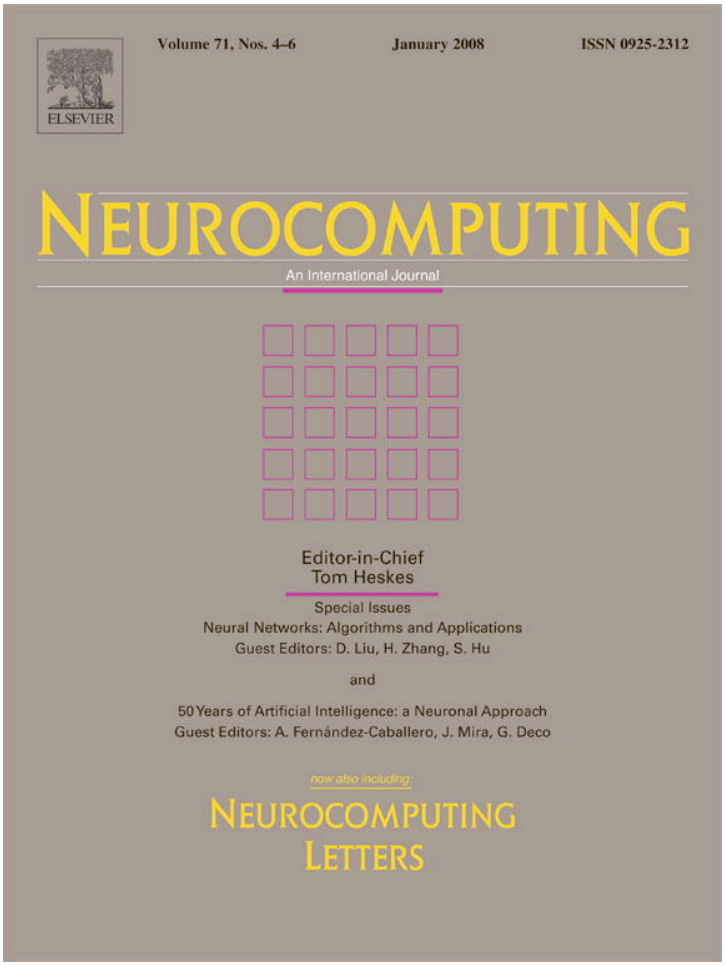
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 71 (2008) 681–692
NEUROCOMPUTING
www.elsevier.com/locate/neucom

On the nature of neural information: A critique of the received view 50 years later

Xabier Barandiaran*, Alvaro Moreno

[a]Department of Logic and Philosophy of Science, University of the Basque Country, Post Box 1249, 20080 San Sebastian-Donostia, Spain

Available online 12 November 2007

Abstract

We offer a critical review of the concept of neural information, as received within mainstream neuroscience from Artificial Intelligence. This conception of information is constructed as a conditional probability of a stimulus given a certain neural activation, a correlation that cannot be accessed by the organism and fails to explain its causal organization. We reconstruct an alternative conception of neural information: a pattern of signals that is selected by the organism (as an autonomous system) to contribute to its self-maintenance in virtue of its correlation with external conditions, a correlation that might further be evaluated by the very system.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Philosophy of neuroscience; Neural information; Autonomous systems; Neurodynamic self-organization; Embodiment; Error detection

1. The problems of the inherited conception of information in neuroscience

The concept of information is, without doubt, one of the basic pillars of computer science, Artificial Intelligence (AI) and cognitive neuroscience. The conceptual and methodological feedback between AI and cognitive neuroscience has been continuous (especially as regards the concept of information) ever since the advent of cybernetics and the development of AI from the 1950s onwards. In fact, neuroscientific literature is full of terms which go beyond the descriptive molecular or biochemical framework, with concepts such as *signal*, *code*, *information* or *content* being introduced since researchers first began to talk about nervous impulses as vehicles for messages in the 1920s and 1930s [1]. However, it was not until the development of Shannon's concept of information and, above all, until the generalized use of the concept of representation and information in the field of AI (at the end of the 1950s), that the use of the term information became common in neuroscientific literature. There is to remind, however, that Shannon's work on information, together with its many developments in computational-algorithmic aspects [14,30], remains silent about how to specify the semantics of a given chain of signals (binary or otherwise). It is precisely this feature of the mathematical theory of information that has permitted it to be widely applied. Yet the (implicit or explicit) reference to semantic information is constant and of fundamental importance when applied to cognitive and neural processes. However, in Victor's words: "the Shannon theory does not attempt to describe the relationship between a sensory or motor domain and neural activity (i.e., the nature of the neural representation) but merely provides an index of how faithful this representation is" [45].

Since the 1950s, Shannon's mathematical theory of information received a great interest among many authors—especially among philosophers—that tried to incorporate in it the semantic dimension. It is worth to mention the early attempt by Carnap and Bar-Hillel [13] to include the semantic dimension within Shannon's frame, and most recently, Dretske's [17,18] influential work. However, other approaches were also explored. To mention probably the most important, in the 1960s, MacKay [34] formulated a quantitative theory of semantic information alternative to Shannon's theory. However, 50 years later, some of the most basic characteristics of the concept of semantic information (semantics understood as intentionality

*Corresponding author.
E-mail addresses: xabier@barandiaran.net (X. Barandiaran), ylpmobea@sf.ehu.es (A. Moreno).
URL: http://www.ehu.es/ias-research/barandiaran (X. Barandiaran).

0925-2312/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2007.09.014

Author's personal copy

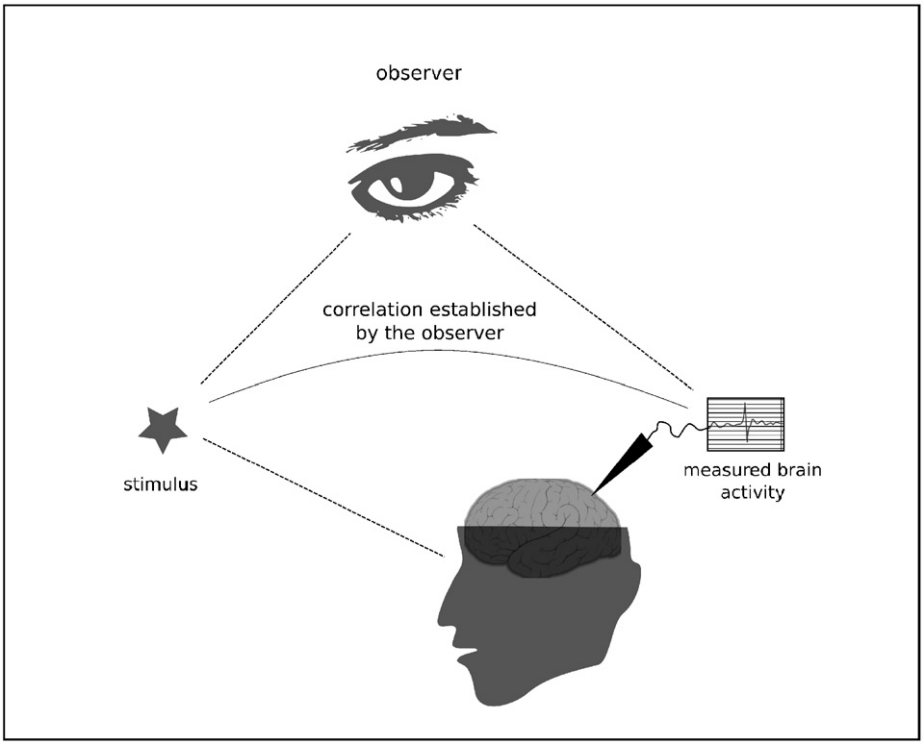


Fig. 1. The most widespread accounts of information in neuroscientific research are based on a correlation between a given stimulus and the activity of certain brain areas. However, this correlation is only accessible to the observer scientists and only for her can the measured activity bear semantic information about external stimuli. Copyright 2007 Xabier Barandiaran. This figure is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/ (includes images by Lumen Virtual Studio and by JimHutchins with GPL and GFDL licence, respectively and published in Wikipedia.org (permission is granted by their respective licences)).

or representational content) still generate problems that continue to have no agreed solution within the computationalist and representationalist paradigms, common both to neuroscience and classic AI. This key problem has taken different philosophical expressions such as the Chinese room argument ([41]), the symbol grounding problem [24], or the problem of intentionality and mental content.[1]

Today, the use of the terms information, code, content, meaning, etc. in theoretical neuroscience is ubiquitous and generally associated with the correlation between a given stimulus and an action potential (or a set of action potentials) in a specific area of the nervous system (NS hereafter). This causal correlation is generally expressed as a conditional probability of the occurrence of a stimulus, given a measure of specific combination of impulses [16,39]. This perspective, at most, enables us to predict (*a posteriori*—after a long series of experiments) the probability of a neuron firing in response to a specific stimulus. Up to this point, information fulfills a descriptive role in the form of conditional probability. However, either implicitly or explicitly, the use of the term is generally associated with a referential semantic value: that of the object that acts as a stimulus and which therefore appears to correlate, with a greater or lesser probability, with the activity of diverse areas in the NS. A characteristic example is the work for which Hubel and Wiesel received a Nobel prize in 1981 in which strong correlations were found between the movement of light spots and bars and the activity of certain neural assemblies in the primary visual cortex of cats; so that different areas of the visual cortex were said to convey information about the position or movement of certain stimuli.[2] This use of the term information is generally associated with the (implicit or explicit) presupposition of a homunculus, which acts as an interpreter of the semantic content. This, in turn, leads to statements such as: "(…) we can ask how the homunculus should best use the spike train data to make a decision about which stimulus in fact occurred." (Ref. [39], page 14). However, the aim for a neuroscientific understanding of cognitive causality should not be so much to infer "the reality" from a set of signals located in some part of the NS, but rather to explain how informational signals are integrated into the production of behavior and why they appear, for the organism, to be

[1]One might be tempted to solve the problem by distinguishing between the concepts of information and knowledge, implying that the concept of knowledge refers to the processing and attribution of content to informational units. However, this distinction cannot but displace the fundamental problem that we want to address here, since the attribution of semantic content to information bits remains problematic even if the distinction is made.

[2]Similar accounts of information can be found in the current literature and neuroscience textbooks; e.g. "The pons, which lies above the medulla, conveys *information about* movement from the cerebral hemisphere to the cerebellum." ([26] italics added).

Author's personal copy

associated with an intentionality or semantics about the world, thereby becoming informational (Fig. 1).

It turns almost inevitable that we make semantic, intentional, projections onto an informational treatment of neural signals. It is correct to say that signals encode, transmit or convey meaning about stimuli. This use of the term information (and all neuroscientific research carried out on this basis) is, however, faced with a fundamental problem. The measured conditional probability between the stimulus and neural activity cannot explain the mechanisms by which the NS works. In fact, the correlation (thus explained, i.e. as a conditional probability) can never be accessed by the NS itself. It is only accessible to the observer who accesses both the stimulus and its dynamic consequences separately, in such a way as to render the process informative for the observer, although not necessarily for the system being studied (see Fig. 1). Maturana was among the firsts to note that a crucial distinction should be made between an observer dependent perspective and an operational or mechanistic perspective [35]. Only within the observer dependent perspective can an informational talk make sense as a higher level descriptive shortcut used among observers. Therefore (for Maturana) information cannot really have an ontological status and, once a full mechanistic explanation of the processes involved is available, the observer-dependent informational description would be dispensable. And yet, there is a continuous habit, which still pervades today, to introduce semantic information as a causal explanatory component of cognitive neuroscience.

Thus, if the semantic content is assured by the conditional correlation, but this correlation cannot be accessed by the cognitive system, in what sense does the conditional probability constitute a semantic explanation of the behavior of the subject? On what basis are these correlations created and transformed (corrected, discarded, modified)? What are the internal mechanisms that retain the correlations and in virtue of what are they retained (since comparisons between environmental states and nervous activity are, in principle, off limits for the NS)? These questions suggest that we should attempt to explain the semantic content of an informational process from the perspective of the organism, based on the specific way in which it fits into its biological organization, rather than from the perspective of an external observer whose privileged access to both the origin of the stimulus and the nervous activity enables him to establish a correlation between the two.

For these reasons, we believe the causal-correlational conception of neural information to be unsustainable. What happens is that in order to explain a natural phenomenon, researchers tend to rely on a concept (that of information) that they cannot reconstruct in a naturalized manner (i.e. making the *explanatory* value of the informational semantics depend on an external observer who does not form part of the phenomenon under investigation). The naturalized perspective, on the other hand, opts for an ontological conception in which the term information refers to a specific type of causality in the NS, and which furthermore enables us to explain the semantic content of an informational process referring to the actual organization of the cognitive agent (and, specifically, of living organisms). The idea is to specify how a physical or biological process becomes informational not for an external observer, who is presupposed to already possess cognitive capacities, but rather *in the actual framework* of the natural system in which the process takes place, playing a specific causal role by constituting a new form of organization within such system.

The rest of the article is divided as follows. First, we shall sketch a reconstruction of the appearance of the NS in order to specify its function and structure within the framework of biological organization. Next, we shall analyze what forms of causality and self-organization are characteristic of the NS and bear it close to an informational description. Then, we shall propose a candidate definition of semantic information that is rooted on biological organization and which does not fall under the homuncular fallacy and where correlations are integrated in the self-maintenance of biological organization. We shall illustrate how our definition works with a neurobiologically informed case study of conditioned learning in Aplysia. Finally, we will outline how semantic information is created and integrated into the development of the NS leading to cognitive semantics. The goal of the paper is that embodied and situated accounts of semantic information be integrated on neuroscientific research, which has mostly focused on observer dependent causal-correlational accounts of information with widespread semantic connotation that are, however, untenable within the traditional framework.

2. The appearance of the nervous system in the living organization

All organisms are *autonomous* systems: i.e. far from equilibrium dissipative structures that recursively generate and regenerate their organization. [40]. As such they need to remain thermodynamically open and engage in interactions with their environment as a fundamental part of their own processes of self-maintenance. This ability to engage in interactions sustained by its internal organization constitutes the living being as an *agent*. With the exception of some types of bacteria which live in an extremely homogeneous and stable environment, all living beings have internal mechanisms which compensate, in somatic time and within certain margins, for the different conditions found in their environments. In other words, they are *adaptive* agents. Organisms have the ability to detect those changes in their environments that are relevant for their maintenance, and to set in motion certain internal and interactive processes, which help reestablish and improve the conditions of that maintenance.

In single-cell organisms and plants, the adaptive agency is supported by the metabolic organization, i.e. by the bioenergetic production and repair infrastructure. However,

Author's personal copy

the metabolic organization cannot support a fast and versatile range of responses for multi-cell organisms whose life form is based on movement. We know that the problem was resolved by the appearance of the NS. This event was probably the result of a combination of two factors: (1) the appearance (as a result of a processes of cellular differentiation) of a new type of cell (the neuron) capable of connecting sensor and motor surfaces plastically, rapidly and (metabolically speaking) economically; and (2) the appearance of some type of metazoan whose body plan enabled the recruitment of these cells to sustain different forms of sensorimotor agency. The strengthening of motor behavior which must presumably have occurred, served as the foundation of an evolutionary process which selected those animals that possessed interconnected neuron networks that acted as a support for functional sensorimotor behaviors of increasing complexity.

However, in addition to enable the development of motility in multi-cellular organisms, the appearance of the NS paved the way for new (qualitatively different) forms of adaptive interaction, much more complex than anything that had gone before. Even the first evolutionary stages of the NS featured rudimentary forms of learning, categorization and memory [3]. This potential ability to support agential complexity resides in the capacity of the NS to generate a domain of its own with an enormous number of configurations, whose dynamics cannot be specified by the organism's basic organization [38].

3. The structure and function of the nervous system

Neurons present a series of specific characteristics that are absent from the cells, which make up other tissues or systems within the organism. The recurrent interactions within the most basic dynamic domain gives rise to higher level dynamic patterns (such as synchronizations at different temporal and spatial scales) which enables relatively few neurons to generate a great dynamic complexity. These characteristics, made possible by the recurrent connections between neurons, enables the creation of an internal *selective* process that happens at interaction frequencies that are much higher than those found in any other of the organism's control processes. This lends the interneural network a series of special characteristics within the organism as a whole: no other intercellular system even comes close to having the NS's capacity to functionally correlate so many units and, at the same time, to selectively modify their states so quickly. *The peculiarity of the NS is, therefore, its ability to generate an enormous variety of states (configurations) per unit of time, and to coordinate an immense number of state transformations simultaneously*. As we shall see, all these characteristics are conducive to the generation of self-organizing regulating processes. In other words, neurons are an extremely peculiar type of cell which have been selected due to their ability to maintain a complex, plastic and rapid rate of interaction, thereby minimizing interference with both local metabolic processes and all other organs and metabolic biological systems (circulatory, respiratory and digestive systems, etc.).

This last factor (the minimization of interferences with local metabolic processes) is of vital importance, since it defines the *decoupled* aspect of the NS: e.g., its dynamics are sub-determined by the purely metabolic (self-constructive) processes of the supporting biological structure (the entire brain). In addition, the NS also has a reticular, cohesive and recurrent internal dynamics capable of maintaining a set of patterns in the face of both internal and external perturbations. As such, its internal dynamic complexity is greater than the interactive dynamics established with its environment. This asymmetry of complexity means that its dynamics appear to a large extent to be *self-determined*. If we wanted to predict the evolution of neural dynamics for a specific moment in time, a knowledge of the metabolic state of its isolated components would not be of any particular help. On the contrary, we would have to pay attention to the corporal and environmental context in which the organism is located and, above all, to the internal dynamics themselves that are recursively generated in accordance with its holistic networked properties, over and above local cellular interactions.

We can now respond properly to the question of the function of the NS within the organism as a whole: its task is to control interaction cycles (the interactive closure necessary for the self-maintenance of the organism) through a network decoupled from metabolic dynamics but nevertheless inserted into a corporal interface (muscular–skeletal and sensory systems). On the one hand, we can observe the material and thermodynamic flow between the ingestion and digestion of food and oxygen, its secretion and the dissipation of heat from the body, together with all the organism's self-constructive machinery and the infrastructure required for its maintenance and repair: circulatory, digestive and respiratory systems, etc. (see metabolic cycle in Fig. 2). And on the other, we see a sensorimotor cycle of adaptive behavior governed by an infrastructure (sensorimotor system and NS) that is decoupled from (although ultimately constructed by) metabolic processes. The functionality of the NS is based on the way in which both processes are connected at a higher level. This functional relationship demands that the interactive dynamics controlled by the NS satisfy the organism's self-maintenance conditions (i.e. that it maintains the thermodynamic and material flow required for the ongoing functioning of metabolism). From the perspective of the dynamic working of the NS, the satisfaction of these conditions surrounding the thermodynamic flow appears as what Ashby [4] termed *essential variables*: a set of variables (temperature, nutritional input, etc.) that must be maintained within given viability limits. It is around the homeostatic maintenance of these variables that the cohesive dynamics of the NS and its internal organization is generated (although, as we will see later on, it is never

Author's personal copy

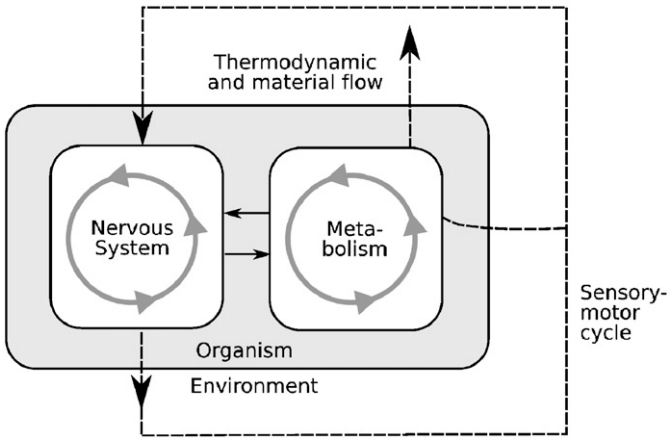


Fig. 2. The basic organization of life is that of a self-maintaining metabolic cycle that remains materially and thermodynamically open. The NS appears as a mechanisms to regulated behavior independently of the ongoing processes of metabolic construction and repair. The NS generates its own dynamic domain coupled through sensory and motor surfaces to the environment. From an organismic point of view the function of the NS is to control the sensorimotor cycle towards the satisfaction of metabolism's needs. The arrow that goes from the sensorimotor cycle to metabolism indicates this fundamental function that shall come to be a crucial component for error detection and, ultimately, for the notion of semantic information. (Copyright 2007 Xabier Barandiaran. This figure is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/).

completely determined by these adaptability constraints). The dynamics of the NS is ultimately assessed on the basis of its contribution to the self-maintenance of the organism, at both phylogenetic and ontogenetic scales. The embodiment of the NS is thus understood in the two-fold sense of an interactive interface with the world (the musculoesqueletal system) and a biological rooting, through which the metabolism constructively sustains the NS and the NS contributes to satisfying the interactive needs of the metabolism (thermodynamic and material exchange with the environment).

4. Self-organization and dynamic causality in the NS

The self-organization of neural impulses on different scales results in the generation of intermediate structures (patterns), which interact between themselves and help organize the global dynamics of the system in a broader sense than just local impulse interactions. This forces us to distinguish between micro- and macro-dynamic levels (and between diverse hierarchies of levels) and enables us to talk about an emerging functionality (e.g., neural microscopic processes whose contribution to the self-maintenance of the system is marked by the emerging patterns they either constitute or in which they participate). On the one hand, the organization of neural impulses into massively parallel and self-organized processes means that the effective causality of each individual impulse is diluted; and on the other hand, the transformation of global patterns regenerates a functionally effective causality at a higher level. This enables us to talk about micro-dynamics from which functional neural macro-dynamics emerge. The main consequence of this causal emergence is that it becomes extremely difficult to establish a localizationist strategy in the study of cognitive processes. Localizationism operates by structurally and functionally decomposing the system into component parts and then establishing a mapping of functions over structures. The aim is to provide a causal explanation in which structural component parts (with their associated functions) are linearly aggregated to provide a full picture of the system's working [7]. In addition, if we establish a computational functionalist decomposition and we find representational correlates in some brain areas it might be possible to reconstruct the full picture of the brain as a computational machine. However, the self-organization of the dynamic working we have seen does not permit a straightforward implementation of this strategy, since the functionality emerges from the non-linear interaction between the components.

The view that cognition is the result of the large-scale integration of the activity of distributed neural groups is becomingly increasingly widespread among many neuroscientists [20,22,42,44]. The more classic view (inherited from traditional or representationalist AI) assumes that the brain operates on the basis of functionally specific modules and that cognitive behavior is the result of the intra-modular processing of information and its inter-module communication. Large-scale integration models, on the other hand, claim that cognitive behavior is the result of global patterns of oscillation that emerge from the reciprocal dynamic interaction between multiple, disperse ensembles of neurons in the NS. Together with the embodied and dynamically situated behavior, this has led many cognitive scientists to prematurely dismiss the use of the concept of semantic information and/or representation in the NS [8,11,21,28,43].

5. The appearance of informational causality in the NS

We can now see that action potentials (together with some neuromodulators) and, above all, the higher-order patterns they generate when they self-organize, constitute a new type of specifically neural observables (local attractors, higher level synchrony, etc.). When a network is formed by various working neurons, we can describe its dynamics using these observables. While we view a neural network as an isolated system (for example, analyzing its behavior in a Petri dish), these observables represent nothing more than a useful way of providing an abbreviated description of a much more complex set of underlying molecular dynamics. However, in the framework of the living activity of an organism located in a specific environment, these neural observables generate a new functional dynamic level (i.e. this higher level description is not just a shortcut description but a causally relevant pattern that contributes

Author's personal copy

to the self-maintenance of the system). This new level has two special characteristics:

1. The non-linear nature of the nervous impulses enables their stable combination to create high-dimensional states, something which, along with the characteristics of the connectivity matrix network and the action of neuromodulators, gives rise to a set of characteristic properties: (a) compositionality (impulses can be organized sequentially in time), (b) recurrency (the network structure enables circularity) and (c) recursivity (impulses can operate on themselves through the neuromodulators they activate[3]). In turn, these observables can self-organize themselves into stable patterns and generate compositional domains at higher levels.
2. The fundamental property of neural observables is their *potential* for a *formal causality* decoupled from the energetically determined causality on which these same signals are propagated. This is the same as when, for example, electrical impulses traveling through computer network cables provoke changes in the terminals not as a result of the electrical energy they carry, but rather due to the sequence of changes in amplitude and frequency.

The NS is thus a dynamic domain that is decoupled from the ongoing metabolic functions of the organism and appears constituted by a recurrent dynamic in constant self-organization with capacity for formal causality. This brings us inevitably to the concept of information. Great part of current cognitive neuroscience research that focuses on codes and information is devoted to the quantitative study of features of neural activity (population codes, interspike intervals, frequencies, etc.) independently of their semantic content. These studies are mainly focused on making explicit which of these features are causally relevant for propagating dynamic variability within the NS. Although often charged with a semantic load, research on neural codes and information (understood as dynamic variability and correlation) needs not necessarily involve reference to semantics or informational content and thus, might not be charged with the paradoxes and problems mentioned at the beginning of this article. It is in the sense of Shannon's information theory that this approach remains useful and save of paradoxes. However, it still remains to be explained what else could the concept of information add to that of propagation of variability and correlations; i.e. to the concept of dynamic and formal causation. The key factor here is the assignment of semantic content to the notion of information.

In order to avoid entering into homuncular and externalistic paradoxes of a non-naturalized concept of semantic information, we should resist the temptation of using metaphors derived from computer and communication technologies beyond what is strictly necessary and naturalizable. The point of union is, without doubt, the possibility of a formal causality, as noted above. However, properly informational causality in communication technologies always depends on a natural cognitive system that interprets the signals. Our aim is to sketch a conceptual framework capable to explain the functioning of the NS making use of the notion of semantic information. A framework that halts the regression to an endless list of interpreting subjects (in the form of cerebral homunculi) to justify such informational causality by rooting it in the dynamic organization of neural activity within an organism.

6. Towards a naturalized definition of neural information

The only way of ensuring that the concept of information be naturalized is (following Bickhard [9]) to elaborate a notion of information (although this author talks of representation) that enables the detection of errors by the system that handles the information. This way, that the semantic content can be evaluated by the very system (not merely ascribed by an external observer) and justified according to the norms generated by the very system (and again not imposed or interpreted from outside). This in turn requires that we naturalize the notion of *normativity* in such a way as to ensure that it emerges from the organization of the organisms and becomes accessible for it; i.e. that the origin of the *norm*, according to which the semantic content is to be evaluated, be centered on the organism.

The problem of normativity, as Kripke put it (although in a different context, that of Wittgensteinian account of rule-following and language games), is that "the relation of meaning and intention to future action is *normative*, not *descriptive*" [32]. So, if we are to justify that something (a signal of the optic nerve) bears meaningful information about a certain environmental condition (about the presence of food) a *description* of the action that the information triggers (presumably eating) does not suffice to ground its meaning or informational content. This insufficiency shows itself when trying to account for the possibility of error or misinformation. If the environmental condition is the presence of poison, rather than of food, a signal might still carry the information that there is food in the environment although this information be erroneous under that particular condition. In such a case, we need to be able to justify that the organism *had* to do something else under such condition and that the information was thus erroneous. Here, the mismatch between descriptive and normative levels appears evident: by describing that the organism has responded to a signal with the action of eating (because it had the erroneous, informational content "presence of food") we can certainly not justify that it had to avoid it. The concept of informational content presupposes that there is a norm according to which semantic content can be evaluated, i.e. a norm that permits

[3]For this reason, the basic difference between NS observables and other types of signals, such as hormones (molecules that act as vehicles for various forms of inter-cellular coordination), besides their much greater speed of transmission, is their recursive 'processing' capacity (signals operating on signals), something which lends the NS a "quasi-syntactic" nature in relation to the dynamics of metabolic processes.

Author's personal copy

to establish that the information, if it is to carry any content at all, is correct or incorrect. But where does this norm come from?

The question has been answered in many different ways, but two major approaches are relevant for our discussion. The first is due to philosopher Ruth Millikan who proposed to solve the problem of normativity with reference to the selective history of a mechanism [37]. The normative function of a mechanism (it could well be a neural structure) is determined by the function for whose performance the mechanism has been selected. On her account, a signal represents whatever it is its normative function to represent. Thus, for instance, if a food detector signals, its informational content is that of "presence of food" because it is by performing that function that it has been selected. This account is thus rooted on the past, on the selective history of the mechanism that generates such signal. The signal could then be false if the mechanism generates it in conditions other than those that produced its selection on the past. However, this account is not without its own problems. For instance, as Mark Bickhard has repeatedly argued [9,15] the dependence of a past history to qualify a function as normative renders it epiphenomenal since it involves reference to the past events to explain the norm here and now, independently of its current causal organization. For instance, her approach does not permit to say that a mechanism had a proper function before it be subject to a selective history. The alternative is to ground the norm or normative function of a mechanism on the organizational context in which it participates. It is here where the autonomous (i.e. recursively self-maintaining) organization of a living organism comes to rescue. Organisms are endowed with implicit norms, which are those required to assure its continuing existence. The normativity arises from the dynamic conditions of possibility inherent in a far from equilibrium organization [15]. Thus, for instance, we may affirm that the normative function of a process, within the organization to which it belongs, is defined by the way in which it contributes to its self-maintenance. For example, the normative function of the active pumping processes of a cell membrane is to regulate the internal concentrations of metabolic reactants, thus keeping the network of metabolic reactions within the concentration limits required for the maintenance of the network itself. In this sense, normativity is inherently circular: a functional process becomes normative because it contributes to the far from equilibrium maintenance of the organization that, in turn, creates/maintains that functional process. The normative function of the NS being, as we proposed above, the interactive maintenance of certain essential variables under viability constraints, it appears now subject to a norm that allows to ground the notion of semantic information.

Normativity, thus, becomes the principle on which *error assessment* is established as a *causally* effective process in the maintenance of the identity of the system. In this way, a naturalized concept of information requires an understanding of the origin of the regulatory values of an organism. Only in the context of the biological interaction between system and environment, in which the regulatory values of agency are generated through the self-organizing processes of signal selection, can we understand the concept of information in natural cognitive systems. Informational processes in the NS are neither mere correlation with states of affairs in the world nor are they mere constraints imposed on self-organizing processes.

Thus, the key factors underlying a naturalized notion of information in the NS are, therefore, closely related to: (a) how information is causally integrated into its organization (in relation to its creation and capacity for constraining neural dynamics) and (b) how these processes acquire a semantic (referential) content in relation to a norm that is internally generated by the system (which enables errors to be detected and corrected). We propose that the activity of the NS becomes informational in nature when its functionality becomes normative through its correlation with an external state of affairs, i.e. when the process contributes to the self-maintenance of the organization it belongs through its correlation with certain environmental conditions. We can, thus, offer the following definition of neural information:

> A pattern of signals **S** contains information, about an environmental condition **E,** within a far from equilibrium organization (**O**) iff: **S** is selected by **O** to fulfill a normative function (i.e. it contributes to self-maintenance) within **O** through its correlation with **E**.

Where **O** is a self-maintained structure of a higher order than **S. O** may be a self-organized pattern of signals, the NS as a whole, the entire organism or even the species as a collective structure (thus, the selection of **S** may occur at different scales: within circuits of the NS, during the development of the NS or at an evolutionary timescale). **E** is not meant to be an observer independent, absolute or objective states of affairs (otherwise inaccessible on its objectivity to an organism) but an environmental condition that is coupled to the self-maintenance of **O** so that its selection can ultimately be evaluated by the very system (either on phylogenetic or ontogenetic scales).

Let us now look at a specific example, which illustrates this definition of information in neuroscience. Let us suppose that an Aplysia is capable of learning that the presence of red light is regularly followed by food, i.e. associating the conditioned stimulus (CS) "red light" with the unconditioned stimulus (US) "food". We know, in some detail, the neurobiological processes involved in conditioned learning in Aplysia [10,12,27] for a similar processes as the one described above.[4] Now, let us imagine that we find an Aplysia in our lab and we want to

[4]What follows is a fictional mechanism based on previous research but where yet unknown mechanisms were invented for the purpose of a mechanistic illustration of the concept of information defended here. In particular, the connection of the optic nerve to neuron B51 is unreal and artificially introduced into the example, the rest of neurobiological details are real.

Author's personal copy

understand its neural activity in terms of information. After a detailed experimental process, we observe that the activity of the optic nerve is correlated with a certain red light intensity, which is, in turn, correlated with the presence of food in the environment of the Aplysia (Fig. 3a). At this point, from the point of view of some computational neuroscientists, we could describe the situation neurobiologically as stating that the optic nerve activity contains information about food in the environment. The conditional probability of a certain optic nerve firing occurring in the presence of food (when food and red light are correlated) holds, according to the orthodox view, this informational content. At this point we can ask… in which sense does the conditional probability provide an informational explanation of the Aplysia's neural activity or behavior? We have argued that such an approach is inadequate or, at least, incomplete. The problem is clearly illustrated in the following situation. We can change the correlation of red light with food to one of red light with poison (Fig. 3b). Under this condition and, according to the orthodox framework, the activity of the optic nerve of the Aplysia would be, in fact, correlated with poison; thus, its informational content would be "poison". If our informational description is to have any explanatory capacity we should conclude that the Aplysia, when receiving a signal with "poison" as informational content, will not swallow. But, obviously, the Aplysia will not stop swallowing because it has no access to the new correlation, only us, as external observers, can have it. If the activity of the optic nerve bears some information for the Aplysia (as it is to be expected from its behavior), it is that of "food". But the informational content is wrong since, in fact, there is poison in the environment, something that the Aplysia will be able to suffer as a consequence of its mistake and, eventually, correct for future interactions. The fact that there be an informational content that is not the result of the actual causal correlation and that it be wrong and corrected by the very organism is something that cannot be addressed under the notion of semantic information as causal correlation.

Only by having a closer look at how the optic nerve activity is integrated into Aplysia's behavior and NS can we go beyond the paradox and provide a full naturalistic explanation of neural information. By focusing on the mechanistic details we might note that Aplysia's optic nerve connects with neuron B51 which, located at the bucal ganglia, has the capacity to modulate the central pattern generator (CPG) that produces swallowing patterns. When there is a correlation between a red light source and the presence of food in the aqueous environment, the Aplysia might be able to do the association between the US (food) and the CS (red light) due to the dopamine release of the anterior branch of the esophageal nerve on neuron B51 (this dopaminergic release acts as a reinforcement signal). The sensorimotor closure of the loop through the esophageal nerve is crucial (see Fig. 3c). Neuron B51 is responsible for the association of US with CS due to three main connections it has: (i) input signal from the optic nerve (that receives signals from the red light), (ii) input signal from the esophageal nerve (that brings information of presence of food from the stomach) and (iii) output signal to the bucal CPG which generates swallowing patterns that leads in turn to new signals from the esophageal nerve (as a result of swallowing). In normal conditions, when looking for food, Aplysia swallows spontaneously with a given frequency. When, as a result of this swallowing, the Aplysia encounters food the esophageal nerve signals to B51. In turn, if this signal from the esophageal nerve is correlated with signals arriving from the optic nerve to B51, B51 will change its behavior: under the presence of red light alone (CS) B51 will, in the future, increase the frequency of swallowing by modulating the bucal CPG. As a result, if red light and food are still correlated in the environment, the signal from the esophageal nerve will further reinforce the association, increasing the swallowing frequency, and so on.

We can analyze now, how this circuit embedded on the Aplysia, as an autonomous organism, might satisfy our definition of neural information. We shall first analyze the case of the esophageal nerve signal bearing information about food. According to our definition the pattern of signal from the esophageal nerve to B51 contains information about food iff: the signal of the esophageal nerve is selected by the organism to regulate its behavior in relation to its effective correlation with food in the stomach. This is indeed the case and the correlation is fixed by evolution (instead of learning) and evaluated at the evolutionary scale. If the esophageal nerve were to signal to B51 under the presence of poison in the environment (rather than food) the Aplysia would not survive due to the way in which signals from the esophageal nerve to B51 are integrated in the behavioral organization of Aplysia (which would accelerate the ingestion of poison and kill the Aplysia). Equally, the signal from the optic nerve bears information about the presence of food in the environment iff: the organism selects that signal to initiate swallowing patterns that presuppose a correlation between optic nerve signal and food. Here, the correlation is fixed, reinforced and evaluated by the information that the esophageal nerve signals to B51. Thus, the informational content of the optic nerve activity in relation to food presence only exists when selected by the organism to trigger the swallowing action that enables, in turn, an assessment of the correlation. It may be the case that the information is false (see Fig. 3c) and, in such a case, either the Aplysia is capable of detecting the error and correct its response or its metabolic homeostasis will be severely disturbed (ultimately leading to death). In either case, we can assert both that the information was there (as interpreted by the Aplysia) and that it was wrong. The optic nerve bears information only to the extent that it is functionally and normatively integrated into the organism (thus, permitting the organism to effectively regulate its behavior accordingly).

Author's personal copy

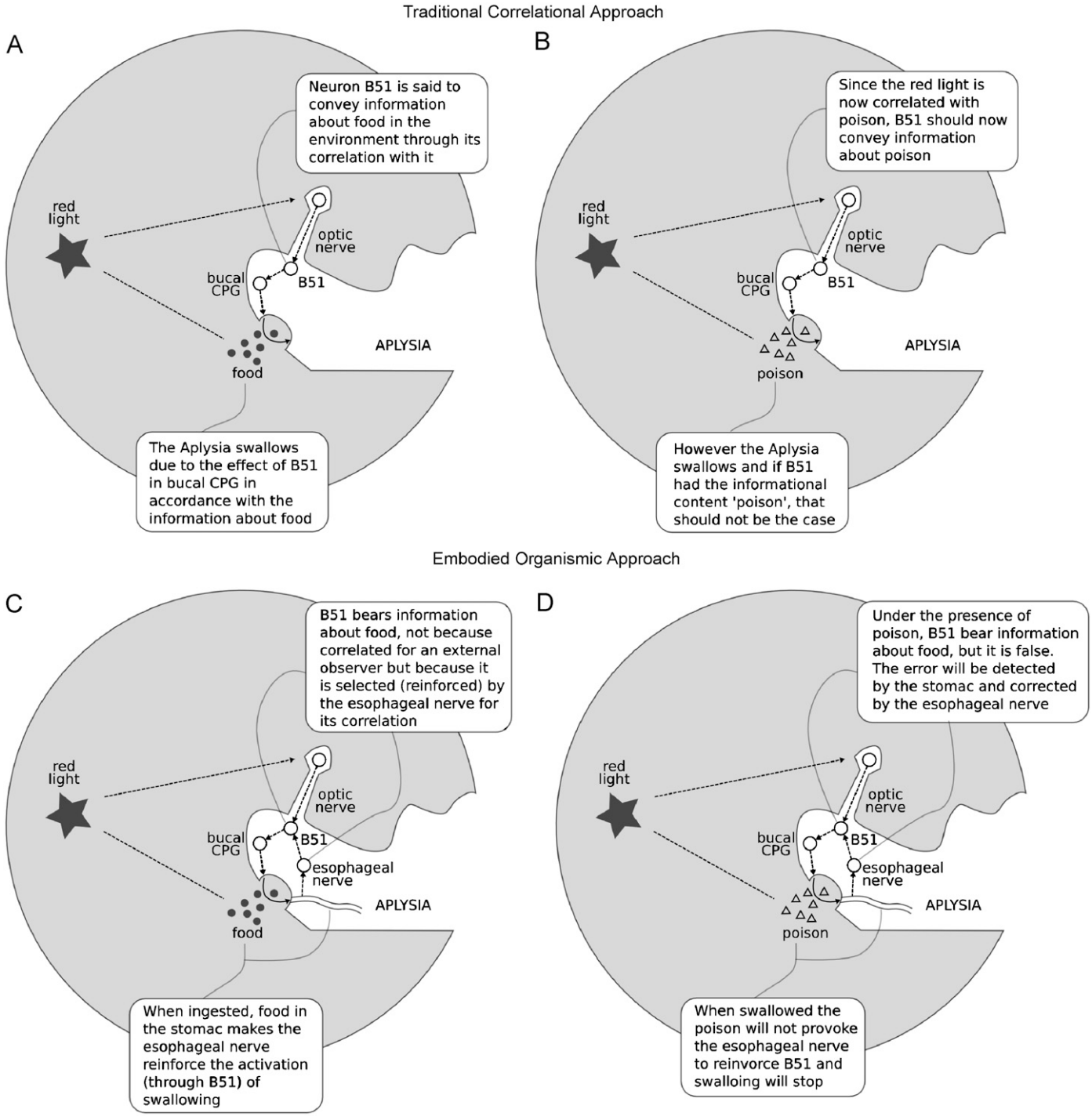


Fig. 3. A case study of the mechanisms underlying conditioned learning and information in the Aplysia. Just by mere correlational studies of the presence of the red-light stimulus, food and the activity of neuron B51 it is impossible to account for semantic information. In case A, an observer could establish the correlation between red light, food and activity in B51 and thus conclude that B51 conveys information about the presence of food in the environment. However, as shown in case B, where food has being substituted by poison, the correlation now hold for poison, thus B51 shall have the semantic content "poison", but the Aplysia swallows under this condition. From the merelly correlational point of view, it cannot be said that B51 (in case B) has the semantic content "food" but that it is incorrect. However, if the whole circuitry involved in the process is considered (cases C and D) the difficulties are solved. Feed-back from the esphageal nerve serves to maintain or destroy the correlation between the optic nerve activity and the swallowing patterns through neuron B51. If the correlation is hold between red light and food, the B51 will maintain it through signals coming from the esophageal nerve. Otherwise (case D) the error will be perceived and corrected. (Copyright 2007 Xabier Barandiaran. This figure is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/).

Author's personal copy

In this way, a signal carries information regarding the presence of food for an organism iff: the actions triggered by this signal in the organism contribute to its self-maintenance *through* the presence of food. This means that the information may be false; since it may be the case that something other than food triggers this signal. This in turn triggers the subsequent interaction whose result *does not* contribute to the self-maintenance of the system (through the presence of food). Furthermore, this error may be detected if the result of the interaction is accessible to the organism, and may even be corrected if the organism possesses behavioral modulatory mechanisms based on error detection. This example illustrates the way in which we can talk about perceptive information in an organism without falling into the trap of the fallacy of the homunculus, e.g. without offering a description of stimulus–impulse correlations which (indefinitely) shifts the problem of meaning back to a system which in turn has to *interpret* the impulse.

The example shown bears a significant resemblance to the Reinforcement Learning paradigm in AI [25] where different methods are used to modify the input output function of a decision making algorithm to maximize a reward signal as it explores, through trial and error, the input–output space (e.g. the space of possible solution to the encountered problems). Yet, there is a crucial difference on the case of Aplysia, a difference that illustrates the problem of normativity explained above. Whereas for the algorithm the reward signal and the error is contingent (an ultimately justified by the intention of the designer) for the Aplysia the consequences of the error are of direct concern for its own existence; the norm is defined from within and it is intimately linked to its own preservation and therefore informationally meaningful. Thus, the definition presented above enables us to talk about neural processes which operate *effectively* as information *for* an organism (since the organism is capable of assessing the semantic content of the informational process), rather than from the perspective of an external observer who establishes the link (either in the form of causal correlation or in any other way) between the signal and the informational content.

7. Learning and development: the creation of information in the nervous system

Unlike genetic information in the ontogeny of the organism, neural information is constantly created in the NS, and it is precisely this creation of information through the interaction of the organism with its environment that enables the gradual "self-determination" of the NS. The creation of information is necessary because it cannot be genetically specified completely. The creation of information *through self-organizing processes* is one of the NS's permanent tasks.[5] This is due to the fact that the information required to specify neural circuits far exceeds the capacity of the genetic information. In the case of human beings, if neural circuits were genetically pre-specified, this would require that 10^6 genes (of which only 20–30% participate in the construction of the NS) would have to store enough information to codify 10^{14} synapses, which, in turn, may take on quantitative values from a fairly wide-ranging spectrum. This reduction in variability (from the set of possible parameters and connections to those which are functional for the organism) can only be achieved through self-organizing processes which select and stabilize specific configurations, e.g. through the generation of information during the ontogeny of the organism.

The concept of information defined here is intrinsically linked to a dynamic process of trial and error, in which functional correlations are generated and destroyed. Apart from the simplest sensorimotor dynamics, which are genetically determined to a high degree, all new interactive functionalities arise as a result of the recurrent and networked (self-organized) interactions of a whole set of neurons (although some raw genetic constraints might apply that canalize self-organizing processes to lead to adaptive behavior). In such systems, no linear and unidirectional sensorimotor trajectory determines the behavior of the organism. From this point of view, perceptive stimulation appears as an undefined set of perturbations (spike trains) in the dynamic state of the neural network. Some of these spike trains are selected or recruited by other interactive patterns of a higher order. Perceptive simulation is therefore modulated from within: sensorimotor cycles, internally generated and regulated, stabilize a sensory pattern in order to integrate it into the sensorimotor cycle. In turn, new motor patterns are generated through CPG structures to maintain a series of internal invariants. These patterns are stabilized (selected or activated) in the form of attractors, resonances, etc. when they contribute to the self-maintenance of the system. It may also occur that, as the result of action–perception–action cycles, the global neurodynamic pattern changes and the organism adopts another type of interaction with the environment. The interactive embodiment of the NS causes this invariant maintenance process to close in sensorimotor cycles with the environment, thus integrating them functionally for the maintenance of the global patterns. At an even greater scale of complexity, we find that the dynamics of the network itself generates learning processes. The most basic form of functional learning processes is that of reinforcing the connections that give rise to behavior that is beneficial to the organism. In this mechanism, neuromodulator activation processes are essential, since they set the functional routes for impulses,

[5]Unlike genetic information, neural information is generated in somatic time. There is no genetically defined neural information because neural information is constructed in the environment of the NS. Through

(*footnote continued*)
development processes, genetic information specifies the mechanisms (including the basic architecture of the NS), which in turn enable the creation of neural information (and, naturally, the construction of the organism as a metabolic system).

Author's personal copy

thus generating an internal *selection* process [19].[6] As shown by comparative EGG studies in adults, infants and newborns [2,36], the correlation dimension of human brain activity increases with maturation, i.e. new correlations are created through development. This increase in the correlation dimension of brain activity cannot arise but through continuous interactions with the environment where different signals are recruited by the NS to sustain new sensorimotor correlations.[7] This "recruitment" can be understood as stabilization of certain sensorimotor loops through the action of internal regulatory mechanisms (which are, in turn, coupled to the self-maintenance of the organism).

Since the notion of information proposed in this paper heavily depends on system–environment interactions, neurobiologically informed robotic research becomes a major tool for exploring the creation and maintenance of informational organization in the NS. An example of such neuroscientific research is provided by Edelman and co-workers' work on how organism's regulatory mechanisms interactively shape neurodynamic organization. A series of robots called DARWIN have served to integrate the action of the so called "value systems" into behavioral neurodynamic organization: by feeding back to the sensorimotor controlling architecture (inspired on human neurophysiology) reinforcement pathways select and stabilize those neurodynamic patterns that lead to "successful" behavior [23]; see Ref. [31] for a recent development. These robots provide adequate models for our definition of neural semantic information if we interpreted that behavioral evaluation is linked to the organism's survival conditions (which is in fact an implicit assumption on the robotic design).

We have argued somewhere else [6,5] that cognitive semantics and genuine mindful intentionality arises when it is the self-maintenance of neural organization itself what is being preserved, i.e. when not only the organism as a metabolic system becomes autonomous but the very organization of the NS. The creation of increasingly complex correlations within the NS through the ontogeny of the organisms leads to a situation where the coherency of neural organization itself becomes a major regulatory principle. This implies a new level of selection and normativity and thus a new level of information: certain perceptive neural patterns and sensorimotor correlations will start to get selected by the global activity of the NS for the maintenance of its organization. In other words, the continuous process of development and learning within the NS leads to selective processes which might not be based anymore on adaptive criteria but on the preservation of the coherency of experience itself; thus, leading to what might be understood as properly cognitive semantic information.

8. Conclusions

The orthodox use of the concept of information (received from 50 years of AI research on syntactic computational processes) is faced with a series of problems which do away with precisely those (intentional and semantic) properties that make the brain one of the most complex and fascinating systems in our universe. We have seen how, on the basis of the semantic or referential contents that is embodied in the adaptive needs of the organism, a whole network of self-organizing processes is generated in the NS: higher level forms of neurodynamic organization select lower level processes creating a looped hierarchy of informational causality. In this way, the degrees of freedom of a system (the NS) capable of a huge number of state transformations (through an enormous recurrent variability between diverse types of observables) becomes organized informationally throughout ontogenetic development. Learning processes, based on the fundamental normativity of the self-maintenance of the system, establish a series of sensorimotor relationships, which, in turn, stabilize other higher-order processes, thus creating a network of sensory and motor relationships that can always be revised. This is how the NS becomes informationally organized for itself and by itself, and not just for an external observer capable of establishing the semantic content through its privileged access to the world and internal neural states.

Acknowledgments

Funding for this work was provided by grants: 9/UPV 00003 230-13707/2001 from the University of the Basque Country and BMC2000-0764 and HUM2005-02449/FISO from the Ministry of Science and Technology. In addition, X.B. has the support of the doctoral fellowship BFI03371-AE.

References

[1] E.D. Adrian, The basis of sensation; the action of the sense organs, first ed., Christophers London, reprint Hafner N.Y., 1928–1964.
[2] A.P. Anokhin, N. Birbaumer, W. Lutzenberger, A. Nikolaev, F. Vogel, Age increases brain complexity, Electroencephalogr. Clin. Neurophysiol. 99 (1) (1996) 63–68.
[3] P. Arhem, H. Liljenstrom, On the coevolution of cognition and cosciousness, J. Theor. Biol. 187 (1997) 601–612.
[4] W. Ashby, Design for a Brain. The Origin of Adaptive Behaviour, 1978 ed., Chapman & Hall, 1952.
[5] X. Barandiaran. Mental Life: conceptual models and synthetic methodologies for a post-cognitivist psychology. In: B. Wallace, A. Ross, J. Davies, T. Anderson (Ed.), The World, The Mind and

[6] The general form of this process is that extrinsic neuromodulators (such as dopamine) regulate the stability of those circuits (or connectivity matrixes) that generate interactive processes with an adaptive value.

[7] There are non-interactive means by which such correlations could be generated, for instance by means of non-supervised and non-reinforced learning that extract correlations from repeated input patterns (e.g. Kohonen self-organized maps [29]). We know, however, that such modes of learning cannot explain but a very limited set of neural organization and that reinforced learning is an important mechanism of brain development. Thus, the increase on the correlation dimension needs to be explained in terms of sensorimotor interactions and their internal (and possibly external, i.e. supervised) evaluation.

Author's personal copy

the Body: Psychology After Cognitivism, Imprint Academic, 2007, pp. 49–90.
[6] X. Barandiaran, A. Moreno, On what makes certain dynamical systems cognitive. A minimally cognitive organization program, J. Adaptive Behav. 14 (2) (2006) 171–185.
[7] W. Bechtel, R. Richardson, Discovering Complexity. Decomposition and Localization as Strategies in Scientific Research, Princeton University Press, 1993.
[8] R.D. Beer, Dynamical approaches to cognitive science, Trends Cogn. Sci. 4 (2000) 91–99.
[9] M.H. Bickhard, Information and representation in autonomous agents, J. Cogn. Syst. Res. 1 (2) (2000) 285–333.
[10] B. Brembs, F.D. Lorenzetti, F.D. Reyes, D.A. Baxter, J.H. Byrne, Operant reward learning in *Aplysia*: neuronal correlates and mechanisms, Science 296 (2002) 1706–1709.
[11] R. Brooks, Intelligence without reason, Artif. Intell. 47 (1991) 139–159.
[12] T.J. Carew, Understanding the consequences, Nature 417 (2002) 803–806.
[13] R. Carnap, Y. Bar-Hillel, Semantic information, Br. J. Philos. Sci. 4 (1953) 147–157.
[14] G.J. Chaitin, Algorithmic Information Theory, Cambridge University Press, Cambridge, 1987.
[15] W.D. Christensen, M.H. Bickhard, The process dynamics of normative function, Monist 85 (1) (2002) 3–28.
[16] P. Dayan, L.F. Abbott, Theoretical Neuroscience, MIT Press, Cambridge, MA, 2001.
[17] F.I. Dretske, Knowledge and the Flow of Information, University of Chicago Press, 1981.
[18] F.I. Dretske, Explaining Behavior, MIT Press, 1988.
[19] G.M. Edelman, Neural Darwinism: The Theory of Neuronal Group Selection, Basic Books, 1987.
[20] W. Freeman, How Brains Make Up Their Minds, Columbia University Press, 2001.
[21] W.J. Freeman, C.A. Skarda, Representations: Who Needs Them?, in: J.L. McGaugh, et al. (Eds.), Brain Organization and Memory Cells, Systems and Circuits, Oxford University Press, New York, 1990, pp. 375–380.
[22] K.J. Friston, The labile brain (I, II and III), Phil. Trans. R. Soc. Lond. B 355 (2000) 215–265.
[23] K.J. Friston, G. Tononi, G.N. Reeke, O. Sporns, G.M. Edelman, Value-dependent selection in the brain: simulation in a synthetic neural model, Neuroscience 59 (2) (1994) 229–243.
[24] S. Harnad, The symbol grounding problem, Physica D 42 (1990) 335–346.
[25] L.P. Kaelbling, M.L. Littman, Reinforcement learning: a survey, J. Artif. Intell. Res. 4 (1996) 237–285.
[26] E.R. Kandel, J.H. Schwartz, T.M. Jessell, Principles of Neural Science, fourth ed., McGraw-Hill, 2000.
[27] E.R. Kandel, The molecular biology of memory storage: a dialogue between genes and synapses, Science 294 (2001) 1030–1038.
[28] F. Keijzer, Representation and Behavior, MIT Press, 2001.
[29] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybern. 43 (1982) 59–69.
[30] A.N. Kolmogorov, Three approaches to the quantitative definition of information, Probl. Inf. Transm. 1 1 (1) (1965) 1–7.
[31] J.L. Krichmar, G.M. Edelman, Brain-based devices for the study of nervous systems and the development of intelligent machines, Artif. Life 11 (2005) 63–77.
[32] S. Kripke, Wittgenstein on Rules and Private Language, Harvard University Press, 1982.
[34] D. MacKay, Information, Mechanism and Meaning, MIT Press, 1969.
[35] H.R. Maturana, Biology of Cognition, in: Maturana & Varela (1980), Autopoiesis and Cognition, Reidel, 1970.
[36] A. Meyer-Lindenberg, The evolution of complexity in human brain development: an EEG study, Electroencephalogr. Clin. Neurophysiol. 99 (5) (1996) 405–411.
[37] R.G. Millikan, Language, Thought and Other Biological Categories, MIT press, Cambridge, MA, 1984.
[38] A. Moreno, A. Lasa, From basic adaptivity to early mind, Evol. Cogn. 9 (1) (2003) 12–30.
[39] F. Rieke, D. Warland, R. van Steveninck, W. Bialek, Spikes. Exploring the Neural Code, MIT Press, 1997.
[40] K. Ruiz-Mirazo, A. Moreno, Searching for the roots of autonomy: the natural and artificial paradigms revisited, Commun. Cogn.—Artif. Intell. 17 (3–4) (2000) 209–228.
[41] J.R. Searle, Minds, brains, and programs, Behavioral and Brain Sciences 3 (3) (1980) 417–457.
[42] G. Tononi, G.M. Edelman, O. Sporns, Complexity and coherency: integrating information in the brain, Trends Cogn. Sci. 2 (12) (1998) 474–484.
[43] F. Varela, E. Thomposon, E. Rosch, The Embodied Mind: Cognitive Science and Human Experience, MIT Press, 1991.
[44] F. Varela, J.P. Lachaux, E. Rodriguez, J. Marinerie, The brain web: phase synchronization and large-scale integration, Nat. Rev. Neurosci. 2 (2001) 229–239.
[45] J.D. Victor, Approaches to information—theoretic analysis of neural activity, Biol. Theor. 1 (3) (2006) 302–316.

Xabier Barandiaran is a Ph.D. student at the University of the Basque Country. He is a graduate student in Philosophy from the University of Deusto (Bilbao, Spain) and obtained an M.Sc. in Evolutionary and Adaptive Systems at the University of Sussex (Brighton, UK). His main research areas are the philosophy of Artificial Life, neurophilosophy, enactive cognitive science and naturalized epistemology.

Alvaro Moreno is a Professor of Philosophy of Science at the University of the Basque Country, in Spain, where he has created a research group specialized in Complex Systems, Philosophy of Biology and Artificial Life. He is author of over a hundred publications and organizer of several international workshops on the Artificial Life, Philosophy of Biology and Cognitive Science.