Commentary on Pentti O. Haikonen's
The Cognitive Approach to Conscious Machines

Giorgio Marchetti

In my opinion, Haikonen's *The Cognitive Approach to Conscious Machines* is one of the clearest and most outstanding examples of the beneficial and positive effects that can be achieved in studies on the mind and consciousness by adopting what I call an "operational approach". By "operational approach" I mean the kind of engineering approach that tries to analyze and define our mental abilities – perception, memory, attention, consciousness, language, reasoning, learning, thought, etc. – in terms of operations, that is, as the working of some organs or systems, and the products of our mental abilities – percepts, memories, conscious states, thoughts, etc. – as the product of the working of such organs or systems (Benedetti, 2006, Ceccato, 1972, 1980, Marchetti, 1993, Vaccarino, 1974).

The operational approach does not satisfy with those kinds of theories such as Emergentism which do not try to understand how a given object works, what makes it function, how its constituent parts interact in order to produce the object's properties and so on, but simply depict the object and its properties as they are. According to Emergentism, an object has a certain property because the combination of the elements that compose the object yielded that specific property. The emergent property cannot be predicted *a priori* from the separate properties of the elements themselves, and is supposed to be something more than the sum of the elements[1]. Indeed, a kind of proposal such as the one put forward by Emergentism can be applied to all objects and situations, and consequently does not explain anything specific, but simply describes a given situation *a posteriori*. As Haikonen states:

[1] A typical emergentist position is that held by Libet: "My view of mental subjects function is that it is an emergent property of appropriate brain functions" (Libet, 2004, p. 86). On Libet's work, see my commentary (Marchetti, 2005a).

the concept of emergence may not necessarily be the best of explanations, maybe not an explanation at all, because it would seem to involve a statement of the style: “In big systems it is suddenly like this and there is nothing more to it” As an explanation, this is more like a leap of faith (Haikonen, 2003, p. 145).

Neither is the operational approach satisfied with those theories which, instead of explaining how an object functions, describe the circumstances that favored the appearance and development of the object’s properties from an evolutionist point of view. Generally speaking, an evolutionist account of how an object acquired its properties explains this acquisition on the basis of a trial-and-error-process which, starting from a casual, random variation in the object’s structure, selected precisely those properties. Patently, the evolutionary explanation of *why* an object acquired certain properties is not a functional explanation of *how* these properties are currently produced or generated, that is, of *which* mechanisms underlie and make the current production and generation of such properties possible. Moreover, Haikonen observes that some authors rely on an evolutionary explanation to account not only for how human beings acquired certain properties, but also for which mechanisms underlie such properties: this, for example, is how some authors explain fundamental human abilities as language (Haikonen, 2003, p. 130), creativity and imagination (Haikonen, 2003, pp. 96-98), and perception and music (Haikonen, 2003, p. 120). As Haikonen clearly shows, none of these fundamental human abilities can be adequately explained through an evolutionary mechanism. As regards creativity, for example:

Random variation theories propose that creativity arises from random excitations of information and the filtering of the most suitable combinations of these. (…) Can we make the random variation method work if we apply evolutionary principles here? We could produce a number of initial schemes by random combination and variation, apply the principle of “the survival of the fittest”; identify then the best alternative and take this as the basis for further random variation while rejecting the rest. (…) Is the scheme that works best at the moment really the best alternative for further development? Good realization of a bad principle may be better than an initial bed realization of a good principle. Engineers know that many times a new, eventually superior technology produces initially results that are inferior to those produced by the old technology. (…) In evolutionary terms the transistor should have died a quick death, the simple “survival of the fittest” rule would have guaranteed this, (…) We humans can do better than evolution by setting goals and predicting results, having insights about the possible usefulness of presently inferior solutions. (…) Artificial creativity by “evolutionary” computing schemes that seek to find the one and best result by selecting only the instantaneously best solution for further development while rejecting the rest will most probably fail (Haikonen, 2003, pp. 96-98).

Nor does the operational approach satisfy with those theories and proposals that take mental abilities and their products for granted, purposefully leaving them unexplained: “purposefully” in the sense that these theories and proposals imply not so much that there are not enough resources and time to analyze mental phenomena and products, as that these phenomena either do not need to be explained because they are self-evident and self-explaining, or worse, can never be explained, or even worse must simply be ontologically trusted and relied upon. The operational approach does not consider any mental ability and product as a given thing-in-itself: on the contrary, it continuously strives to trace the origin of mental phenomena back to the operations performed by some system or organ. Paraphrasing Paul Valéry (1973), the operational approach aims to substitute "a mechanism, a definite act for anything". For example, as regards the problem of the existence of "free will", Haikonen's opinion is clearly against any ontological position, conceiving "free will" as an autonomous reality, entity or phenomenon existing independently of the conscious machine's (or system's) cognitive processing:

It can be said that no actual free will exists here, the outcome is determined by the external conditions and internal needs, drives and emotional status (…) However, this does not exclude the possibility of the perception of "free will". This "free will" may be perceived because the different options can be recognized in retrospect and the possibility of choosing differently is perceived. In this way the real "free will" may be illusory. However, because the choices depend on the status of the system's needs and drives in respect to the prevailing external conditions, it can be said at least that what is reflected here is the system's own will. Therefore it would seem that the impossibility of genuine "free will" would not exclude a system's "own will" that would guide the system's actions towards its goal and requirements (Haikonen, 2003, p. 156).

Since no phenomena can be considered completely independent of our mental activity (in the sense of not being in some way also the result or product of our mental activity, perceptual processes, cognitive processes, etc.: whatever phenomenon we see, perceive, conceive of, imagine, analyze, remember, etc., unavoidably assumes the form that our mind, senses, memory, cognitive abilities, etc. give it), nor can the operational approach satisfy with those theories that consider not only mental phenomena but phenomena in general as a given thing-in-itself, that is, as something independent of the activity of a cognitive system capable of conceiving, perceiving, elaborating, thinking about, and imagining it. This is why for example, I think, Haikonen rejects any idea of "music" as being something independent of cognition:

Is there universal music independent of cognition? I dare to answer swiftly: No universal music exists, "music" is only a categorical name that we have given to artificially produced sound patterns that we find perceptually interesting (Haikonen, 2003, p. 120).

The operational approach is therefore an effective antidote against any kind of metaphysics or metaphysical attitude, that is that kind of attitude which was denounced and criticized by authors such as Ernst Mach[2] and Vaccarino (Vaccarino, 1988), and which aprioristically admits, and conceives of, the existence of entities and phenomena that cannot be explained: entities and phenomena whose existence must simply be accepted as it is, for which no description in, and reduction to, some operational terms can be provided, and which cannot therefore be artificially produced or reproduced.

[2] Mach's denunciation of metaphysics is well explained and illustrated by Campelli (1999).

It is in this spirit that Haikonen tackles, in an operational way, a number of mental phenomena and cognitive processes: sensation, perception, visual perception, auditory perception, touch perception, multisensory integration, attention, perception of self, learning, memorization, perception of time, deduction, reasoning, intelligence, creativity, imagination, emotions, pain, pleasure, beauty, ugliness, humor, laughter, music, motivation, needs, drives, goals, language, grammar, syntax, thought, consciousness and self-consciousness. For some, he just sketches a simple operational analysis, for others he offers an exhaustive description of the mechanisms responsible for their production, putting forward a functional model that can serve as the basis for their artificial reproduction. For all of them, anyway, he strives to give an operational account, without leaving anything undefined or unanalyzed: this is very apparent from the beginning, that is, from the questions he poses about the mental phenomena and cognitive process he is about to investigate. The kinds of questions he poses and the way he poses them do not allow any doubt about his intentions. Here are just few examples:

> What happens in the brain when I have a thought or try to solve a problem? (…) Could a thinking machine be built? Could I do it? What would it take? (Haikonen, 2003, p. IX).
>
> What should it take to design a conscious machine and how should we approach this task? (Haikonen, 2003, p. 1).
>
> On what mechanisms is intelligence based? (Haikonen, 2003, p. 94).
>
> If the pain signals were indeed similar to the other neural signals, like those originating from the eyes or ears, then why would the pain signals be felt as painful, why don't the other, similar signals feel like anything? What could then be the mechanism for pain, the cause for the specific *feel* of pain? (Haikonen, 2003, p. 103).
>
> How can we perceive our own thoughts? What kind of trick is needed here? (Haikonen, 2003, p. 149).
>
> How do we make a machine perceive something? (Haikonen, 2003, p. 181).
>
> The cognitive architecture provides the perception process that is able (…) to produce mental content that is devoid of information about the carrier medium. Thus we can have the flow of "immaterial representations", but the question remains: How does the machine become aware of these representations? To whom are these presented? Do we still need a higher level observer and supervisor, a discrete machine self? (Haikonen, 2003, p. 250).

Undoubtedly, Haikonen is not only a good scientist and technician, but also a good philosopher, since he is fully aware of the philosophical principle that a well-asked question, which already contains the seed of its answer, is already half of a good answer. His questions get to the heart of the problem immediately: how can a certain phenomenon be operationally explained and described? Which operations and mechanisms produce the phenomenon? Can we artificially reproduce the phenomenon?

Obviously, given the limited scope of the commentary, I cannot give a full and exhaustive account of all the analyses put forward by Haikonen and do full justice to these. Therefore, I will focus mainly on what I consider to be the more significant aspects, namely, the mechanisms responsible for perception, consciousness, and language. Let us begin with perception.

Perception

As psychological data and subjective experience show, perception is not a simple, passive process, nor can it be assimilated to a straightforward pattern recognition. Our senses provide only some raw – or somewhat preprocessed – sensory information, and it is up to further processes to interpret and make sense of it. This is also evident from the fact that: a) we only need a few cues in order to recognize an object; b) we are able to interpret the same object in different ways according to the specific context and situation; c) generally speaking, attention, memory, expectations and predictions can affect what will be perceived. Therefore perception can rather be considered a complex, active process that combines the effects of sensory information and the system's inner information: a process based not only on sensory information, but also on expectation, attention, and contextual information[3].

In this view, Haikonen proposes the system shown in Fig. 1 as the basic system for perception.

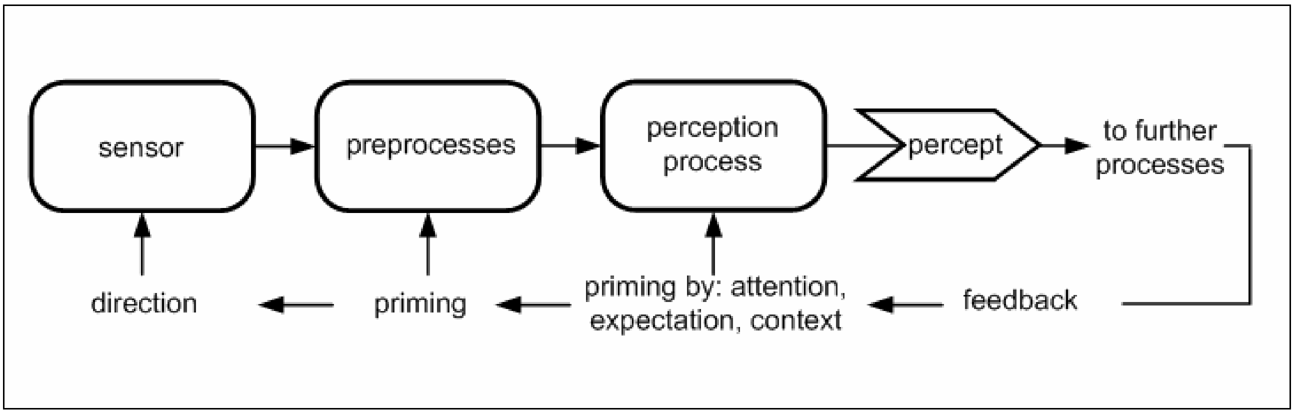

Fig. 1. Sensory and perception processes (adapted from Haikonen, 2003)

In this system, the raw sensory information from the sensors undergoes some kind of preprocessing, such as filtering, transformation and feature extraction. The actual perception process combines the preprocessed sensory information and the system's inner information: in this way, the common part of the sensory signal and the system's feedback can be added together, and

[3] This view is also put forward by some other distinguished authors: among them, I particularly recommend reading Berthoz (1997) and Freeman (1999), above all for the vast evidence they provide.

therefore amplified or "primed". The product of the combination of sensory and feedback information is the "percept". Priming can also be extended to the preprocess, where it adjusts the preprocessing parameters, and to the sensors, where it determines their direction when applicable. Priming:

> will, so to say, help the perception process to find the relevant part of the total sensory information. It is like a teacher giving cues to an unsure pupil. The feedback reflects the system's expectations and therefore what we get now is a percept that should be consistent with the cognitive state of the system. This percept should now fit the puzzle provided by the sensory modalities, context and system's accumulated knowledge (Haikonen, 2003, p. 46).

As Haikonen observes, the mechanism of priming by expectation lies at the heart of, and can explain, phenomena as disparate as:

- illusion, dreams, hallucination, imagination, inner ideas; in the case of illusion for example, Haikonen says: "It can be directly seen that illusions are possible in the proposed perception mechanism. In fact, every percept is a kind of illusion, created by the priming process" (Haikonen, 2003, p. 47);
- misspelling: "Misspelled words are hard to detect because we know and expect the correct words and therefore we will also perceive them as such" (Haikonen, 2003, p. 46);
- the perception of inner states, and inner linguistic thoughts, that is, introspection: "The perception process facilitates also the system's access to its inner states. This is *introspective perception* and in that case the system input, the perceived representation, consists of the feedback representation only, there is no sensory input or the sensory input is suppressed" (Haikonen, 2003, p. 184),
- rhythm (Haikonen, 2003, pp. 61-62);
- recalled memories: "I am proposing here that recalled memories are perceived by the same process that is used for sensory perception. This can be achieved with a feedback that projects the evoked neural activity patterns back to the sensory perception area" (Haikonen, 2003, p. 83);
- short-term memories: "The feedback loop can be made to circulate its signal for a while so that the instantaneous percept is sustained temporarily. Thus the perception/response modules can also act as reverberating short-term memories" (Haikonen, 2003, p. 187; but see also pp. 83-84);
- and attentional mechanism (Haikonen, 2003, p. 183).

By slightly modifying the circuit in Fig. 1, the potentialities of the system can be improved to make it able to perform comparisons. According to Haikonen, the improvement could be achieved through a very basic hard-wired mechanism that detects the relation between sensory signal patterns and feedback signal patterns (Haikonen, 2003, p. 48, pp. 182-183). The comparison between the two patterns can yield either a “match” signal, if the feedback signal and the sensory signal depict the same entity, a “mismatch” signal, if the feedback signal and the sensory signal do not match, or a “novelty” signal, whenever a sensory signal appears without a corresponding feedback signal (Fig. 2).

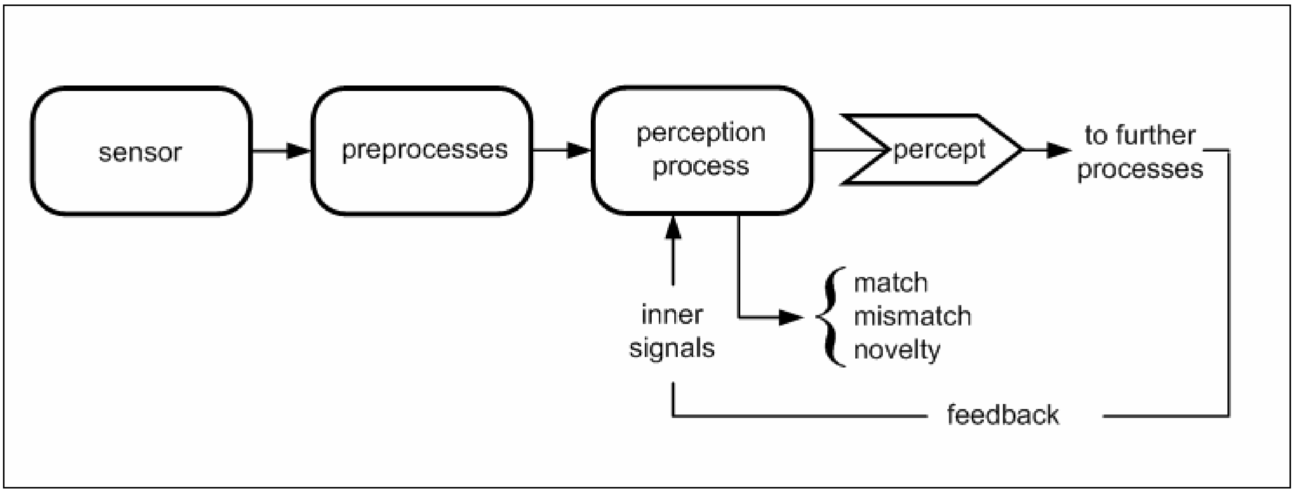

Fig. 2. Match, mismatch and novelty detection (adapted from Haikonen, 2003)

Empirical findings on neural match/mismatch functions in the brain were recorded by Näätänen et al. (1990, 1995), who recorded match/mismatch/novelty related EEG signals.

As Haikonen observes, the mechanism that allows us to perform comparisons should be a very basic hard-wired one, because we are able to make comparisons easily and usually without any conscious effort: we just need to look at, listen to, taste, or touch something, to know that it differs from, or is equal to, something else.

Therefore, the system shown in Fig. 2 allows us to answer questions such as “Is this a book?” or “Have I seen this before?”, that is, generally speaking, to make evaluations and deductions. But it also makes two very common perceptual processes possible: predictive perception and searching perception.

Predictive perception occurs when the feedback signal is a prediction of the input. “This prediction may arise due to associations to previous percepts. In predictive perception match/mismatch signals will indicate the accuracy of the prediction” (Haikonen, 2003, p. 184).

Searching perception occurs when the feedback signal is an internally evoked representation of a desired entity that has to be found or detected. In this case, match/mismatch signals will indicate the successful/unsuccessful outcome of the search.

According to Haikonen, the match/mismatch mechanism is also involved in attention control, in the sense that the mechanism can detect the sensory signals that are relevant for the system (in terms of its needs, evocations, desires, etc.) and pilot the system’s attention towards them (Haikonen, 2003, pp. 68-69, and p. 217). For Haikonen, attention is “a biological neural system’s basic way of favoring the strongest signals, a process that is present already in the simplest central nervous system” (Haikonen, 2003, p. 70). There would therefore be no need for any special “attention box”, since the attention mechanism is distributed within the neural system. Attention operates mainly with signal strengths and thresholds (Haikonen, 2003, p. 217), in the sense that it favors or hinders signals though these: “The signal intensity of distributed signal representations determine which representations will pass the various thresholds and will thus become the focus of operation: the focus of attention” (Haikonen, 2003, p. 213). The system’s attention may be guided by situations and needs that are relevant for the system: indeed, the importance and relevance of a given event, situation, need, representation or inner evocation is one of the main causes of attention modulation and control. How do modulation and control occur? Sensory match/mismatch detection plays a role here. If, for example, a sensory percept matches an internally evoked percept, then match-condition occurs. The match-condition can be used either to amplify the sensory percept through the feedback signal, or to elevate the related thresholds so as to gate out unrelated lower level signals. In this way, whenever a match-condition occurs, the processing of the sensory percept is favored either by the amplification of the sensory signal or by the threshold limits (or by both mechanisms): that is, using Haikonen’s definition of attention, the focus of attention will be maintained on the sensory signal. A mismatch-condition, on the contrary, would lead the system to refocus its attention, whereas a novelty condition would lead the system to focus on the novel sensory signal.

By making another slight modification to the circuit in Fig. 1, it is possible to have a system that can handle and associate representations of different sensory modalities. The kind of circuit that we have considered until now is only able to handle representations of its own sensory modality, and can only associate its own types of representations with themselves via the feedback loop. A real cognitive system would instead include a number of sensory modalities (visual, auditory, tactile, etc.), which can interconnect and give rise to representations symbolizing multiple affairs (for example, a sensed sound pattern can be made to signify an associate visual representation.).

These associative cross-connections between different perception modules can be achieved via in-loop neuron groups as shown in Fig. 3. A neuron group (Haikonen, 2003, pp. 177-180) is a circuit element that can learn the association between an original signal and a number of other signals by repeated coincidences. In a neuron group, an original input signal or representation is associated through repeated coincidences with the signal that is to be associatively connected, or, to use Haikonen's terminology, the "evoking representation". After learning, the evoking representation alone can evoke the same output representation as the one that the original input representation (with which the evoking representation is associated) would produce. A special application of the neuron group is auto-association, by means of which the original input representation of a given perception module X is also used as the evoking representation of the same perception module X.

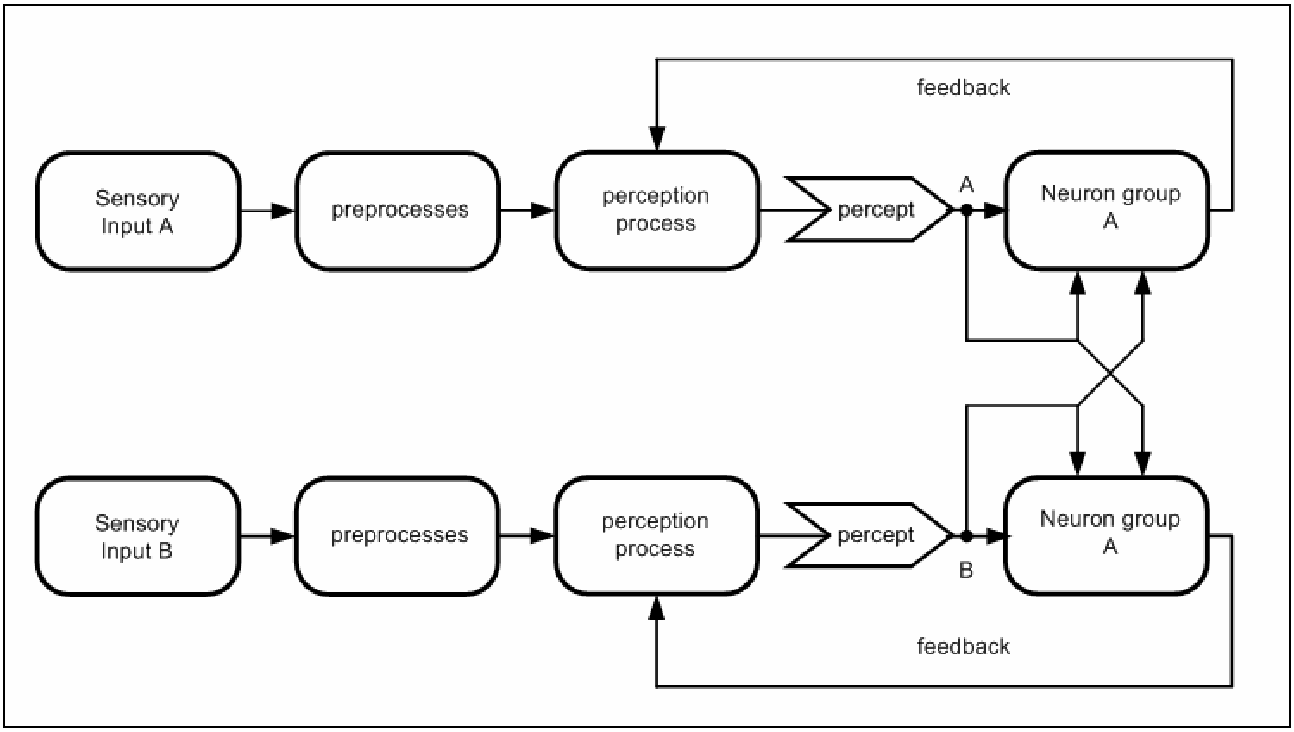

Fig. 3. Cross-connected perception/response modules (adapted from Haikonen, 2003)

In Fig. 3, the two perception modules transmit their percepts to each other's neuron groups:

These cross-connected modules will now be able to learn sensory input entity pairs. If a certain percept A and another percept B appear together repeatedly, they will be associated together. Thereafter, the percept A will be able to evoke

the constituent signals of the percept B at the neuron group of module B.` These signals are forwarded to the perception process of the module B and will thus become the percept B. Likewise the percept B will be able to evoke the percept A at the module A. The associative evocations amplify each other and a self-sustaining closed loop is established; the percepts A and B are bound together (Haikonen, 2003, p. 188).
.

This module can then realize direct perception, primed perception, introspective perception, and cross-associative evocation.

Consciousness

Haikonen observes that our conscious mental life is characterized, among other things, by: a) intentionality and meaning: it can represent, and be about, another thing without being that thing itself; therefore it can give meaning to things and interpret them; b) qualitative properties: our mental experiences are not only about something, they also have the qualities of that something.

In order to tackle the issue of consciousness, he distinguishes the *weak* problem of qualia from the *strong* problem of qualia.

He formulates the *weak* problem of qualia in the following way:

How can things be represented by something else and how can these representations be differentiated from each other; why red and green look different, why different tones sound different, why different odours smell different why the sensation of red is different from the sensation of a sound, etc. (…) How neural signs, each superficially similar to every other, could possibly be about something and carry properties like “redness”, “sourness” or what ever. (Haikonen, 2003, p. 147).

According to Haikonen, each individual signal derives its meaning from the “point of origin”, that is, from its own property detector. A given signal can depict for instance “redness” simply because its causal origin is “redness”, that is, reflected light’s respective spectral properties in outside world. The signal itself does not have, and need not have, the property of “redness”: it simply tells whether the designated property is present or not. The meaning is bound to the physical signal lines and their physical point of origin. Signals are causally connected to corresponding sensor outputs: in this way the basic meaning of each signal is fixed to a detected feature of the outside world. Therefore, technically speaking, all that is needed to solve the weak problem of qualia is for the system to be configured so that the causal point-of-origination meaning is preserved. On this subject, Haikonen observes:

This is not a very weighty restriction and very complicated signal processing operations can still be performed. If you proved otherwise, your mobile phone, video and colour TV would no longer work and autopilots could no longer fly planes from London to New York for example. I am not proposing that such equipment is conscious, I am only illustrating the point that signals can be about something for their system and they can carry the properties of that something (Haikonen, 2003, pp. 147-148).

The architecture of the machine suggested by Haikonen is such that the machine perceives the signals themselves rather than the material substrate that carries the signals: transistors and circuit elements that carry the signals would be “transparent” to the machine. As he observes:

> Circuit transparency is by no means unknown in electronics, in fact in many applications it is a desired feature. Radio, television, telephones all contain the principle of transparency. The operation of the distributed signals in the cognitive machine can be compared to radio transmission where a carrier signal is modulated to carry the actual audio signal. The carrier wave is what is received, yet what is detected is the modulation, the actual sound signal that is in causal connection to the original physical sound via a microphone (Haikonen, 2003, p. 248).

The signals carry information about the sensed features of the external world as an on/off modulation. What the machine “sees” is the carried modulation. In this way, Haikonen observes, the machine also realizes the mind-body effect, by virtue of which we humans perceive our thoughts and consciousness as immaterial, without any perception of the underlying material brain processes. The apparent immateriality of our thoughts would then be caused by omission: “The inability to perceive the actual signals and machinery that carry the perceived information” (Haikonen, 2003, p. 249).

The very same principle of circuital transparency applies not only to the signals and machinery that carry the perceived information, but also to the *location* of the neural sensory points-of-origin of the percepts. According to Haikonen:

> nerve signals do not carry any position information about their actual end-location even though they are rigidly wired. From the brain’s point of view the nerve signals just appear, there is no built-in point of origination. Therefore the perceived point of origination can be whatever we associate it with (Haikonen, 2003, p. 72).

This explains why we do not perceive percepts as originating at the senses or at the related sensory nerve endings, but as being located outside in the world, and why when we scan a rough surface with a rigid stick we perceive the grooves on the surface more than the vibrations of the stick against our fingers. Indeed, since nerve signals do not carry any information about their point of origination, we can freely associate a certain sensation (for example, a visual or a tactile one) of a certain object to the motor commands that enable us to reach such an object. This allows us to reach out for the object (rather than to our eyes or ears!). Even in this case, the possibility of perceiving the point of origination as located outside in the world is caused by the omission of nerve end-position information, rather than by built-in design.

In summary, weak qualia allow us to separate sensations from each other, and perceive various signals and patterns as different: this ability is the direct result of the style of information representation. But weak qualia do not yet explain our capacity to really feel something, to subjectively experience anything, for example the “redness” of “red” or pain. This is exactly what Haikonen defines as the *strong* problem of qualia: “Why some percepts really do feel like something, especially like pain and pleasure, and to some extent beauty and ugliness, too” (Haikonen, 2003, p. 148). As he observes, unlike weak qualia, strong qualia cannot be explained by the preservation of a causal point-of-origination path, that is, by the built-in circuit arrangement of the system. Indeed, feeling pain or pleasure, for example, cannot be explained by means of a sensor detecting the property of an external entity: pain and pleasure are not representations of things and objects of the outside world. They are not properties of a sensed entity. The non-representational nature of pain and pleasure is further exemplified by the fact that we cannot memorize pain and pleasure, and evoke them afterwards.

But if pain and pleasure cannot be explained by means of a sensor detecting the property of a sensed entity, how can we account for them? According to Haikonen:

> Pain sensors do not sense pain, the sensed entity is cell damage and the caused signal indicates only that pain is to be evoked. Thus the feel of pain is not a representation, instead it is a system reaction. The pain signals themselves do not carry the feel of pain, instead the feel arises from the effects that these signals have on the system and this in turn depends on the way the signals are connected to the system (Haikonen, 2003, p. 103).

Therefore, Haikonen considers feeling pain as a “system reaction”. What does this “system reaction” consist of?

> Pain “demands attention”; it disrupts any attention that is focused on any ongoing task. Obviously pain signals are transmitted to every modality in the frontal cortex and the message, so to say, is “stop whatever you are doing and try something else so that this signal might stop!”. This is because the pain signal itself does not know who should do what to stop the damage and therefore it has to broadcast its message to everybody and thus disrupt the attended processes within each modality. Pain does not allow the other modalities to relax, instead it tries to stop their present activity and start something else. (…) I consider this disruptive broadcasting as a fundamental property of pain and *I would like to go as far as to propose that the subjective feeling of pain is indeed caused by attention disruption* especially in the frontal cortex area (Haikonen, 2003, p. 104).

In Haikonen’s view, feeling pleasure is a also system reaction: “Pleasure, like pain is not a property of a sensed entity. There is no pleasure to be sensed and represented, instead pleasure is a system reaction that can be evoked by various sensations” (Haikonen, 2003, p. 105). More specifically, pleasure entails: a) continuing the pleasure-causing activity to sustain the feeling of pleasure; b) focusing attention on the pleasure-causing activity, and excluding attention on other stimuli; c) memorizing pleasure-causing things and acts, so that they can be identified and repeated in the future. For Haikonen, pleasure, like pain, is connected to attention as well, but in a different way from attention:

While pain uses brute force to disrupt the attention within modalities pleasure tries to sustain its attention focus by having non-related circuits and modules relax. In this way only the pleasure evoking activity will be continued while other activities are suppressed (Haikonen, 2003, p. 105).

In addition to pain and pleasure, Haikonen lists some other elementary sensations ("elementary" as opposed to more demanding sensations from a processing point of view, such as the visual and auditory ones) that would elicit a basic system reaction, namely:

(a) good taste and smell, which elicit positive responses, such as acceptance of, and approaching, the source of the good taste and smell;
(b) bad taste and smell, which elicit negative responses, such as rejection of, and withdrawal from, the source of the bad taste and smell;
(c) match, which implies sustained attention;
(d) mismatch, which implies refocused attention;
(e) novelty, which implies focused attention.

According to Haikonen, system reactions are direct and rather automatic pre-wired responses to elementary sensations: they do not require a complicated cognitive evaluation of the stimulus or of the situation, thus enhancing the prospects of survival. The various combinations of system reactions give rise to various emotions: curiosity, astonishment, caution, fear, anger, desire, love, happiness, sadness, envy, horror, etc. (Haikonen, 2003, pp.114-116).

Having explained feelings and sensations (such as pain, pleasure, taste, smell, match, mismatch and novelty, but also beauty and ugliness) in terms of system reactions, Haikonen defines strong qualia as: "temporal behaviour patterns of inner attention caused by system reactions (…) We have strong qualia because we have system reactions that affect attention" (Haikonen, 2003, p. 148).

This way of dealing with the problem of the phenomenal aspect of consciousness is very similar to mine (Marchetti, 2001, 2006). Indeed, in my opinion the phenomenal aspect of consciousness can be also explained in attentional terms. I believe that conscious experience is the product of attentional activity. More specifically: attentional activity can be performed thanks to a special kind of energy, nervous energy, which is supplied by the organ of attention; in order to perceive, think, speak, move, and more in general act, we have to perform attentional activity; when we perform attentional activity, we use our nervous energy; attentional activity directly affects the organ of attention, causing a variation in the state of nervous energy; it is this variation that constitutes the phenomenal aspect of consciousness. Haikonen's proposal, basing strong qualia on attention, and more precisely on its variations ("I have proposed that 'pain', 'pleasure', 'good' and 'bad' get their specific feel via the attention affecting mechanisms", Haikonen, 2003, p. 116), therefore seems to tackle the problem of phenomenal consciousness in the same way as I have done.

However, as far as I understand, there is a fundamental difference between his model of consciousness and my model of consciousness. While in my model, the phenomenal aspect of consciousness is explained with the modification of the energetic state of the organ of attention which is determined and induced by the use of attention itself (in other words: qualia *are* the modifications of the state of nervous energy supplied by the organ of attention), in Haikonen's model, attentional variation causes subjective feelings, but *is not*, and *does not coincide with*, them. In fact, Haikonen's model resorts to some additional components (namely, "system state sensors" and their "related perception process") to explain the phenomenal aspect of consciousness, and how system reactions can be consciously felt:

In order to perceive and internally represent system reactions the system needs *system state sensors* and their related perception process (…) pain, pleasure, taste and smell (good/bad) sensations as well as match/mismatch/novelty states are able to initiate the basic system reactions such as those listed before and also other physiological reactions. These system reactions *are perceived by system sensors and their respective perception process* (Haikonen, 2003, p. 113) (italics are mine).

The block diagram of Fig. 4 shows how system reactions are connected to system sensors and their perception process.

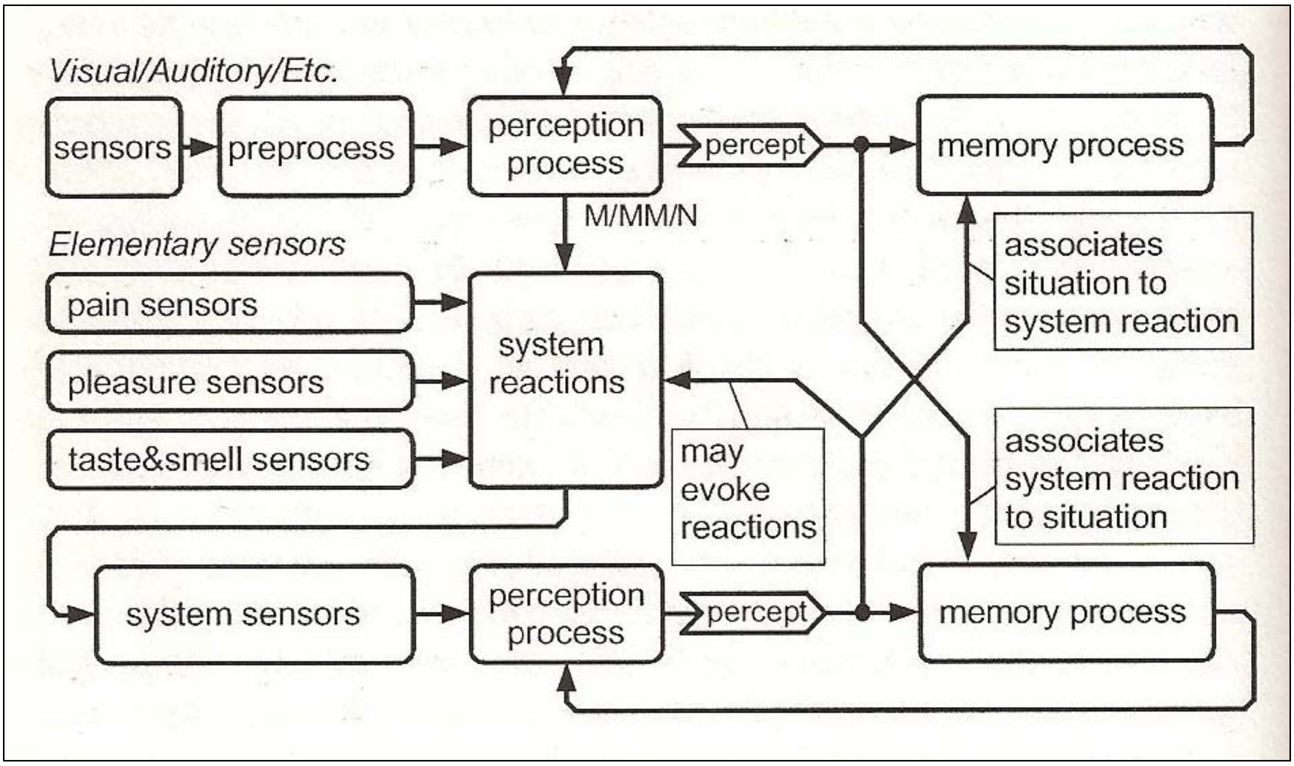

Fig. 4. The connection of system reactions to perception and cognitive processes

The explanation of how a machine can experience strong qualia is therefore shifted to the perception process (indeed, Haikonen clearly states that: “perception processes are a necessary prerequisite for consciousness”, 2003, p. 149, and that: “Consciousness arises from perception, without percepts there is no consciousness”, 2003, p. 271), but not only: in fact, Haikonen, recognizing that not all percepts reach consciousness even though they may affect behaviour, admits that “the perception process alone is not sufficient to explain consciousness” (Haikonen, 2003, p. 250).

So, what is involved in conscious perception? What is it that makes a percept conscious? According to Haikonen:

> the difference between conscious and non-conscious operation would be the level of active cross-connections and binding between modalities; the cross-modality reporting and learning of related associative connections and thus the establishment of episodic memories of the event. In non-conscious operation the cross-connections are minimal and the operation of the different modalities is not unified, it is not about the same topic, there is no binding. In conscious actions the operation of the different modalities would be unified; the inner attention of each modality would be focused on the same topic (Haikonen, 2003, p. 254).

As we can see, Haikonen resorts to the notions that are very well-known in consciousness studies of binding (see for example: Singer, 2001) and widespread brain interactions (see for example Baars’ global workspace theory: Baars, 1988). However, this does not yet constitute an explanation of what happens inside the perception process module that makes a percept conscious. This certainly represents an explanation of what happens outside the perception process, of how the various perception modules interconnect, and how all the system focuses on the same topic, but does not explain which mechanisms and operations occur inside the perception module to turn system reactions or any other sensations into strong qualia.

Another interesting similarity between Haikonen’s model and the model of consciousness I have put forward is the importance that Haikonen assigns to the interactive process between the system and its environment in forming and processing meanings. Haikonen rightly points out that one of the hallmarks of a conscious machine is its ability to produce and process meaningful actions, responses, and communications. While a digital computer manipulates symbols (binary code words, strings of ones and zeros) that do not have any attached meanings for it, but whose meanings are solely determined by the programmer, human cognition is able to autonomously produce and process symbols that have an attached meaning: “It is precisely this processing with meaning that makes human cognition superior to any digital computer. Therefore a cognitive machine should also incorporate symbolic processing with meaning” (Haikonen, 2003, p. 166). But how are these "attached meanings" realized by the cognitive machine? In Haikonen's view, a system's actions and responses can be considered meaningful when they reflect the goals and needs of the system, as well as the limitations and demands set by the environment. Consequently, the requirement of attached meanings calls for tools that bind symbols to "external and internal entities, actions and relationships, that is, sensors and the processes of perception and learning" (Haikonen, 2003, p. 167). The complexity of the interactive processes between the system and the environment are shown by Haikonen in Fig. 5.

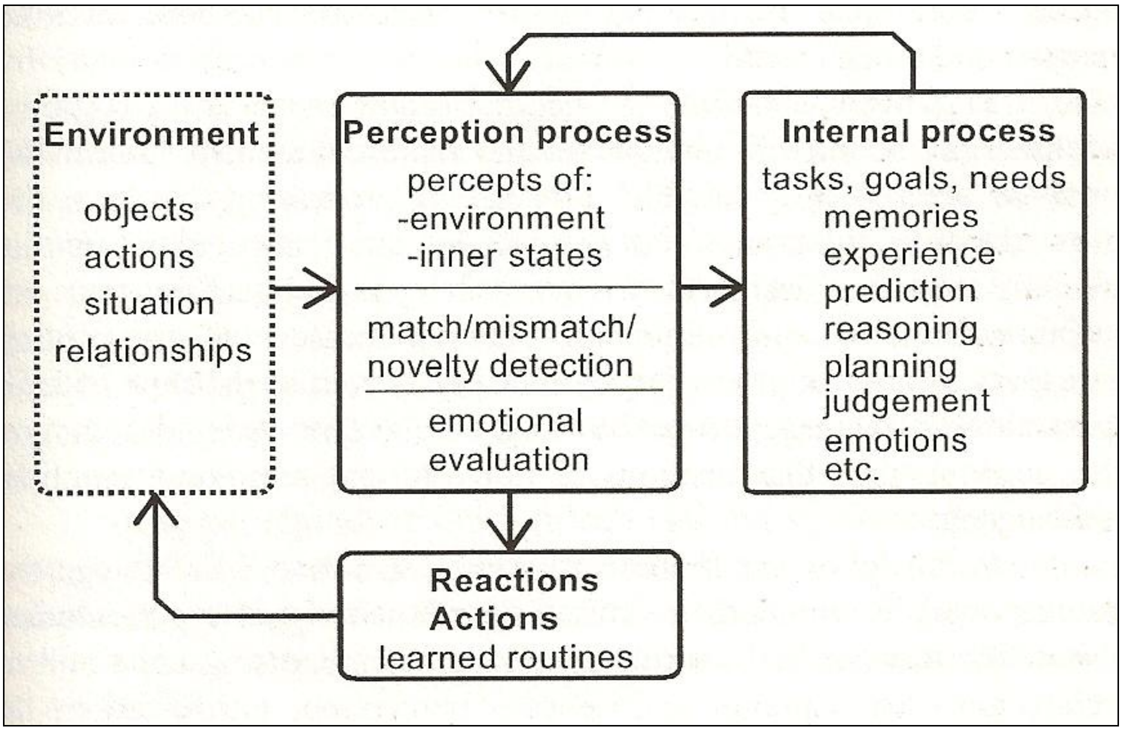

Fig. 5. The complexities of a cognitive system

This is very similar to the one I proposed in my model of consciousness (Marchetti, 2001). In my view, studying consciousness means taking a new perspective that considers how a *subject* emerges from an organism's continuous use and application of its nervous energy. This perspective implies seeing the subject as an active agent that is personally and directly involved in constructing not only itself but also its own knowledge. It becomes a subject because it acts, and, by acting, it differentiates itself from the environment and other beings, thus getting to know them. An object becomes an object and acquires a *meaning* for the subject only if the subject can relate the object to itself in some way. Therefore, every object can be defined in terms of the subject's activity, where "subject's activity" means the activity a subject has to perform in order to emerge as such. An object exists and has a meaning because there is a subject that gives it a meaning, and, conversely, a subject exists and has a meaning because by acting it has been able to differentiate itself from that object (on this point, see also Marchetti, 2005b).

In my model, the process leading to the production of the subject can be divided into three main steps or components: the schema of self, the actions it performs and the conscious perceptions it has consequent to its actions (Fig. 6). These three components are very similar to Haikonen's three blocks: respectively, "Internal process", "Reactions, Actions", and "Perception process".

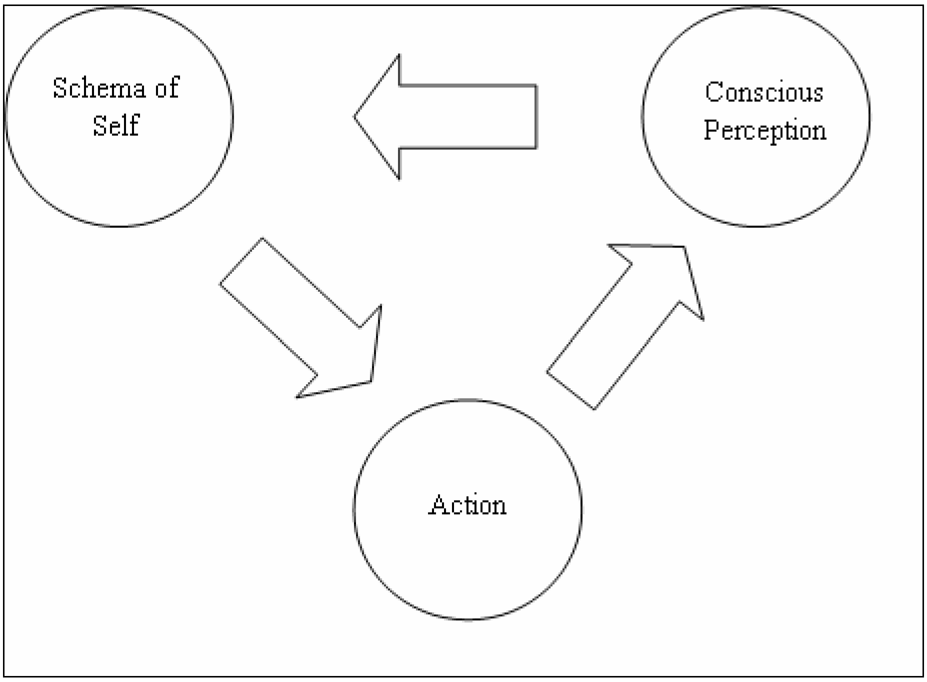

Fig. 6. The process that leads to the production of the subject

The schema of self and all the other schemata it incorporates and coordinates embody all kinds of competence and abilities - linguistic, social, physical, and so on – that the organism innately possesses or has acquired during its life up to that time. It regulates the activities of the organism according to the hierarchy of principles and goals that it incorporates, and the rules specific to each kind of competence. Every action the organism performs is caused by the goals of the schema of self – at the top of which is the principle of survival, which can be expressed as follows: "operate in order to continue to operate" -, and generated and structured by the rules expressed by each kind of competence.

Every action of the organism - whether a single movement, a coordinated sequence of movements, the production of a sound or a word, an inner silent speech, or other - as well as the consequences of the action, can be perceived by the organism. The core parts of the perceptual system are the organ of attention and the somatosensory system and sense-organs. Since every action unavoidably entails sensations that concern the body and the environment, the organism is able, through them, to understand and define its limits and the limits of the objects of the environment. Once the organism has consciously perceived its action or the consequence of its action, the information concerning its body, the objects of the environment or the relation between its body and the objects, becomes available for the schema of self, and can be adequately used to update it and adjust the rules of the relevant competence.

The stream of consciousness (Fig 7) is the result of the uninterrupted interaction of the schema of self and the perceptual system. Every conscious perception affects the schema of self, modifying and updating it. Every modification of the schema of self implies a new particular instruction to the perceptual system, and in general to the organism. The uniqueness of each single pulse of consciousness is determined by the particular instruction that the schema of self gives to the perceptual system each time.

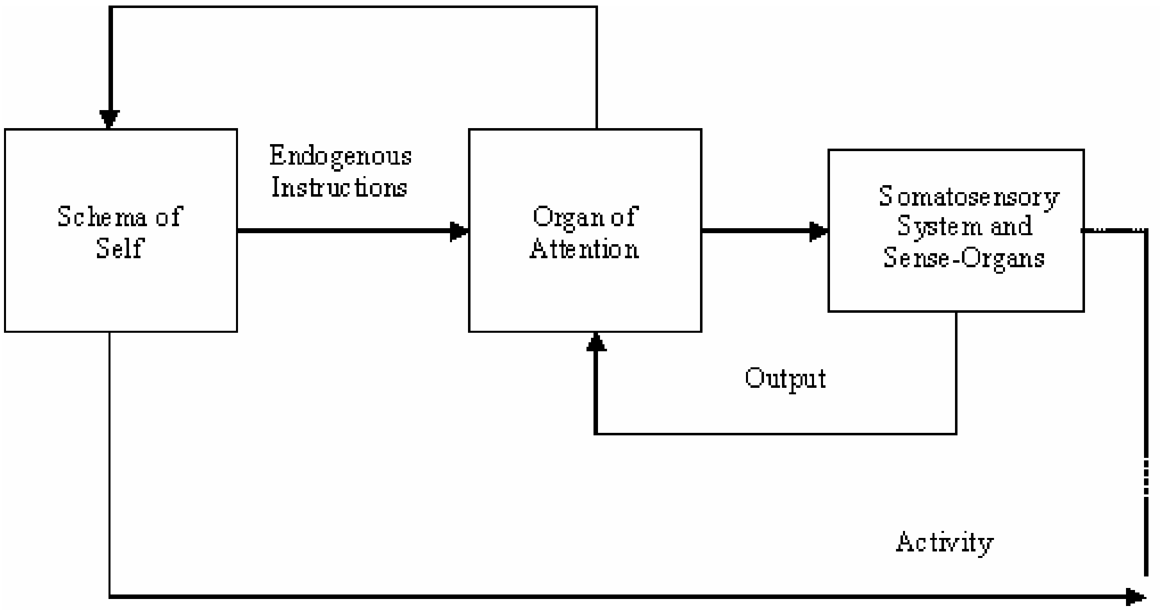

Fig. 7. The circuit that generates the stream of consciousness

Language

According to Haikonen, neither a pure horizontal approach to language – that is, one in which the meanings of words are defined by their relationship to other words – nor a pure vertical approach to language – that is, one that assumes that the meanings of words derive from entities in the outside world, or from our percepts of those entities – are sufficient for the purposes of cognitive machines:

The truth may be that both vertical and horizontal approaches are needed. The meanings of concrete words must be grounded in the real world, but on the other hand the linguistic apparatus must be able to work by itself the rules that it has extracted from the real world (Haikonen, 2003, p. 127).

Haikonen therefore suggests combining the horizontal and vertical approaches: he calls this approach the “multimodal model of language” (Fig. 8).

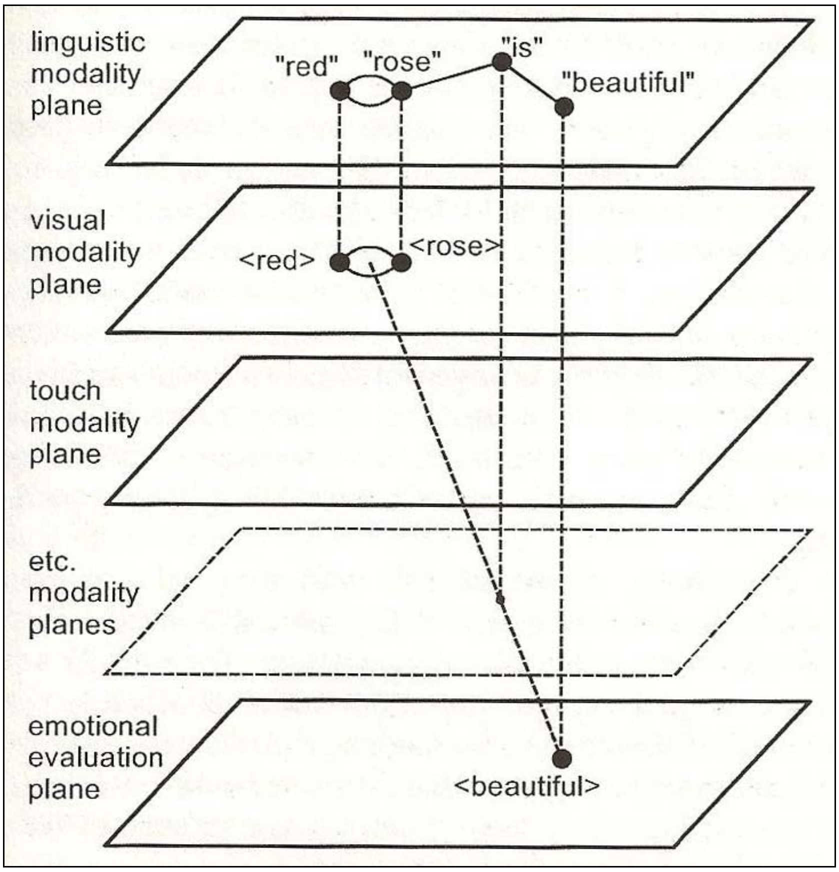

Fig. 8. Haikonen’s multimodal model of language

The multimodal model of language is based on the assumption that each sensory modality acts as a “plane” that stores and associatively manipulates representations of its own kind. The representations within a plane can be associated with each other and also with representations in other planes; the associations can change over time.

These planes learn and acquire their representations via the perception process. Therefore, the basic meanings of representations are causally grounded in the external world entities and sensed bodily conditions.

The plane for spoken language resides in the auditory modality. Due to the associative connections, words may be associated with other representations in the same or other modalities.

Therefore, words no longer depict their sound pattern alone, but also representations belonging to other modality planes: consequently, the percepts of the other planes can be linguistically described.

Here is an example of how the multimodal model of language actually works. Fig. 8 represents the sentence: “Red rose is beautiful”. The meanings of the words are vertically grounded to the visual plane representations of <red> and <rose> and the emotional plane representation for <beautiful>.

This applies to entities related to the external world and to system states, but what about syntactic and grammatical relationships? According to Haikonen:

> Here the binding of <red> and <rose> on the visual plane take place due to the spatial coincidence, both properties are detected at the same position. The idea “<red rose> is <beautiful>” corresponds to the shifting inner attention from the percept <red rose> to the simultaneous percept of <beautiful> and the storage of the link. The binding and linking are conveyed to the linguistic plane and after a few examples like this a pattern emerges, the linguistic plane will bind and link the words of heard sentences in a similar way even if no grounding to other planes exist. This is the basic mechanism for syntax acquisition. *Thus syntax is seen here to arise from the real world relationships between entities*; the syntax reflects these. As I see it, the real world relationships are globally more or less the same, the grammars of natural languages must have syntactical structures to represent these and this is the reason behind the functional equivalence of these structures, not any proposed innate universal grammar (Haikonen, 2003, pp. 132-133).

Syntax would therefore emerge from the recurrent perception of the real world relationships between entities. While this proposal is certainly appealing under certain aspects, it does not seem to be so uncontroversial under other aspects. In fact, any physical situation can be described syntactically in more than one way. A situation that can be described as “Red rose is beautiful” can also be described as “A rose that is red is beautiful”, “A red-coloured rose is beautiful”, “Beautiful red rose”, “If a rose is red it is beautiful”, “I like a red rose”, etc. (moreover, in Italian you can say not only “rosa rossa” but also, inverting the order and slightly changing the whole meaning, “rossa rosa”). In my opinion, this seems to show that explaining syntactical (or grammatical) words as deriving from situations which are repeatedly encountered is not sufficient, and that some other kind of explanation must be taken into account.

I think that the major misunderstanding is caused here by the idea of “real world relationships between entities”. This idea can certainly be reflected by and in syntax and linguistic expressions sometimes: for example, no one will believe you if you say that if you put your hand in the fire you will not burn your fingers, or if you say that a pear can be called an apple, or if seeing a cat eating a mouse you say that it is the mouse that is eating the cat.

Other times however, this idea can hardly be considered the basis of any syntax. Consider for example the phenomenon of “temporal displacement”: given a sequence of very brief stimuli, say a-b-c, subjects often perceive a different sequence, say A-C-B. The phenomenon, which had been noticed by astronomers in the early 19th century, and was largely investigated by Wundt (1902), Rubin (1949), and Vicario (2005), clearly shows that phenomenal time does not correspond, and cannot be reduced, to physical time.

A similar but even more astonishing piece of evidence showing that the experiential world is not the same as the physical world is provided by the phenomenon of *continuous displacement* or *stream segregation* experimentally described by Bozzi and Vicario (1960): when subjects listen to a sequence of stimuli composed of the four tones shown in Fig. 9a that is repeated cyclically, they will hear a single sequence of low and high sounds if each stimulus lasts about 200msec (Fig. 9b), and *two different synchronized sequences* of sounds (a low trill and a high one) if each stimulus lasts about 50msec (Fig. 9c).

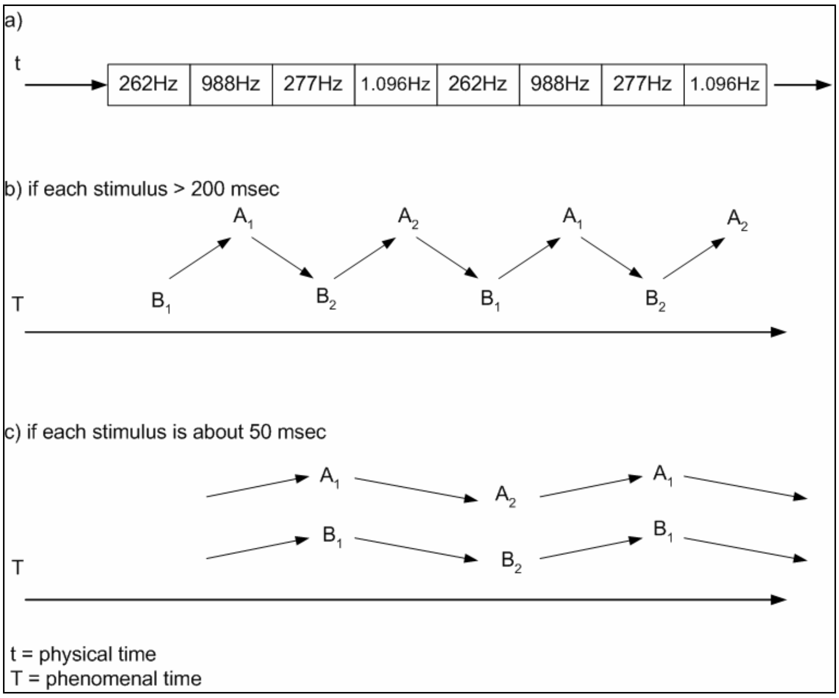

Fig. 9. Bozzi's and Vicario's experiment on continuous displacement

Or, finally, consider the phenomenon of "prior entry": when a person attends to a stimulus, he or she perceives it as having occurred earlier in time than it would if he or she was not attending to it (Shore et al. 2001, Shore and Spence, 2004).

All these strange phenomena clearly show that relationships between entities depend not only on what is really happening out there, in the "real world", but also on what is happening in the observer, in ourselves. This consideration obviously opens up a completely different scenario from the one offered by the idea of "real world relationships between entities". Indeed, if relationships between entities also depend on the attitude, attention, thoughts, opinions, and education of people, they will always be changing and modifying, and therefore can hardly, or at least “not always”, constitute a basis of recurrent, regular experiences on which to build the whole syntax.

Moreover, the idea of “real world relationships” would imply a denial of the productive and discovering power of our mind: if the relationships between entities were already all known and experienced, it would not be possible to always describe events and entities in new and alternative ways. In my opinion, it is possible to overcome these difficulties by approaching the problem of syntax and grammar (but, more in general, of language) with a linguistic or semantic theory that specifically incorporates within it mental operations as the building blocks of the meanings of words. An outstanding example is offered in this sense by Benedetti’s theory (2005a, 2005b, 2006, 2008), which, while not denying the importance of the physical and psychological relationships between entities, provides a systematic basis for performing semantic analyses in terms of mental operations.

Anyway, apart from the problem posed by the proposal of the origins of syntax, it must be noted that Haikonen’s primary concern[4] is that of explaining syntax as a conventional means (whether it is expressed by word order, word endings or inflections) of linguistically conveying the information that words alone cannot convey. Consider for example a case where a person sees a cat eating a mouse. How can he communicate this to another person? With vocabulary only he can utter the words “cat”, “mouse” “eat”. There is nothing in the words only that can convey the information about the relationship between these entities: who eats whom. It is precisely syntax that conveys this information. As such, for Haikonen syntax is not out there in the real world but in speech, even if it could be initially occasioned by real world situations and relationships.

[4] Haikonen’s personal communication.

References

Baars, B.J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press. Cambridge.
Benedetti G (2005a) "A presentation of Operational Methodology". www.mind-consciousness-language.com
Benedetti G (2005b) "Basic mental operations which make up mental categories". www.mind-consciousness-language.com
Benedetti, G. (2006). "Operational Noology as a new methodology for the study of thought and language: theoretical aspects and possible practical applications". *Cognitive Processing, 7*: 217-243.
Benedetti, G. (2008). "A semantics 'outside' language: Operational Semantics. A new semantic theory, based on the nature and structure of thought". www.mind-consciousness-language.com
Berthoz, A. (1997). *Le sens du mouvement*. Odile Jacob. Paris.
Bozzi, P. and Vicario G.B. (1960). "Due fattori di unificazione fra note musicali: la vicinanza temporale e la vicinanza tonale". *Rivista di psicologia, 54*: 235-258.
Campelli, E. (1999). *Da un luogo comune. Elementi di metodologia delle scienze sociali*. Carocci. Roma.
Ceccato, S. (1972). *La mente vista da un cibernetico*. Eri. Torino.
Ceccato, S. (1980). "Intervento di Silvio Ceccato", In Linguiti, G. L.: *Machine e pensiero. Da Wiener alla terza cibernetica*. Feltrinelli. Milano.
Freeman, W. J. (1999). *How Brains make up their Minds*. Weidenfeld, Nicolson. London.
Haikonen, P. O. (2003). *The Cognitive Approach to Conscious Machines*. Imprint Academic. Exeter, UK.
Libet, B. (2004). *Mind Time. The Temporal Factor in Consciousness.* Harvard University Press. Cambridge, Massachusetts.
Marchetti, G. (1993). *The Mechanics of the Mind.* Espansione. Roma
Marchetti, G. (2001). "A Theory of Consciousness". www.mind-consciousness-language.com
Marchetti, G. (2005a). "Commentary on Benjamin Libet's *Mind Time. The Temporal Factor in Consciousness*". www.mind-consciousness-language.com
Marchetti, G. (2005b). "The importance of non-attentional operations for Attentional Semantics". www.mind-consciousness-language.com
Marchetti, G. (2006) "A presentation of Attentional Semantics". *Cognitive Processing, 7*: 163-194.
Näätänen, R. (1990). "The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function". *Behavioral and Brain Sciences, Vol. 13, No. 2*: 201-233.
Näätänen, R., Alho, K. (1995). "Mismatch Negativity – a Unique Measure of Sensory Processing in Audition". *Intern. J. Neuroscience, 80*: 317-337.
Rubin, E. (1949). *Esperimenta psychologica*. Munskgaard. Copenhagen.
Shore, D. I., Spence, C. and Klein, R. M. (2001). " Visual prior entry". *Psychological Science, 12*: 205-212.
Shore, D. I. and Spence, C. (2004). "Prior Entry", In: Itti L., Rees G. and Tsotsos J. (2004) (eds.): *Neurobiology of attention*. Elsevier. North Holland.
Singer, W. (2001). "Consciousness and the binding problem". *Ann N Y Acad Sci., 929*: 123-46.
Valéry, P. (1973). *Cahiers* (ed. by Judith Robinson-Valéry). Gallimard. Paris.
Vaccarino, G. (1974). *La mente vista in operazioni*. G. d'Anna. Messina-Firenze.
Vaccarino, G. (1988). *Scienza e semantica costruttivista*. Clup. Milano.
Vicario, G. B. (2005). *Il tempo. Saggio di psicologia sperimentale*. Il Mulino. Bologna.
Wundt, W. (1902). *Grundzüge der physiologischen Psychologie.* Engelmann. Leipzig.