The Consciousness of Non-Natural Systems

Pedro Medina Martins
Universidade Tecnica de Lisboa
Private Address: R. Cor. Marques Leitao, 27-1°-esq
1700-124 Lisbon, Portugal
p.medinamartins@mail.telepac.pt

Abstract. The construction of non-natural systems endowed with a consciousness similar to that of human beings is one of the most challenging problems that science presently faces. Manifold reasons justify this situation, amongst which one is prevailing: that in spite of the countless papers written about it, our knowledge of the origins of human consciousness is practically null. This paper ventures a fresh way of regarding these questions following two unconventional approaches: (i)Pask's Conversation Theory (CT) which is introduced here as a means of providing a new epistemological/computational tack to the whole question, and (ii)a scrutiny of the relations baby/mother that will work as a prototype for the emergence of consciousness in artificial systems. The convergence of CT with the 'mechanisms' underlying such relations frames a theoretical and practical *construct* - an experimental robot named ACB2 - in which features belonging to a self-conscious being can already be witnessed.

Keywords: Primary/reflective consciousness, CT, psychoanalysis, emulation, non-natural systems, neuro-fuzzy approaches, ACB2.

1 Introduction

1.1 Objectives

The ultimate objective of this paper is to arouse conviction that my attempt at endowing non-natural systems with a consciousness similar to that belonging to human beings is not a task doomed beforehand to failure but, rather, a tenable and factually achievable goal.

Doubtless, such a controversial objective is rather daring but it meant to be so. Indeed, it is almost meaningless to remark that the question of consciousness in human beings has been a conundrum since the time of the Greek philosophers Alcmaeon of Croton and Hippocrates, through Plato and Aristotle, and later from Descartes until present ages.

Recently, the question has acquired a new dimension, when the countless attempts to give some light to it have been extended to the realm of non-natural systems. In fact, if *human* consciousness remains as one of the most challenging however unsolved problems of present-day science, then how can a non-natural system (ultimately, a 'machine') acquire something that we do not *previously* know what it is?

How surprising the answer to this question may be, it is here that the reasoning on which this approach is based begins. For the sake of simplicity, hereafter I shall highlight its major standpoints under two types of (somehow artificial) designations, respectively denoted by "axioms" and "postulates". Whenever necessary, comments about the

condensed statements which such "axioms" and "postulates" contain, will be inserted into the text. In this sense, the first of such statements claims that

Postulate I.1:
(i)No theoretical or practical artificial 'conscious' construct can be built, without an as far as possible deepened understanding of the origins of human consciousness, (ii)Assuming the cogency of the preceding assertion, and taking into account that human beings by occasion of their birth behave (just like present-day machines) essentially in terms of phylogenetically transmitted reflexes, then one main conclusion can be immediately stressed: that, human consciousness is primarily a temporal, constructive process, *(iii)This process is rooted either in biological or silicon-based architecture hereafter named "Id" (by analogy with Freud's psychic instance). This "Id" (of which its owner is thoroughly unconscious) embraces a set of innate* reflex 'mechanisms' *and a host of* potential *processes waiting for* actualization, *(iv)Since human beings are* altricial animals, *this actualization is crucially dependent on the* interaction baby-mother, *the latter being just the 'object' who will assign* meanings *(and, consequently,* existence*) to the baby's inner and outer worlds, (v)It follows from the preceding items* - and this is the thesis of this approach - that the emergence of consciousness both in human beings and in non-natural systems is, ultimately, *a* **social process**.

For the reader to whom the consciousness problem is familiar, the assumption that individual consciousness is the *ultimate* by-product of a *social process* must, certainly be rather difficult to accept. On the one hand, because it cuts across practically all the theses so far followed by neurophysiologists, cognitive psychologists, philosophers, etc. On the other hand, because consciousness has been regarded as one of the utmost *individualized* features owned by each one of us. Indeed, I can only be aware of *my own* consciousness, those belonging to other people being illations, set up by means of extensional analogies. No one denies this, *myself included.* But in spite of its apparent contradiction I still hold the aforementioned statement. And the reason for this attitude lies either in the *meaning* assigned to such an 'individual consciousness' or in the *period* which it refers to.

1.2 The Approach

I cannot get bogged down with a minute analysis of these questions since their content is far beyond the immediate objectives of this paper. However - in order to avoid the inextricable muddle of confusions, contradictory opinions, endless and (sometimes) meaningless discussions with which the consciousness problem is presently entangled - two research lines will henceforth be followed.

The first deals with the meanings of consciousness that hereafter will be adopted. From the 2400 references that Chalmers mentions in the WWW, or from those quoted by Schater [5] in *The Cognitive Neurosciences* - I decided to follow a slight modification of Farthing's definition who has considered the existence of (i)a *primary consciousness*, which "deals with simple perceptual awareness of external and internal stimuli", and (ii)a *reflective consciousness,* which entails 'thoughts about one's own conscious experiences per se' ". Although *generally* agreeing with this distinction, once scrutinized with a

greater detail these meanings entail however three additional questions *that a great deal of researchers in this area tends to neglect*. Briefly, they consist in
-presupposing the existence of an *adult agent*, who is 'aware' of the distinction between (i) and (ii) above,
-distinction that, in turn, subsumes the existence of an "I" (the aforementioned 'identity'),
- which, finally, frames the ubiquitous *distinguo - also practically neglected by all who have devoted their attention to the problem* - between 'awareness' and 'consciousness'. For example, the Longman Dictionary of English Language and Culture (ed. 1998) or the Webster Dictionary (ed. 1999) both define 'consciousness' as 'the awareness of ...' and 'awareness' as 'the consciousness of...'.

From these distinctions the first stresses the importance ascribed to the *period of time* to which one's 'individuality' is applied. No doubt, *adult* human beings (should) own it. But, certainly, a new born baby has not gotten it yet. Likewise, the sense of "ownership", of "in"/"out", or even of the several objects of its inner and outer worlds, are thoroughly out of court. Under these caveats, the preceding questions can take another overall aspect: *it consists in knowing how consciousness (in both senses) can emerge - or is being set up - from such a biological silicon-based 'instance' that in* Postulate I.1 *I have named "Id".*

Once stated in this form, the second of the aforementioned research lines follows almost immediately. For, if we assume that
(i)in its earliest times of life, the baby is *unaware* of the preceding distinctions,
(ii)it is through the *interaction* with its mother that it progressively begins with the setting up of the first differentiations,
(iii)such interactions are primarily *affective*,
(iv)the 'mechanisms' underlying these affections are *potentially* innate and their *actualization* depends on the type of relationships that, throughout time, the baby will have with its overall 'inner' and 'outer' environments,
then three major subject matters will be entailed in my approach: the tenets of Pask's Conversation Theory (CT), psychoanalysis and robotics, all of them being based on a neuro-fuzzy formalism [2, 3].

A minute analysis of the reasons underlying the choice of these (seemingly disclosed) areas - CT and psychoanalysis, on the one hand, and the relation that the latter has with computing/robotic systems, on the other - is thoroughly out of court. In the aim of this paper it is enough to say that they are the by-product of a personal research project that has started about twelve years ago when (at the occasion of the celebration of the half-centenary of Freud's death) I presented in Vienna of a paper entitled "On the Simulation of Depth Psychic Processes" [11a]. There, the links amongst those matters were sketched by the first time. However, for the reader to whom CT's standpoints are unfamiliar (since the relationships between psychoanalysis and affections belong to the public domain), the following "axiom" provides a cursory survey of the reasons why their use is crucial in this approach

Axiom 1.1:
The adoption of the standpoints of Pask's CT [13] *as one of the scaffolding frames for the consciousness problem relies, essentially, upon the following main aspects: (i)The*

philosophical idealist approach *where they are ultimately based. Amongst other consequences this means that the observation of the world is always regarded as being* self *or* other-referenced, never it-referenced *as it happens with practically all the approaches to consciousness so far followed, (ii)This determines, in turn, the use of* relativistic/ normative paradigms *instead of the mechanist paradigm on which the broad mainstream of present-day-science rests, (iii)The preceding items lay emphasis upon the* semantics *and* pragmatics *of conversation together with the role that* interpretation *plays, (iv)This yields, in turn, the (crucial) distinction between* M-individuals *(someone like you and I) and* P-individuals *(where 'P' stands for* perspective), *(v)The role that our existential uniqueness plays in the setting up either of the* meanings *that each one of us (according to our personal experiments experiences) assigns to the 'world' or of* conversation *and* mutual understanding *(vi)The influence of* affections *in the construction both of such meanings and interpretations, as well as the underlying psychoanalytic approaches which provide justification for them, (vii)The use of Pask's* computing means *for* knowledge-representation processes *together with their* neuro-fuzzy *translation, both of which ascribe* a formal basis *to all the preceding items.*

Without going into details on these assertions, it is easy to see that their adoption determines a radically different way of looking not only upon the interactions between/ amongst human beings - of which the relationships baby/mother are an example - but also between human beings and machines. Also, the features of the so-called 'normative paradigms' justify the reason why the conventional neurophysiological theories that treat the origins of (human) consciousness either from a *pure private perspective or according to the standpoints of the physical paradigm*, have been practically excluded from my considerations. This does not imply any kind of judgment about their cogency but the simple *factual* recognition that such theories have left aside so important questions as, for instance, those dealing with our uniqueness, with our interpretations, with our meaning-assignment process, etc.,. Briefly with everything that the classical approach has been unable to justify. On the other hand - and I emphatically stress this aspect - CT's tenets also provide a strong support to the thesis that Postulate I.1 condenses. Particularly, that 'consciousness' is, ultimately, the by-product of a social, conversational process baby/ mother, in which, together with maturation and development, the role played by the latter becomes crucial in the organization, meaning-assignment and coherency ascribed to its 'inner' and 'outer' worlds, (the reader recalls what I said about meanings as being always 'self' or 'other'-referent, never 'it-referent').

For those to whom psychoanalysis is familiar these statements are not surprising. Since, however, CT has also a strong computational component entailing such interactions - a fact which, insofar as I know, does not have happened with practically all the psychoanalytic trends - *then this way of looking upon the problem also opens the door for the setting up of a* construct *embracing a theory of consciousness (natural and non-natural) different from those so far followed by the classical approaches.* Hence, the emergence of the second part of Postulate I.1, stating that

Postulate I.2:
(i)This temporal construct *is deeply influenced by the* earliest relationships *baby/ mother, (ii)It is a basic contention of this paper that such earlier interactions - which work as a*

prototype *for the emergence of a robotic consciousness - are essentially* affective *and their (human) time-transformation obey the steps of a genetically impressed 'program', (iii)This 'program' is centered in Freud's* experience of satisfaction [4], [11d], *and (altogether speaking) it also satisfies the emergence of Spitz's* organizers [14], *(iv)In computing/ robotic terms such a 'program' obeys a* personal *interpretation of Sugeno's complementation* [8] *that has transformed it in Verhulst's equation; every bifurcation is then equivalent to a further step in the development both of the baby and of an* experimental robot *named* ACB2 *(acronym of Artificial Conscious Being 2000), (v)From the whole of these steps, two general stages are especially striking: the first embraces the transformations that occur between birth and the setting up of a sense of 'awareness'; the second, the changes that arise between this 'awareness' and the emergence of a reflective consciousness.*

The importance of these statements will be progressively understood as far as their scrutiny is analyzed. For the moment, let us pay attention to the two aforementioned stages.

2 From Birth to the Ascent of Primary Consciousness

2.1 An Overall Simplification Hypothesis

Two of the most puzzling problems that cyberneticians and/or computing/robotic experts have to solve whenever the consciousness problem is being scrutinized, deal with the fact that the psychoanalytic jargon used to describe, for instance, 'unconscious' and 'conscious', 'I'/'not-I', 'inner'/'outer worlds', 'body'/'mind', etc., commonly entailed in the conjectural processes that babies undergo, are not only hardly translated in formal/ computational terms but also referred to 'realities' thoroughly alien to the children themselves. Furthermore, using them, we are (often unconsciously) induced in reasonings and conclusions which, although at first sight seem to be correct, subsume distinctions and presuppose developments that surely the baby has not yet set up whenever regarded from a more detailed scrutiny. In order to avoid this situation, a simplification hypothesis will be followed hereafter. Let us denote it by Hypothesis I. It states that

Hypothesis 1:
Whenever we have to deal with terms and distinctions that the baby has not yet made, we shall use the symbol of an asterisk (as, for instance, in the expression "the awareness of the distinction mind* body* in a new-born baby...."). As far as a meaning is assigned to such terms distinctions, the symbol will disappear.*

2.2 The Very Beginnings: Freud's Experience of Satisfaction and what it entails

In Axiom 1.1 and Postulate I.2, I pointed out the importance played by conversation (not necessarily verbal) in the mutual understanding of human beings. Doing this, I was also laying a great deal of emphasis upon the earliest social experiences/experiments that a baby undergoes. It is here that our biological features as altricial animals play their most crucial role. As a matter of fact, such characteristics imply that the survival of the

new-born baby depends on the existence of a mother able to provide satisfaction* to its instinctual needs. From such needs (assimilatory, excretory or both) that one related to hunger and thirst - hereafter included in the general term hunger* - will be regarded as the most significant. This is due not only to the complex bodily and mental relationships it entails (and give rise as well) but also because its appeasement (through a nourishing mother) will later frame a host of (unconscious) metaphors embracing social areas and motivations far beyond its original source. Part of the 'mechanisms' underlying the robotic emulation of this hunger* have already been scrutinized in previous works [11b, c]. A condensed version of the results already achieved in this area, is the content of

Axiom 2.1:
From all the instinctual needs which the new born is endowed with, hunger (and its corresponding appeasement/satisfaction*) is hereafter assumed to be the most fundamental. However, due to the extremely complex physiological process that* human *hunger* entails, some simplification hypotheses have to be admitted, particularly whenever its robotic emulation is endeavoured. In this sense it will be hereafter assumed that (i)The decrement of glucose percentage is tantamount to a corresponding alteration in a voltage* v_H *supplied by a main power source, (ii)This voltage is periodically activated by a* biological clock *and its time-variation obeys an* exponential law *(although its time constant may change according to the activity that* ACB2 *may exhibit, (iii)The maximal value* H_{max} *that this curve reaches corresponds to* satiety *and values of the voltage tending to surpass it give rise to avoiding behaviours ("negation" gesture), (iv)There exists a pre-determined reference voltage threshold* H_0 *such that below it and an extreme minimal limit* H_{min}, ACB2 *is said "to be hungry",* (vi)*If* v_H *reaches* H_{min} *and all the remainder power sources of* ACB2 *are already exhausted, then the robot will become irreversibly damaged, (vi)'Hungry' behaviours trigger the (initially* automatic*) activation of* sound circuits *and* motor effectors, *the intensity of which is proportional to the difference between the instantaneous value of* v_H *bellow* H_0 *and this reference voltage, (vii)These alert sounds are supposed to call the attention of a mother who, sooner or later, will "nourish"* ACB2 *with a voltage* v_S *until that satiety is (in principle) reached again, (viii)Due to the different durations (lactation lasts usually 12 times less than the period elapsed between satiety and the re-activation of hunger), the curve translating the nourishment itself is supposed to be a straight line, (ix)According to Freud* [4a], *the whole cycle that begins with satiety, reaches a peak of hunger and ends with the appearance of a mother, her nourishment and (eventually) with a new satiety state, will be named* **experience of satisfaction**.

Two major aspects of this axiom deserve further attention: (a)the robotic emulation of hunger* and (b)the experience of satisfaction itself. Regarding (a), this axiom seems, at first sight, to bring no robotic novelties to what has been built so far. As a matter of fact, machines that behave in the preceding way already exist for long. For instance, Grey Walter's "Turtles" or more recently, the "Cybug" of Craig Maynard, exhibit 'hungry' behaviours hardly distinguishable from those witnessed in natural insects. Furthermore, the 'circuitry' on which such behaviours are based, is - to some extent - an actual emulation of part of our own hunger/nutrition system. Indeed, both have, a 'hunger' and

'satiety' centers working alternatively, both send impulses with a frequency proportional to the difference between the instantaneous state of hunger* (or nutrition) and a reference frequency (corresponding to H_0), both react to hunger* or satiety either by 'crying' or by the use of avoiding behaviours, etc., [7], [14]. In spite of these similarities, ACB2 differs clearly from these machines in several - and crucial - aspects. *Part* of these aspects deals just with the 'experience of satisfaction' itself.

The first reference to the such an 'experience' was made in the *Project* (1885) where Freud condenses in some few lines the psychic evolution that a child undergoes from a pure reflex, not directed or purposeless stage, to a situation in which crying is, for instance, *intentionally* used for some kind of communication (in the case, calling for the help that the mother provides). Five years later, Freud returns to the same subject in his *Interpretation of Dreams* where he adds some important remarks, particularly on the further development of the baby's psychic dynamics. The emulation of this dynamics was the objective of formal scrutiny in several of my papers [11, 12]. Thus, in order to avoid a repetition of the matters there contained, both of the preceding aspects - the robotic and the human - the results so far reached are condensed in

Postulate II.1:
Although fitted with an electronic and computer hardware similar to that used by machines emulating biological behaviours - hunger in this particular case -* ACB2 *presents crucial differences regarding such machines. The most striking deals with the* affective consequences *that the successive repetitions of the experience of satisfaction engender (about 1000 interactions in the first 3 months). As a matter of fact, the initial reactions (apparently) confined to reflex states either of* unpleasure* *or of* quietness *(sleeping)* satisfaction*, *soon undergo deep transformations. These changes consist, essentially, in an enormous increment of the number of affections followed by the emergence of novel meanings to them attached. For the sake of simplicity, I shall divide the formal results already reached in two classes, one dealing with personal achievements, the other associated with interactions themselves (although in reality, both of them always converge towards individual developments). So,*
A)In personal terms
(i)Chronological speaking, the first of such changes deals with the assumed existence of a need's tension *the mathematical expression of which can be found in* [11c], *(ii)the curve that translates the time-variation of this tension is null between satiety and* H_0 *and next increases positively until the maximal peak of hunger is reached, (iii)from this point onwards (corresponding to the beginnings of nourishment) the curve decreases until satiety is (in principle) reached again and the need's tension becomes null, (iv)the transformation from the positive to the negative derivatives of this curve determines the emergence of* relief [11c,d], *(v)relief is always referred to the transformation from a state of* unpleasure* *to a state of* less unpleasure*, quietness *or* pleasure*/satisfaction*, *(vi)according to Freud's Nirvana principle, there exists in human beings an innate tendency of lowering the value of the need's tension, particularly that one related to unpleasurable states. Part of this decrement is made through an* innate connection *between it and* **motor discharge** *through screaming, kicking, etc. In the earliest times of life, this discharge is* involuntary *and* purposeless.

B)In interactive terms,
Since the need's appeasement through the mother is formally unpredictable, then it gives rise to the emergence of interactive indices *like (i')*Affective Balance, Expectation, Frustration, Adequacy, etc., [11b,c,d] *(ii')the repetition of the experiences frames the emergence of the function of* attention* as well as *of a primal concept of* wish [4a,b], *(iii')This, yields in turn the setting up of two affectively 'weighted'* Mother's Images (MIs), either 'Good' (GMI) or 'Bad' (BMI), *(iv')Likewise, the distinction* recollected actual *(deeply associated with wishing) also begins to be established. Indeed, while* GMI *is related both to the actual stimulation of* adequate *exteroceptors* and *to an* effective *appeasement of the need,* BMI *stresses the role of inadequacy, frustration, etc., (v')This fact underlies the substitution of the initial pleasure pain principle by the reality-principle* [3b] *(vi')This rises, in turn, the substitution of the initial, purposeless motor discharge of the need's tension by* purposive movements *tending either to reinforce (pleasurable)* attachment ties *with* GMI, *or (if allowed) used as a discharging path for its* aggressive impulses *towards* BMI.

There are some aspects of this Postulate requiring further deepening, of which the most striking deal with two major questions: (a)the meanings assigned to pleasure*, unpleasure*, quietness, satisfaction*, etc., and (b)the emulation of these and other affections in ACB2. Both are the content of the following axioms:

Axiom 2.2:
A)Pleasure
The answer to (a) obeys a line of reasoning similar to that presented by Szasz [15]. *This author assigns 4 meanings to* pleasure*which, briefly, correspond to (i)* "relief" *(of pain), (ii)* "contact and communication" *as soon as affective ties are set up with the human object who appeases the need, (iii) a "substance" which,* **added to** *something felt as lacking, relieves it, (iv)a* symbolic extension *of the preceding notions in which its physiological part is strictly put aside. This is tantamount to identifying pleasure with the* **acquisition of** *more money, more power, more knowledge, etc., which added to the body or mind, fulfills the sensation of 'lacking of...'.*
B)Pain/Unpleasure
A slight different reasoning is followed regarding the concepts of pain *and* unpleasure. *Indeed,* "Unpleasure" *is the English translation of the German* "Unlust" *used by Freud in his writings and embraces several meanings as, for instance, pain, distress, sadness, etc. This is another aspect also exhaustively mooted by Szasz which has given rise to the consideration of* three *meanings, possible to be assigned to* pain**:** *(i')an* excessive stimulation *felt in some part of our body and working simultaneously as some sort of 'warning message' expressing the existence of some danger or malfunctioning, (ii)together with (i)* a means of asking for relief, directed to someone thought to provide the desired help, *(iii)a* symbolic extension of (ii) *in which the* physiological part of pain is thoroughly left aside and 'pain' *works as a means of expressing a* grief, complain, (disguised) anger *against one who did not provide the expected help or, even, against someone who has nothing to do with the original 'aggressor'. Often, this significance may even lead to actual bodily pains, although now this is not due to some actual medical cause. As a matter of fact, in the absence of an 'outer' object who, really or*

imaginary, is supposed to work as a possible discharging object, it is the patient's body which ultimately will play such a role.

C)Satisfaction

There are two major perspectives from which the term may be regarded. The Webster's New World Dictionary (ed. 1999) presents two synonyms, dealing respectively, with

- "gratification, fulfillment, achievement, compensation, etc.",
- "comfort, pleasure, well-being, content, joyness, etc."

Briefly: while in the former, 'satisfaction' deals with fulfillment, *in the latter it becomes related to a sensation feeling of* pleasure.

D)Adopted Meanings

(i")Regarding pleasure, a modified version of (iii) above will be the one adopted hereafter. *Detailed psychoanalytic reasons justifying this choice can be found in* [15] *but, briefly, they stress two factual truths: on the one hand, that whenever pleasure is identified with something that must be* **added to** *(causing relief, appeasing hunger, representing a wished acquisition, etc.), then it can also be regarded as an extension of the notions mooted in (i) and (iv). On the other hand, as far as expectations regarding such acquisitions are directed to human beings (so, able of fulfilling what is 'lacking'), it enfolds too the object relations referred either to (ii) or to (iv), (ii")Under this proviso, it is obvious that* satisfaction *embraces both of the preceding significances, (iii")Finally, as regard pain and unpleasure, I shall begin with the adoption of the* primal concept of pain *(item i'); as far as affective ties are being set up with other people, then the remainder significances will progressively acquire more importance (although holding the primal sense). Hence, the use of the general word* "unpleasure" *whenever these earlier stages of life are referred to.*

The adoption of these hypotheses will render easier the answer to question (b) above concerning the robotic emulation of the foregoing affections. *Part* of this answer is the content of

Axiom 2.3:

(i)Since young babies (younger than 3-4 months) have not yet set up a clear distinction between what is 'in' or 'out', what is the 'Other' and the 'I', etc., the fundamental significance assigned to unpleasure *is that of* primary pain, *(ii)From the hunger perspective - and although not yet conscious of the bodily places where its cognate' pains' arise - it will be assumed that* ACB2 *owns* two particular sensorial regions: *an 'oral cavity' confined to a* 'mouth' *fitted with pressure sensors and a* 'stomach' *(which may be a simple capacitor), being hunger itself tantamount to the stimulation of* both of *these systems (mouth AND stomach), (iii)Such a stimulation corresponds, in turn, to the emission of impulses with a frequency proportional to the difference between their instantaneous value and corresponding* in-built thresholds. *Whenever such differences are negative (yielding* negative pressures*) they lead to a 'sensation' of* 'emptiness'/lacking of, *to an increasing unpleasure, to the rising up of a need's tension and to active* motor behaviours *(screaming, etc.) tending to lowering it, (iv)These later yield, in turn, the appearance of a mother, followed by an energetic supplying and its consequent relief*

and satisfaction - where satisfaction corresponds here not only to pleasure but also to a situation of high positive pressures in the 'stomach', say, of 'fullness', (v)Terms referred to each one of the foregoing affections are expressed by fuzzy linguistic variables. *These variables are assumed to lie in an* Affective Unit, *those proceeding from sensorial stimulation in a* Sensorial Unit *and those ruling motor behaviours in a* Motor Unit, *(vii)*associations *between amongst sensorial activations, affections, motor behaviours, etc., are supposed to be made in an* Associative Unit [6], [13], *(ix)signals proceeding from these Units are sent to and or received from a* Central Unit *and a* Memory Unit (similar to Braitenberg's Mnemotrix [1], [6]).

An overall, albeit rather simplified scheme of these inter-relationships, is depicted in Fig. 1 (see Postulate II.3). However, before to advance to an explanation of these connections as well as to a further stage of ACB2's emulations, some commentaries about the preceding assertions are required. Firstly, it is worth noting that the interpretation assigned to pleasure, unpleasure, etc., in Axiom 2.2, not only satisfies their psychological definition *but also allows its robotic translation.* This is partially due to meanings adopted for such affections and partially to (ii) and (iii) of Axiom 2.3. Secondly, the supposed existence of an almost permanent (albeit mild) *negative pressure* in the robot's mouth and its corresponding appeasement with the sucking of its fingers/thumb, will play a crucial role in the setting up of some of the first *purposive behaviours* that the baby/ACB2 will exhibit. Thirdly, the assumed existence of a 'stomach' - ultimately an *inner organ* - will also work as another important element, necessary (later on) to set up the earlier distinctions between what is 'in' and 'out' to the body's surface. However, before to scrutinize these later results, some other developmental steps have to be accomplished, particularly with those that entail the setting up of the mother's image(s) referred to Axiom 2.2, now not only from an *affective* and *sensorial* perspective but also from the point of view of its robotic emulation. This will lead us to look at two of the interactive indices pointed out in Postulate II.2 from the more detailed viewpoint, expressed by:

Axiom 2.5:

It was said beforehand that, at the end of the first three months, there are about 1000 possible 'experiences of satisfaction'. However, (i)It is not sure that all of such experiences are able of reaching a state of satiety. Therefore, it makes sense that a new variable - denoted by affective balance (AfB) [11c,d] - *is introduced in the approach, so that it measures the overall degree of pleasure unpleasure achieved either in every experience (instantaneous value) or after a sequence of experiences (average mean value), (ii)By the same token, it is also uncertain that the mother* is, in the beginnings, able of understanding the alert sounds emitted by the baby when this later is hungry. From here comes the introduction of another variable - named* adequacy (Ad) [11c,d] - *which measures the degree of similarity existing between the sensors that effectively appease the baby's hunger and those the mother* stimulates (for instance, caressing a new born is not enough to appease its hunger), (iii)After a certain number of repetitions the baby or* ACB2 **learn** *to bind the vision of the mother*'s face, the abolishing of the sensation of 'emptiness' in its mouth and stomach, the sounds 'she' emits during nourishment or 'her' caresses, etc., so that these elements constitute a* **coherent cluster**

(a sensorial pattern) in Pask's sense, (iv)The construction of this pattern is followed simultaneously by the setting up of a new relationship between it and the relief *provided, together with a specific affective balance, (v)The association between them (sensorial pattern and 'weighting' affection) in the* Associative Unit *determines a new pattern which, ultimately, constitutes the* mother's representation *hereafter named* mother's image (MI). *If the mean value of AfB (and/or other affections to be mooted later) is* positive, *then this image determines the* 'Good' mother's image (GMI); *otherwise, it will define the* 'Bad' mother's image (BMI), *(vi)It is a fundamental contention of this paper that the formation of* MI *obeys an outstar/instar neural model similar to that based on the* neural network for Pavlov's conditioned reflexes [2], [3] *in which the 'trigger' signal comes from the hunger re-activation and the coefficients of the characteristic equations are supposed to depend either on AfB and/or other affections. Particularly important is the assumption concerning the 'forgetting coefficient' supposed to be not only time-variable but also different from zero - say 'stored' in the Long Term Memory Unit (LTMU) - only and only when the* maximal *pleasurable (or unpleasurable) value is reached in some experience* [11d].

This axiom entails crucial assertions deserving further attention. On the one hand, it is worth noting that, in the setting up of MI ('good' or 'bad'), the difference between mean and instantaneous values of AfB is especially striking. As a matter of fact, it is almost certain that 'bad' experiences arise as far as repetitions are occurring. However, if the mean value of AfB so far reached is greatly positive, then the baby's capability for *tolerating* the (occasional) unpleasure must suffice to delay the beginnings of motor discharging (crying, kicking, etc.). Briefly: the mean affective balance works as some kind of 'inertial wheel' that not only may be regarded as one of the 'measuring indices' of the relationship baby/mother but also tends to color *anticipately* its expectation regarding the possible (dis)satisfaction that the new interaction will (or will not) provide. On the other hand, this expectation embraces other psychological 'mechanisms' that the following axiom summarizes:

Axiom 2.6:
Insofar as MI *('good' or "bad") is being reinforced and the baby's cognitive capabilities (like memory and attention*) are being improved, the re-activation of hunger is followed by a host of other processes, non-existing in its earliest times of life: (i)as soon as such a trigger signal re-appears, it is a direct consequence of the neural model adopted that the former* MI *also re-appears, albeit* no actual face has still arisen. *Such an image is ultimately a* recollection *of* MI (RMI), *(ii)and since the baby has not yet established the distinction between recollections and real images, then* RMI *may even acquire a hallucinatory character* [4a, b] *as though the 'mother' were actually there. Since, however, this does not correspond to reality, then its hunger, need's tension and unpleasure continue to increase, (iii)This yields a host of other processes which, briefly, consist* (a)*in the emergence of* expectation (Exp) *and (later) of* frustration (Frst) *if the* expected *appeasement is not achieved,* (b)*in the case of successive frustrations, in the eventual collapse of* GMI *which - if this happens so - is followed by its substitution by* BMI *whether already existing (or in the formation of the latter whether non-existing so far) and* (c)*in the emergence of* aggressive impulses (Ag). *At the same time that these*

'mechanisms' are triggered, the urgency for the appeasement of its unpleasure continues to be active, (iv)This is translated by (a)*the* sensitization *of the receptors related to preceding pleasurable experiences (exteroceptors, interoceptors and proprioceptors) or, which is equivalent, to the emergence of an* attention* *directed to such sensors,* (b)*The re-activation of* GMI *is followed by an emergence of a* wish [4a,b] *for "her" re-appearance yielding, in turn,* (c)*The reinforcement of this image together with the rising of an (embryonic)* intentionality *and* purposiveness *vent by (still incipient)* motor actions directed *either to deepening the* attachment (attractive) ties *with* GMI *or towards* BMI (if allowed) *as means of discharging* aggressive impulses.

There are several aspects of this axiom requiring further deepening essentially because they determine the emergence of a new stage in the baby's development to which primary consciousness will be explicitly related.

2.3 The Ascent of Primary Consciousness

I said in the preceding axiom that as far as GMI tends to become progressively reinforced, also the baby's attention tends to be directed to its sensorial receptors, namely exteroceptors (those immediately associated with lactation itself), although not excluding interoceptors and proprioceptors. This behavioural change - establishing a deep contrast with the earliest behaviours, practically confined to the cycle hunger/sleep - can, in healthy babies, be easily witnessed either in the rhythm of lactation or after it has already finished, since the number of times during which the baby tends to 'look around' increases significantly. This alteration raises new questions that, in turn, will entail new meanings to some of the variables already focused. These, in turn, will frame the emergence of that sense of 'awareness' referred to beforehand as well as the distinction between mind-actions and bodily-actions, etc. *Part* of these achievements is condensed in

Axiom 2.7:
(i)Insofar as GMI *tends to become time-invariant, also the primal concepts of unpleasure and pleasure (formerly associated either with the 'emptiness' lack of pressure in the 'mouth' and 'stomach' of* ACB2 *or with its 'fulfillment fullness') tend to acquire new meanings. So, 'pleasure' no longer corresponds to the simple appeasement of the need* **only** *but also acquires novel features proceeding now from the* communication *(visual, auditory and tactile)* with GMI, *(ii)Likewise, the primal concept of* wish - *automatically triggered whenever hunger is reactivated - is no longer a* passive *wish for a concrete 'object' but, instead, it becomes transformed in an* **active (motor) searching** *for* GMI's **approval** *(already more or less independently from the need's arousal) - as though such gratifications were some sort of pleasurable 'substance'* [15] *that the baby* **'adds to'** *it.*

This axiom stresses three puzzling robotic questions that deserve particular attention: (a)the foregoing *motor searching,* (b)its cognate *purposiveness,* both of which presuppose (c)the *awareness of the actions performed.* Part of these questions will be treated in

Postulate II.2:

(i)It was asserted in Axiom 2.6 that as far as GMI *tends to become time invariant, the baby's expectations regarding the need's appeasement also tend to decrease and its attention begins to be directed not only to its outer environment (via exteroceptors) but also to a specific 'object' which, later on, will be called its own 'body' (interoceptors and proprioceptors included), (ii)Indeed, if the mean value of the Affective Balance has been positive, then the pleasure which the baby has memorized during previous interactions with the mother tends to be reactivated with its own bodily sensations eventhough she is absent, (iii)This motivates the baby towards further repetitions of such experiments (tangible, proprioceptive and kinesthetic)* [4], [11d], *(iv)The consequences of this trend are crucially important for they engender* (1)*Two types of distinctions between* (a)"touching itself" *and* "touching"/"grasping" *(other objects),* (b)*between (a) and* "to be touched", *all of which are* neural patterns [2], (2)*All the experiments entailing (a) yield the progressively erection of a* bodily invariant *altogether associated with pleasurable bodily sensations.* **Formally,** *this invariant is a group of clusters obeying both a* **closure requirement** [10] and *my own definition of* **wholeness** [11b]. *So, satisfying Pask's definition of* P-individual *here looked upon as an* invariant part of the M-individual whom the baby has always been traditionally regarded. *This invariant (of which its owner is altogether unaware) will be named* **Proto-Ego** (PE), (3)*Together with (b) above, the erection of* PE *rises, in turn, the emergence of the* **"Other"** *(as an* **invariant outer entity** physically separated *from the baby),* (4)*This* construct *frames, in turn, the distinction between* **motor processes** *and* **thinking processes** [11d] - *the latter being regarded as something specifically located in a place that (using adult terms) may be denoted as 'inside the head'. Although entailing a rather restrict concept of thinking, it already enfolds autistic thinking, affections, remembering, associative processes, purposiveness and perhaps phantasy, all of them 'thought actions' clearly different from bodily (motor) actions,* (5)*All these changes are followed by the overwhelming ascent of* **primary consciousness**, *a fact which, due to its complexity, requires a distinct*

Postulate II.3:

The emergence of awareness in non-natural systems is supposed to rest on a physiological silicon-based 'architecture' part of which emulates that one actually *existing in human beings and part is simply* conjectural. *This 'architecture' - that deals both with sensors and effectors - will be scrutinized in the following items. Thus, (i)In* sensorial terms *it will assumed hereafter that there is an* actually *innate connection between peripheral sensors and the brain sensorial motor 'homunculi'. Because well-known, the functioning of these 'homunculi' will be not scrutinized in this paper* [6], *(ii)However, for the sake of clarity, every sensor effector of such 'maps' will be named* sensor-image (S'_n) *or* motor-image (M'_n). *The connections between amongst these sensors and effectors are depicted in Fig. 1, (iii)Regarding the* basic conjectures, *I propose two, one of which sets up the distinction between the sensor's* activation *itself and the* data *it sends, and the other assumes that every* S'_n *is activated not only by peripheral sensors* S_n *as traditionally accepted but also by the Attention, Memory and Associative Units. Also, every* S'_n/M'_n *is supposed to send a reference voltage* X_0/M_0 *(linguistically translated by* 'Null Value')*both to a* **comparator** *and to a* **differentiator**.

The comparator determines the difference between the received outlying signals and X_0/M_0*; the differentiator shows if the preceding difference is increasing, decreasing or null, (iv)The information coming from these two devices is sent to* part *of the Central Unit where it becomes associated with the foregoing achievements, particularly those stressed in* Postulate II.2. *This architecture not only embraces all the questions previously stressed but also provides a* **sense of awareness** *to a 'smaller' part of the Central Unit which, taking* (3) *of* Postulate II.2 *into account, has said to play the role of* PE. *This 'awareness' can be easier understood looking at Fig. 2 where the simplified scheme of a robotic* effector *(the emulator of a human arm) is depicted.*

The system is supposed to comprise a motor M, a voltage generator M' both connected through a (simplified) gear box, a linear potentiometer P whose axis rotates whenever M and/or M' are activated and two 'change detectors', each one of them embracing a *comparator* and a *time-differentiator*. It is also assumed that the whole effector may be activated by means of two totally distinct sources: one, named 'Reflx.' is mastered by an automatic/reflex 'mechanism' which is triggered whenever the intensity of the data received by some sensor reaches the neighborhoods of its maximal limits. Briefly: 'Reflx.' is a typical *protective mechanism* which activates the effector to which the 'threatened' sensor is coupled, in a direction that tends to decrease its 'dangerous' stimulation. In the picture, the 'arm' is a simple bar linked to one of the gears that not only activates M/M' but also P. In this sense, when P turns of α degrees between two positions α_0 and α_1, also two corresponding voltages $V_\alpha(t)=\alpha_0$ and $V_\alpha(t+1)=\alpha_1$ are either compared or (if this is the case) *memorized* in the Memory Unit. This type of working in which the initial and/or final positions of some movement are stored in that Unit (at least in the beginnings, before that habituation starts) is typical of a voluntary movement ('Vol'.).

The system stands for a general representation of a human effector. As a matter of fact, the relay 'm' when activated represents the baby's *will* (often unconscious) of moving the muscles (the motor M) of the arm/bar; P stands for one of the proprioceptors/kinesthetic sensors placed at two adjacent articulations in the (human) joints and the remainder elements are part of the circuitry that provides answer to the question of the motion's awareness. Three of these elements are here crucially important: the generator M' that henceforth, for the sake of terminological uniformity, we shall name *motor-image*, the proprioceptive/kinesthetic potentiometer P and the change detectors (differentiators). *It is worth noting that, apart minor differences, this structure entailing a motor element, its 'image' and a comparator/differentiator is practically the same as that used in Fig. 1 about sensors, sensor-images, comparators, etc.*

Once this is assumed, a pertinent question can now be posed: "How does it may provide solution to the problem of non-natural awareness?". To answer the question, we have first to deepen what 'being aware of' entails . In this sense, let us imagine that one of us has made the assertion 'I am aware of seeing the sun' (an example which is extracted from Bertrand Russell's "The Problems of Philosophy").

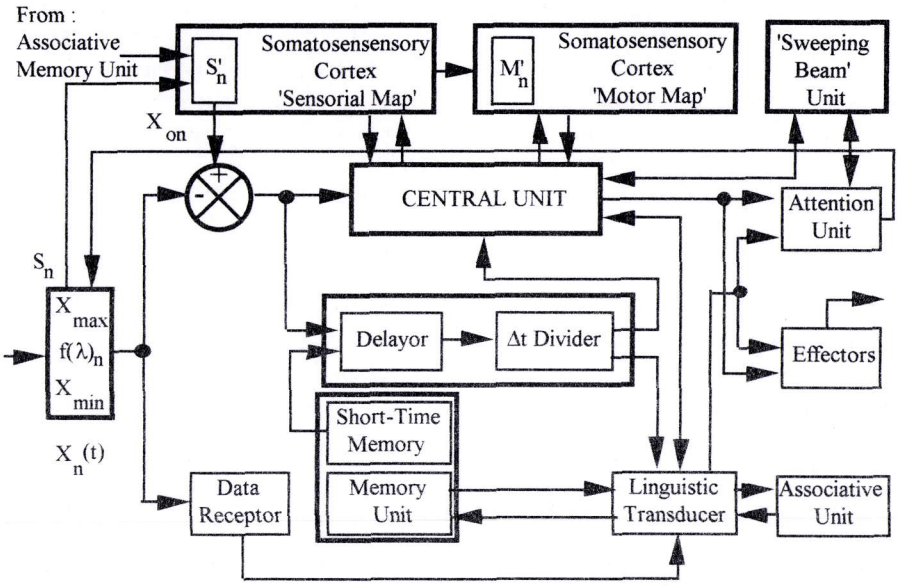

Fig. 1: The overall 'architecture' of ACB2

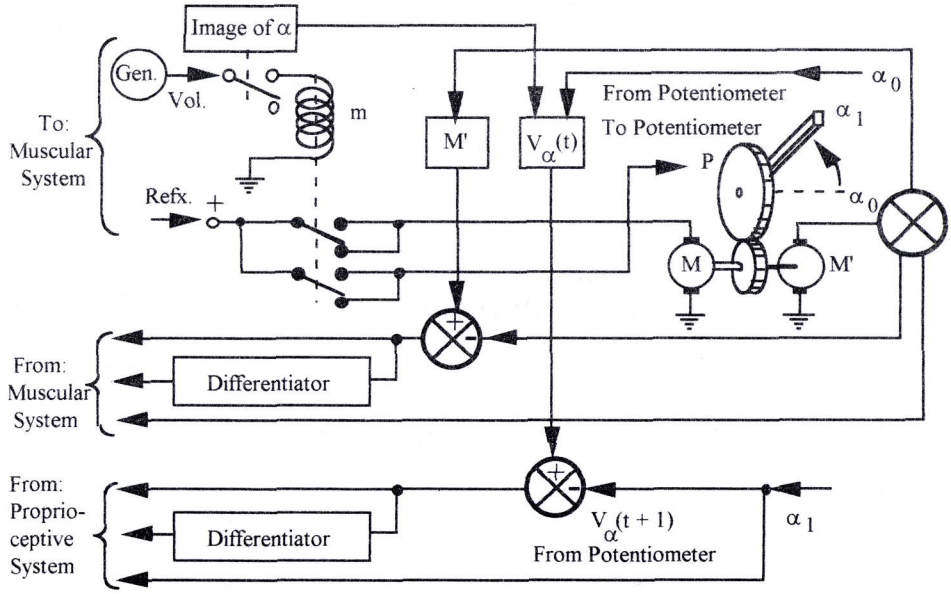

Fig. 2: The awareness of ACB2's arm motion

In this statement we have three elements: an 'outer object', the *sun* (the meaning of which, because irrelevant in the present analysis, I shall not discuss), a *subject* (the

reader or me), and the sensation/feeling (the 'awareness' itself) that each one of us has of *seeing something* - in this case the sun itself. Let us call this 'seen-sun' the *sun-image*. Then, *one* of the conditions for 'being aware of' is that each one of us - I, for instance - must have, inside my head, an image of the sun, such a sun-image referred to above. But this is not all. For, the sun-image may be a momentary illusion, not the image of something that, *at that exact moment*, is effectively happening *outside* my head. Under this caveat, a second condition for 'being aware of', is that *no difference must exist between the actual stimulation proceeding from my (activated) eyes - to which I have ascribed the name 'sun' - and the sun-image inside my head.* In other words: some sort of *identity* must exist between the sun/sense-data and the sun-image/seen-sun. The detection of the existence/ non-existence of this identity requires a *comparator* the output of which must be null whenever we are treating of *static* observations. In the case of time-variable situations, the differentiator is needed for detecting if the possible error is increasing, decreasing, etc. - a type of information particularly important whenever the effector circuit is working automatically (reflex functioning). Obviously, situations in which a constant error exists, yield null derivatives - psychologically translated by habituation. Finally, the third condition for 'being aware of' demands the existence of some Central Unit that must be able not only of receiving such data but also of associating the sun/sense-data, the sun-image and the results of the comparator/differentiator. In this sense, let us assume that we have two patterns, say A and B such that

	Sun	Sun-Image	Comparator
A =	1	1	0

	Sun	Sun-Image	Comparator
B =	0	1	1

Then, while A corresponds to a situation in which the Central Unit compares the stimulation from its 'eyes' (actual sense-data) with the sun-image and faced with the result of the comparator concludes that they both match, B yields a simple recollection of the sun-image. Needless to say, these associations (that may entail matricial patterns/data transmitted by video cameras, etc.) need the using of neural networks.

When the preceding considerations are applied to the baby's development, all the practical requirements for the awareness of (parts of) its own body are satisfied. Suppose, indeed, that an energetic supplement is supplied to the muscles that control the movement of its fingers. This means that, in Fig. 2, a voltage (the 'energetic supplement') is applied to the relay 'm' and consequently that the 'finger-bar' coupled to the axis of M/gear box (the 'muscle') turns of α degrees ($\alpha=\alpha_1-\alpha_0$). The awareness of the muscular tension underlying the finger-bar's displacement is given by the voltage generated by the motor-image M' together with the voltage coming from the (proprioceptive sensor)/potentiometer P. Finally, the certainty - from the baby's side/part of the Central Unit - that an association exists between such a muscular tension and the observed movement, results from a (supposed) null output-errors proceeding both from the comparator/differentiator.

Before to finish, three comments must be added to these considerations: (i)the motor-images M' and the comparator (just like the sensor-images/comparators in the sensorial realm) work as the true upholders of M's functioning. *Indeed, the voltage generated by M' (which is nothing but an exact 'reflex' or 'image' of the voltage applied to M) and the consequent null error, are the elements which assign to the Central Unit the certainty (technically understood) that M has been actually activated*, (ii)the foregoing idea entailing some element, its image and the comparator not only nullifies the approach to consciousness based on the so-called 'homunculus' but also reduces the whole question of 'primary consciousness' to a matter of operations upon (more or less) complex computational patterns, that is to say, computational vectors or matrices, (iii)Finally, the third aspect deals with the role that learning plays in all these processes. As a matter of fact, the scheme of Fig. 2 - albeit general - is an extreme simplified representation of a robotic system constituted by an arm/forearm /hand /fingers. We believe that in babies (perhaps by economical reasons dealing with the amount of efforts to be spent), the first experiments in the realm of primary awareness begin with the observation of the fingers's movements. This requires that small amounts of energy (the activation of the relay 'm' of Fig. 2) are sent to their controlling muscles M. In terms of ACB2, this entails several operations amongst which the role played by the 'Central Unit' is especially striking. Indeed, this Unit cannot have a simple passive role but, on the contrary, it must not only activate this or that sensor and/or effector *but also learn to what sensor(s) effector(s) should it send the needed energetic supply*. In the very beginnings, these supplies must be more or less randomly sent. We, adults, translate this randomness by saying that "there is a lack of muscular coordination". But soon, (3-5 months old), many of such movements have already reached a level of coordination that renders possible the grasping of many objects, particularly those which - using a neural net terminology - correspond to the so-called 'supervised learning'. The 'supervisor' here is, obviously, something that awakens his/her attention (either a brilliant object, a toy, or moreover, everything related to the mother, her face, etc.), briefly, *something that works as a goal to be reached*. This implies a *conscious intentionality* ; consequently, the emergence of a **reflective consciousness**.

3 The Second Stage: The Ascent of Reflective Consciousness

If we compare the initial, reflex reactions that a baby or ACB2 exhibit with the achievements that they have already gotten according to Postulate II there is no doubt that throughout the first months of life a great deal of changes has occurred. Indeed, the primitive meaningless world of the earliest days, that sometimes was unpleasurable and sometimes became pleasurable, has progressively been substituted by a more complex 'world', altogether mastered by an increased number of sensations and feelings already associated either with a 'good' or 'bad' "(M)other" or with the awareness of several distinctions as, for instance, recollected/actual, motor action/thinking, of its affections (although unknowing where they do come from), of its own body - part of which it can altogether control *in order to* express its attachment/aggressive impulses towards the mother or to reach/grasp this or that 'interesting' object other than her (a colored toy, the rubber nipple, etc.). However, at this stage of the baby's development, two dramatic behavioural changes occur: *its first vocalizations begin* (an aspect which I shall not

discuss here) and - crucially important - the increasing coordination between its vestibular system and the muscular maturation of its legs allows that the *primitive universe of 'reachable things', so far thoroughly dependent on the mother's transport, becomes henceforth attainable by itself.* In the aim of this paper two immediate consequences can be stressed from this last modification: (a)the number of (real or imagined) 'pleasurable' objects to be reached, increases enormously and, (b)the number of times in which the baby's objectives are frustrated becomes also proportionally increasing. Item (a) yields a further separation in the aforementioned "Proto-Ego" that henceforth (taking into account the meaning of 'pleasure' focused in (ii) of Axiom 2.5 and the foregoing sense of awareness) is transformed in a **conscious** "Proto-Mine" (PM), say, some sort of (pleasurable) 'image' of that (unconscious) Proto-Ego, to which **a sense of ownership is attached**. Item (b), on the other hand, contributes not only for reinforcing the wish for accomplishing such frustrated goals but also for the emergence of strong aggressive impulses directed towards the 'frustrator ' (the mother, the father, relatives, etc., who have to supervise and protect the restless curiosity of those ages). In order to deal with these impulses (tending to destroy inter-relationships between the elements of PM) several defense mechanisms are used. Essentially, they consist in *projecting* such impulses upon the mother's image (robotically, a transformation of GMI in BMI) followed by an *identification* with her (that in robotic terms corresponds to change part of PM, so that it becomes *formally similar* to BMI). This new structure, say PM*, will acquire later on, the characteristics of GMI (forming a unified MI) which, henceforth, will work as an **observer of PM itself,** i.e., as **self-observer**. The awareness of this self-observation obeys the same 'architecture' described in Postulate II, being just the **reflective consciousness** we have been erecting. Once it has been established two (amongst others) crucial consequences come to light: (i)the initial **non-existing** distinction between *thinking* and *bodily* processes tends to be deepened, and (ii)as far as the sense of **"Mine"** is reinforced, the toddler - due to its interactions with its parents, relatives, other children , etc. - learns to associate it with its own name ("*John* makes this!", "*Peter* wants that!") so that, by 4/5 years old, the "I" emerges explicitly. This is the final stage of the transformation from the earliest altricial **socialization** I have ventured to the (relative) **individualization** that almost all the followers of the consciousness theories have defended so far.

Selected References

(Numbers are solely referred to the bibliography quoted in the text).

1-Braitenberg, V.: *Vehicles*. MIT Press, 4th ed. 1998.
2-Carpenter, A., Grossberg, S.: *Self Organizing Neural Networks*. MIT Press, 1982.
3-Caudill, M., Butler, C.: *Understanding Neural Networks*. MIT Press, 1996.
4-Freud, S. :
a)*Project for a Scientific Psychology* (1895). Standard Ed. 1933.
b)Formulations of the Two Principles of Mental Functioning (1911). In *On Metapsychology: The Theory of Psychoanalysis.*, The Pelican Freud Library, 1964.
c)Negation (1925). In *On Metapsychology: The Theory of Psychoanalysis.*, idem.
5-Gazzaniga, M.: *The Cognitive Neurosciences*: Bradford Books, MIT Press, 1997.

6-Hassoun, M. : Associative Neural Memories. Oxford University Press, 1993.
7-Kandel, E., Schwartz, J., Jessel, T.: Essentials of Neural Science and Behavior: Prentice Hall, 1995.
8-Klir, G., Folger, T.: *Fuzzy Sets, Uncertainty, and Information*: Prentice-Hall, 1988.
9-Klopf, H., Leemon, B.: Extensions of the Associative Control Process (ACP) Network: in "From Animals to Animats", MIT Press, 1993.
10-Maturana, H., Varela, F.: *Autopoiesis and Cognition*: Reidel Publishing, 1980.
11-Medina-Martins, P. R. :
a)On the Simulation of Depth Psychological Processes. Xth European Meeting on Cybernetics and Systems Research, Wien. (1990)
b)An Abridge of a Genetic Psychology of Non Natural Systems. CCAI, Vol. 12 (1995)
c)The Dynamics of Entailment-Meshes. AIP, 1998.
d)*Design for a Being*. In preparation.
12-Medina Martins P. R. and Rocha L.: The In and the Out. XIth European Meeting on Cybernetics and Systems Research, Wien, (1992).
13-Pask G.: *Conversation, Cognition and Learning*. Elsevier.(1976)
14-Spitz R.:
a)*No and Yes: On the Genesis of Human Communication*. IUP 1957.
b)A Genetic Field Theory of Ego Formation - Its Implication for Pathology. IUP, 1959.
c)*The First Year of Life*. IUP, 1963.
15-Szasz T.: *Pain and Pleasure: a Study of Bodily Feelings*. Basic Books Inc.(1975).