Using modular neural networks to model self-consciousness and self-representation for artificial entities

Milton Martínez Luaces , Celina Gayoso, Juan Pazos Sierra and Alfonso Rodríguez-Patón.

Abstract- Self-consciousness implies not only self or group recognition, but also real knowledge of one's own identity. Self-consciousness is only possible if an individual is intelligent enough to formulate an abstract self-representation. Moreover, it necessarily entails the capability of referencing and using this self-representation in connection with other cognitive features, such as inference, and the anticipation of the consequences of both one's own and other individuals' acts.

In this paper, a cognitive architecture for self-consciousness is proposed. This cognitive architecture includes several modules: abstraction, self-representation, other individuals' representation, decision and action modules. It includes a learning process of self-representation by direct (self-experience based) and observational learning (based on the observation of other individuals). For model implementation a new approach is taken using Modular Artificial Neural Networks (MANN). For model testing, a virtual environment has been implemented. This virtual environment can be described as a holonic system or holarchy, meaning that it is composed of autonomous entities that behave both as a whole and as part of a greater whole. The system is composed of a certain number of holons interacting. These holons are equipped with cognitive features, such as sensory perception, and a simplified model of personality and self-representation. We explain holons' cognitive architecture that enables dynamic self-representation. We analyse the effect of holon interaction, focusing on the evolution of the holon's abstract self-representation. Finally, the results are explained and analysed and conclusions drawn.

Keywords- holons, modular neural networks, self-conciousness, self-representation.

I. INTRODUCTION

Understanding consciousness has been defined as "the ultimate intellectual challenge of this new millennium" [10]. Since ancient cultures, consciousness has been discussed by philosophers, jurists and religious leaders. The word "consciousness" comes from Latin *conscientia*, a word used in juridical Roman documents by writers like Cicero [33]. Literally, *conscientia* means "knowledge (science) with", that is, shared knowledge. Historically, it was first used to refer to moral conscience, as in Christian Codices [24]. From the very beginning *conscientia* was associated with responsibility (moral or legal) . Now, conciousness constitutes the basis of modern legal guilt-penalty systems [31]. In this sense, consciousness is a kind of self-awareness; it is a condition for cognition.

More recently, consciousness has been focused by modern disciplines such as Psychology, Neuroscience and Artificial Intelligence (AI). Especially in AI, an important aim is the definition an later implementation of a model for consciousness. In this line of work, the first step is finding an answer to the main question: "Where does consciousness reside?" Is it immaterial, like "the soul", or is there a physical support - a neural correlate - for consciousness? [29]. A neural correlate of consciousness (NCC, according to [6]) are "neural systems and properties of that systems, which are associated with conscious mental states" [14]. Another definition of a NCC, which may perhaps be commonly accepted, is "a neural correlate of consciousness is a neural system (S) plus a certain state of this system (NS), which together are correlated with a certain state of consciousness (C) [10]. The existence of a NCC is widely accepted in scientific community, but unfortunately "how these neural correlates actually produce consciousness, is left untouched" [14]. This is not surprising, because the study of consciousness is not an easy task, taking into account the "complexity of the neuronal architectures involved, it seems risky to draw conclusions simply on the basis of intuitive reasoning" [10]. Due to this complexity, Francis Crick opted to defer even a consciousness definition to avoid precipitation [8].

Consciousness can be divided in two important categories. The first category is similar to self-knowledge, which has to do with the ordinary notion of being conscious. Many people think that this kind of consciousness is the same as knowledge. Actually, though, it is a way of developing declarative memories. Declarative memories are memories that can be recalled and told to others. The second category, called "qualia", refers to the idea that the feelings associated with a sensation are independent of the sensory input. As this is a more metaphysical category than the first one, it will not be considered in this paper. Qualia are frequently formulated in questions like, "Why is the colour red, red?" "Does the colour red appear to be the same colour to you?" Rita Levi Montalcini, the Nobel Laureate for Medicine, pointed out that the three main lines of research into the consciousness problem were: the neurosciences, cognitive science and AI. This paper is concerned with the two last lines, and especially cognitive science.

Another important point that is present in the approach we use in this paper is that consciousness research must focus on both cognitive processes and beahaviour. The

 Manuscript received December 7, 2007; Revised June 2, 2008

essential idea in AI, proposed by Turing in his test (as a measure) and his machine (as a medium), can be established as follows: "The brain is just another kind of computer. It doesn't matter how you design an artificially intelligent system, it just has to produce human like behaviour". Nevertheless, this behaviourism is the main problem in the classic AI field. The Turing test, which takes intelligence and human behaviour to be equivalent, limits the vision of what is possible: from a connectionist or a symbolist point of view, it focuses on behaviour and ignores other relevant aspects. In fact, one can be intelligent by thinking and understanding without acting. Ignoring what happens in the brain and focusing only on behaviour has been and is the greatest obstacle to understanding intelligence.

Of course, such profound questions are quite difficult to answer because our knowledge of the human brain and cognitive processes is still poor. Despite the limitations we have in this field, some psychologists have made considerable advances by observing cognitive features in connection with human – and sometimes animal – behaviour. In this paper we intend to analyse cognitive features and their relation to the learning process and behaviour. From a cognitive science viewpoint, we base our research on an analytical approach to consciousness, focusing on the self-consciousness feature. We propose a cognitive architecture for *self-consciousness* using Modular Artificial Neural Networks (MANN). We implemented a virtual environment with intelligent virtual holons to test the proposed model. Finally, we analyse the results are and draw some conclusions

II. CONSCIOUSNESS FEATURES

Because it is impossible to understand consciousness as a whole, the most common approach - as is usual in science - is analytical. This means that consciousness is defined injectively, that is, based on the features habitually associated with consciousness or the features in which consciousness is believed to play a role. Bernard Baars [5] and Igor Alexander [1] have suggested several cognitive features of consciousness beings. From these and other researchers, we can extract several cognitive features that must be present in the consciousness phenomenon. These features can be divided into three abstraction levels: basic, intermediate and advanced features.

As we consider consciousness as a holonic system, each feature can be viewed as a whole and, at the same time, as a part of the holonic system. Viewed individually, as a whole, these features are not basic at all. However, viewed as parts of consciousness, they can be described as the building blocks of consciousness. This level encompasses *reactivity*, *adaptability*, *associative memory*, *learning ability* and *optimisation*. A lot of successful research has been done into modelling and implementing these features.

Intermediate features are the result of a composition or interaction of basic features (level 1). They include *abstraction, prediction, anticipation, generalization, inference, emotion*, *motivation* and *imagination.* Some research has been done focusing on these features with patchy results.

Consciousness also include advanced features. These are complex and require a cognitive architecture composed of features from levels 1 and 2. They include *free will*, *moral judgement* and *self-consciousness.* As we have already mentioned, these features are the hardest to model and to find a suitable technology for implementation. In this paper, we focus especially in *self-consciousness*.

III. A NEURAL CORRELATE OF SELF-CONSCIOUSNESS

It is sometimes said that consciousness does not have its own neural correlate, but it is just the sum of all the features listed above [5]. Contrary to this, other researchers postulate that consciousness is not merely a sum of cognitive features. They claim that, once all these features are present in an individual, they interact with each other, generating new features at a higher abstraction level. As a result of this *emerging behaviour*, "the whole is greater than the sum of the parts" [19]. The fact is, in any case, that consciousness is always associated with these features. Therefore, a lot of research work has been done proposing models for each consciousness-related feature and also possible implementations in the field of artificial intelligence and cognitive science have been essayed [16] [17] [38] [39]. As we have already said, in this paper, we focus on the last feature listed above: *self-consciousness*

There are different ideas, and consequently different definitions, of self-consciousness. Some researchers [22] [27] make a distinction between *self-awareness* (knowledge of oneself as an entity) and *self-consciousness*. Self-consciousness has been defined as "the possession of a concept of identity, as well as the ability to use this concept to think about oneself" [26]. In some animal species, we can observe earlier states on the path towards *self-consciousness*. Most mammals and birds can recognize other individuals of their species as being similar. This means they have a sense of belonging in terms of their species [40]. A few superior mammals not only have a sense of *self-belonging*, but also demonstrate *self-awareness*. Self-awareness means they can distinguish their own image from that of other individuals, which is one of the signs that confirms they have this cognitive feature. This select group now lists chimpanzees [42], dolphins [35], a recent addition, elephants [34], and of course, human beings.

In the artificial intelligence field, several researchers are working on the implementation of *self-awareness* and *self-consciousness* in robots [25] or even in software holons or soft-bots [15]. Most are focusing on self-image recognition, and robots were recently equipped with this feature. It is noteworthy, however, that although recognition of one's own image implies *self-belonging* or even what has been called *self-body awareness* [30], this does not necessarily prove that the entity (natural or artificial) has *self-consciousness*. To be able to say this, the entity would also have to be able to build an abstract *self-representation* and also be able to use it as essential information for properly interacting with other individuals and with the environment [27] [37] [9].

There is no doubt that *self-representation* is a key component for *self-consciousness*, because "how can anyone have knowledge of you that you cannot represent?" [22]. Conscious individuals have internal representations of things, but *self-representation* is different from this "primary representations". It has been considered as a case of "secondary representation", which are "cognitions that

represent past, future, pretended, or purely hypothetical situations in prepositional form" [4]. It is evident that *self-representation* must be a secondary one, because it is "a constructed mental model of oneself that can be manipulated in fantasy" [4] This cognitive structures are closely related with perspective-taking because "self-recognition and spontaneous perspective-taking develop in close synchrony because both require a capacity for secondary representation" [4].

This *self-representation* must necessarily be abstract to support abstract inference processes. It also needs to be dynamic and flexible enough to adapt to both changes to its own self and changes in the environment. Obviously, this would be impossible with a static self-representation. Contrariwise, an individual needs to learn about itself - like humans do –, and its *self-representation* would undergo changes induced by a learning process throughout the individual's whole lifetime. In this process, individuals interaction has a great influence. The poet Arthur Rimbaud said "I is some one Else" ("Je est quelqu_un d_autre"), suggesting that we conceive ourselves through the eyes of others" [36]. Indeed, other individuals influence our self-representation because we not only build a secondary representation of the self, but also of the others. This other individuals representations are also a case of secondary representation because "it is not a perception of a situation but rather a constructed mental image of another person's perception of this situation" [4].

By this interaction, the individual construct relations with other individuals, and as a result "each individual has an overall repertoire of selves, each of which stems from a relationship with a significant other", This becomes "a source of, the interpersonal patterns that characterize the individual. Each self is keyed to a mental representation of a significant other" [3]. This source of information becomes a sort "narrative center [...] of all subjective experiences and memories in a given individual" [11]. Taking this facts into account, we consider that first of all, a self-conciousness model must include both self and other individual representations and the close relation between these cognitive features must be also defined. On the other hand, because of the importance of individual interaction in self building process, we considered that a simulator that includes interaction between modelled systems would be an adequate testing strategy for *self-consciousness* models.

Another important and essential feature is to be able to reference this abstract information and apply it in connection with other cognitive features. One such feature is *self-imagination*. Self-imagination implies the ability to "see" one's own representation, a certain conception of what one is like. Another is *self-inference*, meaning the ability to infer information and reason inductively and deductively about oneself. Finally, *anticipation* is another related feature. Anticipation is the ability to foresee results taking into account knowledge about oneself.

Clearly, *self-consciousness* is a complex cognitive feature. It includes an abstract and dynamic *self-representation*, a mechanism for using this representation and interaction with other cognitive features to evaluate this representation for inference and anticipation. This suggests a modular cognitive architecture. Taking these points into account, we chose ANN to provide a neural correlate of *self-consciousness* in intelligent individuals.

Information cannot be addressed without taking into account both natural and artificial information processing devices, because information is an abstraction that is only materialised when it has a physical representation. In particular, self-information has a representation, which, in this paper, is called self-representation. This makes it possible to use and process this information. Cognitive capabilities like self-consciousness and abstraction can be implemented to provide devices with intelligent behaviour, which is the goal of Artificial Intelligence. In this paper, self-consciousness, and abstraction, or the ability to separate the essential from the secondary, are built into the holons. Abstraction is necessary for recognizing other individuals, because these representations are an abstraction of reality, which is useful for each holon's behaviour [2] [13] [32].

The term informon is used in this paper to designate the basic component of information. Indeed, an informon is an information entity. Information can take the form of data, news or knowledge. Information is produced when some degree of uncertainty exists. As Sanders [41] suggested, information is produced as a result of an uncertainty reduction process. Denning [12] defines information as the meaning that someone attaches to a data set. Brook [7] gave another definition making a distinction between "knowledge", as a structure of linked concepts, from "information" which he defines as a small part of "knowledge". Following on from this, Mason indicates that "information can be viewed as a collection of symbols […] that has the potential of changing the cognitive state of the decision-making entity" [23].

If we lump all these definitions together, information can be defined as "a difference, caused by an underlying process, almost always driven by interest, able to transform a cognitive structure through a collection of symbols that has the potential of changing the cognitive state of a [holon]". In a holonic System, holons are immersed in a medium. A "medium" is defined as any environment that can transmit signals or phenomena. Phenomena appear as information to perception. The perception of phenomena is certainly a form of information. Signals are represented by a code of signs. Signals can be coded to produce signs. Signs are the way in which signals are coded. Sign study and analysis is called semiotics.

Data are signs organized in a certain pattern. Data are representations of phenomena, that is, they present phenomena again, hence re-present. When data is interpreted, that is, given a meaning, structure, relevance and purpose, you get news. News can be defined as messages that cause changes in receptor perception. News is transported between systems that have the ability to understand, assimilate and use it. News that is combined with action applied becomes useful information.

Knowledge and wisdom are two higher level cognitive concepts. On the one hand, knowledge can be defined as "news plus action plus application": ideas, rules, procedures, models, laws, theories that guide decisions and actions. On the other hand, wisdom is "knowledge plus experience, plus principles and ethical and aesthetic constraints, judgements and preferences". Wisdom can be individual or collective.

From a formal viewpoint signs have three aspects: syntax, semantics and pragmatics. In this paper, from a syntactic viewpoint, each holon's state, growth and self-confidence is represented by a numerical value. Each numerical value represents a state, a growth and a self-confidence level.

Finally, from a pragmatic viewpoint, each holon decides its actions based on the values of other holons. On the strength of their "representational" basis, there is no way of telling data, news and knowledge apart, as they actually use the same signs and signals. Instead, we can identify how and for what purpose these structures are used. This way they can be categorized. This connects with the problem of the "reference framework" for interpretation.

As stated above, information is out of the question without an information processing device. Therefore, we use the term *holon* to denote the basic information processing element [21]. This term is used then to refer to entities that behave autonomously and, at the same time, as part of a bigger whole. A holon then can be defined as an independent element that behaves autonomously and is self-organizing, recursive and cooperative. A holon must contain information processes, and possibly physical processes. In addition, a holon must be able to cooperate, because it behaves autonomously and acts as part of a whole. Note that holons are not self-sufficient. Nevertheless, they are part of a whole. This is why they need to be able to cooperate, a process by means of which a set of such entities develop commonly accepted plans that they implement in a distributed manner. As explained above, the ability to cooperate is a must. It must be possible to add new entities, and delete and modify others in such a holonic system. Additionally, each holon can self-replicate, which provides the functionality of recursion, self-organization and self-production.

All holons have four impulses: action, communion, transcendence, dissolution. Holons can be classed by the following levels:

Instruction: this level contains the primary holons, cooperative entities that process data. They produce new data and simple news. They are specialized and are able to perform primitive operations.

Component: component holon emerges when the elementary instruction-level holons are structured hierarchically (holarchy); its functionality is greater than the sum of its instruction holons and it is capable of outputting more sophisticated news and/or knowledge.

Entity: entity holons are formed by means of hierarchical relationships between component holons. They have beliefs, motivations and intentions and are able to change their behaviour based on previous experience.

Organization: collaborative entities are called holonic organization.

In this paper, holons are composed of instructions (level 1), and their final cognitive architecture has several components (level 2). They are, as a whole, entities (level 3) because of their data, news and knowledge processing level and their ability to change behaviour according to previous experience. However, viewed as part of a whole, the whole, that is, the system, represents an organization (level 4). A holonic structure should consider the cooperation and collaboration domain. Each holon, with its own goals within the domain, operates and communicates with other holons, providing the context in which they can locate, contact and interact with each other.

IV. EXPERIMENTAL PROCEDURE

How could *self-representation* be modelled in a software system? One might think at first glance that it is quite easy for a software system to know its own state, as any system is able to read its own variables at any time. But that is not really *self-consciousness*. If we apply direct self-knowledge, what we get is simply a reading of the system state, which has nothing to do with self-consciousness. Take human beings, for example, the idea we have of ourselves (meaning our qualities, strengths and weaknesses) does not come directly as information provided by our own body, but it is built as a result of a learning process. When we are very young we have an unrealistic idea of what we are really like, but the longer we live – provided our learning process works properly - the more realistic our self-representation becomes.

Therefore, from a cognitive science viewpoint, system variables must be separated from *self-representation*. This means that, on the one hand, we would have variables concerning holon features (which means its personality if we think of it as a feature vector with a different level of development in each variable) and, on the other hand, the holon's perception of itself. As already mentioned, an abstract representation is needed of this personality, as is a learning process for changing this representation. Furthermore, *self-consciousness* is out of the question without the ability to continuously sense the environment and the *self-representation* and then adapt actions accordingly. For this reason, a process for using this self-information in connection with other cognitive features, such as inference, anticipation and optimization through a learning process also needs to be implemented. As the psychologist Phillip Johnson-Laird said, "Planning and action control requires a self model, including its goals, abilities, options and an actual state" [20]. This learning process would not be possible if the conscious entity is isolated. The self-consciousness learning process includes interaction with other individuals. Many research works in the field of psychology have shown that interaction is essential for developing consciousness [28]. Additionally, this process also has to be dynamic to allow learning process optimization. Because of the above features of *self-consciousness* and *self-representation* we considered ANN to be a good implementation choice. On the one hand, ANN are an adequate representation for a neural correlate of consciousness as they are biologically inspired. Incidentally, brain processes are quite different from traditional (algorithmic) computation. There are no explicit algorithms in biological neural systems. Contrariwise, intelligence, and consciousness, resides in neuron connectivity. Taking this into account, an ANN is suitable for modelling consciousness, as it does not incorporate problem-solving algorithms, and cognitive features reside in the weight configuration. Furthermore, as ANN are modular, they are adequate for implementing cognitive architectures. Being dynamic, they provide for dynamic *self-representation*. Finally, ANNs are learning trainable by definition. This allows the *self-representation* to evolve and be optimized throughout the learning process.
In the case of human beings, *self-representation* is not confined to an individual having a standard internal representation of him- or herself as a human being, as opposed to some other species (*self-belonging*), but also extends to the abstract representation of his or her self with his or her unique personality. Therefore, in our virtual environment, we equipped holons with features that determine their abilities and behavior. First, we defined holons that had a particular size and shape. Depending on these features, the holon has a bigger or smaller chance in

competitions with other holons. A holon's size grows from a random initial size as time passes. After a period of time, they disappear, and a new holon appears in their place. These features were added to prevent the virtual environment reaching a state where whole holon population was in a terminal status, as this would make it difficult to test the evolution of self-representation and associated processes. After testing with a different number of growth levels, the number of possible growth levels was finally set at ten, because this extended the holon life cycle, facilitating learning process. These levels are represented in the virtual environment by increasing the holon's diameter. Another feature, which can be defined as holon "state" is dependent on factors that we will explain later. Ten "states" (0 to 9) were also defined and represented as different colors: violet, dark blue, light blue, dark green, light green, yellow, orange, magenta, light red and dark red. Fig. 1 shows holon interaction.

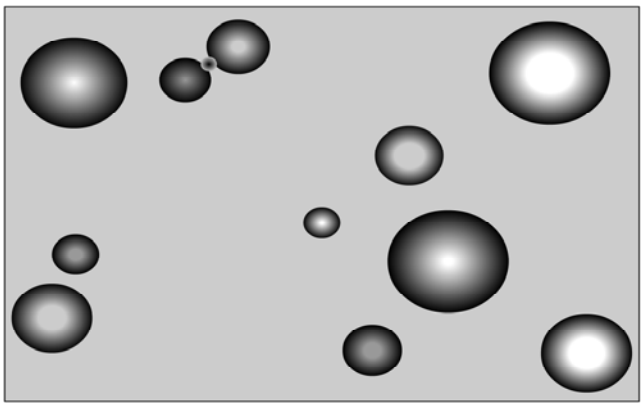

Figure 1. Evolution of relative feature weighs.

In this paper, we consider consciousness as a result of social interaction with an internal learning process. Therefore, we created a virtual environment, with a certain number of interacting holons to test the proposed model. The interaction was defined as a competition between holons, where each holon competes with another (one at a time), for example, in a contest. In the virtual environment, the holons sometimes attack, and sometimes flee other holons, depending on how they rate themselves (*self-consciousness*), meaning their evaluation of the perception they have of their own qualities (*self-representation*). These holons were also defined with the aim of observing other individuals' behavior to optimize the accuracy of their own abstract representation by both learning from their own experiences and observing other holons' experiences (*observational learning*). Throughout this learning process, the abstract representation the holon has of other individuals evolves, but, more importantly, it also improves its *self-representation*. This improves its evaluation and anticipation of its future actions.

Holons perceive growth true to its real value, but state is perceived with some error, depending on the individual. Initially, these values are set. Therefore, the holon focuses first on learning the relative importance of each quality (growth and state) for competition through a learning process. As a result, a neural network module represents some kind of "competition function" in each holon. In a second phase, when holons have an approximate notion of how to evaluate their own qualities, the accuracy of their perception of others and themselves also tends to improve. This means that *self-representation* evolves in this second phase and becomes more realistic. Finally, a *self-confidence* feature was added. This feature is defined as the length of the random error factor added for *self-representation*. This way we could generate different self-representation tendencies and test their effect on holon activity.

In the first learning phase, observational learning is very important because it allows holons to learn the representation function. In the second phase, direct learning allows each holon to learn its own qualities and to improve its self-perception.

As our goal is to build a NN implementation of self-representation and self-consciousness, we define the initial conditions as follows:

1. *Representation Function*: This function means the contribution of each holon's features to its global value. This function is initially unknown. Assuming this function is the same for all holons; it is only present in self-representation. In the initial state, this function is unknown and therefore randomly set.

2. *Other holon global values*: These values are unknown in the initial state. Nevertheless, they are initialized with approximations (as a result of an imperfect perception). We used a random error function uniformly distributed across a range of 10%. We also assumed that while the global values are unknown, the individual holon features are known.

3. *Global own value*: In the initial state this value is unknown and the first approximation is the self-representation neural network output. We also considered the holon feature values as unknown, and therefore randomly initialized.

A learning process is also needed to evolve self-representation and self-consciousness. This learning process includes two different ways of learning:

1. *Self-Experience Learning*: When a holon has a confrontation with another, it forecasts the possible outcome. If the forecast is wrong (the result is the opposite of what was expected), the holon adjusts the representation of the other holon according to reality and also adjusts its self-representation to include the result function.

2. *Vicarious Learning*: When the holon observes a confrontation between two other holons, it also makes a forecast of the possible outcome. If the forecast is wrong, the holon adjusts the representation of global values of both observed holons and also adjusts the result function in the implicit neural network of each holon. As ANN have just two layers in these cases, a Delta-rule algorithm was implemented to adjust neurons.

Diagram in Fig. 2 illustrates the main steps in learning process:

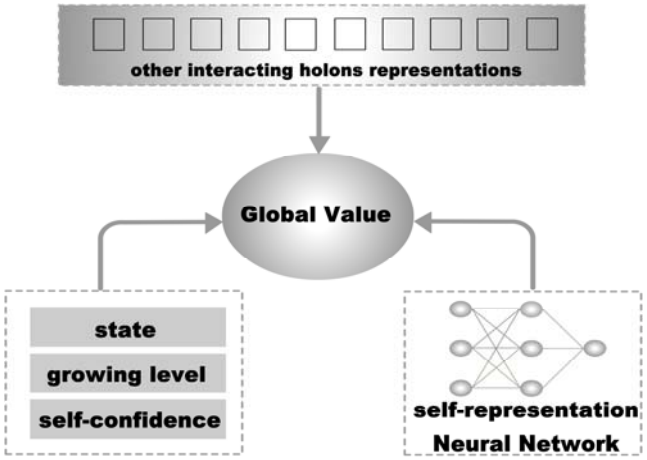

Figure 2. Learning process.

Neural Networks were used for abstract self-representation, representation of other individuals and also function evaluation. This means that it represents the process of using self-information to anticipate and decide future actions. Fig. 3 shows the topologies used for each module.

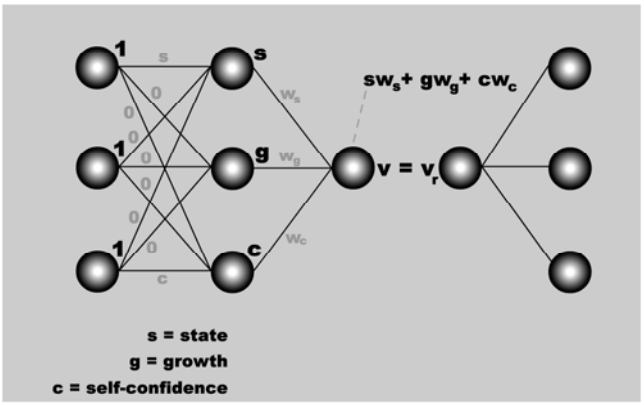

Figure 3. Neural Networks topology

Clearly, multi-layer perceptrons were used (they were the preferred option as they have proved to be universal approximators [17], but any other kind of ANN can possibly be used). Each holon is equipped with a certain number of ANN. The system is, therefore, a modular-ANN (MANN). The main ANN module contains the self-representation (including feature values and evaluation function), and other modules have representations with feature values of other holons.

Self-Representation Module

Each holon has a self-representation (the ANN topology on the left) containing the holon features, and the global value is the ANN output.

The relative impact of each feature is represented by the weights that connect the hidden and output layers. The hidden layer input values (feature values) are used as weights in one of the connections between the input and hidden layers. The other weights in this layer are set at 0, and input values from the input layer are set at 1. As a result, processing this multi-layer perceptron returns an output value that represents the global value of each holon from its own viewpoint (self-representation).

When this global value changes, all the weights can be adjusted by back-propagation. Nevertheless, in case of connections between the input and hidden layers, these weights are used to calculate new s, g and c for hidden layer inputs. Later, these weights are set as mentioned above.

Note that both the feature values and the evaluation function (based on NN-weights) are represented in this self-representation module.

Other Holon Representation Module

As we assume that the evaluation function is the same for all holons, we only represent feature values for other holons. The global value of these holons is calculated as a weighted sum of these feature values. This is represented by the net on the right in Figure 3.

As a result of this M-ANN architecture, each holon will be able to recognize other individuals' capabilities. Additionally, each holon will have a self-representation. Self-representation means how the holon views itself. This information is used by the holon's central process to evaluate its possibilities compared with other individuals in social interaction, as it can be seen in Figure 4.

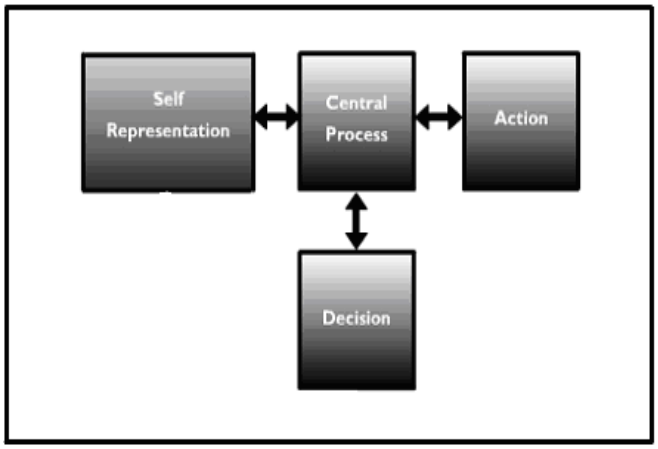

Figure 4. Cognitive Architecture

V. RESULTS

After implementation, we tested the system with different initial configurations where we primarily varied the number of holons and perception error range. As a result, we observed how self-representation evolved (in each holon) and its influence on later holon behaviour.

First, we will analyse the evolution of relative feature weighting. Self-representation converges at the relative contribution of each holon feature to global value. This means that the individual not only learns more about itself globally (global value) in a second phase of learning process, but also learns more about the relative importance of each of its own features. This is shown in Fig. 5.

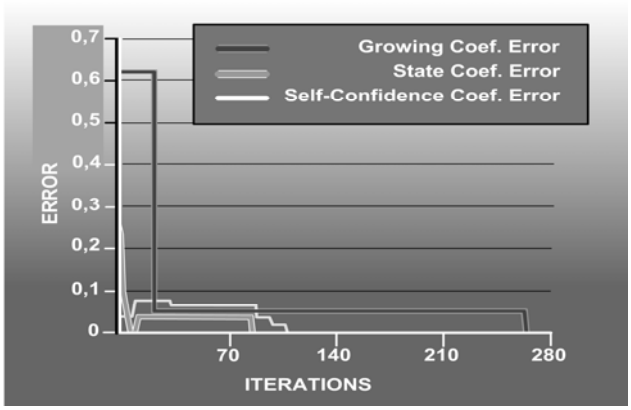

Figure 4. Features representation evolution.

This chart shows the convergence of the result function for the three implemented features, and, as a result, the convergence of self-knowledge for each of the three features of a holon. Fig. 5 illustrates how the error level decreases in a few steps to an acceptable level of about 0.05, and then converges to an almost exact perception of each feature in a second phase. Fig. 6 shows the relative perception error of three holons after consecutive contests. Because the first holon (in white) avoided contests after the 6th iteration, learning was unsuccessful in its case. Anyway, all holons tend to minimize their perception error, and also improve their forecasting accuracy.

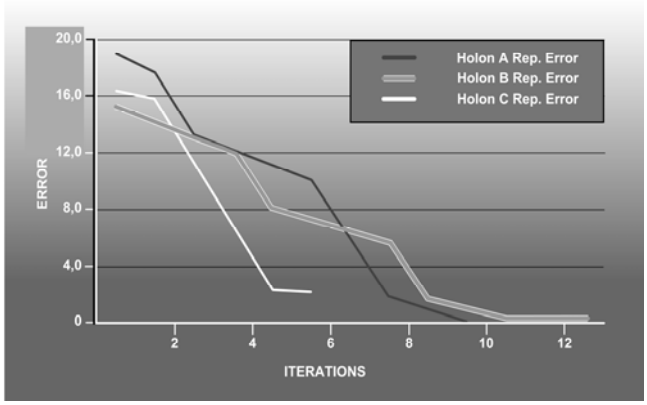

Figure 6. Self-representation evolution.

Fig. 7 shows how self-representation evolves throughout the process. Again, there are three holons, plus their global values (from their own viewpoint). In these cases error is minimized after an initial period of instability, product of the interaction with differently valued holons.

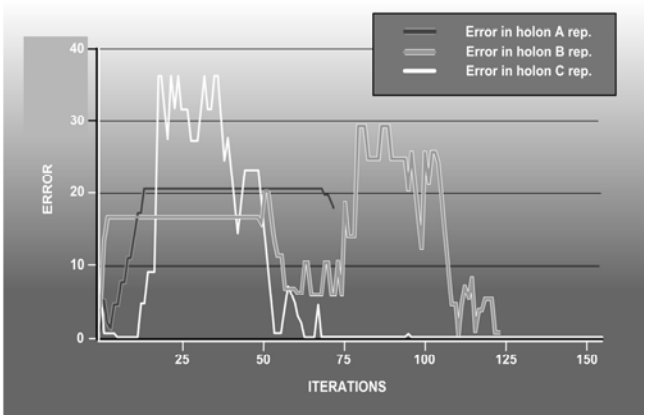

Figure 7. Other individuals representations.

VI. CONCLUSIONS

As discussed in this paper, we analysed the relation between *self-consciousness* and *self-representation*. Our focus was that conscious individuals constantly modify their behaviour depending on the representation they have of other individuals, but more importantly, depending on the use they make of the information provided by their *self-representation.*

With the model proposed and implemented in this paper, we were able to observe *self-representation* implemented with holons and found that was useful for representing:

1. Time-dependent evolution of self-representation
2. Influence of self-confidence on self-consciousness
3. Relation between level of interaction and self-consciousness development.

We can conclude that the use of ANN is suitable for implementing cognitive features, particularly in the case of self-consciousness and self-representation, for several reasons:

1. Biologically inspired

The human brain is a physical organ, and its thinking part is based on neurons. The proposed model must ape this. ANN imitates physical neuron structure, their connectivity and mechanisms. As ANN are biologically inspired systems, they are suitable for modeling consciousness.

2. Non-Algorithmic

In a physical brain, there are not any algorithms; intelligent beings' thought processes are completely different from the way computers traditionally operate, which is algorithmic. As consciousness resides in weights configuration there is a neural correlate.

3. Modularity

Modularity is essential for modeling in cognitive science. We have seen that there are many different levels of cognitive features. Some features are composed; others interact with each other. Furthermore, if a module is damaged, the functionality degrades, but the system continues operating.

4. Adaptability

Cognitive architectures with ANN are also flexible and adaptable through a learning process. In the approach taken in this paper, self-consciousness and self-representation are not innate features, but are the result of an interaction process. In this process the individual interacts with the environment and acquires capabilities of self-consciousness and self-representation through a learning process.

The interaction among perception, anticipation and decision processes and self-consciousness has been thoroughly analyzed by psychologists, neurobiologists and engineers working in cognitive science. In this paper, we saw how MANN-equipped holons in a simplified cognitive

model interact with each other and how self-representation and self-consciousness evolves as a result.

As we could analyze in this paper, self-consciousness is a complex cognitive feature. Despite is not feasible to design a realistic model in the current state-of-art, it is possible, by an abstraction, to focus on some aspects of this cognitive feature. In this paper, we focused on how self-consciousness is based on self-representation. Particularly, we focused on how self-representation is not an inherent feature of conscious entities, but it develops as a result of a learning process. An important conclusion, is that this learning process depends essentially on interaction between conscious entities, and it can include both direct and observational learning. Of course, the self-representation model and the learning processes described in this paper are quite far from a realistic model. Nevertheless, they ilustrate that it is possible to model a dynamic self-representation in artificial entities that evolves as a result of a learning process based on interaction. Moreover, it also shows that according to some consciousness' properties such as modularity, dynamic nature and learning-based development, Modular Neural Networks appear to be suitable structures for model implementation.

AKNOWLEDGDMENTS

We wolud like to thank INAP (National Institute of Public Administration) for funding project DISTIC-AD P07105113, and Rachel Elliott (CETTICO: Center of Computing and Communications Technology Transfer), for her help in translating this paper. Our thanks also go to Salomé Garcia, form acting a intermediary between the two universities.

REFERENCES

[1] Alexander I et al.. *How to Build a Mind. Mapping the Mind Series.* Columbia University Press, New York, 2000.

[2] Alkins, P. *El Dedo de Galileo. Las Diez Grandes Ideas de la Ciencia*. Espasa-Calpe, S.A. Madrid, 2003.

[3] Andersen, S.M., et al. *The unconscious relational self. The new unconscious* (pp. 421-481). New York: Oxford University Press, 2005

[4] Asendorpf, J. et al. *Self-Awareness and Other-Awareness II: Mirror Self-Recognition, Social Contingency Awareness, and Synchronic Imitation*. Developmental Psychology, 1996, Vol.32, No, 2,313-321.American Psychological Association, Inc, 1986.

[5] Baars, B. *A Cognitive Theory of Consciousness*. Cambridge University, Cambridge, 1988.

[6] Block, N. *Two Neural Correlates of Consciousness*. Trends in Cognitive Sciences, vol (9), 2, 2005.

[7] Brook A., De Vidi, R. *Self-reference and self-awareness.* John Benjamín Publishing Company, 1980.

[8] Crick, F. *The astonishing Hypothesis: The Scientific Search for the Soul*. Touchstone Ed. New York, 1996.

[9] Decity, J.; Chaminade, T. *When the self represents the other: A new cognitive neuroscience view on psychological identification*. Science Direct, 2003.

[10] Dehaene S. & Changeux, J.P. *Neural Mechanisms for Access to Consciousness*. The Cognitive Neurosciences. Third Edition, 2003.

[11] Dennet, D. *Consciousness Explained*. Boston: Little, Brown and Co., 1991

[12] Denning, P. The profession of IT: The IT schools movement. CACM, Vol.44:8, 2001, pp. 19-22.

[13] Dossey, B. *Core Curriculum for Holistic Nursing*. Jones & Bartlett Publishers, Santa Fe, NM, 1997.

[14] Fell, J. *Identifying neural correlates of consciousness: The state space approach*. Science Direct. Available online at www.sciencedirect.com, 2004.

[15] Franklin, S.; Graeser, A. *Modeling Cognition with Software Agents*. Proceedings of the Third International Conference on Cognitive Modeling, Groeningen, NL, ed. N. Taatgen. Veenendal, NL: Universal Press, 1999.

[16] Haikonen, P. *The Cognitive Approach to Conscious Machines*. Exeter, UK., Imprint Academic, 2003.

[17] Haikonen, P. *Conscious Machines and Machine Emotions*. Machine Consciousness Models Workshop, Antwerp, BE, 2004.

[18] Haykin, S. *Neural Networks . A comprehensive Foundation*. Second Edition. Pearson Prentice Hall and Dorling Kindersley, India, 2006.

[19] Hopfield, J. *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*. Proc. Natl. Acad. Sci. USA 79: 2554-2558, 1982.

[20] Johnson-Laird, P. *Mental Models: towards a cognitive science of language, science and consciousness*. Harvard Cognitive Science Series. Vol 6. , Cambridge, 1983.

[21] Koestler, A. *The ghost in the machine*. Hutchinson Publishers, London, 1967.

[22] Levine, A. *Conscious Awareness and (Self-) Representation.* Consciousness and Self-Reference, ed. Uriah Kriegel, MIT/Bradford. Ohio, 2002.

[23] Mason, R. *Measuring Information Output: A communication Systems Approach*. Information and Management 1, 219–234, 1978.

[24] Mathew 5:3, *New World Translation of Holy Scriptures*. Presbyterian and Reformed Publishing Company, Phillipsburg, New Jersey, 1982.

[25] McCarthy, J. *Making robots conscious of their mental state*. Working Notes of the AAAI Spring Symposium on Representing Mental States and Mechanisms, Menlo Park, California, 1996.

[26] McGaughey, William. *Rhythm and Self-Consciousness: New Ideals for an Electronic Civilization*. Thistlerose Publications, Minneapolis, 2001.

[27] Menant, C. *Evolution and Mirror Neurons: An introduction to the nature of self-consciousness*. TSC, Copenhagen, 2005.

[28] Menant, C. *Evolution of Representations. From basic life to self-representation and self-consciousness*. Tucson consciousness conference, Arizona, 2006.

[29] Metzinger, T. *The Neural Correlates of Consciousness.* Cambridge, MIT Press, 2000.

[30] Nielsen, M. et al. *Mirror Self-recognition beyond the face*. Child Development. V 77. Blackwell Publishing, Oxford, 2006.

[31] Nietzsche, F. *On the Genealogy of Morals*. Oxford University Press, Oxford [1887] (re-print), 1998.

[32] Pazos, J. et al. *Informones y Holones*. Levi Montalcini, R.: La Galaxia Mente. Editorial Crítica, S.L. Barcelona, 2000.

[33] Pina Polo, F. *Marco Tulio Cicerón*. Ariel S.A. Ed. Barcelona, 2005.

[34] Plotnik, J.M., et al. *Self-Recognition in an Asian Elephant.* Proceedings of the National Academy of Sciences 103: 17053-17057, Washington, 2006.

[35] Raiss, D., Marino, L. *Mirror Self-recognition in the bottlenose dolphin: A case of cognitive convergence*. Proceedings of the National Academy of Sciences of the United States of America, vol. 98-10, Washington, 2001.

[36] Rochat, Ph. *Five levels of self-awareness as they unfold early in life*. Science Direct. Available online at www.sciencedirect.com, 2003.

[37] Rossenberg, G.; Anderson, M. *A brief introduction to the guidance theory of representation*. In Proceedings 26th Annual Conference of the Cognitive Science Society. CAN, 2004.

[38] Sloman, A. *What sort of architecture is required for a human-like agent?* Michael Wooldridge and Anand Rao, editors, Foundations of Rational Agency. Kluwer Academic Publishers, Oregon, 1997.

[39] Stan, F. *IDA: A Conscious Artefact?* Machine Consciousness, Ed. Owen Holland, UK., Imprimnt Academic, 2003.

[40] Wang hui. *The Individual and Modern Identity in China*. Chinese Academy of Social Sciences, China, 2003.

[41] Worthen B., Sanders J. (1973). *Educational Evaluation: Theory and Practice*. Jones, Worthington, Ohio, 1973.

[42] Vergio, S. *Animal Self Awareness.* Available online at http://www.strato.net/~crvny/, 1997