THE ROLE OF THE ARTIFICIAL INTELLIGENCE WITHIN THE CONTEXT OF THE HUMAN FACTORS IN THE NUCLEAR SAFETY

Marco Antonio Bayout Alvarenga. ANGRA-II Licensing Coordinator
Comissão Nacional de Energia Nuclear. Rio de Janeiro. Brasil
E-mail: Bayout@brlncc.bitnet

ABSTRACT

The effective evaluation of a human-machine system depends heavily on a cognitive model of the human behavior. The basic question is: how can we model the human cognition? The response should be found in the five disciplines that form the Cognitive Sciences: Artificial Intelligence. Cognitive Psychology. Neurophysiology. Linguistics. and Philosophy. Among them. the Artificial Intelligence appears as the catalyser of the contributions and discoveries in the other four. trying to realize that cognitive model with the tools of the Computer Science. Sometimes. it seems us if these disciplines spoke different languages to describe the same ideas. It is necessary a holistic treatment of such questions that include the human cognition and its modeling. This becomes more clear when we observe that there are nowadays different methodologies that must be integrated in some way. This is the case of the symbolic approach (artificial intelligence). connectionist approach (neural networks) and the fuzzy logic. This paper makes a review of the available methodologies. showing the problems and the current solutions to answer the following question. How is possible to develop a human-machine system and an intelligent interface based on the Artificial Intelligence that fulfills the following characteristics: human-centered design. cognitive simulation of the human behavior. and dynamic function allocation. This paper concludes with proposals of national projects to be applied to the Brazilian situation.

1. HUMAN-MACHINE INTERACTION: HUMAN FACTORS AND THEIR IMPORTANCE FOR THE UNDERSTANDING AND MODELING OF HUMAN ERRORS

We have availably to us much technological sophistication in our days as well as increasingly automated safety systems. Despite this. we continue to observe serious accidents some of them catastrophic such as TMI. Chernobyl. Challenger. Bhopal. and others in the aviation industry. In most cases. we are dealing with machine-centered designs instead of human-centered designs. This can be explained with an example mentioned by Rouse [1]: "Is the objective of a pilot the transportation of the airplane from the city A to a city B. or inverting the positions. the airplane is a mean by which the pilot will transport human beings with safety and efficiency from the city A to the city B? ". In the same way. the objective of the nuclear power plant operators is not the simple operation of the nuclear reactor. but the nuclear power plant is the mean by which the operators will generate electricity for the society with safety and efficiency. Beside this. the processing of information in the computer does not match the processing of the information in the human brains. Some operation support systems fail because the designers do not consider this critical difference. This does not mean necessarily that the computers must emulate the characteristics of the human brain. although there are researchers that seek this utopia. Nevertheless. the computerized systems should communicate with human beings. considering how they acquire and process the knowledge (cognitive characteristics). We can conclude from the above that it is important two aspects: human-centered designs. and a cognitive model to simulate the human behavior for the design and evaluation of the human-machine systems. The Rasmussen's framework [2] is the best way to examine the processing of information occurred during the execution of the tasks executed by the operators in the control rooms of the nuclear power plants. This model has three levels of behavior: skill-based level. rule-based level and knowledge-based level. This third level is based on the creativity and free association of ideas to solve a problem that extrapolates the limits of the design. The operators have to analyze the situation. to determine a new state for the system. to

define new tasks necessary to achieve it, and to follow procedures that could or not be anticipated in the operation management. The majority of the operation support systems fails in the non-consideration of this cognitive level.

The main reason to keep the operator in the control room controlled by intelligent computers is that they are necessary to deal with non-anticipated emergencies in the designs. The designers could not imagine all the scenarios of the nuclear accidents to develop the safety systems for each contingency. Consequently, the operators are trained in simulators, but the scenarios again do not guarantee that all the operational events will be covered. This forces us to think about an alternative to the classic supervisory control mode: the collaborative control mode, an idea that has been investigated by Rouse [1]. The collaborative control mode requires a dynamic function allocation. Both machine and humans must supervise each other, one of them assuming the control of the situation if a non-recoverable error from the other occurs. This is important in non-familiar and non-anticipated events. The functions allocator must allow that every possible operation function be allocated to the operators. This is coherent with a human-centered design, reserving to itself those functions and systems that must be always automated, as in case of the protection systems. To accomplish this task, the functions allocator needs adequate information coming from an operator cognitive model. It needs also an error monitor to evaluate the operators cognitive state and to detect, classify, and recommend corrective action of possible human errors. It is possible the ocurrence of an insufficiency concerning to operators cognitive resources or the necessity of very fast actions due the catastrophic deterioration of the system. In this case, the functions allocator will take over the actions from the operators to be executed through the automated systems. However, this has to be done after informing the operators about the situation and giving him/her the chances to recover. Now, let's examine in the next item the possibility of developing the human cognition and human errors models.

2. COGNITIVE MODELS FOR THE HUMAN BEHAVIOR: SYMBOLIC MODELS VERSUS CONNECTIONIST MODELS

The cognitive operator model requires a human intelligence simulation. The specific field of the computer science that investigates this matter is called Artificial Intelligence (AI). The AI deals with the following basic questions in this simulation: the knowledge representation, the knowledge manipulation or control, and the capacity of learning new knowledge. These three items must be integrated into a cognitive architecture that reflects the human characteristics. To do this, the AI should work together with other five disciplines: cognitive psychology, neurophysiology, linguistics, philosophy, and anthropology.

Nowadays, we have available three basic types of cognitive architectures [3]. The first one, Model Human Processor (MHP) models the humans as communication channels. The information flows across the perceptual, cognitive, and motor subparts, which are subject to certain restrictions represented by several time delay constants. Many concepts in this model such as long-term memory and short-term memory were conceived along the researches done in the cognitive sciences after the second world war [4,5]. In the second one, the humans are seen as a symbolic processor of logic inferences. Examples of this type are the States, Operators, and Results (SOAR) and the PUPS, together with their predecessors, General Solver Problem (GSP) and the Adaptive Control of Thought (ACT) [6-8]. These architectures are based on content-dependent rules also called production rules, represented by the general form: IF conditions THEN actions.

The AI has produced much "expert systems" based on such rules that capture the knowledge of experts in different fields. They are contrary to the formal logic (propositional or predicative) working with syllogisms. Many investigators have tried to reduce all human knowledge to the syllogistic inference rules: Aristotle (384-322 B.C.), Gottfried-Wilhelm von Leibniz (1661), George Boole (1854), Friedrich-Ludwig Gottlob Frege (1879, 1893, 1903), and Bertrand Russell (1910-1913). Alonzo Church and Alan Mathison Turing proved in 1936 that there is no formal procedure reliable enough to determine the status of an inference in the predicative logic. The solution only came in 1965, when J. Alan Robinson proposed a resolution principle of inferences. This led to the PROLOG (Programmation en Logique) in 1971-1972, which have been adopted in the Japanese project of the fifth generation of computers. As mentioned by Philip N. Johnson-Laird [5]: "Resolution is intelligent, but artificial. It provides cognitive scientists only with a standard of comparisons

because people are hardly likely to translate all premises into a standard disjunctive form and to use only a single rule of inference".

A different approach of cognitive architecture based on production rules is the Theory of Induction (TI) developed by Holland et al. [9]. Holland criticizes the architectures above in several aspects. Concerning the SOAR/GPS, the critics are concentrated in the fact that the basic components of the architecture (initial state, goal state, allowable operators, and applicable constraints) may be only partially known at the moment of the problem solution. A critic driven to both PUPS/ACT and SOAR/GPS is that their production systems allow only one rule to fire (with conflict resolution or chunking processes). Holland's Theory of Induction (TI) [9] consists in the firing of multiple rules through the strengths associated to the rules, exploring the possibilities of parallel competition and collaboration between the rules. To get the same, ACT provides a partial matching of the rules, but this can generate non-valid conclusions. Although ACT uses strength in the rules as in TI, the activation spreading is automatic and independent of the rules execution. In TI, the architecture is based on the coupling of the rules and on the support of one to another. A final critic to both ACT and SOAR is that they use a quasi-linguistic representation of processes associated with the conscious thinking. They ignore the subcognitive level where we find unconscious processes that help the other level in solving problems. In the level of conscious thinking, TI simulates the human cognition through higher order operations, such as planning, within a system entitled Process of Induction (PI). The level of unconscious and subcognitive thinking is simulated by the classifier systems exposed later in this paper. The interface between these two levels is not fixed, leaving as a research how learning at the subcognitive level is integrated with learning at the cognitive level. Induction involves two classes of mechanisms. The first one are the mechanisms for revising the strength of existing rules through the quantification of their uses with the parameters like bid and pay off. They change during the temporal sequence of the rules, which are coupled by matching of their conditions and actions. The second one are the mechanisms for generating new rules through genetic algorithms (GA) and PI of higher order. GAs allow operation of generalization and specialization of rules, applying genetic operators to the rules (e.g. crossover). PI deals with abduction, generalization, specialization, concepts formation, and analogy. The latter is so important in the assessment of scientific discovery processes that Marvin Minsky constructed his particular cognitive architecture called Societies of Mind (SOM) to make easier the reasoning by analogy. SOM is based fundamentally on the knowledge frames concept [10,11].

In SOAR, preferential operators reduce the distance to the goal. In PI, the operators are found through the recategorization of the objects instantiated by synchronic rules to start relevant diachronic rules. PI works with framelike declarative data structures stored in the long-term memory (classifier systems use only rules and their couplings). These structures are linked through concepts. As in ACT, messages sent from the active concepts in the declarative memory produce the clusters of the facts or conditions to fire rules in the production memory. Initially, active concepts are the initial conditions and goals to be achieved. The rule actions generate new concepts to be active or effectors in the environment. The concepts activation spreading is not automated as in ACT, but obeys a processing cycle that uses the subgoaling concept of SOAR. As in ACT, rules change their strengths through the use, but in this case the strength is incremented by the support of the active concept.

'Mental models' are an organized set of simultaneously active or activable rules. They have expectations about future situations through diachronic rules and associations to other relevant events through synchronic rules. To construct 'mental models', the human beings use pragmatic reasoning schemes consisting of a set of highly abstract and generalized inferential rules useful to infer new empirical rules relevant to the problem solving. 'Mental models' are bridge models between two extreme positions: human reasoning using syntactic rules of the formal logic and human reasoning using domain dependent content-specific rules. Holland et al. [9] use a different way from the Johnson-Laird's mental models [12]. The main difference relies on the ways to explain errors. The latter focuses on the working-memory capacity to process several rules at the same time (what explains some skill-based level errors such as forgetting and slips). The other focuses the capacity of mapping concrete situations into pragmatic schemes, generating inferences according to the focused rules (what explains errors in the knowledge- and rule-based level) [4].

Human beings can reason by using any of these three schemes mentioned above: formal logic rules, expert systems rules and mental models [5]. In any case, the AI poses the point of view of a dualism. This dualism is characterized by the existence of a strong algorithm. It is independent of how it is embodied in the physical brain or how could it be simulated on the electronic computer. It does not take into account a particular culture (anthropology) either. This dualism (mind and body completely disconnected) emphasizes that the physical embodiment of an algorithm is totally irrelevant. It is to be supposed consequently that the algorithm has some rather disembodied "existence" [13]. Another question concerns the problem of the non-algorithm nature of the human mind. We have to face the following mathematical discoveries: the Bertrand Russell's Paradox (1902), the Kurt Gödel's Incompleteness Theorem (1931), and the Alonzo Church and Alan Mathison Turing's Thesis (1936) for the David Hilbert's 33rd problem. This carries us to two opposing schools of thinking [13]. One is the Platonic view (which is the Gödel's point of view), the absolute existence of mathematical entities and the acceptabilities of infinite sets. The other school is that from the intuitionism (or finitism) initiated by Luitzen Egbertus Brouwer in 1924. He refused the existence of any infinite set (which is the reason for the arising of such statements like the Russell's Paradox). This reminds the Aristotle's thinking. However, Brouwer rejected "The Law of the Excluded Middle" of the classical logic. The dualist conception is due to René Descartes who was dualist-idealist, contrary to Plato who was monist-idealist, the same view of Immanuel Kant. Georg Wilhelm Friedrich Hegel looked for a synthesis between the two directions, Kant and Descartes. The internal contradictions in the Hegelian idealism were explored by Arthur Schopenhauer, Friedrich Nietsche, and Martin Heidegger. The Philosophy could not give us yet a consistent view of the body-mind dichotomy problem. Plato was a metaphysical idealist, Kant was an epistemological idealist and Hegel was an absolutist idealist whose ideas were materialized in the state power.

The problem of how the algorithm is embodied is treated within the cognitive architecture type called connectionist models or neural networks. In this approach, the syntactic rules of the symbolic models are not explicit. They are encoded in a parallel and distributed network consisting of nodes and connections between them, in analogy with the nets in the physical brain. The neural cells and their connections (synapses) are modeled as if they were capacitive-resistive elements of an electrical net. The importance of neural networks lies on the fact that they can embody unconscious behavior. There are dissociations between conscious and unconscious mind. This implies that unconscious mind may depend on some form of distributed representation similar to the connectionist method. This is contrary to the symbolic representation with rules that are used by the conscious mind to accommodate those dissociations [5]. However, conscious mind contains also non-symbolic thinking: feelings and sensations. Some critics are done concerning the neural network approach [14]: the lack of hierarchical structures appearing in the human brain and the lack of selectional learning mechanisms instead of an instructional one. The first deficiency can be surpassed by a deeper knowledge of neurophysiology, the second one, through the use of genetic algorithms (see in the fourth item).

3. FUZZY LOGIC: THE LINK BETWEEN THE RULE-BASED LEVEL AND THE KNOWLEDGE-BASED LEVEL OF THE RASMUSSEN'S FRAMEWORK

Despite the critics of Johnson-Laird [12] and Holland et al. [9] claiming that Fuzzy Sets Theory is not totally adequate for the concepts combination in the generation of the new rules in the symbolic conscious processes, Fuzzy Logic (FL) can be used in the non-symbolic unconscious processes to answer the question posed by Holland et al. [9]:

"Perhaps the most basic issue to be addressed in future computational work on induction concerns mechanisms that might account for the emergence of high-level cognitive processes from more elementary subcognitive ones. Given an initial set of feature detectors, hard-wired response patterns, inductive operating principles, and other innate system components, how do abstract concepts and inferential rules eventually arise from experience? Current work on classifier systems and neural networks may help in addressing this complex question".

In fact, Fuzzy Sets Theory has been used by Hunt and Rouse [24] to show how is possible for an expert like a nuclear operator to change from a familiar situation when he/she uses rule-based (RB) symptomatic diagnostic search to the unfamiliar situation when knowledge-based (KB) topographic diagnostic search has to be used in a topological way. If there are no more fuzzy rules to apply in the KB level, then a process of induction is triggered (in the Rouse/Hunt approach, fuzzy rules are used in the knowledge-based level). Fuzzy Logic violates the Law of Noncontradiction ($A \cap A^c \neq \emptyset$), the Law of the Excluded Middle ($A \cup A^c$ = universe of events), and the Law of Identity ($A=A$), the Aristotle's Three Laws of the Thought [15]. Fuzziness is a son of the ambiguity or vagueness. So, we could have

$$A \cap A^c \neq \phi; A \cup A^c \neq universe; A \neq A$$

With this assumption, FL can represent uncertainties better than probabilistic theories that use the randomicity concept. This arises from the fact that FL can represent intermediate events between the mutually exclusive events A and A^c [16]. Probabilities Theory works with the Principle of Bipolarity (A and not-A) and there is no permission for events between A and not-A. A great number of our concepts are fuzzy and can admit intermediate states. For example, the adulthood concept is not a bipolar set: adult (> 18 years old) and non-adult (<18 years old). It has a fuzzy degree according to the age and this fuzziness is maximum at the age of 18 [16]. In the Probabilities Theory, we reason about the probability of the event be an element of the set A. P(x is an element of A), and P(x is an element of A) + P(x is not an element of A)=1. In the Fuzzy Logic, we use the membership function $m_A(x)$ of the element x in the set A: $m_A(x)$ maps x-values into the degrees of membership in the closed set [0,1]. So, we can divide this set into any number of fuzzy regions, associating with them the correspondent fuzzy concept. Empirical rule as those in the theory of induction will have in their conditions/actions parts several fuzzy concepts that can be classified according to their membership functions. Rules can be seen therefore as a mapping of a fuzzy set into another fuzzy set.

4. GENETIC ALGORITHMS: THE NECESSITY OF SELECTIVE AND EVOLUTIONIST STRUCTURES

As Edelman exposed in his book [14], the modern synthesis of the biology or neodarwinism is based on two basic principles. They are: Principle of the Heredity of Gregor Mendel (1822-1884) and the Principle of Natural Selection of Charles Darwin (1809-1882). This synthesis started in the 1940s with Oswald Theodore Avery, who discovered that the DNA was the heredity material. The biological systems differ from other physical systems because they are selection systems that work with the notions of evolution and heredity. Edelman proposed therefore a neural Darwinism within a Theory of the Neuronal Group Selection (TNGS). It is based in three tenets. The first one says that the neuroanatomy (neural topology) of a given specimen is found by developmental selection. The second one does not change the anatomical pattern but the synaptic connections are selectively strengthen or weakened. The third tenet says that perceptual categorization is formed by reentrant mapping of paralleled selection of inputs between neuronal groups. Primary consciousness (found in animals) arises when reentrant loop in the memory triggers a conceptual categorization of concurrent perceptions. The high-order consciousness is linked to the language symbolic capacity and arises with the specific area of memory reserved to the semantic structures leading to a conceptual explosion, due the linguistic experience.

Studying the problem of language acquisition by children who can deal with complexities of language even in low ages, Noam Chomsky postulated a view of a generative grammar in which the rules of syntax are independent of semantics. Language can be understood therefore as a formal system or an algorithm. However, some researchers have pointed that children make sense of situations and of human intentions and then of what is said. According to the cognitive grammar of G. Lakoff, conceptual embodiment occurs before the language. Meanings arise because concepts are embodied, and the syntax rules arise from the linguistic experience. In any way, therefore, a cognitive architecture must demonstrate a linguistic cognitive competence [17].

It is a great challenge to construct neural networks according to the first tenet. However, we have an option to construct selection mechanisms according to the second and third tenets to strength or weaken some synaptic constructions in the neural architecture. This mechanism is the genetic algorithm [9]. The Theory of Induction is modeled in the subconscious level by three algorithms: the classifier systems, the bucket brigade (rule supports, bids, and payoff), and the genetic algorithm. The classifier systems are a set of rules consisting of conditions and actions. The conditions are fixed by a set of detectors and the actions by a set of effectors. For example, IF[pressurizer level is low, rapidly decreasing] THEN [starts the second charging pump] has two detectors and one effector. The detectors and effectors are represented by a combination of {1,0,#} notation, where 1 and 0 are binary representations. The symbol # represents a "don't care" effector or detector. In the rule above, the detectors and effectors not included would be represented by #s. These rules compete each other according to their strengths to pass messages forward. The changes of the strengths are done by the brigade algorithm. The strength increases through the continuous uses of the rule: new strength=old strength-bid+payoff, where bid is paid for the predecessor rules and payoff is gained by the next rule. If there is no rule to be applied to the conditions, new rules could be generated by the genetic operators. The main operator is the crossover operator. It selects a random position *i* of the string from two classifiers and exchanges the segments to the left. Then, it replaces the two lowest strength strings with the two new strings. The strings to be combined are chosen according to a probability distribution proportional to their average strength. Instead of recompiling (chunking) the existing rules as in SOAR, TI hypothesize new rules.

5. COGNITIVE MODELS OF NUCLEAR OPERATORS : THE STATE OF THE ART

In the nuclear industry, we can detect the following cognitive models [18]. They are: Cognitive Simulation Model (COSIMO), Cognitive and Action Model of an Erring Operator (CAMEO), Integrated Reactor/Operator System (INTEROPS), Cognitive Environment Simulation (CES), MIT model [19-23].

All these models are similar concerning the processing of information in the operators' mind and the sequence of diagnosis executed by them. They follow basically the Rasmussen's framework, divided in skill, rule, and knowledge levels of processing and using the symptomatic, topographic and hypothesis-and-test diagnostic searches. Some models have more details about the steps done by the operator, while others make a macro-division. All of them agree that the operator reason with familiar situations represented by frames that are triggered by adequate symptoms, formulate hypothesis, confirm them, and choose those used frequently in the past (those having more strength). Known procedures are followed according to the final decision making. The knowledge representation in this rule-based level is the same as in the ACT and SOAR. However, instead of production systems with conflict resolution or chunking, a procedural attachment embodied into frames of production rules is used with small variations. Sometimes, utilizing Blackboard Architectures (BBAs) and agenda of tasks (COSIMO), or programming the basic sequence of the operator tasks (as in the Rasmussen's framework) as general production rules. The models differ just in the aspects of unfamiliar situations in the knowledge-based level (KBL) and in the formation of human errors. In the case of KBL, there are several alternatives: induction processes (to be implemented in COSIMO), topographic rules (T-rules) hierarchically constructed (CAMEO), and qualitative reasoning using means-ends analysis (INTEROPS and CES).

In my point of view, the questions coming from the causal ordering discussion in the qualitative reasoning put an enormous obstacle in the generation of a reliable set of rules to be used in the KBL [24,21]. On the other hand, T-rules are yet an option to S-rules in the RBL(Rule-based Level). What we need are some kind of NEW rules generated during the diagnosis process. This kind of rules only can be achieved through an induction process that is different from a deduction process used in the RBL. Concerning the human errors, the following methods are adopted: formation of slips and lapses through stress functions proportional to a time-pressure stress function, the number of rules processed at the same time (the limited capacity of the working memory), and the exponential decay of facts in the working memory (COSIMO, INTEROPS, HUANG/SIU); computation of the capacity of dividing attention among the information channels: visual, auditive, cognitive, motor (CAMEO). While it is necessary a combination of these two tendencies, nothing has be done concerning the errors formation in the KBL (mistakes), for example, misapplication of good and

strong rules and application of bad rules. Nevertheless, the induction processes to be used in COSIMO will force necessarily the use of such errors treatment to represent the KBL errors [19].

6. THE NECESSITY OF HYBRID COGNITIVE MODELS: FUZZY NEURAL NETWORKS COUPLED WITH GENETIC ALGORITHMS AND INDUCTION THEORY

Using the Kirchhoff's electric current laws, we have (Perkel, 1981) the equations of the neural networks [15]:

$$C_i \frac{dx_i}{dt} = \frac{x_i}{R_i} + \sum_{j=1}^{n} \frac{x_j - x_i}{r_{ij}} + I_i$$

$$C_i \frac{dx_i}{dt} = \frac{x_i}{R_i} - \sum_{j=1}^{n} S_j(x_j) m_{ij} + I_i$$

where r_{ij}=cytoplasmatic resistance between neurons i and j.

$$R_i = \frac{1}{R_i} + \sum_{j}^{n} \frac{1}{r_{ij}}$$

is the lumped resistance.

$$m_{ij} \quad \frac{1}{r_{ij}}$$

and $S_i(x_i)=x_i$, a linear signal function. The Hopfield Circuit (1984) arises when the synaptic connection matrix is symmetric ($M=M^T$).

It is usual to define two field of memories F_x and F_y and the associated equations (continuous additive Bidirectional Associative Memories-BAM):

$$\frac{dx_i}{dt} \quad A_i x_i \quad \sum_{j}^{n} S_j(y_j) m_{ij} + I_i$$

$$\frac{dy_j}{dt} \quad A_j Y_j + \sum_{i}^{n} S_j(x_i) m_{ij} + J_i$$

The Cohen-Grossberg activation dynamics (1983) [15] is a multiplicative model of the Hopfield circuit to avoid saturation problems.

$$\frac{dx_i}{dt} = a_i(x_i)[b_i(x_i) - \sum_{i}^{n} S_i(x_i) m_{ij}]$$

Adaptive Bidirectional Associative Memory (ABAM) is a generalization of the Cohen-Grossberg model [15]. If Brownian diffusions or 'noise' processes perturb the ABAM models, we get stochastic differential equations, with random processes as solutions (RABAM) [15]. If we scale the independent Gaussian white-noise processes with the square-root of annealing schedules or 'temperatures', we have the simulated annealing of a RABAM. Neural Network can be considered therefore dynamic systems represented by stochastic differential equations.

The neural network architectures can be classified in feedforward or feedback (BAM, ABAM, RABAM) models and they need to be trained with external data representing our knowledge about a specific subject.

There are two types of learning: unsupervised (like the Pavlovian training) and supervised (renewal-punishment system). There are four types of unsupervised learning: 1) Signal Hebbian Learning law-SHL (Hebbs, 1949); 2) Competitive Learning Law-CL (Grossberg, 1969); 3) Differential Hebbian Learning Law-DHL (Kosko, 1988); 4) Differential Competitive Learning Law-DCL (Kosko, 1990).

SHL $$\frac{dm_{ij}}{dt} = -m_{ij} + S_i(x_i)S_j(y_j)$$

CL $$\frac{dm_{ij}}{dt} = S_j(y_j)[S_i(x_i) - m_{ij}]$$

DHL $$\frac{dm_{ij}}{dt} = -m_{ij} + S_i(x_i)S_j(y_j) + \frac{dS_i(x_i)}{dt}\frac{dS_j(y_j)}{dt}$$

DCL $$\frac{dm_{ij}}{dt} = -\frac{dS_j(y_j)}{dt}[S_i(x_i) - m_{ij}]$$

In particular, CL means that F_x neurons compete for the activation induced by the signal pattern S(x) generated by the field F_x. The synapses learn only if the their postsynaptic neurons win ($S_j(y_j(t))=1$). DCL means "learn only if change".

These laws have also random versions, RABAM and ABAM theorems, and they guarantee adaptive global stability [15]. Competitive ABAM (CABAMs) are equivalent to the Adaptive Resonance Theory (ART) of Grossberg (1982). ART means "learn only if resonate". The Boltzmann machine is a particular case of simulated annealing with periods of Hebbian and anti-Hebbian learning (negative signal).
The Adaptive Vector Quantization (AVQ) Centroid Theorem establishes that

$$E[\mathbf{m}_j] = \text{average}\{\mathbf{x}_j\} \quad [15]$$

for linear CLs and DCLs with $S(x_i)=x_i$ and indicator functions $S_j(y_j)=I_{D_j}(x)$.

The Supervised Competitive Learning (SCL) uses the reinforcement function r_j defined by known decision-class indicator function:

$$r_j = I_{D_j} - \sum_{i \neq j} I_{D_i}$$

Other supervised learning models are the LMS and the BBA. The Least-Mean-Square (LMS) Algorithm uses the gradient-descent algorithm

$$m_{k+1} = m_k - c_k \nabla_m E[e_k^2]$$

with $E[e_k^2]=e_d^2$, the LMS assumption for the mean-squared error, which simplifies to $\mathbf{m}_{k+1}=\mathbf{m}_k+2ce_k\mathbf{x}_k$. The backpropagation algorithm (BBA) propagates the instantaneous squared error backward from F_y through the hidden fields to F_x at each iteration (which biologically was not confirmed yet).

Models of the competitive Cohen-Grossberg equations have been applied for the pattern recognition problems. The BBAs and the ARTs have been applied to the problems of coding and categorization. The competitive Cohen-Grossberg equations and the BBAs models have also been used for speech recognition and synthesis. Robotics control uses either BBAs or DHL. In the knowledge representation, ART (words and letters) and Boltzmann Machines (to represent schemas like frames) are used. For the causal inferences, we have the Fuzzy Cognitive Maps (FCMs) using the DHL [15].

The Adaptive Vector Quantization (AVQ) theorem [15] allow us to capture all the rules that someone needs to control a nuclear reactor. These rules are embodied in matrices called Adaptive Fuzzy Associative Memory (AFAM) [15] that map a set of input fuzzy variables read in the control room to a fuzzy set

consisting in variables that should be changed to put the reactor in a safe and steady-state condition. If we train a neural network with some competitive learning using the data of simulation of a transient in a nuclear power plant, we can extract cluster of rules in regions defined by our fuzzy variables that have membership functions fixed according to our fuzzy sets. They are fixed with our knowledge of the physical phenomena. Regions that enclose a great number of associated rules will have more strength than others. These are preferential rules used frequently by the operators. Empty regions can be fulfilled with new rules that are discovered by genetic algorithms when there are no more applicable rules to the unfamiliar situation. Preferential rules are responsible for the errors committed when someone applies good rules to the wrong situation. Bad rules could arise if there is no time to induce new rules with the GA. The algorithm stops without achieving an optimal recombination of the rules with the crossover operator. Time pressure can be simulated through the forgetting of the rules with less strength in the fuzzy regions and of some input variables when the visual-auditive channels enter into conflict with the cognitive processing, pushing the cognitive workload index above a fixed limit.

7. CONCLUSION: THE HUMAN FACTORS PROGRAM AT CNEN, INCLUDING THE COGNITIVE MODELS FOR THE EVALUATION OF HUMAN-MACHINE SYSTEMS

At CNEN, we intend to develop a Human Factors Program for the ANGRA-II Licensing that includes a revision of the chapter 18 of the USNRC Standard Review Plan. A cognitive model will be used to simulate the human-machine interface as well as to simulate the human errors to be used in a Probabilistic Safety Assessment (PSA). We have a system to collect data [25,26] from the ANGRA-I operational incidents that can be attributed to human failures, but we have neither a protocol to investigate the root-causes nor an error classification based in cognitive models. We think that the following initiatives will be useful, an improvement of our data collecting system, a development of a protocol, and, the most important, a cognitive model based in the previous discussion to evaluate the ANGRA-I/II NPPs human-machine interface and the operators training systems, to choose a method of human reliability [27], and to diagnose the NPP operation during emergencies [28]. We know that, above all, the IAEA is giving increasing attention to these questions and we would like to follow the developments in this area through the international cooperation with countries that have more experience in this area.

REFERENCES

[1] ROUSE,W.B., 1991. *Design for Success - a Human-Centered Approach to Designing Successful Products and Systems*, John Wiley, New York, N.Y.

[2] RASMUSSEN,J., 1986. *Information Processing and Human-Machine Interaction: an Approach to Cognitive Engineering*, North-Holland, New York, N.Y.

[3] ELKIND, J.W., CARD, S.K., HOCHBERG, J., HUEY,B.M. (Eds.), 1989. *Human Performance Models for Computer-Aided Engineering*, National Academic Press, Washington, D.C.

[4] REASON, J., 1990. *Human Errors*, University Press, Cambridge, London.

[5] JOHNSON-LAIRD, P.N., 1988. *The Computer and the Mind*, Harvard University Press, Cambridge, Massachusetts.

[6] NEWELL, A., 1990. *Unified Theory of Cognition*, Harvard University Press, Cambridge, MA.

[7] ANDERSON, J.R., 1983. *The Architecture of Cognition*, Harvard University Press, Cambridge, MA.

[8] CARBONELL, J., 1990. *Machine Learning -- Paradigms and Methods*, MIT Press, Cambridge, MA.

[9] HOLLAND, J.H., HOLYOAK, K.J., NISBETT, R.E., THAGARD, P.R., 1986. *Induction: Processes of Inference, Learning, and Discovery*, MIT Press, Cambridge, MA.

[10] MINSKY, M., 1993, "Allen Newell, Unified Theories of Cognition", *Artificial Intelligence*,59,343-354.

[11] MINSKY, M., 1986, *The Society of Mind*, Simon & Schuster, New York, N.Y.

[12] JOHNSON-LAIRD, P.N., 1983, *Mental Models -- Towards a Cognitive Science of Language, Inference, and Consciousness*, Harvard University Press, Cambridge, Massachusetts.

[13] PENROSE, R., 1991, *The Emperor's New Mind*, Penguin Books, New York, N.Y.,

[14] EDELMAN, G.M., 1992, *Bright Air, Brilliant Fire on the Matter of the Mind*, Harper Collins Pub.

[15] KOSKO, B., 1992, *Neural Networks and Fuzzy Systems*, Prentice-Hall, Englewood Cliffs, N.J.

[16] KOSKO, B., 1993, *Thinking Fuzzy-The New Science of Fuzzy Logic*, Hyperion, New York, N.Y.

[17] FEHLING, M.R., 1993, "Unified Theories of Cognition: Modeling Cognitive Competence", *Artificial Intelligence*, 59, 295-328.

[18] ALVARENGA, M.A.B., 1993, "Models and Evaluation of Human-Machine Systems", *MITNE-304*, MIT, Cambridge, Massachusetts.

[19] CACCIABUE, P.C., DECORTIS, F., DROZDOWICZ, B., MASSON, M., NORDVIK, J.-P., 1992, "COSIMO, a Cognitive Simulation Model of Human Decision Making and Behaviour in Accident Management of Complex Plants", *IEEE Transac. on Systems, Man and Cybernetics*, 22, 1058-1074.

[20] FUJITA, Y., YANAGISAWA, I., NAKATA, K., ITOH, J., YAMANE, N., KUBOTA, R., TANI, M., 1993, "Modeling Operator with Task Analysis in Mind", in *Topical Meeting on Nuclear Plant Instrumentation, Control, and Man-Machine Interface Technologies*, Knoxville, Tennessee, 505-512.

[21] SCHRYVER, J.C., 1992, "Object-oriented Qualitative Simulation of Human Mental Models of Complex Systems", in *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-22, 526-541.

[23] DANG, V., HUANG, Y., SIU, N., CARROLL, J., 1993, "Analyzing Cognitive Errors using a Dynamic Crew-Simulation Model", in *Proceedings of 1993 IEEE Fifth Conference on Human Factors and Power Plants*, Monterey, California, 520-525.

[24] IWASAKI, Y., SIMON, H.A., 1986, "Theories of Causal Ordering Reply to de Kleer and Brown", *Artificial Intelligence*, 29, 63-67.

[25] ALVARENGA, M A B., BARCELLOS, R.A., COSTA, E.M., SOARES, H.V., DE LIMA, J.M.,1993, "ANGRA-I, Operational Incidents -- Ten Years of Evaluation", in *Proceedings of the SMIRT-12 Conference*, Stuttgart, Germany.

[26] ALVARENGA, M.A.B., 1991, "Statistical Analysis of Components Failures based in the Operational Incident Annual Reports -- The Brazilian Case", in *Proceedings of the SMIRT-11 Conference*, Tokyo.

[27] ALVARENGA, M.B.A., GUIMARÃES, A.C.F., 1992, "Safety Holistic Analysis for Advanced Nuclear Power Plants", in *Proceedings of the International Conference on Design and Safety of Advanced Nuclear Power Plants (ANP'92)*, Tokyo, Japan.

[28] ALVARENGA, M.A.B., GUIMARÃES, A.C.F., 1991, "Applications of Artificial Intelligence and Expert Systems in ANGRA-I Emergency Preparedness -- The Brazilian Case", in *Proceedings of the International Conference on Fast Reactors and Related Fuel Cycles (FR'91)*, Kyoto, Japan.