future—a *plan* [14] (what is imprinted in the human genes is the capability of *making* blueprints). A newspaper represents a *repository* or storage of information (example 4, section 1) deliberately planned for later use by other humans [15]; and a cruise missile is an artifact deliberately constructed by a human with sophisticated information processing capabilities (recognition of terrain and target) for its ultimate mission. Finally, to a list of inanimate products of life systems we may add all genetically engineered organisms. All products of life systems are based on information from the past; products of human thinking can also include information about the future [5].

5. BIOMOLECULAR INFORMATION AND MEMORY

Let us recapitulate some key points. In an information-based interaction, the presence of a given pattern in complex system *A* causes a specific change in complex system *B*. No energy transfer is involved, although energy must be supplied to the different stages of the interaction process. The pattern can be any set of things, forms or features that bear and maintain a certain relationship in space and/or time (remember that we are *not* considering patterns made by humans or programmed machines.) We may represent this pattern in simplified form as a function of space and time $P_A = P_A(\mathbf{r}, t)$. *P* represents a physical variable that can be simple (e.g., a color, sound intensity or frequency, or the coordinates of a discrete set of objects) or extremely complex (the spatio-temporal distribution of electrical pulses in brain tissue). In particular, if the pattern consists of a spatial or temporal *sequence* of objects or events from a finite repertoire (e.g., the bases in the DNA helix or the action potentials fired by a neuron), we call it a *code*. In general, spatial patterns do not require a local supply of energy to persist in stable form, but temporal patterns do (see next section). The complex system *A* (the sender) could well exhibit many different patterns which, however, from the energetic point of view are all equivalent (*a priori* equally probable). In that case, changing a pattern would change the information-based interaction (i.e., the response of *B*) even if this would not imply any change of the initial thermodynamic state (energy and entropy) of the system.

It is tempting to introduce right here the minimum number of bits needed to describe the function $P_A(\mathbf{r}, t)$ as the algorithmic measure of the information content of pattern P_A. That, however, would only represent the information content of the pattern itself as an *object*, but not of the information actually involved in the *interaction* $A \rightarrow B$. In other words, it would *not* represent any correspondence and it would be devoid of the meaning we have given to the concept of information in section 3 [16]. Indeed, algorithmic information (structural information in this case) is a subjective concept because it refers to a descriptive formalism; following our initial commitment, it will not be discussed in this article.

Let us now consider two broad categories of fundamental information-based interactions. The first comprises cases in which the change in *B* triggered by pattern P_A consists of the appearance of *another pattern* $P_B = P_B(\mathbf{r}, t)$. We call P_B the *image* of P_A and say that in this interaction "*B* has received information from *A*, processed it and transformed it (or encoded it) into a pattern

P_B". A condition is that the correspondence between the two patterns be unique: P_B is triggered by P_A and *only* by P_A (for multiple correspondences and ambiguous responses in neural systems, see further below). If the pattern elicited in *B* is physically identical to that in *A*, we call the process a *replication* or *copy*. If it is only partially identical, the process is a *transcription* or *partial copy*. In the general case, if there is no point-to-point, instant-by-instant or feature-to-feature correspondence between both patterns (the uniqueness of the correspondence remaining, however), we call the process *mapping* (with the image, i.e., pattern P_B, a *map* of P_A). Sometimes we may also call it a *translation* or *transformation*. What is being imaged, mapped, replicated, translated or encoded is information: it is what remains invariant throughout the whole interaction process, thus embodying the uniqueness of the correspondence.

In the second broad category of fundamental information-based interactions the change in *B*, triggered by pattern P_A, is some specific *action* or chain of actions (e.g., the assembly of "molecular machines"—see below). In that case we call the process an *instruction*, and the information mediating the correspondence "pattern→action" an algorithm or *program*.

The processes described in the preceding paragraph represent the possible types of information-based interactions between biological macromolecules. Indeed, the formation of large organic molecules reacting specifically to patterns, not just forces, was the primordial condition for the emergence of life on Earth; the prebiotic ocean-atmosphere system provided the appropriate medium. Chemical reactions are governed by physical interactions in the quantum domain, in which, according to our previous discussion, information as such plays no role—these interactions just happen. But in the process of molecular synthesis large and complex polymer-like macromolecules emerged as *code-carrying templates* whose effect on other molecules in their environment is to bind them through a catalytic process into conglomerates according to patterns represented by the code. Some of these polymers also served as templates for the production of others like themselves—those more efficient in this process would multiply. Replication and natural selection most likely already began in a pre-biotic chemical environment.

Concerning these macromolecules, we can say that beyond a certain degree of complexity, *information as such* begins to play the decisive role in organizing the chemical environment. Because of the role of information as the controlling agent, these molecules can trigger chemical reactions that would be highly improbable to occur naturally under the same environmental conditions and energy sources. In summary, the first "grand moment" in the evolution of life occurred when carbon-based molecules of sufficiently high complexity appeared for which the information expressed in their structural patterns began to take a highly selective control of their interaction with the surrounding medium [17], leading to an explosion of self-organizing complexity that characterizes today's Earth environment.

The information-bearing patterns in all this are sequences of chemical radicals in a polymer, the molecular codes. Coding of information in molecular chains, and the translation of this information into a different kind of molecules, is indeed the mainstay of molecular genetics. But information about *what* is it? For each species of living beings the conditions for survival and reproduction are keyed to a particular kind of environment, the so-called biological niche. For bacteria, viruses, fungi and plants, different niches require different types of metabolism, symbiotic relationships and defenses; for animals, in addition they require different kinds of sensory organs and different strategies for food gathering, social behavior and defense. All this information is encoded in the genetic molecules and determines the structure and, in animals, the instinctive behaviour of the organism.

Organisms indeed are partial images of the environment, with the imaging process involving complicated transformations such as mapping soil structure into expected root network, food stuff into appetence, danger into avoidance, cavity into shelter, etc. [18].

Controlled by the genetic information, this is accomplished at the intracellular level through the construction of "*molecular machines*" or apparatuses that carry out physical-chemical tasks which collectively make up the functioning organism. But there is no "transient" or "quick" storage of environmental information in molecular genetics: it is all the result of a slow process of *evolution* that shapes the informational content through chance mutations and elimination of the unfit. When we say "information on the environment has been stored in the genetic structures of an organism" we are really describing a process of information extraction (example 3 in section 1) and storage "by default" (for lack of a better term). By this we mean that the information content in a molecule such as DNA is not derived from physical laws and deliberately stored, but that it emerges gradually in the course of a long Darwinian process of selective adaptation [19] (see pertinent comment in example 3, section 1). It could be called an "unreflected learning process by trial and error" [10], in which memory storage occurs via the non-proliferation of DNA structures that lead to poorly adapted organisms, and efficient multiplication of the better adapted ones. It is through this evolutionary process that the necessary "accord" between sender and recipient—an essential part of information-based interactions (*Fig. 2*)—is gradually built into the information handling components of an organism. There is no pre-existing information or purpose in adaptive evolution—the latter will happen whenever the circumstances are right. Note that this process is possible only in an organized, non-chaotic environment bearing regularities and features that do not change appreciably over many generations of a species; to function, it calls for certain properties of matter that are only possessed by nucleic acids.

6. NEURAL INFORMATION, MEMORY AND COGNITION

For faster complex information processing in multicellular organisms, and to enable memory storage of current events, a nervous system is necessary, as well as sophisticated sensory and motor systems to couple the organism to the outside world in real time. At the earliest stage of evolution the nervous system merely served as a high-speed information transmission device in which an environmental signal, converted into an electrical pulse in a sensory detector at one end, is conveyed to a muscle fiber at the other end, where it triggers a contraction. Later, neural networks evolved that could analyze and discriminate different types of input signals and send specific orders to different effectors. Finally, memory circuits emerged, that could store relevant information for later use.

There are two basic ways in which information is represented or encoded in the nervous system [20]: (1) a dynamic, transient form given by the specific *spatio-temporal distribution of electrical impulses* in the neural network (analog postsynaptic potentials and standardized action potentials); (2) a static form represented by the spatial distribution of synapses (inter-neuron connections and their efficiencies) or *synaptic architecture* of the neural tissue [21]. Type 1 is a dynamic pattern, changing on a time-scale of tens of milliseconds and involving millions of neurons, that requires a continuous supply of energy to be maintained; type 2 is static and in principle can subsist with no extra supply of energy. In general, even in the simplest neural network, there are several mutually interacting levels at which information is represented or mapped. First of all, there is always an *input level* at which physical input patterns are displayed as sensory signals and then converted into neural patterns (environmental information in the exteroceptive sensory system and information about the physical and chemical state of the organism in the interoceptive system). This is followed by several levels of neural processing in the afferent nervous system and the brain, depending on the complexity of the organism and the task involved.

Concerning information-processing operations in the nervous system, we first consider the process of memory storage and recall in general terms. Let us assume that we have an input (sensory) system S in which the patterns are physical expressions (maps) of sensory stimulation, and at least three consecutive stages of neural information processing which we designate as systems A, B, C, … (some of which may be part of the same neural circuitry). We also assume that patterns in A are of type 1 above (spatio-temporal distribution of neural signals) whereas the patterns at the higher levels B, C, … may also include those of type 2 (changes in the synaptic architecture). When a pattern P_A is activated in A by a pattern P_S in the input channel (such as the physical signal representing the image on the retina of a visually sensed object, or the oscillation pattern of the basilar membrane in response to a musical tone), it triggers a specific pattern P_B in B, P_C in C, etc. If these patterns persist ("resonate") for an interval of time longer than the duration of P_S, we say that P_A, P_B, P_C, … are *short-term* or *working memory* representations or images of P_S. We further assume that there is a *two-way* information-based feedback relationship between the consecutive neural levels A, B, C, such that when, for instance, pattern P_B in B is evoked in some independent way at a later time, the original pattern P_A to which it corresponds, is elicited in A (*Fig. 3*). If this is the case, we say that information on P_A, and hence also on the corresponding sensory input P_S, was *stored in long term memory* (this long-term image would have to be of type 2, i.e., a change in the actual hardware). The reverse feedback process, in which the original pattern P_A , P_B, etc. are being reconstructed, is called a *memory recall*.

A most fundamental neural processing operation is the formation of *associative memory*, to which many different types of more complex memory forms can be related. Suppose that there are two nearly simultaneous input patterns P_S and Q_S (e.g., the visual image of an object and that of its written name), triggering the responses P_A and Q_A, respectively, at the primary neural level A, P_B and Q_B at the secondary level B, and so forth. We now assume that when these input patterns are repeated several times, changes occur in the neural architecture or "hardware" of level B such that (see sketch in *Fig. 4*): (i) a *new* pattern $(PQ)_B$ emerges that is a representation of the simultaneously occurring stimulus *pair P* and Q; (ii) this new pattern appears even if either *only P* or *only Q* is presented as input; (iii) whenever pattern $(PQ)_B$ is independently evoked, *both* P_A and P_B will appear at the lower level A. As a result (*Fig. 5*), a sensory input P_S will trigger an additional response Q_A at level A even *in the absence* of input Q_S , and an input Q_S will trigger P_A (as feedback from level B). This is the essence of what is called an *associative recall* (there is a close equivalence with Fresnel holography here [22]). The formation of a new image of the pair PQ at the secondary level B represents the basic neural *learning process*; the change in hardware (synaptic architecture) that made this possible represents the *long term storage* of information on the correlated inputs P and Q.

If a short-term memory mechanism is operating, P_S and P_Q need not be simultaneous—the new combined pattern can establish itself on the basis of the respective temporarily stored patterns (conditioning). Notice that P_S and Q_S could be inputs from two different sensory modalities (e.g., the visual image of an object and the spoken name of it); or they could be two different parts of the same object (e.g., the face of your dog and the body of your dog). One of the two inputs could be "very simple" compared to the other; in that case one calls it a *key*; a very simple key as input can thus trigger the recall of a very complex image consisting of the superposition of many different component patterns [23]. This means that the replay of a specific neural pattern P_A can be triggered by cues other than the full reenactment of the original sensory input P_S—a *partial* reenactment may suffice to release the full image P_A. This also goes for a partial input from the higher stages, and represents a fundamental property of the mechanism of associative memory recall (see section 6). I would go so far as to say that this represents the most basic algorithm for "intelligent information processing". Today all these processes can be simulated with neural network computer software [24].

When there are many levels or stages of neural information processing, as happens in the brain, many different inputs that belong to correlated environmental objects or events can lead to the formation of one single representation, image or map at one of the uppermost information-processing levels. In turn, because of the two-way coupling between most of the intervening levels, elicitation of that common higher level image can feed back to the lower stages and trigger maps that correspond to one or several individual components of the set of original input images. When that happens, we say that the system has *cognition* of the common properties that characterize the group of individual objects or events, thus defining a specific category for them. Elicitation of the common image by any single input component represents the process of object *recognition* (for instance, elicitation of the common image corresponding to the category "apple" by the smell of a bowl of apples, the sound of the word "apple", a picture of an apple orchard, or an apple pie).

In highly concatenated networks, one and the same input pattern can trigger different images at the highest level, depending on circumstances such as the current state of information processing and previously stored information (think again of the Rorschach test). This points out the fact, mentioned earlier, that a common code must exist between sender and recipient, to ensure a unique, one-to-one correspondence between source pattern and response. In a neural system, the common code between environment and a sensory system's lower stages (*Fig. 3*) evolved during the slow course of evolution. In the associative memory process (*Fig. 5*), on the other hand, the neural network hardware changes during the learning process in real time, and so does the "accord", or common code (think of learning a new language). In genetically prewired neural circuits as those dealing mainly with information from the organism (see section 7), that common code appears during the course of evolution.

The reader must have realized that, according to the preceding discussion, there are two and *only two* distinct types of natural (not made) information systems: biomolecular and neural information systems. Bacteria, viruses, cells and the entire flora are governed by information-based interactions of the biomolecular type; their responses to physical-chemical conditions of the environment are ultimately controlled by molecular machines and devices manufactured and operated according to blueprints that evolved in the long-term past. In plants they are integrated throughout the organism by a chemical communications network. It is not our intention to describe in this article the actual machinery that carries out the operations of biomolecular information transfer, replication, transcription, translation and instructtion for the construction and operation of these molecular machines. Let it just be said that much progress has been made in recent years in their understanding. The mechanism of protein manufacture in ribosomes is now pretty well understood, and the detailed mechanism for the transcription of parts of the genetic code in the synthesis of messenger-RNA by an enzime called RNA polymerase II has recently been unraveled [25] (example 5 in section 1).

Concerning neural networks, the human brain is the most complex of all—as a matter of fact, it is the most complex yet at the same time most organized system in the Universe as we know it. Again, it is not our purpose to describe its "hardware"—except for a few aspects important for our discussion. The brain (see next section) evolved in a way quite different from the development of any other organ. Separate layers appeared, with distinct functions, *overgrowing* the older structures but not replacing them, thus preserving "older functions"—the hard-wired memories or *instincts* represented by synaptic patterns "frozen in time" (type 2 storage) that a species has acquired during evolution. The outermost layer, or neocortex, executes all higher order cognitive operations, based on information acquired in real time during the life of the organism. Subcortical structures such as the "limbic system" are phylogenetically old parts of the brain carrying information acquired during the evolution of the species. We thus have a system in which two basic "modes" of information coexist and cooperate. As we shall see in the next section, this "cortico-limbic cooperation" (we should say, more precisely, this "coherent, cooperative mode of interaction") leads to *consciousness* in higher mammals.

In the *human* brain, there is a third mode—not based on any "new" network hardware but representing a new information processing *stage* at which the neural activity of both lower processing modes, the instinctive and the cognitive one, are combined and mapped together in tight coherence and synchrony into a "representation of representations", giving rise to the awareness of one's own brain function and *self-consciousness*. In summary, in an animal brain we have information from the long term past of the species acquired during evolution, originally stored in the genes and expressed in prewired circuits, and information on the ontological past and the present acquired through the senses and actively stored in the long-term memory of the brain. Humans, as we shall discuss in more detail in section 8, have the notion of future time and the ability of making long-term predictions—in informational terms, their brains also handle *information on the future*.

7. REPRESENTATION OF INFORMATION IN THE BRAIN [26]

How is information actually represented in the brain? In this discussion we are mainly interested in the neural representation of higher-level cognitive information—maps that involve many processing stages in the brain and hundreds of 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 data. Concerning the dynamic mode of encoding neural information (section 5), there is now a convincing ensemble of data that show that this encoding indeed appears in certain areas of the cerebral cortex in the form of a *specific spatio-temporal distribution of neural impulses*. For instance, the mental representation or image of an object (visual, acoustic, olfactory or tactile) appears in certain areas of the cerebral cortex in the form of a specific distribution of electrical signals *that is in one-to-one correspondence with the specific features sensed* during the perception of this object. According to this result, "cognition" (see section 5) implies the occurrence of a specific neural activity (neural activity pattern, type 1) that is in one-to-one correspondence with the object or category of objects that is being recognized, remembered, imagined or dreamed about.

We could describe this pattern of neural activity mathematically with a function $\psi(\boldsymbol{r}, t)$, representing the neural firing rate at point $\mathbf{r}$ in the brain tissue at the time t (a specific form of pattern $P(\mathbf{r}, t)$ described in section 5). Unfortunately, this function would be *ultra-discontinuous*: two neighboring neurons may participate in totally different tasks, and thus fire causally unrelated electrical impulses. We could use a vector function $\psi = [f_1(t), f_2(t), \ldots.. f_m(t), \ldots.. f_n(t)]$, where $f_m(t)$ is the firing rate of the m^{th} neuron in an ensemble of n cells. But each cognitive task involves millions if not billions of neurons. In any case, if we *could* determine ψ, there would be *one specific* vector function for each mental image; a typical imaging event would last only 50-200 ms (unless kept active in short-term memory—section 6). 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 ⇒ apple tree ⇒ apple, all having ψ_{apple} in common). Recent experiments on categorical percep strongly confirm the main aspects of this description [27]. Even if it cannot be used in any mathematically tractable form, we must emphasize that ψ is a *physical* quantity—an n-dimensional distribution function. The generation and transmission of electrical impulses in a neuron demand a very large energy supply compared to other functions of the cell; it is precisely the localized increase in oxygen consumption that is detected in fMRI and PET tomographies; their images thus represent an "average ψ" over millions of neurons (about 1 mm resolution)—an algorithm not unlike the definition of macroscopic variables like density and temperature in thermodynamics.

As anticipated in section 5, 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.

On the other hand, during a given learning phase the so-called association areas of the cortex and the pre-frontal lobes display patterns of neural activity triggered by a series of specific stimulus constellations arriving simultaneously through different sensory channels (for instance vision and hearing, when a dog is taught to go to its doghouse). As this happens repeatedly, changes occur in the connections between neurons, the synapses [28]. In particular, new synapses are established between neurons that fire simultaneously during the learning process in such a way that *after* the learning procedure, the stimulus constellation in only one input channel (the conditioned stimulus or "key", section 5) will suffice to trigger the full pattern of neural activity specific to the entire original event (associative recall, *Fig. 4*). The above examples show that one must recognize the neural activity distribution (pattern of type 1, section 5) as the fundamental "physical quantity" that represents real-time information in the working brain. Long-term memory is represented as static type 2 patterns by the distribution of synapses in the neural network, which is continuously being modified as new information is acquired and old one deteriorates.

Let us examine in more detail how the various levels or stages of information processing and representation operate in the brain. For that purpose, we examine the sketch of *Fig. 6* depicting an oversimplified view of the visual system [29]. The neural circuitry in the periphery and afferent pathways up to and including the so-called primary sensory receiving area of the cortex 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, specific chemical radicals in smell; processes of type 3, section 1). 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; or in the auditory system, sorting out those resonance regions on the basilar membrane that belong to one musical tone [29]). 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 (processes of type 5). The recognition process can be "automatic" by associative recall, or require a further analysis of the full sensory input in the prefrontal cortex. As one moves up the stages of *Fig. 6*, the information processing becomes less automatic and more centrally controlled; in particular, more *motivation-controlled* actions and decisions are necessary, and increasingly the *previously stored* (learned) information will influence the outcome.

The paths shown in *Fig. 6* can also operate *in reverse* (section 5 and *Fig. 3*): there is experimental evidence now that the memory recall (or, in humans, also the imagination) of, say, a given object, triggers neural activity distributions at the various lower levels that would occur if the object was actually seen. In the auditory system, "internal hearing" operates on this basis: the imagination of a melody or the recall of an entire musical piece is the result of the activation, or "replay", of neural activity triggered somewhere in the prefrontal cortical areas (depending on the specific recall process), which then feed the appropriate information back down the line of the auditory processing stages creating sensory sensations without any sound whatsoever entering our ears.

The motivational control deserves special attention. As mentioned above, one of the lower, phylogenetically much older parts of the vertebrate brain is the so-called *limbic system*, which works on the basis of genetic information represented in mostly prewired (inherited) networks. We use this term as a short-hand for a system that comprises several deep structures including the amygdala, the ventral tegmental 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), 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. 7*). 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*. Many control and communication operations within the limbic system work on a neurochemical basis [20] (dopamine, serotonin, etc.). Emotion (controlled by the deep structures) and motivation (controlled by the anterior cingulate) 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 [30]. The limbic system constantly challenges the brain to find solutions to alternatives, to probe the environment, to overcome aversion and difficulty in the desire to achieve a goal, and to perform certain actions even if not needed physiologically at that moment (e.g., animal play for the purpose of training in skilled movement, listening to music by humans). 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. As we shall see below, this interplay gives rise to consciousness in higher animals.

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 hurt? Why do we feel pleasure scratching a mosquito bite, eating chocolate or having sex? How would we program similar reactions into a robot? [31] 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 or the species, respectively. Since such anticipation comes *before* any actual harm or benefit could arise, it helps guide the organism's response into the 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 unexpected, genetically unprogrammable, circumstances. Plants cannot respond quickly and plants do not exhibit emotions; their defenses (spines, poisons) or insect-attracting charms (colors, scents) developed through the slow process of evolution (section 5). Reactions of non-vertebrate animals are neural-controlled but "automatic"—there is no interplay between two distinct neural processing systems and there are no feelings guiding the behavioral response.

8. INFORMATION IN CONSCIOUSNESS AND SELF-CONSCIOUSNESS

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 that, even if many well-defined processing levels are involved, there is monolithic *coherence* and *synchronism* [32], and *consistent specificity* with each cognitive act, feeling or motor output. Borrowing a concept from condensed matter physics, we say that this represents a cooperative action creating unity in an ultra-complex system that otherwise would be overwhelmed with irrelevant information and noise. While no doubt ψ involves many "subservient" programs or "subroutines" which never reach consciousness (think of driving and using a cell phone at the same time!) and trigger or control very specific information-processing operations in highly localized "modules" [33], *there is only one "main program"* which leads to unity of perception and behavior, and represents the *basic conscious state* of the animal or human brain (close but not equal to what has been called "core consciousness" [34]). There are operational and evolutionary reasons for having a single state of consciousness. First of all, if several "main programs" with different emotional states 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. Indeed, there is no spatial confinement for ψ; rather, it is in a sort of "functional confinement" imposed by the demand for coherence. 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. 7*).

This brings up a new element in information-processing, namely that of *decision-making*. It becomes relevant when one and the same input elicits in a complex neural network several competing images or responses, forcing the "main program" to choose one and suppress all others. This choice among alternatives demands rapid interaction with many different brain regions that hold previously acquired information—yet another argument for the mode of "monolithic coherence" of brain function. It is important to note that in biomolecular information systems there is no real decision-making—alternatives, if offered, are "chosen" on a statistical, probabilistic basis.

Turning to the *human* brain, from the neurophysiological and neuroarchitectonic points of view it is not particularly different from the brain of a primate like the chimpanzee. It does have a cortex with more neurons (between 10^{10} and 10^{11}) 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 (about 10^{14}). 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 [35] 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 as images or represent-tations (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* [11].

This had vast consequences in human 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 [11]), to a quantitative concept of elapsed time, and to the awareness of future time [36]. Along came the possibility of long-term prediction and planning ("information about the future" [37], [5], i.e., representation of things that have not yet happened, in contrast to anticipation of events based on genetically acquired information), the postponement of behavioral goals and, more generally, the capacity to overrule the dictates of the limbic system (think of a diet) and also to willfully stimulate the limbic system, without external input (think of sexual self-arousal). The body started serving the brain instead of the other way around. Mental images and emotional feelings could thus be created that had no relationship with momentary sensory input—the human brain can go "off-line", as expressed by Bickerton [38]. Abstract thinking and artistic creativity began; it also brought the development of beliefs (unverifiable long-term predictions) and values (genetically untested priorities).

Concurrently came the ability to encode complex mental images into simple acoustic signals (keys—see section 5) 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, and the human thinking process began to be influenced and sometimes controlled by the language networks. Finally, though much later in human evolution, there came the deliberate storage of information in the environment; this externalization of memory led to the documentation of feelings and events through written language, musical scores, visual artistic expression, and science—to human culture as such. And it was only very recently that human beings started creating inanimate systems capable of *processing* information and entertaining *information-based* interactions with the environment—the human-made systems that we have tried painstakingly to avoid in this paper.

At the root of all this is the human capability of making "*one representation of all representations*". There is a higher-order level of representation in the human brain which has cognizance of consciousness and can manipulate independently the primary representation ψ of current brain activity. It can construct an image of the *act* of forming first-order images of environment and organism, as well as of the reactions to them. 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 [32]; again, coherence of brain function assures the latter. The capacity of retrieving and manipulating information from the memory without any external or somatic trigger; the feeling of being able to observe and control one's own brain function; the feeling of "being just one" [39]; and the capacity of making independent decisions [40] and overrule (or independently stimulate) the limbic response, collectively represent what we call *human self-consciousness* (close but not equal to "extended consciousness" [34]). In other words, self-consciousness is far more than a passive feeling—it represents the capability of some very unique information processing *actions*.

There is no need to assume the existence of a separate neural network, some highly localized "seat" of consciousness or some mysterious immaterial entity like "mind" or "soul" for this highest-order representation. There is enough information-handling space in the cortical and subcortical networks that can be shared with the primary representtation ψ (except, perhaps, that there may be a need for a greater involvement of the prefrontal cortex and that the language networks may participate in an important way [41], although this is disputed by some [34]).

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. A useful metaphor for this is the following: consciousness is "watching the movie that is running in the brain"; self-consciousness is the capacity of human beings "to splice and edit that movie". Finally, the fact that consciousness involves both feelings and cognitive functions clearly shows that it emerges from a tight and coordinated interplay between the neo-cortical and the old limbic structures of the brain. In fact, to function, this highest level representation *must* be spread over the entire brain: the historical "homunculus" is not a cluster of just a few "master neurons" but consists of about 10^{11} interacting elements—it is the "monolithic coherence and synchronism" of their cooperative function what makes you the one and only *you*!

ACKNOWLEDGMENT

I am grateful to Professor Valentino Braitenberg of the Max Planck Institute for Biological Cybernetics for enlightening discussions.

NOTES AND REFERENCES

[1] See for instance: W. H. Zurek, editor, *Complexity, Entropy and the Physics of Information* (Addison-Wesley Publ. Co., New York, 1990).

[2] P. Davies, “Physics and Life”, in J. Chela-Flores, T.Owen and F. Raulin, eds., *The First Steps of Life in the Universe* (Kluwer Acad. publ., Dordrecht, The Netherlands, 1990), pp. 11-20.

[3] E. J. Chaisson, *Cosmic Evolution: the Rise of Complexity in Nature*. (Harvard University Press, Cambridge Mass. , 2001).

[4] See for instance: E. Mach, *Science of Mechanics*, (1893; Translation: The Open Court Publ. Co., La Salle, Illinois, 1942). In Ernst Mach’s formulation of the interaction between two classical mass points isolated from the rest of the Universe, one establishes that there is a linear relationship between the acceleration vectors: $m_1\boldsymbol{a}_1+m_2\boldsymbol{a}_2 = 0$, valid for many classical interaction mechanisms (gravitational, electrostatic, a massless compressed spring pushing the mass points apart, etc.). The measurable quantities are the acceleration vectors; the constant ratio of their moduli a_1/a_2 is defined as the “inertial mass of body 2 in units of the mass of body 1”. The term $\boldsymbol{f} = m\boldsymbol{a}$ is called “force on mass m” in the interaction and interpreted as the “agent” responsible for change (acceleration). Newton’s laws become corollaries in this formulation.

[5] J. G. Roederer, “Information, life and brains”, in J. Chela-Flores, G. Lemarchand and J. Oró, eds., *Astrobiology* (Kluwer Acad, Publ., Dordrecht, The Netherlands, 2000), pp. 179-194.

[6] The satellite’s motion is derived by integrating the equation of motion $m\ddot{\mathbf{r}}(t) = \int \rho(\mathbf{r}')(\mathbf{r}'-\mathbf{r})/|\mathbf{r}-\mathbf{r}'|^3 d\mathbf{r}'^3$ (the motion of the asteroid remains essentially unaffected by this interaction because of its much larger mass).

[7] In this article we shall distinguish clearly between “signals” and “messages” (see example 5 in the text): thus defined, a signal has neither purpose nor any information at its origin, it only becomes information when it is detected and used (or deliberately discarded) by an organism.

[8] Think of the “A-ness” conveyed in visual interaction with the symbols *a*, *A*, α , $\aleph$, or the “three-ness” conveyed by any set containing three objects. Note a certain analogy between our definition of information with Cantor’s definition of integer number: “that which all coordinable sets have in common”.

[9] We shall not attempt here a formal definition of “purpose”. This concept has too many subjective connotations, so whenever the term appears in this article, it should be taken with great care.

[10] B.-O. Küppers, *Information and the Origin of Life* (The MIT Press, Cambridge Mass. 1990).

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

[12] J. Bricmont, J, ”Science of chaos or chaos in science?”, Physicalia Mag. **17**, 159-208 (1995).

[13] Consider a vessel of volume V, divided into two parts V_1 and V_2, separated by a partition. Each section is filled with a different ideal gas at the same pressure and temperature. When the partition is opened, part of the gas in V_1 will diffuse isothermally and irreversibly into V_2 and vice versa. If the number of molecules of the two gases is n_1 and n_2, respectively, the increase in entropy, after mixing equilibrium is reached, will be $\Delta S = n_1 \text{ k} \ln (n/n_1) + n_2 \text{ k} \ln (n/n_2)$ (k = Boltzmann constant). This value is minimum for the case in which the volume V is divided into two halves, in which case $\Delta S = n \text{ k} \ln 2$ (where $n = n_1 + n_2$). Gibbs’ paradox is the following: consider the case in which the molecules initially in V_1 are *indistinguishable* from those in V_2. Opening the partition changes nothing from the macroscopic point of view; therefore, the entropy obviously must remain constant. Yet, from the microscopic point of view, we *know* that molecules from V_1 will diffuse into V_2, and vice versa. So, if the previous calculation is correct, the entropy should increase!

We can explain the dilemma using our definition of information (section 3). By saying that “the molecules of V_1 (or V_2) are doing this or that after the removal of the partition”, we are implying that we consider these molecules *tagged*, making them distinguishable from each other (in our mind) during the process [5] (e.g., a cousin of Maxwell’s Demon has painted a tiny zero-energy marker on each molecule before mixing starts!). The information carried by each tag has the value of one bit (for the case $V_1 = V_2$; if $V_1 < V_2$ the amount of information would be larger for a molecule from the smaller volume). It is well known from traditional information theory that adding information to a system *decreases* its entropy by $\Delta s = -k \ln 2$ *per bit*. Taking this into account, the decrease of total entropy in the "mental" tagging process ($n\ \Delta s$) is compensated exactly by the increase of entropy (ΔS) in the "mental" diffusion and mixing process! Note in this example that the concept of information only has to do with the "thinker"—it plays no role in the dynamics of the physical system per se.

[14] Note that the bird of a given species will always build the *same* nest (slightly modified to adapt it to the immediate environment) whereas a human can develop an unlimited variety of blueprints for a house and change them at will before the house is actually built.

[15] Animals, of course, also leave "imprints" of information in the environment for some future use (example 4, section 1). Yet there is a crucial difference. If we were to suddenly destroy all animal imprints (tracks, nests, food reserves, etc.), animal life would still continue essentially as it was today: the key information for survival and preservation is genetically stored in the organism, and relics of animal "cultures" are not used by later generations as a source of information. But as described in many science-fiction novels, if we were to destroy at once all human imprints (our "externalized" memory storage), civilization as we know it would disappear and future generations would be left in a state similar to where humanity was tens of thousands of years ago.

[16] For instance, I could look at a complicated Chinese character whose algorithmic information content as a graphic symbol is high—but for me as a recipient *reader* the information content would be zero!

[17] No doubt that we are dealing here with a "threshold" in molecular size and complexity in which a transition takes place to the capacity of information-based interactions. This domain threshold is conceptually similar to the boundary that separates the domains of quantum and classical behavior of matter, the region of wave function de-coherence—and may even have something to do with it.

[18] V. Braitenberg, "Remarks on the semantics of 'information' ", preprint (Max Planck Institut für Biologische Kybernetik, Tübingen, 2000).

[19] The fast adaptive capability of viruses and immune system response really deserve a separate consideration in this discussion.

[20] There is a third information system—the chemical neurotransmitters—which I will not discuss in this article, because its function is mainly that of information transmission and modulation of global synaptic function (a sort of neural "volume control").

[21] At an individual neuron level, information is encoded transiently in the form of time sequence of action potentials, and in a long-term form in the spines and ramifications of the axon and corresponding synapses.

[22] First discussed in detail in: K. Pribram, *Languages of the Brain* (Prentice Hall, Inc. Englewood Cliffs, New Jersey, 1971).

[23] Note that, in terms of algorithmic information, knowledge of the *functional* $P_A(P_B)$ would allow a great reduction in the information content of the complex pattern P_A (data compression).

[24] T. Kohonen, *Self-Organization and Associative Memory* (Springer Verlag, Berlin, 1988).

[25] See for instance: A. Klug, "A marvelous machine for making messages", Science, **292**, 1844-1845 (2001), and articles described therein.

[26] Excerpts from J. G. Roederer "On the representation of information in life, brains and consciousness", *Lecture notes*, *Borsellino College on Neurophsyics* (The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy, 2001), 13 pp.

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

[28] See for instance: A. Matus: "Actin-based plasticity in dendritic spines", Science, **290**, 754-758 (2000).

[29] For a similar discussion of the auditory system, see J. G. Roederer, *The Physics and Psychophysics of Music* (Springer-Verlag, New York, Berlin, Heidelberg, 1995).

[30] The functions of the limbic system are sometimes referred to as “the four F's”: **F**eeding, **F**ighting, **F**leeing and … **R**eproducing!

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

[32] Chr. von der Marlsburg, “The coherence definition of consciousness”, in M. Ito, Y. Miyashita and E. T. Rolls, eds., *Cognition, Computation and Consciousness* (Oxford University Press, Oxford, New York and Tokyo, 1997), pp. 193-204.

[33] For some recent examples, see P. E. Downing, Y. Jiang, M. Shuman and N. Kanwisher, “A cortical area selective for visual processing of the human body”, Science **293**, 2470-2473 (2001); and J. V. Haxby, M. I. Gobbini, M. L. Furey, Al Ishai, J. L. Schouten and P. Pietrini, “Distributed and overlapping representations of faces and objects in ventral temporal cortex”, Science, **293**, 2425-2430 (2001).

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

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

[36] 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 brain scientists.

[37] E. Squires, *Conscious Mind in the Physical World* (Adam Hilger, Bristol, New York, 1990).

[38] D. Bickerton, *Language and Human Behavior* (Univ. of Washington Press, 1995).

[39] Even in the most acute cases of multiple personality disorder there is only *one* of the personalities in evidence at any given time.

[40] There is increasing evidence this is only a feeling: experiments show that by the time we think we have made a free will decision, our brain has already preplanned all pertinent steps for us!

[41] This does not mean that one always thinks in words.

August 2002

FIGURES

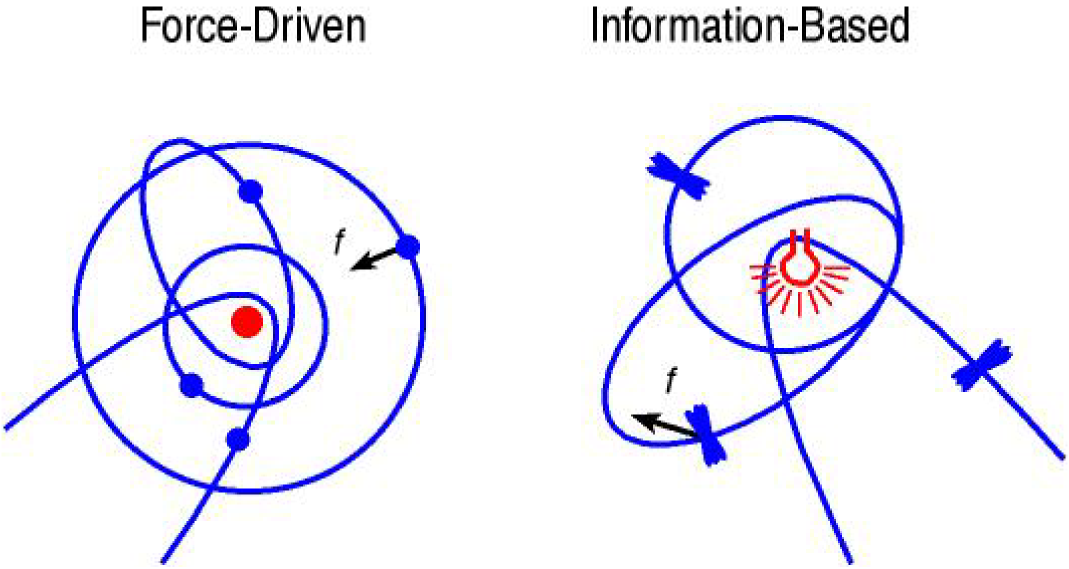


Fig. 1: Examples of the two main types of interactions: force-field driven (satellites in orbit around a central body, left), and information-based (insects "in orbit" around a light source, right). Information and information-processing play no role in the former, whereas in the latter we have the chain light emission → pattern detection → pattern analysis → muscle activation, in which neither force nor energy but information is the controlling agent throughout.

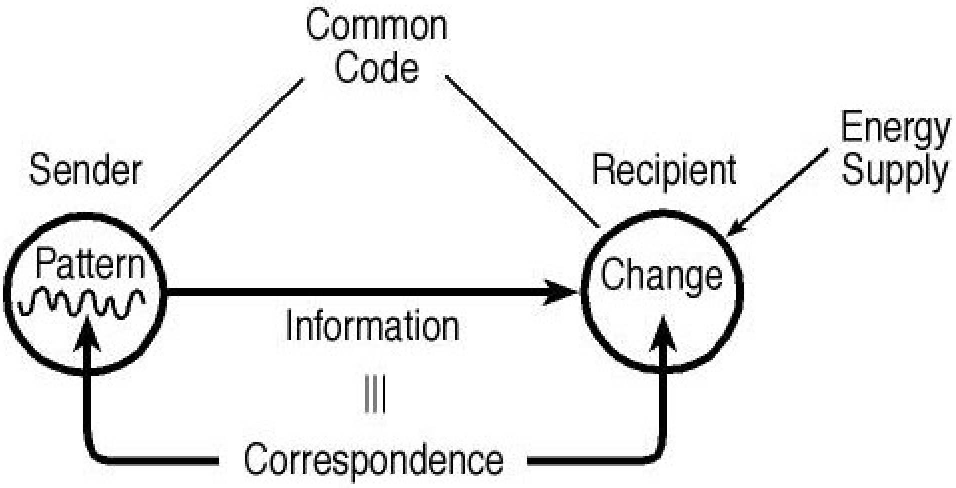


Fig. 2: In an information-based interaction a one-to-one correspondence is established between a pattern in the "sender" and a specific change (e.g., a pattern or an action) in another complex body, the "recipient". Information is the agent that represents this correspondence. The pattern could be a given spatial sequence of objects (e.g., chemical radicals in a molecule), a temporal sequence (e.g., the phonemes of speech), or a spatio-temporal distribution of events (e.g., electrical impulses in a given region of the brain). A natural (not artificially made) information-based interaction must either emerge through evolution or be developed in a learning process, because it requires a common code at both ends that could not appear by chance.

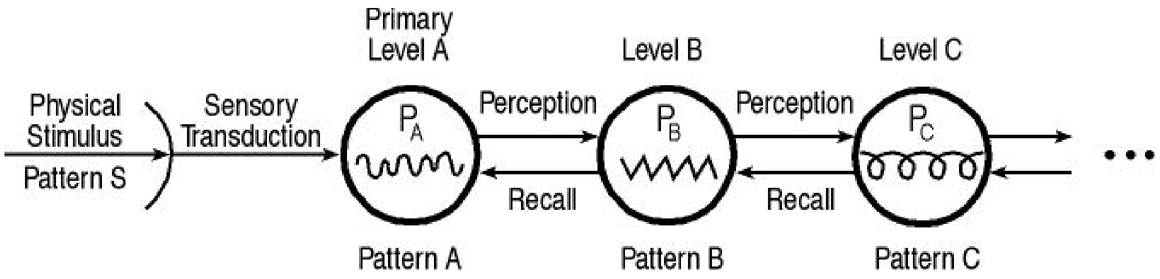
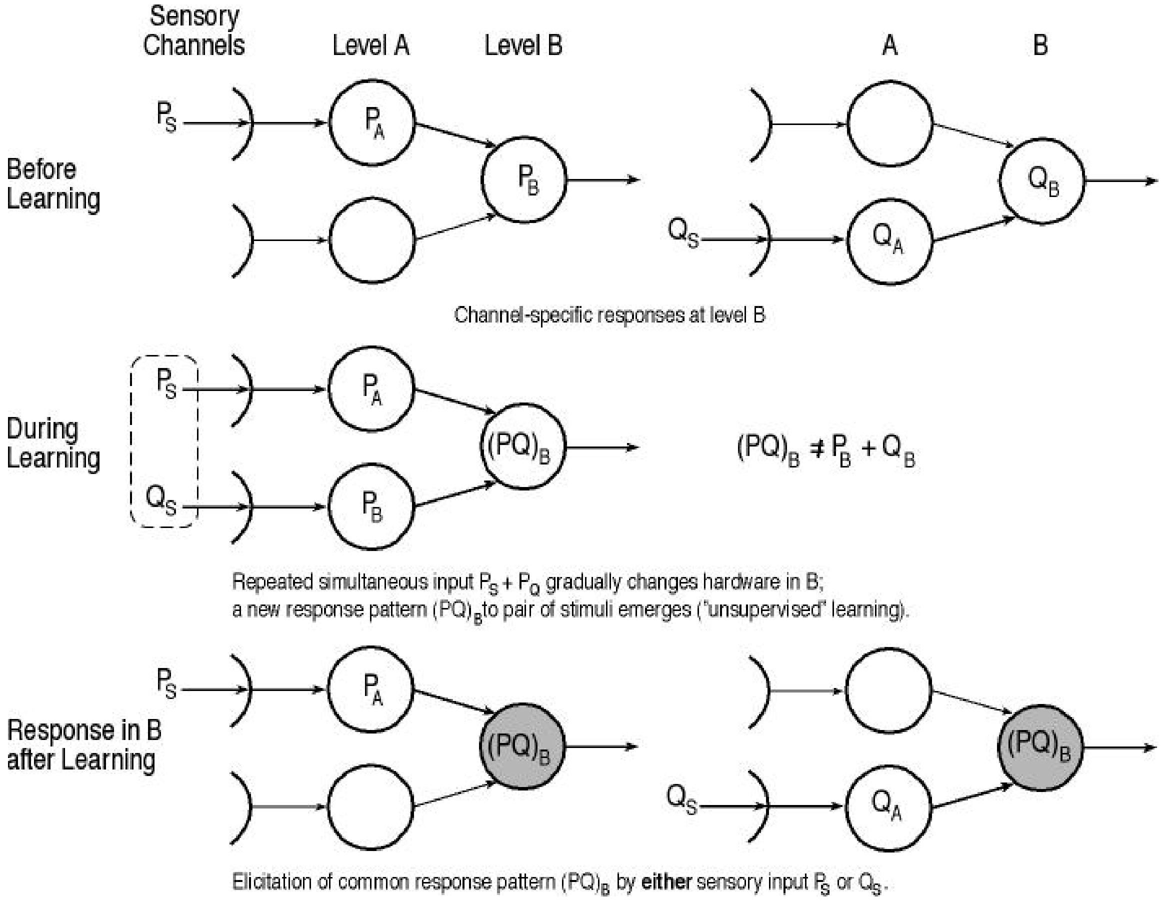


Fig. 3: Consecutive levels for the processing and representation of information in the neural system. The input could come either from the environment or the organism itself. In the process of perception, information is mapped at various higher stages in the form of a specific spatio-temporal distribution of neural activity. Persistence of this activity seconds or minutes after the input signal represents the short-term or "working" memory. Retrieval of information stored in long-term memory requires re-elicitation of the pertinent specific distribution of neural activity at lower levels in a reverse imaging process.



Channel-specific responses at level B

Repeated simultaneous input $P_S + P_Q$ gradually changes hardware in B;
a new response pattern $(PQ)_B$ to pair of stimuli emerges ("unsupervised" learning).

Elicitation of common response pattern $(PQ)_B$ by **either** sensory input P_S or Q_S.

Fig. 4: Sketch of the basic mechanism of learning and associative memory in the neural system. The change in "hardware" necessary to elicit the new, combined pattern (PQ_B) consists of a change in the spatial configuration and efficacy of synapses (neural "plasticity"). This new spatial configuration embodies the representation of information stored in long-term memory. It also represents a change in the common code (*Fig.2*).

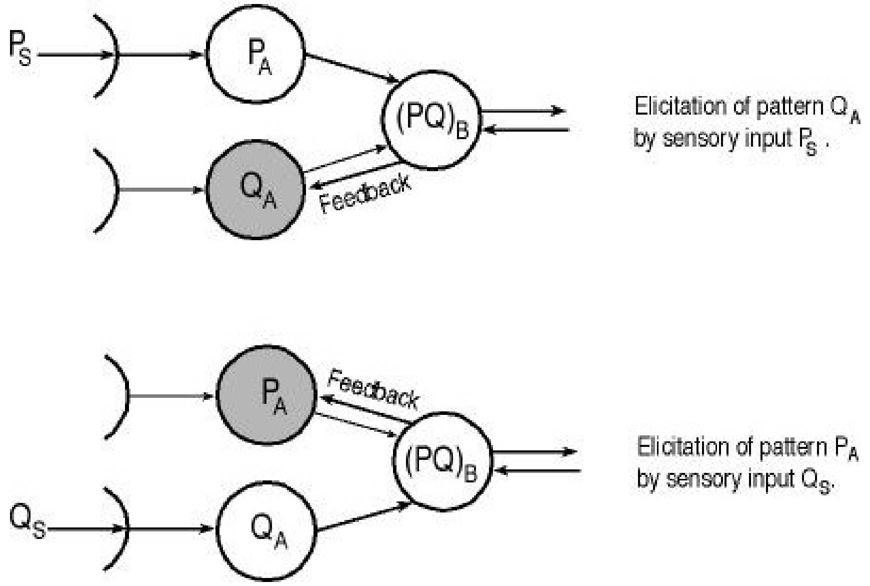

P_S
P_A
$(PQ)_B$
Q_A
Feedback
Elicitation of pattern Q_A by sensory input P_S .
P_A
Feedback
$(PQ)_B$
Q_S
Q_A
Elicitation of pattern P_A by sensory input Q_S.

Fig. 5: Sketch of the associative recall process. The inputs could come from two different sensory modalities, or represent different parts of the same object. One could be a very complex pattern, the other a very simple one, which then would be called a "key" for the recall process (e.g., the visual image caused by a large musical instrument, and its name "piano"). The inputs S_P and S_Q need not be sensory; they could come as feedback from other, higher, stages of neural information processing.

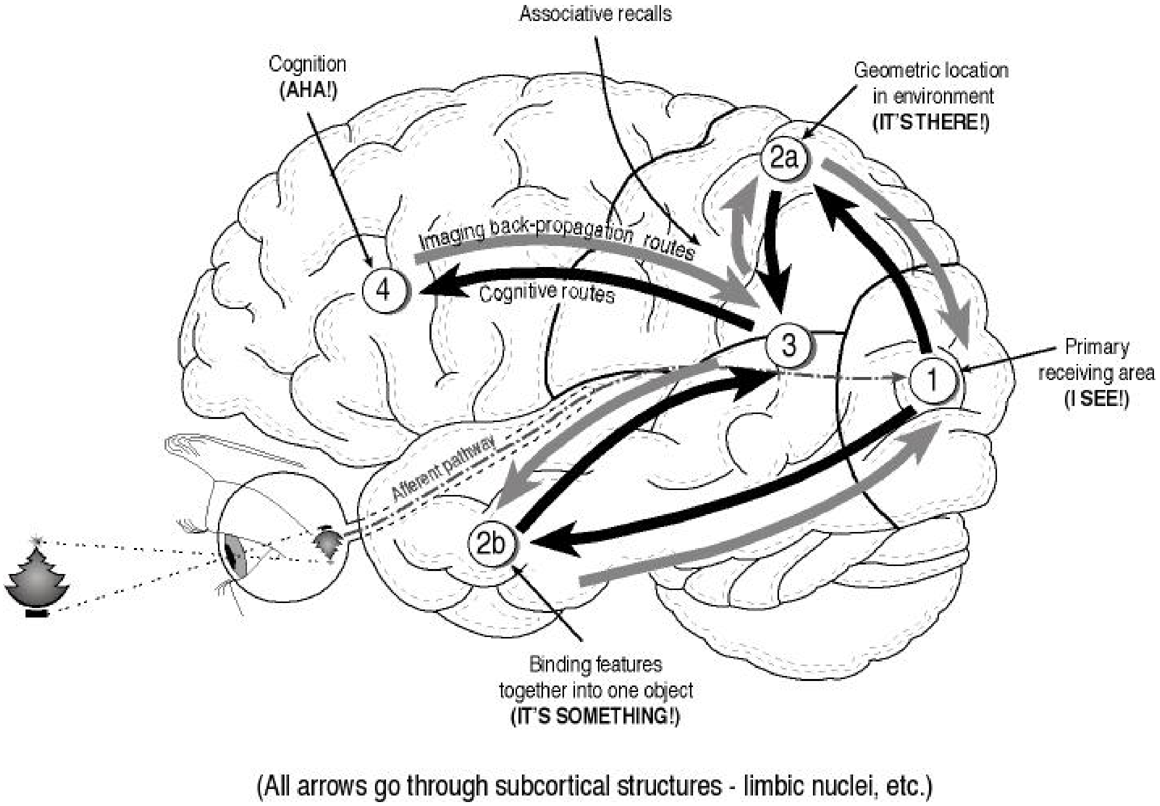


Fig. 6: Ascending information routes and processing levels for visual information. The routing through lower, subcortical levels (not shown) checks on current subjective relevance of the information on its way to the prefrontal cortex. Only humans can override the output of this stage. Through feedback pathways, the imagination of a given object triggers neural activity distributions at lower levels that would occur if the object was actually perceived by the eye.

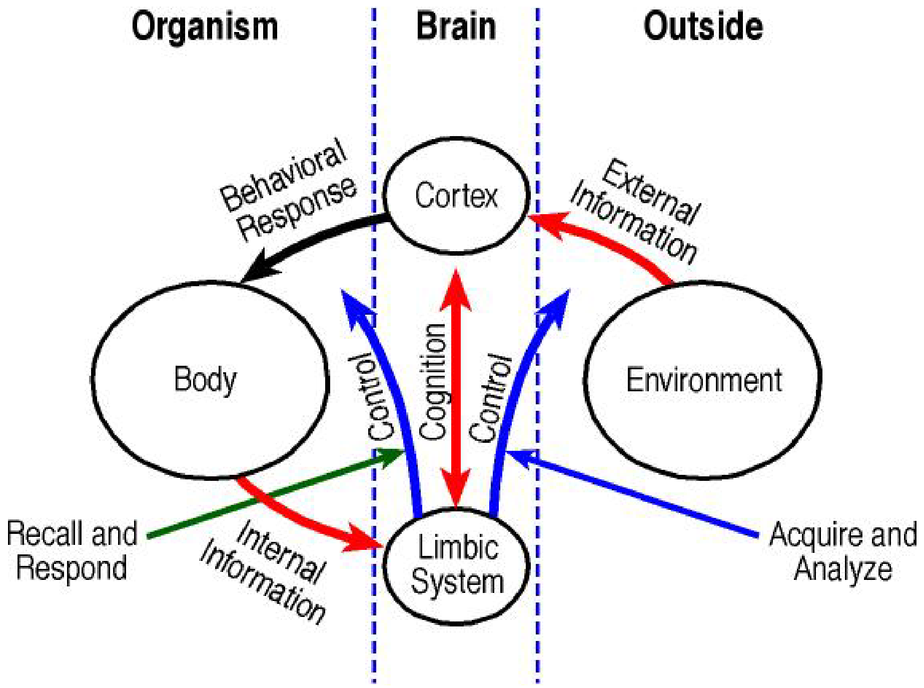


Fig. *7*: Basic functions of brain regions collectively known as the limbic system—the brain's "police station" that assures cognitive functions and behavioral response that are most beneficial for the organism and the propagation of the species. The limbic system controls emotion and communicates interactively with higher processing levels of the cortex, particularly the prefrontal regions, relating everything the brain perceives and plans to the needs of the organism and vice versa. In higher species, the coherent operational mode of the cortico-limbic interplay gives rise to consciousness. The human brain is able to map and coordinate this interplay at an even higher level, which leads to self-consciousness.