A process based architecture for an artificial conscious being

Riccardo Manzotti
LIRA-Lab, DIST, University of Genoa,
Viale Causa 13, I-16145 Genova (Italy)

Abstract

A conscious being is a system that experiences (feels) something. In order to build an artificial conscious being we need to give an account of what it is to experience or feel something. Any project that aims to design an artificial conscious being thus needs to take issue with the notion of experience or feeling. As I argue in the following, for the purposes of robotics this task can be profitably approached if we leave behind the dualist framework of traditional Cartesian substance metaphysics and adopt a process-metaphysical stance. I begin by sketching the outline of a process-ontological scheme whose basic entities are called 'onphenes'. From within this scheme I formulate a series of constraints on an architecture for consciousness. An architecture abiding by these constraints is capable of ontogenesis driven by onphenes. Since an onphene is a process in which the occurrence of an event creates the conditions for the occurrence of another event of the same kind, an onphene-based architecture allows for external events to provoke the repetition of other events of the same kind. In an artificial conscious being, this propensity to repeat events can be considered as a functional reconstruction of motivation. In sum, if we base the architecture for an artificial conscious being on onphenes, we receive a system that experiences (feels) and is capable of developing new motivations. In conclusion I present some experimental results in support of this claim.

1. Is consciousness indeed an irresolvable mystery?

What is consciousness? How is it possible that a part of reality (the conscious subject) has an experience of some other part of reality? Why is the phenomenon of consciousness so elusive within a physicalist frame of reference? One of the main aims of this paper is to suggest that consciousness loses some of its theoretical mysteriousness if one abandons two traditional assumptions: i) that reality and its representation are two different entities, and ii) that reality is constituted by static things (which are occasionally engaged in dynamic interactions). In short, I want to argue here for the claim that the problematic character of consciousness diminishes to the extent to which we succeed in relinquishing (i) a dualistic stance in (ii) the traditional object-ontological (substance-ontological) framework. As I shall try to show in the following, it is possible to challenge both of these assumptions and present an alternative ontological framework in which basic elements of a theory of the mind can be profitably formulated, and from which some guidelines for experimental work in the field of artificial agents can be derived.

The traditional disciplines which examine the problem of consciousness are philosophy, cognitive science and neuroscience (Chalmers 1996; Editor 2000; Crick and Koch 2003; Zeki 2003). However, more recently there is growing awareness about the central role of robotics for the study of consciousness: "to understand the mental we may have to invent further ways of looking at brains [and we] may even have to synthesize artifacts resembling brains connected to bodily functions in order fully to understand those processes" (Edelman and Tononi 2000). Since the construction of a conscious artifact should help us to understand the processes of thought itself, the engineering approach to the problem of consciousness – i.e. the attempt to design and build an artificial conscious being – is receiving increasing attention. (Steels 1995; Aleksander 1996; O'Brien and Opie 1997; Manzotti, Metta et al. 1998; Schlagel 1999; Aleksander 2000; Martinoli, Holland et al. 2000; Togawa and Otsuka 2000; Aleksander 2001; Buttazzo 2001; Manzotti and Tagliasco 2002; Perruchet and Vinter 2002).

Most of the past engineering approaches to the construction of cognitive agents did not address the issue of phenomenal experiences as such, but rather focused exclusively on the analysis of behavior (Brooks 1991; McFarland and Bosser 1993; Arkin 1999). However, phenomenal experience is one aspect of agents which cannot be reduced merely to their behavior. The philosopher David Chalmers argued that there are two orders of problems in the study of the mind: one is the cognitive-behavioral-functional problem which he labels the 'easy problem', the other is the phenomenal problem labeled the 'hard problem' (Chalmers 1996)). The engineering approach to consciousness cannot avoid tackling the 'hard problem' for a conscious being is an entity that *feels*, not an entity that *does* something. The 'hard problem' is even more of an obstacle in this context since engineers have traditionally constructed 'objects' and consequently lack even the conceptual tools required to deal with the design of 'subjects.' Until now, the world of engineering simply had no place for phenomenal experiences.

That consciousness and phenomenal experience exist we know by introspection. We are 'open to the world', we experience the smell of flowers, the color of the sky, the meaning of a sentence. Furthermore we do not have any evidence that our consciousness could not be realized in anything else but a human body. There are no theoretical arguments tying phenomenal consciousness exclusively to DNA-

based organisms. This lack of negative evidence opens up the new research area of artificial consciousness.

Instead of adopting the traditional idioms and tools used in the discussion of the 'hard problem', researchers in artificial consciousness begin best by putting the standard conceptual framework under close critical scrutiny. This is the thrust of my overall argument in this paper, which I present in five steps. In a first step I outline the limitations of the current conceptual framework for the understanding of consciousness. Then I sketch an alternative framework based on a kind of process named 'onphene'. In a third step I present core elements of an account of consciousness based on onphenes. In step four I describe the architecture of an artifical device which implements an onphene-based account of consciousness, i.e. an architecture based on motivations. In the last and final step I explain the design of the experimental setup. To anticipate the overall conclusion of this argument, I will suggest that consciousness as something different from reality is an invention which stems from a impoverished conception of reality itself. The problem of consciousness has been created by the hypothesis of an abstract physical world of purely quantitative objects or substances.

2. Representation and physical world: an unnecessary dichotomy

In order to model conscious experience, we need to concentrate on three characteristics that in combination furnish a working hypothesis about the nature of conscious experience. As conscious beings we know that experience (i) is always an experience of something i.e. every experience has a given content and (ii) experiences can be phenomenally distinguished on the basis of their content. For instance, between the experience of drinking lemonade or looking at a painting by Mondrian there is a difference in their content. Further, setting aside extreme forms of skepticism, it is an introspective part of our natural epistemic stance that (iii) the content experienced represents external events (Dennett 1969; Block 1988; Dennett 1988; Fodor 1990; Chalmers 1996; Bickhard 2001), at least in ordinary perception. In a sufficiently open sense of the term, introspection commits us to the 'representationalist' standpoint. The claim that conscious experience occurs, that it has a phenomenal content, and that phenomenal content is tied to what happens, can be summed up as the thesis that to be conscious of something is to have a representation of that something.

On the basis of these considerations, we can put forth two hypotheses: (1) there are occurrences of events that correspond to contentful phenomenal experiences; (2) these occurrences represent something. According to this point of view, an event has a phenomenal content only if it has a representational role (Fodor 1981; Millikan 1984; Dretske 1993; Dretske 1995; Clark and Tornton 1997; Bickhard 1999). Hence conscious experience is the occurrence of events with phenomenal content and, by implication, with a representational role.

The notion of a representation is commonly taken to be more than an abbreviation for the claims (i) through (iii) above. Rather, it is used as an explanatory notion with a meaning of its own--a representation is something that presents (or re-presents) something else. If we adopt an object ontology and assume that the world is composed only of objects, then a representation is be an object which presents another object. Yet how could we make sense of the kind of 'presentation' involved here, how could we think of an object 'referring' to another? One way to approach this question is by taking representation to be somehow effected by similarity. This approach – termed the 'copy theory of representation' by Nelson Goodman (Goodman 1974) –

does not get us anywhere, however, since the identity principle establishes that an object is just itself, and that no object can be another object at the same time.

From within an object-ontological setting the only other way to understand representation is to view it as a relation. The representing entity and the represented one are linked by some kind of abstract relation (semantics, aboutness, or intentionality). However, this is in conflict with a commitment to physicalism which admits one relation only, namely, causality. Between events there are only causal relations, there are no intentional, teleological, formal or semantic relations. It is not a coincidence that most of the attempts to naturalize semantics, perception and representation are based on some kind of causation (Grice 1961; Armstrong and Malcom 1984; Haybron 2000). However, if we accept that representation is a species of causation then representation becomes an ubiquitous phenomenon.

If we want to hold on to the insight that phenomenal experiences correspond to the occurrence of representations, while at once holding on to a commitment to physicalism, a physical interpretation of representation must be found. Obviously the classical dualist Cartesian model of representation as a relation between physical and mental items is not helpful here. However the dualistic model of representation (the representing item is something different from what it represents) is unavoidable given an underlying ontology that divides the world in separate objects.

Two entities are separate if their existences, in a given instant t, are mutually independent. (If you would destroy your computer, nothing would happen to mine in the same instant). All objects which do not stand in parts/whole relations are separate. If an object ontology is accepted the dualist model is the only way coherently to make sense of the introspective data of conscious experience. As long as we describe ourselves as entities that are separate from what they are conscious of, we have taken on board the supposition that experience is some kind of duplication of the external world inside the internal domain of the subject.

In fact, the dualism of the external domain and the internal domain is independent of the additional qualification of the external as physical and the internal as mental. It is a contingent detail that in the XVIIth century the internal domain could only be assessed as mental while nowadays we also have a neurophysiological description of this domain. On the basis of an object ontology representation always implies the existence of a dualist counterpart (either a copy or a relatum) of the world, no matter how that counterpart is characterized. In this respect current neuroscientists assign to the brain the same role Descartes attributed to the *res cogitans*. However, since it is quite clear that the brain taken as a physical object cannot contain copies or isomorphic relational counterparts of the external world (i.e of something with completely different physical properties). Thus it remains a mystery how the brain can represent the external world.

For an object-based account of representation the only viable strategy is to leave physicalism behind. One might suppose, then, that in the brain there are qualia or pure phenomenal qualities, the modern version of the secondary qualities, which can only be identified in terms what they represent and thus are a label for a problem rather than a solution. Alternatively, one might embrace a functionalist stance according to which brain states are representation of something in the world due the fact that they a certain functional role for the organism. The functional domain is an abstract domain of input/output relations built on top of, and always extraneous to, physical events. However, functional roles do not account for the introspective difference between conscious and subconscious perception.

In short, traditional object or substance ontologies are committed to explaining representations (and hence the mind) in a dualist fashion. Conversely a Cartesian dualist standpoint is naturally compatible with a substance ontology (even though it does not entail the latter). Consider the following three claims:

a) the world consists of separate substances or objects
b) the mind represents the world (or the mind is equivalent to a set of representations of the world)
c) representations are different from what they represent (dualism)

As worked out consistently in XVII century metaphysics, a commitment to a) and b) entails a commitment to c). If the mind represents the world, and the world and the mind are made of separate entities, the mind must be a separate entity from the world; thus representations must be separate (and hence different) from what they represent.

However, is claim (a) indeed an *a priori* truth? We can reject the assumption that the world made of separate substances for at least the following three reasons. First, there is enough evidence from microphysics to militate against classical objects or substances as a type of fundamental entities (Cramer 1988; Zohar 1990; Stapp 1998; Auletta 2000). Second, the claim is no logical necessity—as we shall see presently, a different ontology can be formulated. Third, besides the problem of representation there are a number of fundamental *ontological* difficulties arising for object ontologies (Seibt 1990).

If (a) is rejected, it is important to settle for the right type of alternative ontology. An event ontology, for example, gets us from the frying pan into the fire. From a scientific point of view, the idea of a single event is a nomological absurdity. In contemporary science we cannot speak of anything which is not the object of an experiment, the result of a measurement, the target of an observation, or a postulated interaction whose results we observe. If something is not directly or indirectly observed, it is not part of what is empirically known. However, in order to be observed an event must be in relation with other events. It must in some fashion 'present' itself to other events. But then we must admit that in science there are no singular events—events are derived entities, namely, interactions of processes. Singular events or autonomous static objects are abstractions: like the Euclidean point or line, which are not part of the real world. Unfortunately these abstractions have been misunderstood as the real world, in the sense of Whitehead's 'fallacy of misplaced concreteness' (Whitehead 1925).

In the following I will argue that once we replace the traditional object-ontological framework with a suitable process-based alternative we can deal with the relation between mind and the world without falling in the dualistic trap. I advocate the following alternative set of assumptions:

a) the world is an assembly of processes which are not necessarily separate
b) the mind represents the world (or the mind is equivalent to a set of representations of the world)
c) representations are not different from what they represent (monism)

Assumption (b) has remained unchanged, while (a) has been changed and, as a result, (c) also. In the following I will try to show how a process ontology permits us to: i) account for representation without dualism, ii) treat the mind as a set of representations; iii) ensure true knowledge of the world without solipsism or dualism.

3. A process ontology for representation: reciprocal causation or 'onphene'

In order to introduce the basic notions of our process ontology let us start out with some simple observations about everyday physical processes in which the absence of any form of dualism, or even of duality of domains, is intuitively clear. Subsequently we will use these illustrations to introduce a new categorization for conscious experience.

The canonical illustration for a physical phenomenon in which the physical continuity between the represented object/event and the representing object/event is quite evident is the rainbow (Insert Figure 1). When the sun is sufficiently low on the horizon and sheds its rays at a right angle on a sufficiently big volume of water droplets suspended in the air, an observer (either a human being or a camera) can see a colored arch. All drops of water reflect the sunlight in the same manner, yet only those that are in a particular geometrical relation between the observer position and the direction of the sun rays are seen as part of the rainbow. The rainbow cannot be defined in any meaningful sense other than from the point of view from which it is seen. In this sense the rainbow, although constituted by a set of physical entities (drops of water in space reflecting light in a certain way), cannot be defined without knowing where and how it will be seen. For instance, it is not possible to see oneself as flying under or around a rainbow. Furthermore, the rainbow is a private physical phenomenon because two different observers always see two (however slightly) physically different rainbows. Since two separate observers occupy two different positions in space, they select different rays of light and accordingly different drops. Rainbows thus cannot be said to exist independently of the act of observation. In fact, if there were no eyes (or camera) looking at the rainbow, the relevant set of droplets of water would not produce any effect and the phenomenon called 'rainbow' would not exist at all. Even if it were possible to argue that an expert physicist who knew the drops' position, the sun's position and the observer's position might be able to calculate the projection of the rainbow on the observer's retina, such a calculation would require the knowledge of the observer's position as an essential prerequisite. The rainbow occurs only when it is seen. The cause (the arch of drops) is not there as a distinctive whole until it produces an effect (the projection in the observer's retina). Here we cannot separate the cause from the event, in the same way in which we cannot separate the events from their relation. The effect is responsible for the existence of the cause.

This is quite familiar from conscious representations, of course. For instance, if there are six points on a wall, in a hexagonal arrangement, these do not exist as an arrangement or a whole as long as do not produce any effect as a whole. If there were no human observer to see them, it would be extremely improbable that they would produce any effect as hexagonal arrangement in purely causal interaction with physical entities. Thus we can say that the hexagonal arrangement comes into existence only when it is *seen* for the first time. Generally speaking, it is impossible to define the physical existence of a structural arrangement without referring to a cognitive system which processes it as that arrangement. (Of course, the cognitive system in question does not need to be a human observer. The arrangement of six dots can be processed by much simpler systems as well. More complex arrangements, however, like a face or a word, require interaction with more complex systems, and in general there is a continuum of representational unities created by interacting agents.)

The unity of the arrangement, its 'objecthood', is a consequence of its capability of producing an effect as a whole. In fact, many objects can produce a joint effect only by means of the interaction with a cognitive agent. The whole, the content of the

observer's perception, is at the same time 'inside' and 'outside' of the observer. The whole 'arises' or comes about because of the act of observation and its occurrence is the very act of observation in itself. There is no inner event and outer event: the physical process that is responsible both for the six dots and their recognition as hexagon. The six dots as a whole arise when they produce a joint effect.

Even though most obvious in conscious representations, the same situation obtains in all perceptual events. Whenever we have a representation, there is no real distinction between the represented event and the representing one. Both occur conjointly as different aspects of one physical process. This process fits the requirements of representation since it contains what it has to represent: there is no need to assume a duplication of reality as in the Cartesian framework. On the other hand, all the problems surrounding traditional versions of identity theories (i.e. theories endorsing the identity between brain processes and mind processes) are undercut since the mind is no longer a neural activity located exclusively in the brain (a property of a substance) and separated from what it should refer to: the physical substratum of representations. Rather, the mind is a process which starts in the external world and ends in the brain. Any process is extended in time and space.

The puzzle of representation does not arise within the new framework since there is no longer any separation between the outside world and the internal world. In this way a mind no longer corresponds to an emergent property of a system duplicating external reality by means of some internal code. The mind *'enlarges'* to cover that part of reality which it represents. In fact, representation is no longer a *re-presentation* but is just a *presentation* (and thus tantamount to occurrence) of reality inside a system of events.

If we weigh the empirical (introspective) evidence concerning experience as well the degree of internal coherence of traditional theories, we might well say that the hypotheses of a purely quantitative physical object which is outside of the domain of our experience and (at the same time) the efficient cause of our perception is not well supported. This holds at least in the sense that there is no good empirical and theoretical reason that would militate against the radical approach of introducing a new account of representation and conscious experience based on processes – spatio-temporally extended dynamic entity. The envisaged new kind of entity is nothing more than the occurrence of reality without the division between "real reality" and "experienced reality".

In order to avoid all unnecessary and misleading connotations I will use a new term: 'onphene' to refer to an entity of this type. 'Onphene' is the contraction of *ontos* (existence), *phenomenon* (representation), and *episteme* (being in relation with) (Figure 2). The choice of words is motivated by the fact that the onphene is (i) a physical process (something that exists or is ontic), (ii) corresponds to a phenomenal content (a representation) and (iii) is in relation with other entities (nonseparate). Similar types of processes have been proposed by a number of authors: Whitehead's prehension (Whitehead 1933), reciprocal causation (Newman 1988; Hausman 1998), intentional relation (Manzotti 2000; Manzotti 2001). There are analogies also with Maturana and Varela's autopoiesis and Merlau-Ponty's circular causation (Merleau-Ponty 1945/2002; Maturana and Varela 1980; Maturana and Varela 1987/1998).

The notion of the onphene allows us to formulate a unified theory of mind, body and environment. Following an onphene-based approach, the distinction between a representing brain and a represented body plus environment is arbitrary and unnecessary. No pure disembodied mind has ever been experienced. On the other hand, instead of postulating that the brain has a dual state, an invisible property

corresponding to a phenomenal experience even more mysteriously linked to the external events, the mind is an activity that reaches beyond the physical area occupied by the brain at a certain time t. The mind is 'enlarged' to cover all those events which constitute the content of the conscious mind – these are physically part of the mind. There is no more dualism: there is just one reality. There are no more representations: there are just events constituted by the interaction of processes. We feel something because the process that we are is extended to comprehend those events we experience. Events that previously had no part in our developmental history become entangled in our internal process. The traditional picture of a boundary between an internal domain and an external domain is replaced by the image of the occurrence of a immensely complex fabric of processes continuously merging and dividing.

In an onphene-based ontology there are also events, but these figure as a derived category. Above I argued that events must be relational entities since they are essentially the result of interactions. The existence of such interactions is guaranteed by the fact that onphenes are by nature entangled entities in the sense that they cannot happen without being in interaction with others. Since onphenes are characterized as transferences with reciprocal causation (the existence of the effect effecting the existence of something that is cause for this very effect) an onphene not interactively entangled with other onphenes would be a contradiction in terms. An onphene-based ontology is naturally projected onto a relational structure where all parts of reality contribute to each other's identity and occurrence to different degrees. Events are abstractions of interactions but they carry the relational nature not always on their conceptual sleeves: onphenes and events are related like the trajectory of a bullet in physical space and the Euclidean points on the metrical counterpart of that trajectory, or like the south pole of a magnet to the magnetic field.

Using again the example of the rainbow, we can abstract two events from the rainbow-onphene: the cause, i.e. the reflection of sunlight from an arch of drops in the cloud, and the effect, the perception of the arch in the observer's retina. These two events belong to the same process (as well as the many other events along the path from the arch-shaped reflection to the perception in the retina which we could abstract). Strictly speaking, an event is a second-order interaction: an interaction of an interaction and a measuring process, since in order to ascertain the identity and existence of an event (the drops in the cloud or the activation in the retina), we need other onphenes, other processes (for instance a probe that measures the density of chemical compounds in the retina). As we shall see presently, even though the names of events denote ontologically derived or secondary entities, terms for events are useful tools in the formulation of an onphene-based framework.

4. The Enlarged Mind: onphenes and motivations

If the world is an assembly of onphenes, why and how do 'subjects' come about? A subject is a 'knot' of onphenes, but why do such 'knots' form and what ties them together? The onphene-based framework allows us to define a particular causal structure in terms of which we can demarcate those onphenes which constitute a conscious agent. The causal structure in question is at the core of what we mean by 'motivation'.

Any onphene has an abstract projection as a causal chain of events. This projection does not exhaust the nature of the onphene but provides a useful simplification. From the causal point of view an onphene is a process that links the occurrence of the cause with the occurrence of the effect in the form of a reciprocal causation. It corresponds to a situation in which the occurrence of an event E_1

produces (a) an effect E_2 and (b) the condition for the occurrence of the causal relation between that kind of cause (E_1) and that kind of effect (E_2) (compare again as examples the rainbow or a pattern). Talk about onphenes in terms of their causal abstractions, i.e. in terms of causal structures on events, provides a convenient tool to formulate a criterion for consciousness, as I shall explain now.

In an onphene-based ontology a subject S is a complex bundle of onphenes which consists of all those onphenes that are S's experiences at every instant of S's conscious life. To explain the emergence of a conscious mind is to explain how onphenes interact together to engender the occurrence of more and more complex onphenes. The model for this process is 'ontogenesis', the formation of complex units by linking together different causal chains into a unified causal process (or, to use the non-causal idiom, the progressive entanglement of more and more onphenes). Ontogenesis is familiar from the origination of planets – just as planets form without there being an a priori center of gravity, conscious beings may form without there being an a priori subject or transcendental Ego. Planets form due to conditions under which a large number of particles mutually attract. Similarly, conscious beings form due to conditions under which more and more phylogenetically induced motivations and goals merge into a giant unified causal process which is the subject. The conscious mind is the product of such a development, but a processual product: a kind of process or a way of events taking place.

A conscious being, then, is a bundle of onphenes linked in the way in which onphenes entangle during ontogenesis. This linkage may be considered as a goal of natural selection – a system capable of incorporating its past causal relations and their relata. If we view ontogenesis as an evolutionarily 'necessitated' developmental procedure, the consciousness, i.e., the ontogenesis of onphenes, also receives an evolutionary explanation. Onphenes entangle in the way of ontogenesis since this proved an evolutionary advantage and in doing so conscious beings were formed.

Motivations play the key role in the ontogenesis of conscious beings. A motivation is an internally produced criterion for the control of developments. In fact, a subject can be viewed as the process resulting from the incremental aggregation of onphenes elicited by motivations. We distinguish between fixed or hard-wired motivations and acquired or ontogenetic motivations. Fixed motivations are a priori coded and hence do not depend on the ontogenetic history of a system. By contrast, acquired or ontogenetic motivations must be the result of the interactions with the environment. In the following we will refer exclusively to the latter. *A (ontogenetic) motivation is here defined as a process whose probability of occurrence is increased* due to its own occurrence, in combination with the existence of certain embedding conditions.

To illustrate, let us suppose you see a face and hear that this is Sigourney Weaver. As a result you will be able to recognize her again and again. A process that has happened just once (the perception of a face *as* Sigourney Weaver's face) has produced itself as a future possibilty. You are conscious of her because her face has become a part of the processes that are your ontogenetic make-up. Your mind has enlarged itself by incorporating a new process that remains intertwined into what you are. The occurrence of such processes of recognition may be a matter of chance–they happen just in case you happen to see that face. But as soon as the occurrence of your recognition of Weaver's face increases the probability (frequency) of the occurrence of such recognitions, motivation has set it. The process of recognizing Sigourney could become a target for your future action – you might buy cinema tickets to repeat the recognition.

Motivations and phenomenal experiences are very similar in their causal structure. However they play a different role with respect to other onphenes. The crucial difference is, however, that motivations are processes which 'call' for their repetition. A phenomenal experience does not produce a modification in the subject's causal structure, while a motivation propagates its effects through time in the subject's history due to a stable modification in the subject's causal structure. A phenomenal experience (onphene) is an event C whose effect E is the cause of a causal relatedness of a cause of kind of C and an effect of the kind of E. Simplifying we might say, a phenomenal experience is an event Q which generates a causal relation of the kind Q instantiates. Motivations are just that with an additional element of causal structure. A motivation is an onphene which generates *future* occurrence of onphenes of this kind. More precisely, switching to the causal idiom, a motivation is an event C whose effect E is the cause of the causal relatedness of a cause of kind of C and an effect of the kind of E, *and of future occurrences of a causal relatedness of this kind.*

To have a motivation for getting something means to be in relation with an event: the target event (what we want to achieve) becomes the cause of all or most of our action. How we act is not caused by a 'final cause' but by our past, i.e., by the processes which entered a past processual constitution of the subject that we are. However, once the specter of final causation is properly replaced, it is admissible to speak of a system's goals. For instance, assume that the system has been exposed to the presence of Susan, and as a result the system aims at having Susan in its field of view. The system will behave with the goal of repeating as much as possible the process of seeing Susan. It was the process of seeing Susan that modified the structure of the system in such a way that among its goals there is 'having Susan in the field of view'. The occurrence of the process 'seeing Susan' has increased the probability of its own repetition. If Susan would not have come into the system's field of vision, the system would not have modelled its criteria for 'preferred visual object' around her visual appearance. The Susan process became entangled into the ontogenetic history of the observing subject by adding a new motivation.

A caveat: what increases its possibility to happen again is not the appearance of Susan (which depends only on Susan), but the process of Susan's being seen and recognized by the system. Motivation is not a causal reflection of conditions in environment–rather, it is a self-reinforcing activity. Playing tennis is a way to become fond of playing tennis.

The view of the motivation I have set out here acquires its full meaning only within the broader process-based approach I have sketched here. Without this background, the suggested account of motivation suggested can still be read in purely causal terms, but in doing so its explanatory power with respect to the constitution of a self will be lost. Let us review the main conceptual elements of this process-based approach encountered in this and the previous section:

Process or onphene	The basic unit of reality
Event	An abstraction from the interaction of processes
Cause	What takes part in a process
Effect	What can be part of other processes after the occurrence of a process

Content of an onphene	The cause of an onphene
Reciprocal causation	When the occurrence of the effect is responsible for the occurrence of the cause or rather when the occurrence of an event C whose effect E effects a causal relatedness of events of the kind of C and E, respectively (an onphene viewed in objectivistic terms.)
Represented event or object	An onphene from the point of view of the cause
Representing event or representation	An onphene from the point of view of the effect
Phenomenal experience	A representation, i.e. an onphene
Motivation	A process whose occurrence creates the condition of its own repetition
Action	An event provoked by a subject as a result of a motivation

Table 1 A list of used terms and their definitions

5. A process based architecture

An architecture is the description of the essential features of an embedded system in order to produce a certain phenomenon. We can now present a general architecture for an artificial device designed to engender the occurrence of (a great number of) the described processes. To this effect the architecture must work with a physical body (sensory and motor systems). The goal is to build something which will become part of the 'external' flow of processes. Thus the body and its sensory and motor equipment are necessary – a computer without the capacity of interacting directly with the physical aspects of its environment would not suffice.

The general idea is the following. The system must be capable of self-organizing the flow of incoming stimuli and at the beginning it must do so driven by pre-defined criteria. This is consistent with what happens during the ontogenesis of a biological being in a natural environment. The fixed goals of phylogenesis put constraints on the complete freedom of ontogenesis (which is conditioned only by the environment and the experience). There is a obviously a trade-off between flexibility and adaptivity (to new situations and potentially unpredictable events) and the control that phylogenesis can exert upon individuals. The primary objective is to design and implement an architecture into which processes get trapped and find a way to repeat themselves over and over.

At the beginning the system merely contains some bootstrapping criteria aimed at orienting its ontogenesis towards specific and useful classes of stimuli. These criteria are the equivalent of instincts in a biological being. Otherwise the system is literally a *tabula rasa*. When something happens it is processed by the system; however the system has still to build its own internal perceptual categories. If the phylogenetic criteria give their approval, the system begins to build its ontogenetic categories. At first they will be just perceptual categories, subsequently some of them can be

selected as newly generated criteria for controlling further ontogenetic development. Generally speaking the architecture is composed of three parts, in the following referred to as modules (Figure 3): a phylogenetic module to bootstrap the system, a module to store new categories, and an ontogenetic module to determine which events have to become new ontogenetic criteria (motivations) for the system.

Whenever something happens and it is part of the sensory experience of the system, it can produce an effect. Further this effect is the condition for the future repetition of the same kind of causal processes. For instance, the system will become capable of recognizing faces because faces have been part of its past. Faces will become objects of perception because the system will have developed structures to recognize them. The process structure of the ontogenetic development of the system is further reinforced by the development of new ontogenetic criteria which will control the future choices of new classes of objects.

To offer a more comprehensive illustration, consider an experimental set up where at the beginning a series of stimuli with different colors and shapes are presented. They are perceived as sets of features taken as a whole (including shape). The stimuli are categorized on the basis of all their properties (shape, size, texture, orientation, spatial frequency) according to (a) criteria contained in the Phylogenetic Module and (b) criteria created by the system. The Phylogenetic Module provides only a color criterion. Yet the system might create a category for stimuli of similar shapes or similar colors or similar texture. By a category we mean a class of different stimuli that are perceived by the system as the same stimulus. In order to build category the complexity of sensory stimulation must reduced. In the first phase of development only those categories are created which have a ready-made criterion in the Phylogenetic Module If a colorless stimulus were introduced, albeit equipped with other properties including shape, the Phylogenetic Module will pass it over. In the second phase the Ontogenetic Module comes into play. The Ontogenetic Module can transform a number of categories into new criteria. Subsequently such criteria supplement the criteria of the Phylogenetic Module. If brightly colored triangles are shown to the system, they will become a new category (triangle). After a while, this category will become a suitable candidate for transformation into a new criterion. From then on the Ontogenetic Module will accept even a colorless triangle in the category triangle (colored and not). In turn, the colorless triangle category shall be transformed by the Ontogenetic Module so as to become a further category in itself. In this way new categories can be formed which are connected only to the new ontogenetic properties (shape) regardless of the phylogenetic properties (color).

The criteria play a fundamental role in the Ontogenetic and Phylogenetic Modules. The criteria are implemented by a 'Relevant Signal' which manages the creation and allocation of incoming stimuli into the relevant categories. For instance, if the incoming stimulus corresponds to a brightly colored object, the system will produce a strong Relevant Signal. If the incoming stimulus corresponds to a dull grey object, the Relevant Signal will be weaker. A criterion depends on the value the system gives to the incoming stimulus with respect to the whole past ontogenetic history. The content of an ontogenetic criterion is given by a category. An ontogenetic criterion is a motivation. Technically a motivation is implemented by selecting a given category. All categories developed during the phylogenetic phase potentially provide the content for the same number of criteria. Only certain categories shall become criteria (motivations).

The described architecture must be implemented by a physical structure that is activated by, and develops motivations, on the basis of incoming events. The

architecture makes use of elementary associative processes, simple Hebbian learning and case-based reasoning. The events occurring nearby the motivation-based architecture become the seeds for motivations. Due to the existence of the architecture the events in its environment become entangled in a growing network of onphenes and so we can say in turn that the physical implementation of the architecture organizes the environment.

In the following I will introduce the architecture in somewhat greater detail to describe how this architecture at once engenders the self-organization of incoming stimuli and uses them both to categorize reality and to develop criteria on the basis of which new categories can be introduced. We will see how the system gradually modifies its structures and overcomes its initial limitations by developing new criteria. The architecture is aimed at mimicking the development of motivations in human beings. For instance, a human might develop an interest for cars even if nothing in his/her phylogenetic code is explicitly directed towards cars. By contrast, an insect cannot develop new motivations but must follow its genetic blueprint: it has no ontogenetic development. One of the issues of this architecture is to divide explicitly the ontogenetic part from the phylogenetic one.

In a nutshell, the architecture's three main modules are: the Category Module, which is basically a pattern classifier; the Phylogenetic Module which contains the *a priori* criteria; and the Ontogenetic Module which applies Hebbian learning and develops new criteria by using the patterns stored in the Category Module. The incoming stimuli are categorized in the Category Module on the basis of the Relevant Signal coming from the Phylogenetic Module and the Ontogenetic Module. At the beginning, the Relevant Signal depends on those properties of the incoming stimuli that are selected by the Phylogenetic Module; later the Relevant Signal is flanked by the new signals coming from the Ontogenetic Module.

4.1 Category Module

The Category Module has the role of grouping in clusters the stimuli coming from the external events. The process of cluster definition is based on an internally built-in criterion for clustering and on the presence of a Relevant Signal (Insert

Figure). Whenever an incoming stimulus is received, a Categories Vector, which is the output of the CM, is computed; the elements of this vector provide an indication of which cluster best represents the current stimulus. The Categories Vector is empty at the beginning and eventually becomes larger and larger adding new categories. Each of its components measures how much the incoming stimulus matches the corresponding category. The CM tunes its activity to the Relevant Signal (the sum of the Relevant Ontogenetic Signal and the Relevant Phylogenetic Signal).
If and only if the Relevant Signal is active, every time a signal is received, the CM performs the following actions:

i) if the stimulus is too similar to the already stored stimuli, do nothing;
ii) if the stimulus is sufficiently similar to one of the previously created clusters, the stimulus is added to that cluster;
iii) if the stimulus is not sufficiently similar to any of the stimuli already stored, a new cluster is created.

By storing a stimulus only if the Relevant Signal is active, the system does not assign new resources to every incoming signal (the first rule is useful to avoid storing equivalent stimuli).

4.2 Phylogenetic Module

This module is the only one that has some built-in criteria concerning the relevant properties of the incoming signal. Functionally, it has the same role as the genetic instincts in biological systems. A Phylogenetic Module autonomously produces a signal on the basis of some external events (the presence of soft or brightly colored objects). For instance, a baby of 2 months looks with more curiosity at brightly colored objects than at dull colorless objects, independently of any past experience. This behavior requires the existence of a hardwired function looking for a relevant property of images (saturated colors). The Phylogenetic module provides criteria that can be used to select correct actions (for instance those actions that maximize the presence of the interesting stimuli). If the system were composed of just the PM and the CM, the system would be a reinforcement learning system.

4.3 Ontogenetic Module

Whereas the Phylogenetic Module has built-in criteria about the nature and the relevant properties of the incoming signal, the Ontogenetic Module selects new criteria on the basis of experience. Functionally it has the same role as the acquired ontogenetic criteria in biological systems. The main goal of the Ontogenetic Module is to transform a subset of categories into criteria. Not all the categories built by the CM will become criteria. For instance, if an infant is exposed to colored stimuli of a given shape, she will develop a particular perceptual sensibility for that kind of colored shapes. After a while, the shapes alone (not colored shapes) will become a category. Under certain conditions, the category (colorless shape) will be transformed into a criterion: the Ontogenetic Module will produce an active Relevant Signal even in absence of colors when the specific shape will be present. If she has spent a lot of time looking at colored triangles, it is possible that she will become interested in triangles, independently of their color. She could eventually be interested in grey triangles. The Relevant Signal gives a measure of how much the incoming stimuli is part of the ontogenetic history. By means of Hebbian learning (roughly: what happens together is reinforced), the Ontogenetic Module communicates to the system to what extent each cluster of the Category Vector has been correlated in the past with the signal produced by the Phylogenetic Module.

The main goal of the architecture is to create a structure that can be changed completely by the architecture's own 'experiences.' In the architecture there is a clear-cut division between the phylogenetic part (the *a priori* section) and the ontogenetic part produced by the interaction with the environment. Whenever an event is capable of being recognized by the CM and then selected by the OM, it becomes part of the ontogenetic structure of the developing agent. The event is responsible for the occurrence of a process, whose occurrence will increase the probability for such a process to occur again. The events that become the content of the system motivation are those events that have been able to modify the agent structure. They are abstractions from processes which have become entangled in the system history and perpetuate themselves by means of the system itself.

The processes made possible by the existence of the system can be considered logically and physically continuous with the environment. Furthermore, they shape and modify what the environment is – in fact they create new kind of objects by creating the conditions in which the new objects can exert their effects. Since they are the result of the environment itself, they can be considered as the result of the self-organization of the environment.

The entire architecture operates under one single general directive: if a process passes through the architecture, the probability for the occurrence of that process has to increase.

4.4 A comparison with Pavlov's classic conditioning

It might be instructive to compare the motivation-based architecture I have sketched in the previous paragraphs with Pavlov's classic experiment of conditioning. There are two good reasons for such a comparison: i) there are strong similarities; ii) there is evidence that many cognitive learning processes could be reduced to Pavlov's associationism (Pavlov 1955/2001; Perruchet and Vinter 2002).

Pavlov focused on modifications of the relation between a given stimulus and a given response. Although Pavlov's test animal was able to select a different stimulus (the ring of the bell), the focus was more on the fact that the animal was capable of linking it to a behavior (the salivary response) than on its capability of selecting a given stimulus from the continuum of the environment. In Pavlov's experiment, there are two hardwired receptors for two different kinds of stimuli (sound of a bell and meat powder): one is a neural structure capable of recognizing the presence of food and another is a neural structure capable of recognizing the ring of a bell. Before the conditioning process, the behavioral response (the salivation) was only connected with the presence of food. During the training the conditioned response became stronger: more drops of saliva were secreted. The learning consisted in the creation of a connection between the conditioned stimulus and the response.

In our case, the conditioned stimulus does not exist before the conditioning process. The machine is not capable of recognizing the unconditioned stimulus (the shape of an object). It only recognizes colored objects. At first sight our experiment might resemble Pavlov's experiment. The Phylogenetic Stimulus and the Ontogenetic Stimulus could be taken to correspond to the Unconditioned Stimulus and the Conditioned Stimulus, respectively. The Developmental Signal could be counted as the Response (first Unconditioned and then Conditioned). However, the analogy is not sufficienlty smooth. In our case, since the color was presented conjointly with the shape of an object, a new ontogenetic stimulus (the shape) is added to the machine's repertoire of stimuli. In other words, the *Umwelt* of the machine is increased and enlarged by a new kind of event. In the case of the motivation-based architecture as described two things happen: i) the machine learns to recognize something which was previously unknown to it; ii) the machine links this new stimulus to a given motor behavior. (Insert Figures 4 and 5)

Briefly, Pavlov's experiment highlighted the fact that the test animal was capable of establishing a new association between an already familiar stimulus to a motor response. The goal of our experiment is to model the development of the capability of recognizing new stimuli.

6. Experimental results: the emergence of motivations

To test the architecture, we conducted an experiment in which a robot implementing the described motivation-based architecture evidently develops a new motivation on the basis of its own experiences. In the experiment, an incoming class of visual stimuli (not coded inside the architecture) produces a modification in the system's behavior that changes not only *what* the system is doing (behavior) but also *why* it does what it does (motivation at the basis of behavior) the system is doing.

Something which happens in the environment (the appearance of a class of shapes) becomes part of the agent's behavior.

The system has, in this preliminary experiment, a single behaviorial choice: to direct or not its gaze towards objects. A series of different shapes associated with colors were presented to the robot. The system was equipped with a phylogenetic motivation that is aimed at brightly colored objects; a colorless stimulus, independently of the shape, did not elicit any response. Since the system has an Ontogenetic Module it develops further motivations directed towards classes of stimuli different from those relevant for its Phylogenetic Module. After a period of interaction with the visual environment (i.e. the presentation of a series of elementary colored shapes), the robot was motivated by colorless shapes also. The category of shape alone had been accepted by the Ontogenetic Module. The system showed the ability to develop a motivation (by directing its gaze towards the stimulus) that was not envisaged at design time and that is the result of the ontogenetic development.

For the robotic set-up a robotic head with two degrees of freedom had been adopted which is equipped with a videocamera capable of acquiring logpolar images (Sandini and Tagliasco 1980; Sandini, Questa et al. 2000), see Figure 6, i.e. images like those perceived in human beings (with a fovea and a periphery). (Insert Figure 6) The robotic head has two degrees of freedom: the camera is capable of a tilt and pan independent motion (Figure 7a). The robotic head is programmed to make random saccades; a Motor Module generates saccades on the basis of an input signal l that controls the probability density of the amplitude r. If λ is low (near to 0), the probability density is almost constant, otherwise, if λ is higher, a small amplitude is more probable (Figure 7b). (Insert Figure 7) This probability schema is to ensure that the motor unit mimics an exploratory strategy. When a visual system explores a field of view, it makes large random saccades. When it fixates an interesting object, it makes small random saccades.

We presented different sets of visual stimuli to the system. A first set consisted in a series of colorless geometrical figures as shown in Figure 8a on the left. (Insert Figure 8) The frequency with which the system was looking at different points was measured. The system spent more time on stimuli corresponding to its motivations by reducing the amplitude of its saccades. At the beginning the system gazed around completely randomly with large saccades since its Ontogenetic Module was unable to catch anything relevant and the Phylogenetic Module was programmed to look for very saturated colored objects, which were absent in the first set. To get a qualitative visual description of how much time was spent by the system on each point of its field of view, we assigned to each point of the visual field an intensity value proportional to the normalized time the system gazed at it. The images in the centre of Figure 6a-b, and c were generated after 10^3 saccades (equivalent to about 500 sec). The brighter a point of the image, the more frequently the system gazed at it. With the first set of visual stimuli, the resulting image is in Figure 6a. The system does not show any polarization towards a specific part of the field of view. Subsequently we presented a different stimulus: a series of colored figures. The difference is shown in Figure 8b. The head spent more time on the colored shapes instead on the white background because of the phylogenetically implanted rule. Finally we presented again the initial stimulus (the set of colorless shapes). The system spent more time on the colorless shapes than on the background (Figure 8c). The behavior of the system changed since the system added a new motivation (shapes) to the previous ones (saturated colors).

7. Conclusion

The outlined theory of consciousness is a theory of phenomenal experience – consciousness as phenomenal experience is here defined in terms of the nature of processes involved in development of a conscious agent. Unlike many theories that address the problem of consciousness and phenomenal experience, the account I have sketched here can be tested in the laboratory - even though gained by philosophical reflection rather than by induction, the hypotheses I have put forth have a counterpart in the design and implementation of an architecture which put to the test in various experiments with machines.

I have suggested here that the subject is the result of the self organization of the environment of which the developing subject is a part. If the world were devoid of subjects, it would be a very different place in causal terms. If shapes or drawings were to be carved in stone by a capricious whirl of wind, their appearance would not increase the probability of their repetition. On the other hand, in a world populated by subjects, every process that becomes part of the structure of a subject increases the probability of the recurrence of that kind of entanglement. Thus the presence of a subject is tantamount with the presence of a certain type of process—the subject is nothing more than a collection of processes of a certain kind.

Due to the existence of subjects processes propagate themselves in a new and interesting way. The phenomenal experience of shapes, colors, behaviors, actions ensures 'its own' repetition countless times wherever subjects are located. According to the account outlined here, the subject is as a set of processes (onphenes) that, by means of motivations, becomes progressively integrated during the subject's development. The content of the conscious mind, i.e. the content of the phenomenal experiences of a subject, is not 'inside the head'; rather consciousness is how the world is organized due to the existence of the subject.

To restate the main steps of my argument, I pointed out that classic physicalistic object ontology is saddled with presuppositions that inevitably lead to dualist accounts of representation. Instead of maintaining such ontology and then trying to justify the existence of the mind by adding new hypotheses (like qualia or dual aspect of the world) in a dualistic fashion, I suggested a wholesale rejection of the traditional object ontological framework. I introduced a new type of ontological entity – the onphene – that contains both 'ontic' and 'phenomenal' aspects and is best understood as a *'presenting'* of the most general sort. The notion of the onphene undercuts the traditional dualism of a conscious mind separated from the physical world. I then offered an account of motivation, often considered the hallmark of conscious agents, in terms of onphenes. Motivation can be ascribed to a system if that system supports onphenes with a certain causal role, namely, self-propagating onphenes. In terms of self-propagating onphenes I defined criteria for the existence of representing subjects—subjects exist if self-propagating onphenes (motivation) exist. On the basis of the suggested definition of motivation I sketched a general architecture in which self-propagating onphenes can be 'induced'. Finally, I reported on experimental findings that confirm the presence of such processes (motivation-based learning), in a simple learning situation.

Given the line of argument presented, the architecture described can be considered as the general recipe for the design of a conscious machine: it must be able to 'catch onphenes', promote and control a large number of them, and to integrate them. The first step is implemented by means of the cooperation between a Category Module and an Ontogenetic Module inside a Motivation-based Architecture Module.

Here I described only the implementation of the first step for a simple test case where the input capability is limited to shapes and colors.

Future work will concern the building of a network of Motivation-based Architecture Modules, the implementation of other different motor and sensor capabilities, and the improvement of the general process-based framework. The implementation of a mechanism of progressive unification and integration of a large number of processes should eventually lead to the development of an artificial subject.

Acknowledgements
I would like to thank to Johanna Seibt for many helpful comments on a first draft of this paper. The work reported here has been supported by the ADAPT Project, IST-2001-37173.

Figure 1 A rainbow. Does it exist without an observer? Does its observer exists without the drops of water? What is the cause and what is the effect?

Figure 2 A visual analogy to explain the role of the onphene. It can be seen under three main perspectives: as a *phenomenon* (what appears?), as *ontos* (what is?) and as *epistemê* (what is in relation with?). The three standpoints, traditionally separated, can be seen as three manifestation of the same underlying principle.

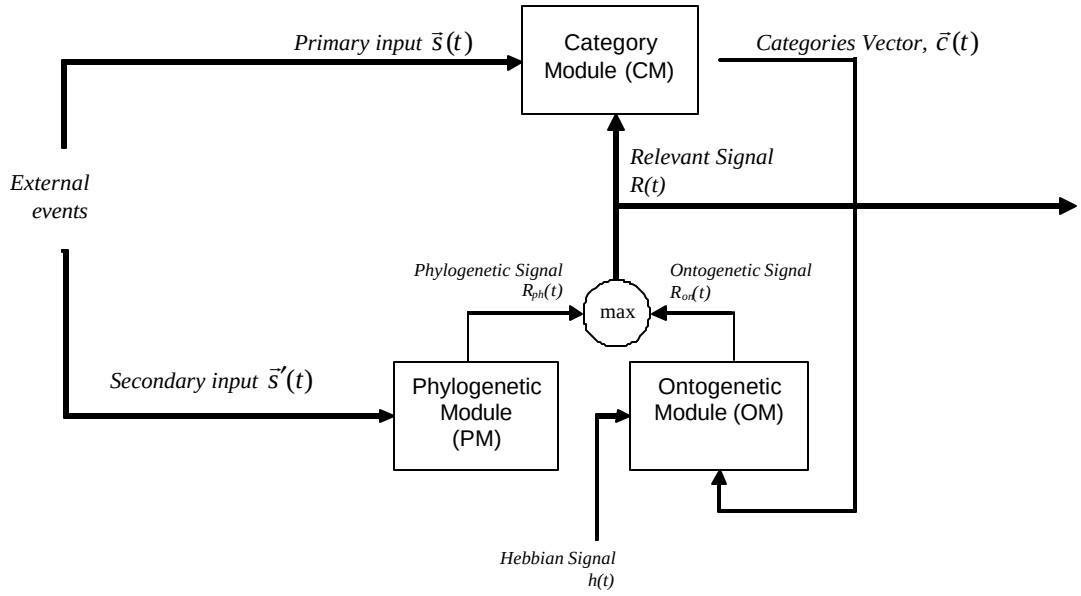


Figure 3 Motivation-based architecture scheme.

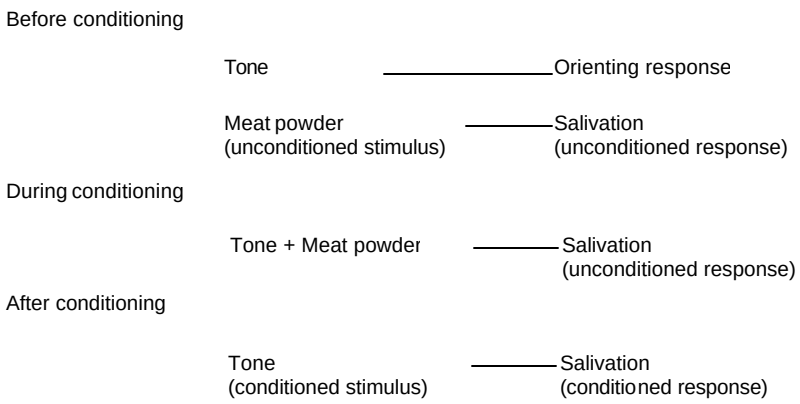


Figure 4 The three stages of conditioning in the classical Pavlov experiment.

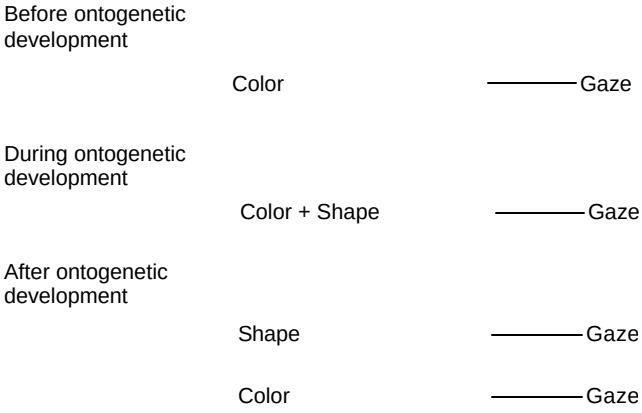


Figure 5 The three stages of ontogenetic development from a process based standpoint (our experiment).

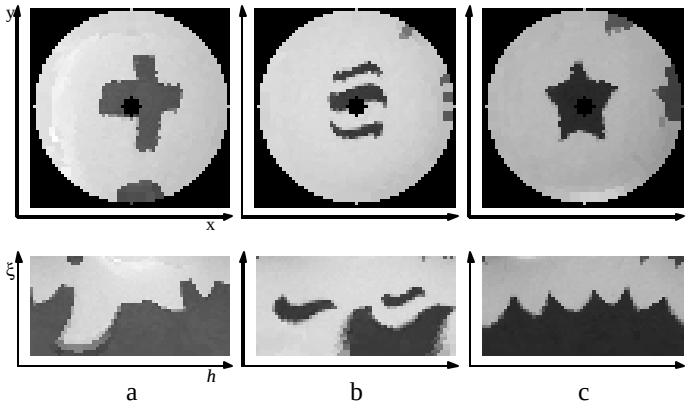
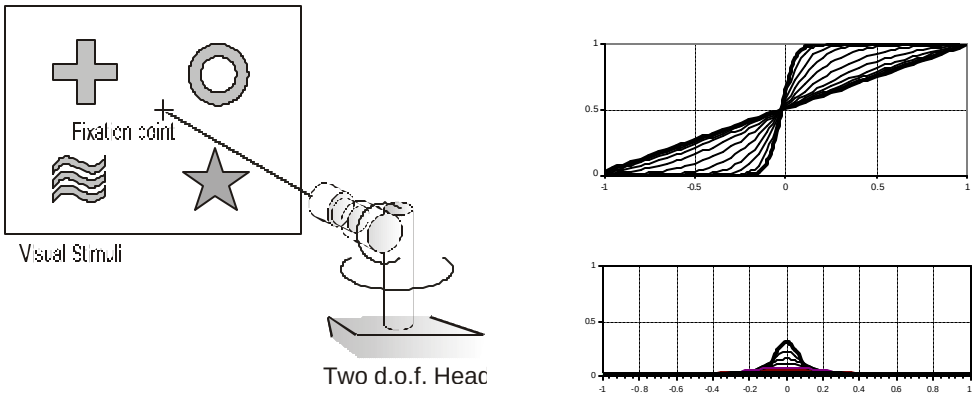


Figure 6 The Cartesian (upper row) and log-polar (lower row) images for a cross a), a wave b), and a star c).



a b

Figure 7 a) Sensory and motor set-up, b) The probability density function on the basis of the control parameter λ.

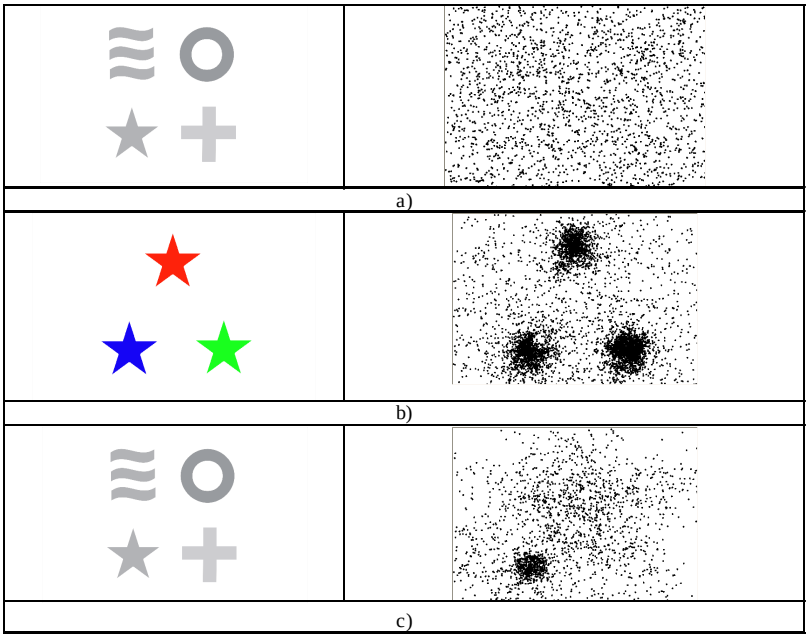


Figure 8 Experimental results.

5. References

Aleksander, I. (1996). Impossible Minds: My Neurons, My Conscio usness. London, Imperial College Press.

Aleksander, I. (2000). How to Build a Mind. London, Weidenfeld & Nicolson.

Aleksander, I. (2001). "The Self 'out there'." Nature **413**: 23.

Arkin, R. C. (1999). Behavior-Based Robotics. Cambridge (Mass), MIT Press.

Armstrong, D. M. and N. Malcom (1984). Consciousness and Causality: A Debate on the Nature of Mind. Oxford, Blackwell.

Auletta, G. (2000). Foundations and Interpretation of Quantum Mechanics. Singapore, World Scientific.

Bickhard, M. (1999). Representation In Natural and Artificial Agents. Semiosis. Evolution. Energy: Towards a Reconceptualization of the Sign. E. Taborsky. Aachen, Shaker Verlag.

Bickhard, M. (2001). The Emergence of Contentful Experience. What Should be Computed to Understand and Model Brain Function? T. Kitamura. Singapore, World Scientific.

Block, N. (1988). What Narrow Content is Not. Meaning in Mind: Fodor and his Critics. B. Loewer and G. Rey. Oxford, Blackwell.

Brooks, R. A. (1991). "New Approaches to Robotics." Science **253**(September): 1227-1232.

Buttazzo, G. (2001). "Artificial Consciousness: Utopia or Real Possibility." Spectrum IEEE Computer **18**: 24-30.

Chalmers, D. (1996). The Components of Content. Philosophy of Mind: Classical and Contemporary Readings. D. Chalmers. Oxford, Oxford University Press**:** 608-633.

Chalmers, D. J. (1996). The Conscious Mind: in Search of a Fundamental Theory. New York, Oxford University Press.

Clark, A. and C. Tornton (1997). "Trading spaces: Computation, representation and the limits of uninformed learning." Behavioral and Brain Sciences **20**: 57-90.

Cramer, J. G. (1988). "An Overview of the Transactional Interpretation of Quantum Mechanics." International Journal of Theoretical Physics **27**(227).

Crick, F. and C. Koch (2003). "A framework for consciousness." Nature Neuroscience **6**(2): 119-126.

Dennett, D. C. (1969). Content and consciousness. London, Routledge & Kegan Paul.

Dennett, D. C. (1988). Quining Qualia. Consciousness in Contemporary Science. A. Marcel and E. Bisiach. Oxford, Oxford University Press.

Dretske, F. (1993). "Conscious Experience." Mind **102**(406): 263-283.

Dretske, F. (1995). Naturalizing the Mind. Cambridge (Mass), MIT Press.

Edelman, G. M. and G. Tononi (2000). A Universe of Consciousness. How Matter Becomes Imagination. London, Allen Lane.

Editor (2000). "In Search of Consciousness." Nature Neuroscience **3**(8): 1.

Fodor, J. A. (1981). Representations: philosophical essays on the foundations of cognitive science. Cambridge (Mass), MIT Press.

Fodor, J. A. (1990). A theory of content and other essays. Cambridge (Mass), MIT Press.

Goodman, N. (1974). Language of Art.

Grice, P. (1961). The Causal Theory of Perception.

Hausman, D. M. (1998). Causal Asymmetries. Cambridge, Cambridge University Press.

Haybron, D. M. (2000). "The Causal and Explanatory Role of Information Stored in Connectionist Networks." Minds and Machines **10**: 361-380.
Manzotti, R. (2000). Intentionalizing nature. Tucson 2000, Tucson, Imprint Academic.
Manzotti, R. (2001). Intentional robots. The design of a goal seeking, environment driven, agent. DIST. Genova, University of Genoa.
Manzotti, R., G. Metta, et al. (1998). Emotions and learning in a developing robot. Emotion, Consciousness and Qualia, Naples and Ischia (Italy).
Manzotti, R. and V. Tagliasco (2002). "Si può parlare di coscienza artificiale?" Sistemi Intelligenti **XIV**(1): 89-108.
Martinoli, A., O. Holland, et al. (2000). Internal representations and Artificial Conscious Architectures, California Institute of Technology.
Maturana, H. R. and F. J. Varela (1980). Autopoiesis and cognition: the realization of the living. Dordrecht (Holland), D. Reidel Pub. Co.
Maturana, H. R. and F. J. Varela (1987/1998). The tree of knowledge: the biological roots of human understanding. Boston, Shambhala.
McFarland, D. and T. Bosser (1993). Intelligent Behavior in Animals and Robots. Cambridge (Mass), MIT Press.
Merleau-Ponty, M. (1945/2002). Phenomology of Perception. London, Routledge.
Millikan, R. G. (1984). Language, Thought, and other Biological Categories: New Foundations for Realism. Cambridge (Mass), MIT Press.
Newman, A. (1988). "The Causal Relation and its Terms." Mind **xcvii**(388): 529-550.
O'Brien, G. and J. Opie (1997). "Cognitive science and phenomenal consciousness." Philosophical Psychology **10**: 269-86.
Pavlov, I. P. (1955/2001). Selected works. Honolulu, (Haway), University Press of the Pacific.
Perruchet, P. and A. Vinter (2002). "The Self-Organizing Consciousness." Behavioral and Brain Sciences.
Sandini, G., P. Questa, et al. (2000). A Retina-like CMOS Sensor and its Applications. SAM-2000, Cambridge, USA, IEEE.
Sandini, G. and V. Tagliasco (1980). "An Anthropomorphic Retina-like Structure for Scene Analysis." Computer Vision Graphics and Image Processing **14**: 365-372.
Schlagel, R. H. (1999). "Why not Artificial Consciousness or Thought?" Minds and Machines **9**: 3-28.
Seibt, J. (1990) *Towards Process Ontology: A Critical Study on the Premises of Substance Ontology*, Ph.D. Diss. University of Pittsburgh, UMI-Publications, Michigan.
Stapp, H. P. (1998). Whiteheadian Process and Quantum Theory of Mind. Silver Anniversary International Conference, Claremont (Cal).
Steels, L. (1995). Is artificial consciousness possible? Consciousness: Distinction and Reflection. G. Trautteur. Napoli, Bibliopolis.
Togawa, T. and K. Otsuka (2000). "A model for Cortical Neural Network Structure." Biocybernetics and Biomedical Engineering **20**(3): 5-20.
Whitehead, A. N. (1925). Science and the modern world. New York, Free Press.
Whitehead, A. N. (1933). Adventures of ideas. New York, Free Press.
Zeki, S. (2003). "The Disunity of Consciousness." Trends in Cognitive Sciences **In press**.
Zohar, D. (1990). The Quantum Self. New York, Quill.