Abstracta 4 : 2 pp.123 – 147, 2008

BRAIN AND BEHAVIORAL FUNCTIONS SUPPORTING THE INTENTIONALITY OF MENTAL STATES

João de F. Teixeira and Alfredo Pereira Jr

Abstract

This paper relates intentionality, a central feature of human consciousness, with brain functions controlling adaptive action. Mental intentionality, understood as the "aboutness" of mental states, includes two modalities: semantic intentionality, the attribution of meaning to mental states, and projective intentionality, the projection of conscious content into the world. We claim that both modalities are the evolutionary product of self-organized action, and discuss examples of animal behavior that illustrate some stages of this evolution. The adaptive advantages of self-organized action impacted on brain organization, leading to the formation of mammalian brain circuits that incorporate semantic intentionality in their *modus operandi*. Following the same line of reasoning, we suggest that projective intentionality could be explained as a result of habituation processes referenced to the dynamical interface of the body with the environment.

1. Semantic and Projective Intentionality

Although discussions concerning the notion of intentionality have predominantly been part of the agenda of philosophers of mind, philosophers of language and cognitive scientists, the interest in the nature of intentional states has recently called the attention of cognitive neuroscientists and neurobiologists. One of the most influential neuroscientists of the XXth century, Walter Freeman, elaborated a philosophical view of brain functions based on an inner disposition for perception and action in the environment (Freeman, 1995).

In philosophical literature, the problem of intentionality can be dated back to Aristotle and Thomas Aquinas, who attempted to understand the inner dispositions of living beings to direct actions toward intended goals, instead of a sheer reaction to causal forces of the environment. The modern formulation of the problem was made by Brentano (1925/1973): "Every mental phenomenon is characterized by [...] what we could call, although in not unambiguous terms, the reference to a content, a direction upon an object (by which we are not to understand a reality in this case) or an immanent objectivity". This view is often characterized as the "aboutness" or "of-ness" of mental states.

The concept of intentionality derived from Meinong, Brentano, Husserl, and Frege is present in the contemporary philosophy of mind and language (Dennett, 1969; Searle, 1983). We call it *semantic intentionality*, since it is related to the meaning attributed to mental states, including the cases of linguistic meaning. For more than two decades it has played a central role in the epistemology of cognitive sciences, as a keyword for discussing the reaching and limitations of computational strategies in the characterization of genuine cognition and mentality. John Searle (1980) has forcefully called attention to the serious obstacle posed by the philosophical problem of intentionality for artificial intelligence. The celebrated "Chinese Room Argument" aims to show that the states of a computational system do not have meaning except when interpreted from the outside, i.e., by those who interact with it. The more recent version of the argument (Searle, 1992) emphasizes the role played by the brain. Semantic intentionality is considered to be a product of the causal powers of the brain; if no such causal powers were replicated, no intentional state would ever be produced.

In the context of the ongoing debate about consciousness and the brain, some authors question why conscious states refer to the world external to the brain, but not to the internal states of the brain assumed to produce them (Velmans, 1993). This is also mentioned as the problem of knowing why "when I observe an object in the world the object as I observe it is entirely a product of my sensory (and conceptual) mechanisms, [but] it appears, however, to exist entirely independent of me; even its color appears as an intrinsic property of the object rather than as a mode of my perception" (Ellis & Newton, 1998: 421). Following Velmans' terminology, we call this phenomenon *projective intentionality*.

What kind of brain function could support mental intentionality? Would a naturalized (neurobiological) account of intentionality meet the demands of a philosophical theory of intentionality? In this paper, our objective is to pursue an account of intentionality in terms of brain and behavioral functions. This account is not a full scientific explanation, but an outline of a possible evolutionary explanation. We review results stemming from neurobiology in order to argue that intrinsically intentional systems in the brain are the evolutionary product of functions that support adaptive self-organized actions. The intentional mark of mental states is therefore conceived of as derived from biological functions that can be accounted for in non-intentional terms. Our account of intentionality begins by analysing two modalities of self-organized action present in living beings, namely *internally generated* and *goal-directed behavior*, evolutionarily leading to the development of specific brain structures, as e.g. networks for visually guided control of movement in the parietal cortex.

2. Internally Generated Behavior

A simple internally generated behavior is ultimately a property of every living cell, but a nervous system is necessary for the control of a multi-cellular organism in the performance of complex behaviors. The notion of an internally directed complex behavior implies that:

a) more than one pattern of behavior is available to the animal;
b) the choice of the pattern to be performed, in a given context, is made by the animal;
c) the operation of the neuronal mechanism that makes the choice is not determined by external stimulation;
d) the animal does not necessarily construct a representation of goals to be achieved.

An example of simple self-determined behavior is the sexual activity of animals. Animals have a behavioral repertoire that contains diverse forms of social behavior. Sexual behaviors occur in situations where other social behaviors are equally possible. It is engendered by electrochemical mechanisms in the nervous system, involving hormones and synaptic facilitation.

In the performance of the activity, the animal does not have to represent a reproductive goal. The activity is perfectly conceivable in terms of a positive feedback loop, where each performed step is rewarded with pleasant sensations. The performance of the sequence of acts does not require the representation of a goal to be attained at the end. Not even a representation of pleasant sensations is necessary, since the reward is closely entangled with the activity itself. The process may be compared to an operant conditioning with a null time delay for the reward.

Simple internally generated behavior is close to the concept of teleonomy (Mayr, 1989). "Teleonomy" means that biological functions are determined by a "program" encoded in the DNA. However, recent developments in molecular biology have shown that the DNA does not determine completely the function of cells; such functions are better characterized as *epigenetic*, i.e. the result of a dynamic interaction of genes, proteins and environment. Neurons, for instance, are highly sensitive to external information and do not develop all their potentialities in the absence of external stimulation. Therefore, in the context of our discussion we replace the concept of teleonomy by the concept a *self-organizing system* composed by the organism, its genes and the environment.

The neurobiological understanding of self-directed behavior is related to the conception of the brain as a dynamical system. In this view, the control of behavior is described in terms of neuronal activity approaching "attractors". In far-from-equilibrium systems, such as living beings, the increase of entropy is blocked by a flow of external negative entropy. The effect of the second law, in this context, is not driving the system to thermodynamic equilibrium, but to self-organizing processes leading to states where energy acquires a stable configuration. "Attractors" are the most stable classes of states; once the system reaches one of them there is a high probability that it will remain there for a reasonable amount of time.

Conceiving of brains as dynamical systems is a methodological simplification, since the physical components of such a system have not been formally defined. The problem lies in the complexity of brain structures, which encompass multiple levels of analysis. The solution found by Freeman (1995) was to assume a simplified but realistic neural network model of the brain. In the dynamical analogy, the number of attractors in the state space of the model corresponds to the number of different kinds of behavior. The structure of the state space can be designed to explain the transitions between such attractors, assuming a stable environment. As the animal switches from one kind of behavior to another, the trajectory of the dynamical system moves from one attractor to another.

An *orienting mechanism* (Sokolov, 1975) is responsible for the choice among stereotyped patterns of response, in the simplest modality of internally oriented behavior. In reptiles, it is possibly related to the function of basal ganglia, as proposed by McLean (1990) and Panksepp (1998). An example of an orienting system with a small number of attractors is found in the study of the behavior of small fishes. The behavioral state space of a fish or reptile, according to the McLean (1990) classification, contains only four kinds of behavior: feeding, fleeing, fighting and reproduction. An interesting model of fish behavior, examining the reaction of fish to their own mirror image, goes back to Tinbergen's studies of the innate behavior of the three-spined stickleback (Tinbergen, 1951).

A current experimental setting for the study of fish behavior makes use of Oreochromis Niloticus males, which compete for territory. One male fish is set in a rectangular aquarium with a partition, where one of the walls is a mirror (Volpato, 1997). The subjects are e.g. young (2 months of age) Oreochromis Niloticus specimens. At the beginning of the experiment the partition is down, and the fish explores the half of the aquarium that does not contain the mirror. After five minutes the partition is lifted. As the fish comes closer to the mirror wall, the image of a male fish is produced. When it perceives the image, it displays two kinds of fighting behavior, frontal and lateral. The behavior is, of course, also displayed by the mirror image. The perception that the "other" (the image) is aggressive increases the aggressiveness of the fish. An excitatory feedback between actions and perceptions occurs.

When two agonistic fish struggle, the consequence is the increase of stress, eventually leading to the death of one of the opponents. In the case of one fish and a mirror, there is also stress, but the outcome is an inhibitory turning point: as stress increases, the fish decreases the signals of aggressiveness; this phenomenon is of course reflected in the image, and influences the fish's next reaction. The inhibitory feedback process leads to an extinction of the aggressive behavior, and the fish begins a new exploratory process. Swimming through the aquarium, it may find food, probably changing from exploratory to feeding behavior. The feeding behavior also has two phases, excitatory and inhibitory feedback. The inhibitory phase begins with satisfaction, leading to extinction. At this moment, the fish begins a new exploratory behavior that leads it to the vicinity of the mirror, where another cycle of aggressive behavior begins.

Lorenz and Tinbergen's model of innate behavior, with hierarchically fixed patterns released by specific types of stimuli, is the first candidate for an explanation of the fish's fighting behavior (Lorenz, 1952). In natural contexts, the release of fixed patterns is related to social companionship. In the experimental case above, stimuli are absolutely constant, so any variation in the fish's perception is a consequence of its own previous action (e.g., if it changes the position of its body, it will see the same environment from a different perspective). The fish has a stereotyped reaction to the image, releasing the fixed pattern of fighting, and after some time (depending on individual variations) the central mechanism that released the pattern inhibits it. After the investigatory behavior, when the fish comes near the mirror for the second time, the fixed pattern is released again, followed by the inhibitory mechanism.

The internal feedback from action to perception presumed to exist in this kind of fish is the corollary discharge (Sperry, 1950). It has only the (hypothetical) role of stabilizing the visual field, and cannot, under normal circumstances, produce a new kind of behavior. In Sperry's experiment, circular swimming was produced, but one eye of the fish was covered, and the other rotated in 180 degrees. Moreover, the cognitive processing in the fish's brain, as in many other species with the same brain architecture, can only relate external signals with categories corresponding to fixed patterns (although it has been argued that it is possible for some fish species to have learning capacities not found in the observed species; see, e.g., Barlow, 1968, and Bitterman, 1968).

In our example of internally generated behavior, we identify a *proto-intentional function* – corresponding to the operations of the orienting mechanism – that only selects an option from a limited repertory, according to the perceived properties of a stimulus. In this case there is no need for "aboutness" of mental states, not even for mental states properly. However, it is necessary for brain activity to be directed towards the stimulus and its dynamical changes that elicit the continuous operation of the orienting mechanism, defining the kind of behavior to be adopted at each moment. This may be the first step in the evolution of intentionality.

3. Goal-Directed Behavior

A second step in the evolution of intentionality was made possible by means of the development of brain structures that provide the basis for a choice of behavior based on evoked emotional responses. In rodents, E. Sokolov (1975) and O. Vinogradova (1975) related the "orienting reflex" with the limbic system. In this expression, the word "reflex" does not mean a response elicited by external conditioning. On the contrary, it refers to an unconditional instinct that guides behavior in the following kinds of situations:

a) when a stereotyped pattern is initiated, without being triggered by a specific stimulus, e.g., a chicken searching for food, in the absence of any specific food stimulus;

b) when a stereotyped pattern is triggered by a specific stimulus, e.g., a predator-like sound inducing the prey to escape;

c) when there is a stereotyped reaction to a change in a habitual environmental setting, e.g., the miller who wakes up when the noisy mill stops working.

One role of the limbic system in rodents would be signaling changes among types of behavior. Sokolov and associates tested this hypothesis experimentally, focusing on cases b) and c) above (see diagrams in Sokolov, l975, p. 226, 233). An information-processing version of his theory is currently used in psychophysiology (see Hughdal, 1995; for the role of the amygdala, see LeDoux, 1994).

The emergence of emotions provides a basis for an inner selection of behavior (LeDoux, 1994; Panksepp, 1998), but in the absence of the representation of possible goals to be attained and their impact on future emotions, the choice process is largely a trial-and-error one. Not surprisingly, the phylogenetic evolution of the limbic system occurred together with the development of frontal cortical areas able to elaborate on the representation of goals. The mammalian brain, evolving over but preserving the limbic orienting mechanism, shows a gradual development of the associative areas of the neocortex supporting primate (and human) typical functions. These areas compose the *executive system*, involving an attentional subsystem, as well as a large network that supports declarative and working memory (temporal, parietal and frontal associative neocortical regions). The executive system (see D'Esposito and Grossman, 1996) was presumably segregated from and superposed to the earlier orienting mechanism, allowing larger flexibility of responses to environmental stimuli. Corresponding to this evolution, a second step in the evolution of self-organized action is *goal-directed behavior*. This is a special case of self-determined behavior when the animal *represents a goal* to be achieved and directs its activity towards the goal. The category applies only to animals that have the capacity of representing goals and planning a sequence of actions leading to the goal.

A goal is a state of affairs that is not available to the animal at the moment when it is represented. Regarding the relative "non-existence" of the represented goal, this modality of representation is not far from Brentano's concept. For animals that represent goals, our simple dynamical analogy above seems to be insufficient to account for their behavior, since the behavioral repertory becomes as large as their representational and implementational capacities, generating a complex behavioral state space. It is also important to consider that the sequences of actions of the animals are referring to the represented goal. This implies that the animal can perform actions that generate unpleasant sensations, if such actions are conceived as a necessary step to reach the intended goal.

The capacity for directing actions to a goal reveals the existence of a second cognitive function, besides the orienting mechanism. It is the *attentional executive function* (see Posner, 1995), responsible for the guidance of action and subsequent perception, according to intended goals of the organism. In primates, the attentional system involves a large network that encompasses the prefrontal cortex, its connections with parietal and temporal associative areas, and the cyngulate gyrus. Therefore, understanding goal-directed behavior would require the consideration of *two dynamical brain sub-systems that complement and oppose each other* – the orienting and the attentional executive systems (see Grossberg and Merrill, 1996).

A dynamical account of goal-directed behavior requires the consideration of such inter-connected dynamical systems, both opposed and complementary to each other. The distinction between the present state of affairs and the intended one (the goal) can be metaphorically conceived of as a kind of "opponent processing" between the systems (for a review of opponent processing, see e.g. Seymour, O'Doherty, Koltzenburg, Wiech, Frackowiak, Friston and Dolan, 2005). In a very simplified view, the excitation of attentional systems (prefrontal cortex and cyngulate gyrus) focusing on the goal to be attained, generates an inhibition of the orienting system. Such inhibition is necessary because the orienting system is evolutionarily tuned for stereotypical responses to immediate stimuli. The pursuit of a goal requires inhibition of automatic responses in favor of the release of actions that are necessary for achieving the goal. It is possible, of course, that both systems become compromised with each other, when a sub-goal happens to be compatible with the demand of the orienting system. On the other hand, strategies that systematically impose sacrifices upon the orienting system may disrupt the overall balance and fail.

The consideration of two coupled dynamical systems could therefore help explaining goal-directed behavior, but it is still insufficient to account for one aspect of the cognitive processing involved in the behavior: how is the animal able to represent something inexistent in the external world? This problem directly addresses the issue of "aboutness" of mental representations.

4. From Self-Organized Action to the "Aboutness" of Representations: the Example of Self-Initiated Locomotion

The "aboutness" of mental representations is a disposition of brain states that does not necessarily arise in direct connection to behavior; i.e., the “aboutness” of those states may exist even if a behavioral pattern is not elicited from them. Of course, such aboutness allows us to assign to those states *possible* or *prospective objects* in the world, and to support goal-directed behavior. Explicit memories, for instance, are endowed with aboutness, although not eliciting behavior or not taking part in an actual behavioral cycle.

In order to understand the possibility of the evolutionary transformation of self-organized action into mental intentionality, we avoid a common interpretation of the problem. Intentionality is frequently conflated with mental representation and, in so doing, some philosophers (following Brentano) claimed that no physical thing could be about another thing. Intentional states could not be brain states. However, there is no *a priori* reason to deny that “aboutness” could happen between physical beings. We are in the world, and so are our mental states. Why do the latter happen *in* the world and, nonetheless, are *about* something?

The riddle becomes less mysterious if we initially consider that an account of self-organized action provides us with an explanation for the advent of a distinction between the "inner" and the "outer". Once we are able show how self-organized action could generate such a distinction and how the sense of externality of our experience may arise from habituation and blockage of brain self-reference, we eschew most of the philosophical conundrums concerning the nature of intentionality.

How to explain the intentional character of memories, beliefs, desires and other cognitive states in terms of brain events, even when they are not connected to some kind of behavioral manifestation? An account of the nature of such brain states is likely to require more than what is proposed in the theory of attractors. It requires taking into account possible evolutionary processes that determined the functional architecture of the brain in such a way that its *internal* states are *directed towards the domain of adaptive interaction* with the environment, even when the present behavior is not being controlled by such states.

Therefore, in attempting to tackle the nature of “aboutness”, two issues need to be addressed: first, how brain states not connected to behavior can have “aboutness”, and secondly, how conscious processing, although internal to the brain, generates episodes which are experienced as occurring outside it.

With respect to the first problem, a central modality of self-organized action is *self-initiated locomotion*, an ancient functional motor skill, one that emerged in the evolutionary process as a life-strategy of diverse living beings in order to cope with the fulfillment of their basic biological needs. Self-initiated locomotion is an adaptive advantage, providing the possibility of seeking out new energy sources by changing to different environmental configurations. The important cognitive consequence of this skill is the possibility of the organism drawing a distinction between the "inner" and the "outer" in terms of "what moves" and "what may not move" relatively to a self-initiated motion.

Intentionality in the sense of “aboutness” includes *representing* objects and processes. The notion of *pragmatic representation* implies that perceptual objects are viewed under different modes of presentation, depending on the actions of the organism (Clark, 1996). Such modes of presentation reveal different properties of the object that would still be considered as ways of presenting the very same object. The different modes of presentation are related to different actions of the organism towards the object, i.e., the action of the organism towards an object influences or determines the means whereby the object is cognitively apprehended. The classical example of an intensional (with an s) context provided by the philosophical literature is the Fregean discussion of the "Morning Star" as contrasted to the "Evening Star". They are, of course, different modes of presentation of the planet Venus.

Current literature on this topic holds that one can have different beliefs about the "Morning Star" and the "Evening Star", although such representations pick out the same object in the world. One can entertain beliefs about the Evening Star that would not withstand if they were extended to the Morning Star. In other words, the possible truth of belief sentences about the Morning Star and the Evening Star differ in spite of the fact that they pick out the same object in the world, namely, Venus. The change in truth value of sentences involving the Morning Star and the Evening Star is affected by the presence of a pragmatic representation that mediates our cognitive relationship to an object in the world (Venus).

Some philosophers consider the generation of pragmatic representations a purely linguistic surface phenomenon. There are also philosophers who attempted to relate the intensionality (or multiple modes of presentation of an object) with the intentionality manifested in behavior. This second approach is worth exploring, since there may be a relationship between intensionality (with an s) and intentionality (with a t) involved in behavior. Some philosophers (Dretske, 1981) have maintained that if we explain how and why such intensional contexts are generated we will also have explained how representations are generated. Thus, we would have an explanation for “aboutness” as well, in so far as the latter is a pragmatic representation. The evolutionary principle at the genesis of such contexts is, we submit, the possibility of apprehending an object by moving around it, i.e., by self-initiated locomotion.

The generation of intensional contexts by moving around objects can be appreciated if we take into account a situation in which an organism is viewing an object "a" in a certain context and acquires the belief that "a" is F. When viewing it in another context the organism acquires the belief that "a" is G. However, it might be the case that such an organism does not acquire the belief that "a" is F and G. Certainly an explanation is needed here for the fact that it could believe that "a" is F without being G, even if in another situation it comes to learn the fact that it was indeed one and the same object "a" it came across in both situations. The reason why the organism could have different beliefs about the same object lies in the fact that it was not capable of realizing *a priori* that it was actually relating to the same object.

So viewed, the phenomenon can be understood as fundamentally rooted in differences between modes of presentation *generated by the organism's self-initiated* *locomotion*. The latter is what generates multifarious modes of presentation of an object while moving around it. Intensionality can thereby be conceived of as something rooted in the more primitive notions of moving around an object and a "failure" in our processes of inspection of objects in the world. Such an inspection does not allow us to equate *a priori* different modes of presentation as being presentations of the same object in a different pragmatic context.

Self-initiated locomotion allows us to understand the origin of intentionality as “aboutness”, i.e. our brain states being about something in different pragmatic contexts or in the perceptual absence of the denoted object. Could it be sufficient to explain representations that do not denote anything, i.e., representations about an inexistent object? For instance, in the case of the intensional representations "the aunt that died twenty years ago" and "the death of the unicorn", the difference lies in the fact that the former, although picking out something perceptually absent, could be re-enacted in some kind of effective behavior, whereas the second cannot.

The main characteristic of the latter kind of representation is the brain making an inductive (and/or abductive) inference from what exists to what does not exist. The possibility of making such an inference may well be supported by the capacity of representing perceptually absent objects. A possible line of research on this issue is relating the way we come to have meaningful thoughts about inexistent objects and the child's apprehension of the hidden sides of objects. Although such experimental research with infants has been developed since Piaget's pioneering efforts, the relation with philosophical discussions of intentionality still remains to be established.

5. Some Experimental Evidences for Brain Systems Characterized by an Intentional Modus Operandi

In the study of visual perception, the organization of different visual appearances of the same stimulus has been related (Ullman, 1984) to the organism's movements. One possible implication of the influence of action on perception is that the very idea of an invariant object would emerge from distinguishing variations in perception due to bodily movement from variations resulting from the object's movement. So viewed, bodily movement in self-initiated motion is what accounts not only for the concocting of visual episodes as forming an object, but also their invariant properties identified from the organism's variation of perspectives.

The same role is played by self-initiated motion in auditory perception. Whenever we hear a sound, we hear not only the pitch and the tone but also localize the source of the sound. Such localization depends on binaural mechanisms and also on the organism's possible shifts in spatial location relative to a sound source. Self-initiated locomotion allows to differentiate whether it was the sound source that moved or the organism that changed place in the environment.

Empirical evidence in favor of the view that self-initiated motion plays a central role in the organism's apprehension of physical objects in the world was advanced by a classical series of experiments carried by Richard Held and Alan Hein at MIT (Held and Hein, 1958; Hein and Held, 1962; Held and Hein, 1963). This work points to the importance of self-motion in several cognitive tasks, including the development of motor skills, organization of perception and the individuation of physical objects. For example, Held devised an experiment with humans in order to show that "the importance of body movement and particularly of self-produced movement derives from the fact that only an organism that can take account of the output signals to its own musculature is in a position to detect and factor out the decorrelating effects of both moving objects and externally imposed body movement" (Held and Hein, 1963, p. 378).

Using a prism, Held and Hein (1963) conducted an experiment where the hand of a subject was moved constantly, and such a movement was perceived as independent of actual motion that could be taking place. The subject could not control the changes in his visual inputs, because of the effect of the prism. The result of the experiment pointed to the deterioration in coordination between eye and hand under conditions of active movement. Furthermore, it showed the importance of the correlation between movement and sensory feedback in maintaining accurate coordination. The series of experiments demonstrate a role of motor-sensory feedback in visual adaptation, as well as the role of active movements in facilitating this process. Such experiments with visual adaptation were also expanded to situations involving movements of the entire body and not just the arm or the hand. They reinforce the conclusion of existence of a link between motor and visual mechanisms in the central nervous system (see also recent studies pointing to the same conclusion, as Gentilucci, Chieffi, Daprati, Saetti and Toni, 1996; Tse, Cavanaugh and Nakayama, 1997).

Since we consider self-initiated motion a biological function selected by its advantages in the fulfillment of basic needs, intentionality would thus arise as a result of primitive non-intentional forms of agency. The circumstances in which such a condition occurs begin with the notion of an organism's attempts to satisfy those biological needs by moving around. As an evolutionary consequence of this habit, the nervous system comes to incorporate a lead to the world that allows the organism to behave successfully. Such interaction between organism and environment could start as a trial-and-error task or as a kind of cycle where the effects of those movements are registered by the nervous system and form a kind of feedback loop that selects actions until behavioral adequacy is attained. The registering of those effects plays an essential role in constraining new cycles of trial-and-error, thus narrowing the set of behaviors to a set of adequate behaviors. Once such adequacy is attained, the trial-and-error cycle may evolve into a disposition to behave in a certain way in the environment.

Initially, dispositions may constitute just a motor skill that defines a trajectory between the informational input and the reaction of the organism – a cycle where informational input triggers a certain pattern of behavior. This is already a gain for the satisfaction of the biological needs of the organism, and the nervous system could pre-empt the trial-and-error cycle by incorporating such a disposition in the form of a reflex arc.

The possible gain in behavioral adequacy is, nevertheless, constrained by a lack in terms of resilience. If the environment changes, the triggered behavior may not be beneficial for the organism. The next step in the evolution of behavior is therefore the induction of brain mechanisms controlling behavior to be *resilient*, i.e., prone to possible environmental changes. Based on the emergence of such resilience, representational "aboutness" can be understood in terms of a *residual directness*, the by-product of an interrupted or incomplete behavior cycle, as classically proposed by Miller, Gallanter and Pribam (1960). These authors suggest that directness is a pervasive property of brain states, ranging from reflex functions to goal directed behavior and, finally, to "aboutness". Our hypothesis – following the classical hypothesis of Miller et al. (1960) – is that representational "aboutness" would evolutionarily depend on more basic dimensions of intentionality.

In the present state of evolution, we can have dispositional states in the absence of interrupted behavior, or even in the absence of any ongoing behavior related to such states. According to our hypothesis, the present state of affairs is a result of an evolutionary process that led to the emergence and maintenance of the resilience of brain states in order to support adaptive actions.

This fact implies that an appropriate explanation of "aboutness" should take into account the evolutionary emergence of new brain structures and functions able to represent possible actions independently of ongoing behavior. Cognitive Neuroscience has progressively studied diverse brain systems that fulfill this requirement: neuronal networks localized in the anterior cingulate cortex, premotor cortex, posterior parietal, inferior temporal and prefrontal cortex, with a possible involvement of the cerebellum and basal ganglia.

Advances in the studies of functional anatomy and the proposal of more realistic theoretical models have led to a distinction between brain areas that represent the intention of performing an action and the areas that control the performance of the action (see Mazzoni, Bracewell, Barash and Andersen, 1996; Colby, Duhamel and Goldberg, 1996; Snyder, Batista and Andersen, 1997; Cisek, Grossberg and Bullock, 1998). These findings corroborate observations made by Held, Hein, Ullman and others (see Jeannerod, 1997) about the relation of perception and action. Such a concept of intentional representation was incidentally refuted by the most influential researcher – J.J. Gibson (1979) – who stressed the importance of action in perception.

The consideration of motor influence upon perception, assumed by relatively few researchers in the past, became one of the main areas of research in contemporary cognitive neuroscience. Commenting on a recent study by Ballard, Hayhoe, Pook and Rao (1997) on "deictic codes", Goodale remarked: "It joins nicely the fields of motor control and cognition [...] It also makes evolutionary sense by suggesting that mechanisms which evolved for the distal control of movement might have been co-opted (in both a literal and figural sense) for the computations underlying the cognitive life of the animal" (Goodale, 1997).

Another field of research has recently been developed, focusing on prefrontal areas responsible for an anticipation of possible future events (Hasegawa, Blitz, Geller and Goldberg, 2000). According to Miller and Cohen, "cognitive control stems from the active maintenance of patterns of activity in the prefrontal cortex that represent goals and the means to achieve them. They provide bias signals to other brain structures whose net effect is to guide the flow of activity along neural pathways that establish the proper mappings between inputs, internal states, and outputs needed to perform a given task" (Miller and Cohen, 2001, following the research made by Rainer, Rao and Miller, 1999). The representation of goals is assumed to involve a cognitive operation called *prospective memory*, defined by Burgess, Quayle and Frith (2001) as "the functions that enable a person to carry out an intended act after a delay".

6. The Problem of Projective Intentionality

The problem of projective intentionality (Velmans, 1993) refers to explaining how processes going on in the brain are experienced as occurring in the external world. For obvious adaptive reasons, brain states should not produce sensations of their own workings. Projective intentionality has an important function of allowing the organism to direct action upon objects in the world, and not towards the neuronal machinery that supports cognition.

The generation of an *externally-referenced content* when a stimulus is received, by the exclusion of the inner brain's workings, gives rise to the "first-person" or "phenomenal world" (Chalmers, 1996). The phenomenal world is generated by the activity of the nervous system, but it is experienced as standing outside. Signals impinging on the peripheral sensors are processed by the central nervous system, and nonetheless their generators are perceived as standing outside.

It is generally accepted in neuroscience – following classical studies made by Muller in the XIXth century – that the localization of a percept is determined by the *afferent pathway*. For example, a touch sensation is localized in the region of the skin where contact occurred, and not in the brain receptors of the signal. This assumption helps to describe the phenomenon but does not attempt to formulate an explanation.

From a physicalist perspective, there is an apparent paradox in the fact that sensations are not localized in the place where they are presumably produced: how can a signal that is physically located in the brain *be perceived as* being outside the brain? A biological fact that could help dissolve the paradox is that the brain does not have internal sensors. In other words, all the input to the central nervous system, in normal situations, comes from the peripheral sensors (eye, ear, touch sensors at the skin, etc.). However, the problem that remains to be explained is why percepts are localized at the location where the peripheral sensors are situated (or "in front of" them), and not *where the signal is sent to*.

A possible solution comes from noting that the influx of external signals to the brain is necessary for the production of percepts. Endogenously produced action potentials and rhythmic patterns do not produce any sensations by themselves (except in dreaming and hallucination, two phenomena we will not discuss here). In the case of direct brain stimulation in patients under surgery, or with chronically implanted electrodes (Penfield and Boldrey, 1937), the external stimulation provoked by the neurologist is necessary to produce "phantom" sensations (i.e., sensations of an inexistent external object). Our hypothesis is that recurrent relations maintained by the active organism with the environment indirectly support the projective relation. This hypothesis may be developed in terms of a process of *habituation* that engendered *a peculiar structure of neuronal networks*, such that only inputs that match patterns from the peripheral sensors are able to reach a certain threshold of activity corresponding to conscious perception.

An exceptional situation, in evolutionary terms, is the ingestion of substances – hallucinogenic drugs – or chemical disbalancements – as in schizophrenia – that produce sensations endogenously to the brain. In these cases, projective intentionality is present, since all endogenously produced sensations are subjectively or phenomenically referred to the external world, but the recurrent interaction with the external world is perturbed at the moment when the sensations are generated. However, it is important to consider that the subjects who experience such endogenously-generated sensations interacted with the external world during their entire life, allowing the brain intentional mechanisms to develop.

An analysis of how habituation works in the ontogenetic scale (see Sokolov, 1975; Gray, 1995) suggests an adaptation of brain perceptual mechanisms for the recognition of distal stimuli based on information carried by proximal signals. For instance, when a person holds a tool, (e.g. a hammer), he/she first experiences contact with the tool, but as he/she uses it, he/she becomes habituated and then the attentional focus turns to the interface between the hammer and the part of the environment that offers resistance to it, e.g. the nail.

The signals generated internally to the brain are the most proximal stimulus, the ones to which we are most habituated. Is it possible that in the distant evolutionary past, or in some artificial situation, the brain could perceive its own workings? The answer seems to be positive, since it is possible for the brain to monitor its own activities, by means of a technique called "biofeedback" (see e.g. Sterman and MacDonald, 1978; Hauri, Percy, Hellekson, Hartman and Russ, 1982; Ayers, 1991). Another example is closing the eyes and pressing the eyeballs with a finger; some phantom imagens can usually be seen, presumably produced by remaining excitatory activity of retinal cells. In both examples, the nervous system perceives a fragment of its own workings, but in normal adaptive situations habituation mechanisms preclude brain self-perception. These examples serve to demonstrate that it is not impossible for the brain to perceive parts of itself, hence suggesting that projective intentionality would be generated by habituation mechanisms that *block* such self-perception.

Discussing computational mechanisms in the cortex, Phillips and Singer (1997) wrote: "the local processors are in effect discovering distal variables and relationships [...] these foundations do not constitute intentional representation proper because such local processors do not distinguish between the signals they receive and the distal causes from which those signals arise" (1997, p. 663). This remark provides one clue for the discovery of the brain mechanisms underlying projective intentionality. It could be supported by inhibitory mechanisms necessary to block the perception of signals internal to the brain. For instance, it is well known that the existence of inhibitory interneurons restrict feedback loops in cortical columns. Cortical layer 4 seems to be protected by inhibitory neurons, so as to receive input only from the thalamus; more precisely, it may be inhibited soon (around 12 ms) after receiving an input, as a way of avoiding excitation by the same signal re-entering from layer 5 to the superficial layers.

Such inhibition could favor the tangential spreading of the excitatory potential to other columns, generating sequential cognitive processing.

An interesting consequence of habituation processes is that the perceptual space of animals is progressively shaped by the space of their actions. If projective intentionality is supported by inhibitory mechanisms related to learning processes, then in a familiar environment the space where animal localize the stimuli is learned to be the same space where they act. For instance, a chained dog, although having the same brain architecture as a stray one, *perceives* the world differently – an observation made by Kurt Goldstein and discussed by philosopher Merleau-Ponty (1945). In other words, the *structure of the action space*, defined by the relationship between body and environment, *shapes the structure of the perceptual space*, both in phylogenetic and ontogenetic scales. As a consequence, when we find ourselves in a strange environment a feeling of disorientation emerges, since the perceived space does not match the action space to which our neuronal networks were previously habituated.

Based on such intuitive observations, as well as on the scientific data reviewed above, we suggest the possibility of a neuroscientific approach to intentionality.

João de F. Teixeira and Alfredo Pereira Jr
Federal University of São Carlos, Brasil
São Paulo State University (UNESP), Brasil

jteixe@terra.com.br / apj@ibb.unesp.br

References

Ayers, M. (1991) 'A Controlled Study of EEG Neurofeedback Training and Clinical Psychotherapy for Right Hemispheric Closed Head Injury', *Los Angeles: National Head Injury Foundation Annual Conference*.

Ballard, D.H., Hayhoe, M.M., Pook, P.K. & Rao, R.P.N. (1997) 'Deitic Codes for the Embodiment of Cognition', *Behavioral and Brain Sciences* 20, 723-767.

Barlow, G.W. (1968) 'Ethological Units of Behavior', IN D. Ingle (ed.), *The Central Nervous System and Fish Behavior*, Chicago: University of Chicago Press.

Bitterman, M.E. (1968) 'Comparative Studies of Learning in the Fish', IN D. Ingle (ed.), *The Central Nervous System and Fish Behavior*, Chicago: University of Chicago Press.

Brentano, F. (1925/1973) *Psychologie von empirischen Standpunkt / Psychology from an empirical standpoint*, translated by A. Pancurello, D. Terrell, L.L. McAlister, New York: Humanities Press.

Burgess, P.W., Quayle, A. & Frith, C.D. (2001) 'Brain Regions Involved in Prospective Memory as Determined by Positron Emission Tomography', *Neuropsychologia* 39 (6), 545 - 555.

Chalmers, D.J. (1996) *The Conscious Mind*, New York: Oxford University Press.

Cisek, P., Grossberg, S. & Bullock, D.S. (1998) 'A Cortico-Spinal Model of Reaching and Proprioception Under Multiple Task Constraints', *Journal of Cognitive Neuroscience* 10, 425-444.

Clark, A. (1996) *Being There: Putting Brain, Body and World Together Again*, Cambridge, MA: The MIT Press.

Colby, C.L., Duhamel, J.R. & Goldberg, M.E. (1996) ‘Visual, Presaccadic and Cognitive Activation of Single Neurons in Monkey Lateral Intraparietal Area’, *Journal of Neurophysiology* 76, 2841-2852.

Dennett, D.C. (1969) *Content and Consciousness*, London: Routledge and Kegan Paul.

D’Esposito, M. & Grossman, M. (1996) ‘The Physiological Basis of Executive Functions and Working Memory’, *The Neuroscientist* 2, 345-352.

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: The MIT Press.

Ellis, R. and Newton, N. (1998) ‘Three Paradoxes of Phenomenal Consciousness:Bridging the Explanatory Gap’, Journal of Consciousness Studies, 5 (4), 419-42.

Freeman, W.J. (1995) *Societies of Brains*, Hillsdale: Lawrence Erlbaum.

Gentilucci, M., Chieffi, S., Daprati, E., Saetti, M. & Toni, I. (1996) ‘Visual Illusion and Action’, *Neuropsychologia* 5 (34), 369-376

Gibson, J.J. (1979) *The Echological Approach to Visual Perception*, Boston: Houghton-Mifflin.

Goodale, M. (1997) ‘Pointing the Way to a Unified Theory of Action and Perception’, *Behavioral and Brain Sciences* 20 (4), 749-750.

Gray, J.A. (1995) ‘The Contents of Consciousness: A Neuropsychological Conjecture’, *Behavioral and Brain Sciences* 18, 659-722.

Grossberg, S. & Merrill, J.W.L. (1996) 'The Hippocampus and Cerebellum in Adaptively Timed Learning, Recognition, and Movement', *Journal of Cognitive Neuroscience* 8, 257-277.

Hauri, P.J., Percy, L., Hellekson, C., Hartman, E. & Russ, D. (1982) 'The Treatment of Psychophysiologic Insomnia with Biofeedback: a Replication Study', *Journal of Biofeedback and Self Regulation* 7 (2), 223-236.

Hasegawa, R.P., Blitz, A.M., Geller, N.L. & Goldberg, M.E. (2000) 'Neurons in Monkey Prefrontal Cortex That Track Past or Predict Future Performance', *Science* 290, 1786-1789.

Hein, A. & Held, R. (1962) 'A Neural Model for Labile Sensorimotor Coordinations', IN *Biological Prototypes and Synthetic Systems, Vol. 1,* New York: Plenum Press.

Held, R. & Hein, A. (1958) 'Adaptation of Disarranged Hand-Eye Coordination Contingent Upon Re-Afferent Stimulation', *Perceptual and Motor Skills* 8, 87-90.

Held, R. & Hein, A. (1963) 'Movement-Produced Stimulation in the Development of Visually Guided Behavior', *Journal of Comparative and Physiological Psychology* 5, 872-876.

Hughdal, K. (1995) *Psychophysiology*, Cambridge, MA: Harvard University Press.

Jeannerod, M. (1997) *The Cognitive Neuroscience of Action*, Oxford: Blackwell

LeDoux, J.E. (1994) 'The Amygdala: Contributions to Fear and Stress', *Seminars in the Neurosciences* 6, 213-237.

Lorenz, K. (1952) 'The Past Twelve Years in the Comparative Study of Behavior', IN C.H. Schiller (ed.), *Instinctive Behavior*, New York: International Universities Press.

Mayr, E. (1989) *Toward a New Philosophy of Biology*, Cambridge, MA: Harvard University Press.

Mazzoni, P., Bracewell, R.M., Barash, S. & Andersen, R.A. (1996) 'Motor Intention Activity in the Macaque's Lateral Intraparietal Area I. Dissociation of Motor Plan from Sensory memory', *Journal of Neurophysiology* 76, 1439-1457.

McLean, P.D. (1990) *The Triune Brain in Evolution*, New York: Plenum Press.

Merleau-Ponty, M. (1945) *Phenomenologie de la Perception*, Paris: Gallimard.

Miller, E.K. & Cohen, J.E. (2001) 'An Integrative Theory of Prefrontal Cortex Function', Annual Review of Neuroscience 24, 167-202.

Miller, G.A., Gallanter, E.H. & Pribram, K. (1960) *Plans and the Structure of Behavior,* New York: Rinehart and Winston.

Panksepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions*, New York: Oxford University Press.

Penfield, W. & Boldrey, E. (1937) 'Somatic Motor Sensory Representation in the Cerebral Cortex of Man as Studied by Electrical Stimulation', *Brain* 60, 389-443.

Phillips, W.A. & Singer, W. (1997) 'In Search of Common Foundations for Cortical Computation', *Behavioral and Brain Sciences* 20, 657:722.

Posner, M.I. (1995) 'Attention in Cognitive Neuroscience: An Overview', IN M.S. Gazzaniga (ed.), *The Cognitive Neurosciences*, Cambridge, MA: The MIT Press.

Rainer, G., Rao, S.C. & Miller, E.K. (1999) 'Prospective Coding for Objects in Primate Prefrontal Cortex', *Journal of Neuroscience* 19, 5493-505.

Searle, J. (1980) ‘Minds, Brains and Programs’, *Behavioral and Brain Sciences* 3, 417-424.

Searle, J. (1983) *Intentionality: An Essay on the Philosophy of Mind.* Cambridge: Cambridge University Press.

Searle, J. (1992) *The Rediscovery of Mind*, Cambridge, MA: The MIT Press.

Seymour, B., O'Doherty, J.P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K. & Dolan, R. (2005) ‘Opponent Appetitive-Aversive Neural Processes Underlie Predictive Learning of Pain Relief’, *Nature Neuroscience* 8 (9), 1234-40.

Snyder, L.H., Batista, A.P. & Andersen, R.A. (1997) ‘Coding of Intention in the Posterior Parietal Cortex’, *Nature* 386, 167-170.

Sokolov, E.N. (1975) ‘The Neuronal Mechanisms of the Orienting Reflex’, IN E. Sokolov & O. Vinogradova (eds.), *Neuronal Mechanisms of the Orienting Reflex*, New York: Lawrence Erlbaum.

Sperry, R.W. (1950) ‘Neural Basis of the Spontaneous Optokinetic Response’, *Journal of Comparative Physiology* 43, 482-489.

Sterman, M.B. & MacDonald, L. (1978) ‘Effects of Central Cortical EEG Feedback Training on Incidence of Poorly Controlled Seizures’, *Epilepsia* 19, 207-222.

Tinbergen, N. (1951) *The Study of Instinct*, Oxford: Clarendon Press.

Tse, P., Cavanaugh, P. & Nakayama, K. (1997) ‘The importance of parsing in high-level motion processing’, IN T. Watanabe (ed.), *High-Level Motor Processing*, Cambridge, MA: The MIT Press.

Ullman, S. (1984) ‘Visual Routines’, *Cognition* 18, 97-159.

Velmans, M. (1993) ‘A Reflexive Science of Consciousness’, IN G. Bock & J. Marsh (eds.), *Experimental and Theoretical Studies of Consciousness*, Chichester: John Wiley.

Vinogradova, O.S. (1975) ‘The Hippocampus and the Orienting Reflex’, IN E. Sokolov & O. Vinogradova (eds.), *Neuronal Mechanisms of the Orienting Reflex*, New York: Lawrence Erlbaum.

Volpato, G. (1997) ‘Fighting Behavior of Oreochromis Niloticus (Video)’, Botucatu: State University of São Paulo (UNESP).