In the Quest of the Missing Link

Guilherme Bittencourt*
Laboratorio de Controle e Microinformatica
Departamento de Engenharia Eletrica - Universidade Federal de Santa Catarina
88040-900 - Florianopolis - SC - Brazil - E-mail: gb@lcmi.ufsc.br

Abstract

This paper presents a generic model for a cognitive agent based on the hypothesis that the cognitive activity has three main characteristics: self-organization, evolutionary nature and history dependence. According to this model, a cognitive agent presents three levels: reactive, instinctive and cognitive. Each level, together with its lower levels, is intended to model a complete agent, each new level just increasing the behavior complexity. The generic model is instantiated into a computational architecture that integrates connectionist, evolutionary computation and symbolic approaches.

1 Introduction

During its forty years of existence, *Artificial Intelligence (AI)* research produced an heterogeneous set of methods adapted to solve problems in some, usually rather specific, domains. Efforts geared towards an unified theory have not succeeded, even the fundamental research is divided among several uncompatible approaches, e.g., *Physical Symbol Systems* [Newell, 1980], *Connectionism* [Rumelhart and McClelland, 1986] and *Evolutionary Computation* [Goldberg, 1989]. Although methods from different approaches have been successfully combined in some systems (e.g., neural networks and expert systems [Fu, 1994]), no general theory of *Hybrid Systems* is presently available.

This paper defines a generic model for a *cognitive agent* and proposes a computational architecture, coherent with this model, that integrates all the approaches mentioned above. The model is based on the following basic hypothesis: (i) Cognition is an emergent property of a cyclic dynamic self-organizing process [Morin, 1991] based on the interaction of a large number of functionally independent units of a few types [Changeux, 1983]. (ii) Any model of the cognitive activity should be epistemologically compatible with the Theory of Evolution. That applies not only to the "hardware" components of this activity but also to its "psychological" aspects [Wright, 1994]. (iii) Learning and cognitive activity are closely related and, therefore, the cognitive modeling process should strongly depend on the cognitive agent's particular history [Piaget, 1963].

The paper is organized as follows. In Section 2, we introduce the generic model for a cognitive agent. In Section 3, we present a computational architecture coherent with the proposed model. In Sections 4, 5 and 6, we discuss some details of the proposed architecture that are relevant to the three adopted hypothesis. Finally, in Section 7, we summarize our proposal.

2 Generic Model

The proposed generic model for a cognitive agent presents three levels: *reactive, instinctive* and *cognitive.* Functionally, these three levels are similar to the *reactive, deliberative* and *meta-management* components of Sloman's architectures for human-like agents [Sloman, 1996]. The model also presents some similarities with the Rasmussens's models [Rasmussen, 1991]. The reactive level consists of an evolutionary environment, whose elements are *patterns,* extracted from perceptive information about some external world, *effector controls* that can produce some action in the same external world and a population of *reactive agents* that tie together perception and action. This evolutionary environment is submitted to a *Natural Selection* process where the *fitness function* is associated with the *emotions* of the agent, defined as a global response reflecting the agent's present state [Kitano, 1995]. The environment complexity and the agent variety are not bounded, phenomena such as co-evolution, arms race, parasitism, symbiosis, etc are expected to occur. In particular, the agents can organize themselves in co-evolutionary groups analogous to the *agencies* proposed by Minsky in his *Society of Minds*

*The author is partially supported by *Conselho National de Desenvolvimento Cientifico e Tecnoldgico (CNPq),* grant No. 352017/92-9.

[Minsky, 1986]. At this level, processing is totally parallel and is characterized by a rapid perception/action cycle. At the end of each cycle, the best agents in the community, according to the fitness function, are allowed to *act,* i.e., to emit control commands to the body. This first level is intended to model simple animals, such as insects. The reactive level definition has many points in common with the *Enactive* theory proposed by Varela et al. [Varela *et al,* 1991].

The instinctive level introduces a *long term memory* into the model. As the evolutionary process at the reactive level proceeds and the situations repeat themselves in the world, it is possible to identify the populations of agents in the environment responsible for a useful action in a given situation. If we take, from these populations, the best and the worst agents, according to the fitness function, it is possible to abstract their properties and to obtain a *general description* of a given population, a kind of "genetic reserve" indexed by situation. We claim that long term memory is composed essentially by these descriptions and that the act of "remembering" corresponds to the introduction at the reactive level of *new populations,* whose agents are *genetically encoded* according to these general descriptions. Once there, these new populations are, up to a certain limit, able to recreate their original enviroment, from which they were generalized. The long time effect of memory in such a model is analogous to the "breeding" and "taming" of reactive agent populations. At this level, processing is less massively parallel and is characterized by a much longer cycle that needs many repetitions of the same situation to be completed. The two lower levels together are intended to model more complex animals, such as mammals.

Finally, the cognitive level is concerned with the manipulation of the general descriptions generated at the instinctive level. Its functions are the usual cognitive functions: *deduction, abduction* and *induction.* The cognitive level is based on two complementary activities: the learning of descriptions of relevant situations - a synthetic, holystic activity - and the generation of new strategies of action - an analytical, local activity. The *contents* of the cognitive activity, i.e., the contents of the *short term memory,* are defined to be exactly the relations between these two complementary activities.

Although the cognitive level could be based on any adequate symbolic logic formalism, it is "embodied" in a strict sense: on the one hand, its symbolic expressions refers to "lived" situation descriptions, giving it a "real" semantic. Moreover, these descriptions can be used to generate specific behavior patterns in a real world environment or in some "hypothetical" environment, simulated through the *sentiments* of the agent, defined as emotions induced through memories generated by cognitive activities. On the other hand, the results of its inference mechanisms are intended to be useful as the seeds for new useful behaviors and not simply true theorems. These two interpretations of the symbolic expressions in the cognitive level - (i) representations of the properties of the instintive level descriptions and (ii) seeds for action skills - axe similar to the *representational* and *functional* roles of Smith's *impressions* [Smith, 1987].

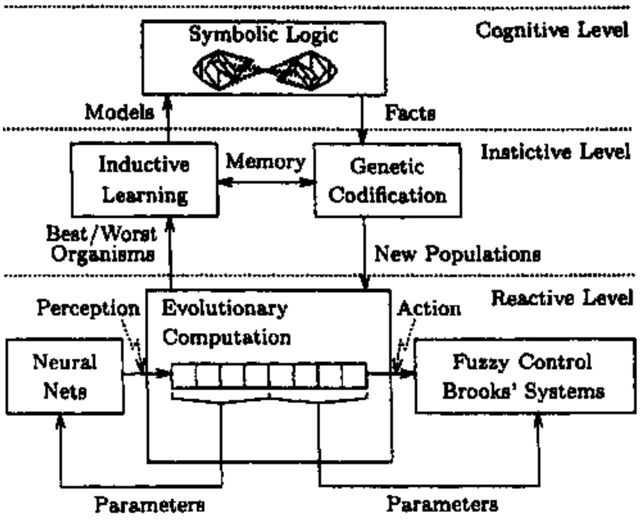

Figure 1: The Architecture

It is interesting to note that the instintive and cognitive levels together can be thought of as a more complex version of a reactive level agent, also functionally conecting perception and action. But in this case, the agent refers to an internal world: the reactive level environment. We claim that a necessary condition for the development of cognition is that this high level agent (which in some sense is the only one that we can really call "cognitive agent") be itself "embodied" in a community of similar cognitive agents. Moreover, we propose that the structure of long term memory reflects a *reconstruction,* at the reactive level, of the relevant structures of the environment in which the cognitive agent is itself embodied, particularly its socio/ structures [Neumann, 1968]. The agents in this cognitive environment would have two communication channels: (i) *experience,* they (partially) agree about the effects of actions in the external world and about the contents of emotions and sentiments, i.e., they believe they have similar reactive levels; and (ii) *language,* they can share, through the representations at the cognitive level, knowledge about experience, specially that knowledge that would give rise to useful skills at the reactive level.

3 Architecture

The instantiation of the above model for the human cognition would rise several questions about its coherence with respect to the known facts about human evolution, physiology and psychology. We avoid these questions and, heeding Sloman's advice [Sloman, 1995], try to explore the *design space* through a computational architecture for cognitive agents based on that model. The proposed architecture, presented in figure 1, has the same three levels and is intended to be a minimal implementation of the generic model.

The central activity of the machine is performed by the *Evolutionary Computation* module at the reactive level. It computes some kind of genetic algorithm where the genotype of the population members encode control parameters of the perception and action mechanisms, represented in figure 1 as the *Neural Nets* and *Fuzzy Control/Brooks' Systems* modules, respectively. The function of the *Neural Nets* module is to account for perception and the *Fuzzy Control/Brooks' Systems* module represents the effectors of the agent. Both these technologies have interesting properties for our purposes: neural nets are able to learn how to extract patterns from raw information, even without an external supervisor [Kohonen, 1987], reducing the complexity of the situation identification process. Moreover, they can be easily coded into a genotype structure through the weights associated with the network links. Fuzzy control and Brooks' systems are able to control quite complex behaviors in a distributed way and can also be easily coded, because their mechanisms depend on a limited number of parameters.

The agents in the reactive level environment can be seen as "atomic potential actions" or "skill fragments" joining together the perception and action mechanisms and the actual functional capability for an effective guided action in the external world. The genotype of each of these agents encodes three types of information: (i) the weights to be used in the input neural nets, (ii) the parameters that define the action mechanism, e.g., the fuzzy set definitions in the case of a fuzzy control mechanism, and (iii) a functional definition of the connection between the neural net output and the effector control input. Such an agent, when it is "plugged" into the appropriate perception/action circuit, uses its encoded weights and parameters to tune the perception/action mechanisms according to its necessity, and applies its encoded function to drive the relevant parts of the external effectors through the desired action according to the present perceptual input.

The model supposes a fitness function that attributes a grade to each reactive level agent and can be associated with some measure of the internal state of the machine. The real *actions* occur at the reactive level: after the genetic algorithm is running for some generations since the last action, the best agents according to the fitness function are activated. There are no central control and there are lots of things happening at the same time. Each action changes the internal state of the agent and the external environment, as perceived through the neural net module, and the cycle restarts. The model does not represent an "information processing machine", the perceptual input represents a restriction on the generated actions but does not determine them. Furthermore, the machine may continue its activities independent of input and output, it is enough to *simulate* a previously learned fitness function to create a hypothetical situation. In fact, we claim that much of the machine activity would be performed during the periods of inactivity, in which it "sleeps" and "dreams".

To learn a fitness function, a memory mechanism is necessary. If we observe the "arena" at the reactive level and, as a situation repeats itself, collect the best and the worst agents in handling such situation, then the genetic material in these agents can be used as examples and counterexamples in an *Inductive Learning* mechanism (e.g., [Salzberg, 1991]). The results of this inductive mechanism are general descriptions of populations, that can be associated with appropriate fitness function values. The complexity of these descriptions is proportional to the generality of the inductive learning mechanism and its underlying representation language. If this language is general enough to be interpreted as a logical description of the world situations, then we can use it to give a semantic account to a symbolic logic module, where the learned descriptions are interpreted as *models* of the external world.

Finally, the cognitive level is defined as a *theorem proving mechanism,* although it presents some unusual features in such systems: its expressions refer to well defined structures, the general descriptions of the instinctive level, and its inference results can be used to genetically encode new populations at the reactive level, and therefore, their value as representations of the world can be tested in "real action" conditions. In our minimal model, the genetic codification module corresponds to a fuzzy control system generator, but in a more general specification it could be any control system generator, reintroducing the *Cybernetics* view that *Control Theory* is an important part of cognitive activity. Coherently with the generic model, we suppose that the machine would be "embodied" in a world where human beings and cognitive machines are joined in communities.

4 Self-Organizing Logic

The connection between a symbolic theorem proving mechanism and the elements in the instinctive level of the proposed architecture deserves a more detailed treatment. The cognitive level activity could be described as follows. Initially, the learning mechanism, at the instinctive level, selects the best reactive level agents. The genetic material of these agents encode situation descriptions and adequate actions to be performed in these situations and it is used to derive a *logical description* of the relevant behaviors in these situations. The process that generates these descriptions is called *conceptualization.* At the cognitive level, the logical descriptions are interpreted as *possible models* of the world. The main activity of the cognitive level is to assume that the learned models are *all* the possible models for these situations and to derive an associated *factual description* that has exactly those models. This mechanism is very similar to McCarthy's *Circumscription* [McCarthy, 1980].

The usual theorem proving methods - *Resolution* [Robinson, 1965] and *Semantic Tableaux* [Smullyan, 1968] - present some properties that make them not adequate for our purposes: (i) they present external, i.e., meta-logical, inference rules, (ii) they are not designed to generate a factual description from a set of possible model descriptions, only to the reverse, and (iii) their proof methods are local mechanisms. To avoid these problems, we propose a new inference method, based on the transformation between conjunctive and disjunctive canonical normal forms, the *dual transformation,* that involves only internal properties of the underlying logical system (i.e., there is no inference rule external to the logical language) and that has a global and concurrent nature [Bittencourt, 1997].

Given a logical formula W, we call a *theory* the two sets Φ and Ψ that contain, respectively, the clauses and dual clauses associated with the canonical normal forms of *W.* The two sets Φ and Ψ are a kind of "holographic" representation of each other. Each clause in Φ consists of a combination of all dual clauses in Ψ and, conversely, each dual clause in Ψ consists of a combination of all clauses in Φ**.**

Intuitively, the proposed inference algorithm consists in the elimination of the contradictory dual clauses, the combination of the associated substitution fragments into a set of independent substitutions, the application of these substitutions to the dual clause set and, finally, the generation, through the dual transformation, of the clause sets associated with these instances of the dual clause set. The clauses in these instances are new theorems which can be added to the original clause set and the cycle may be repeated. Clearly this mechanism does not present external inference rules, all that is done is to exploit the duality of the canonical representations and their semantics.

The inference method is totally symmetric. It does not matter if we begin with clauses, i.e., a factual account of a situation, or dual clauses, i.e., a description of the possible models of the situation. Analogously to the elimination of contradictory dual clauses, in the clause form it is the tautologic clauses that can be eliminated. Therefore, the repeated transformation between canonical representations is not only able to infer new theorems but it is also able to *refine* the theory.

These theoretic properties of the inference method are enough to eliminate the two first restrictions to theorem proving methods. The third one must necessarily take into account implementation concerns. Clearly, the most expensive operation of the proposed inference method is the dual transformation. To improve its efficiency, we developed a concurrent algorithm for the dual transformation, based on a geometrical representation, that, on the one hand, generates only dual clauses that are not subsumed by any other and, on the other hand, is naturally concurrent and therefore easy to parallelize.

According to this algorithm, a theory is represented through a n dimensional *hypercube,* where n is the number of predicate symbols that occur in the theory. Each vertex of this hypercube can be associated with a certain combination of predicates given by those coordinates of the vertex that are not zero. A (dual) clause is associated with the vertex labeled exactly with those predicates that appear in its literals. The literals in each dual clause can be thought to *represent* some clauses where they appear or where there are literals subsumed by them. This information is stored in the geometric representation as integer sets associated with each literal that indicate in which (dual) clauses the literal is present. The result of this construction is that the dual transformation and the manipulations necessary to the inference/refinement mechanism can be done through local communication between neighbor (dual) clauses in the hypercube. The holographic properties of the canonical representation allow for a global effect in one form to be calculated locally in the other form [Bittencourt, 1996].

This representation allows the integration of the instinctive and cognitive levels: the learning mechanism could be thought of as concurrently 'feeding" each separate node of the hypercube with the appropriate dual clause representation of some situation. After some refinement/inference steps the new inferred model descriptions can be directly verified through the learning mechanism. Furthermore, the clause representation can be used to generate new reactive level populations, whose action effects could again be verified through the learning mechanism. It is interesting to note that this kind of integration does not eliminate the *robustness* of the reactive level environment. In the model, the cognitive activity can *influence* the global behavior, but its scope is limited.

Another interesting property of the representation is that it easily supports a special kind of higher-order logic, where the theories, with their predicate and function symbols as parameters, are considered as new atomic formulas. This property has much in common with Peirce's *semiosis* process [Peirce, 1974]. According to Peirce, a *sign (S)* represents an *object (0)* for another sign, the *interpreter (J).* This last sign - called an *habit* or *mental law* - may determine another interpreter I_1, which may determine an I_2, etc. The actualization of the potentially infinite sequence $I, I_1, I_2 \cdots$ is called semiosis. This triadic relation can be found in all three levels of the model. In particular, at the cognitive level, the memorized descriptions are the signs that represent possible models to theories and theories are signs that represent new populations to the instinctive level. Moreover, theories as signs may represent possible (hypothetical) models for other theories, mediated by a meta-conceptuaiization process.

5 Evolution

In the proposed model, we are double committed with the evolutionary approach: on the one hand, the most important component of the model is in itself an evolutionary mechanism - the reactive level - and, on the other hand, the model is designed in such a way that the components at each level could have evolved one after the other, because both, the reactive level and the pair reactive/instinctive levels, are complete models of simpler agents.

The choice of an evolutionary mechanism as the base of the model is inspired by the fact that the function of the central nervous system is to integrate the most different perception and action mechanisms. This integration could be done through a genetically "wired" functional mechanism, but none would be flexible enough to adapt itself to a variable environment. A more robust solution is to use a *simulation* of the relevant characteristics of the external world, under the same natural selection principles, to make more secure decisions about future actions. If the real-time simulation is reliable enough, even without a memory, a reactive level agent would be able to "learn" how to act in a given situation, i.e., the adequate reactive level agents will be (slowly) selected through interactions with the real world. The problem is that, without a memory, the range of experiences is limited to the actually lived situations and that implies that certain situations, e.g., fatal ones, cannot be simulated.

In this sense, long term memory is an *improvement* to a previously existing mechanism. An improvement which allows *off-line* simulations, because memory allows the simulation of the internal fitness function associated with *hypothetical* situations. These hypothetical simular tions greatly accelerate the evolutionary process at the reactive level. The cognitive level is another improvement of the mechanism, because it allows the *creation,* through inference, of models of situations that never occurred in the agents' experience, further accelerating the evolutionary process at the reactive level. But its full strength comes from the fact that it also allows the establishment of a second communication channel between agent and world: language. Language may be used in the description of *abstract* situations, in particular, it allows one to conceptualize social situations and to reconstruct the relevant social relations into a coherent internal theory that can be suitably simulated at the reactive level.

6 Learning

One of the central issues of the model is the fitness function to be provided and its relation with the meaning of the words "emotions" and "sentiments", loosely introduced in the model description. More formally, we call emotions an internally generated feedback mechanism, based on perception and memory, able to *guide* the evolution of the reactive level environment in such a way that the global agent *survives.* Emotions are the keys for long term memory, only emotionally important learned descriptions are memorized and the same emotions are able to fetch them from memory.

According to the model, long term memory existed prior to symbolic cognitive activity and it was in some sense *adapted* to cognitive purposes. As cognitive products, models and factual descriptions, do not have in principle any emotional contents, to memorize them it is necessary to artificially create these emotional contents. These cognitively generated emotions we call sentiments. Because sentiments are associated with cognitive states and therefore to language constructions, they are fundamental for the appropriate learning of social behaviors, mainly transmitted through language. The coherence between sentiments and emotions is the base for the cognitive development. If the descriptions we learn lead in fact, at the reactive level, to the emotional effects we were told they should, we can have the "sentiment" that we learned an useful thing.

Although the theories at the cognitive level are strictly grounded in the learned descriptions, the real "meaning" of these descriptions is different for each cognitive agent, because it is defined as the effect of these descriptions when their associated populations are introduced in the reactive level, and this effect is determined by the particular history of the reactive level environment, i.e., by the history of emotions and sentiments that guided its development.

7 Conclusion

In their book *The Embodied Mind,* Varela et al. [Varela *et al.,* 1991] propose three questions about cognition and answer them from three points of view: the cognitivist research program, the emergence (connectionist) program and their own *Enaction* theory. To summarize our proposal, we answer the same three questions from the point of view of the proposed model:

- *What is cognition ?* The articulation of two parallel evolutionary processes: one that occurs in an internal environment where a history of structural coupling brings forth a (physical) world, i.e., the reactive level, and, another that occurs in an external sentimental/social/cultural environment where a history of symbolic exchanges, mainly through language, between cognitive agents brings forth, in each one, a (sentimental) world, i.e., the cognitive level. The integration between the environments is mediated by the long term memory.
- *How does it work ?* The reactive level is based on an evolutionary process guided by the emotions and the cognitive level is based on a cyclic process that transforms factual descriptions into models and vice versa. The models are interpreted as memorized general descriptions of situations and the factual descriptions are used to generate new reactive level populations. Cognitive theories can be communicated through language. The cognitive level can also generate sentiments, i.e., artificial emotions, that allow it to influence memory contents.
- *How do I know when a cognitive system is functioning adequately ?* When it becomes part of an ongoing existing physical and sentimental/social/cultural world.

Acknowledgments

The author is grateful to all those that helped in the development of this paper, in particular the LCMI team and the anonymous referees.

References

[Bittencourt, 1996] G. Bittencourt. Boxing theories (abstract). *Journal of the Interest Group in Pure and Applied Logics (IGPL),* 4(1):479-481, 1996.

[Bittencourt, 1997] G. Bittencourt. Concurrent inference through dual transformation. *Journal of the Interest Group in Pure and Applied Logics (IGPL), in press,* 1997.

[Changeux, 1983] J.-P. Changeux. *L'Homme Neuronal.* Collection Pluriel, Librairie Artheme Fayard, 1983.

[Fu, 1994] L. Fu. Rule generation from neural networks. *IEEE Transactions on Systems, Man and Cybernetics,* 24(8):1114-1124, August 1994.

[Goldberg, 1989] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley Publishing Company, Reading, MA, 1989.

[Kitano, 1995] H. Kitano. A model for hormonal modulation of learning. In *Proceedings of IJCAI 14,* 1995.

[Kohonen, 1987] T. Kohonen. *Content-Addressable Memories.* Springer-Verlag, Berlin, 1987,

[McCarthy, 1980] J. McCarthy. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence,* 13(l,2):27-39,1980.

[Minsky, 1986] M.L. Minsky. *The Society of Mind.* Simon and Schuster, New York, 1986.

[Morin, 1991] E. Morin. *La M6thode 4, Les Idees.* Editions du Seuil, Paris, 1991.

[Neumann, 1968] E. Neumann. *Ursprungsgeschichte des Bewusstseins.* Kindler Verlag GmbH, Munchen, 1968.

[Newell, 1980] A. Newell. Physical symbol systems. *Cognitive Science,* 4:135-183, 1980.

[Peirce, 1974] C.S. Peirce. *The Collected Papers of CS. Peirce.* Harvard University Press, Cambridge, Mass., 1974.

[Piaget, 1963] J. Piaget. *The Origins of Intelligence in Children.* Norton, New York, 1963.

[Rasmussen, 1991] Steen Rasmussen. Aspects of information, life, reality, and physics. In C. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, editors, *Artificial Life II.* Addison-Wesley, 1991.

[Robinson, 1965] J.A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM,* 12(1):23-41, January 1965.

[Rumelhart and McClelland, 1986] D.E. Rumelhart and J. McClelland, editors. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1: Foundations,* volume 1. M.I.T. Press, Cambridge, MA, 1986.

[Salzberg, 1991] S.L. Salzberg. A nearest hyperrectangle learning method. In *Machine Intelligence 6,* pages 251-276. 1991.

[Sloman, 1995] A. Sloman. A philosophical encounter. In *Proceedings of IJCAI 14,* 1995.

[Sloman, 1996] A. Sloman. What sort of architecture is required for a human-like agent ? In *Proceedings of AAAI-96,* August 1996.

[Smith, 1987] B.C. Smith. The correspondence continuum. *Center for the Study of Language and Information,* January 1987. Report No. CSLI 87-71.

[Smullyan, 1968] R.M. Smullyan. *First Order Logic.* Springer-Verlag, 1968.

[Varela *et al.,* 1991] F.J. Varela, E. Thompson, and E. Rosch. *The Embodied Mind: Cognitive Science and Human Experience.* MIT Press, Cambridge, MA, 1991.

[Wright, 1994] R. Wright. *The Moral Animal.* Vintage Books, New York, 1994.