International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:11, No:1, 2017

Towards a Computational Model of Consciousness: Global Abstraction Workspace

Halim Djerroud, Arab Ali Cherif

***Abstract*—**We assume that conscious functions are implemented automatically. In other words that consciousness as well as the non-consciousness aspect of human thought, planning and perception, are produced by biologically adaptive algorithms. We propose that the mechanisms of consciousness can be produced using similar adaptive algorithms to those executed by the mechanism. In this paper, we present a computational model of consciousness, the "Global Abstraction Workspace" which is an internal environmental modelling perceived as a multi-agent system. This system is able to evolve and generate new data and processes as well as actions in the environment.

***Keywords*—**Artificial consciousness, cognitive architecture, global abstraction workspace, mutli-agents system.

I. Introduction

THE interest in proposing a computational model to access consciousness is multiple; it permits the validation of theories of consciousness by providing an implementation of the latter and, following on from this, by comparing experimental data with actual data. The first problem that jumps out when you try to create a model for consciousness comes from the very word conscious, and its use in describing different situations such as the act of not being asleep or the loss of unconsciousness, as a state that can be modified by taking drugs, of paying attention to a particular stimulus, self-awareness [6] or just as much, moral conscience.

For the philosopher Ned Block [4], conscious phenomena would include four principle aspects: (1) Access Consciousness [8], (2) Phenomenal consciousness, (3) Reflective consciousness and (4) Self-awareness.

In this articli, we will focus on the computational modelling of access or representational consciousness (referred to as consciousness A) [4], this governs intentional properties (attention, propositional attitude, reasoning, intent and control of actions). Access consciousness is the consciousness that permits us to act rationally. During the course of this article we begin by introducing the main works in this field and will briefly describe each of the relevant models. We will focus on a particular model, the "global worksapce" model, which translates into English as the "global workspace" model and which forms the basis of our work. In the aftermath we will present the advantages and disadvantages of this model. Following on from this we will propose our own model which draws heavily on the "global workspace". Finally we will present some comparisons with other models and finish with our conclusions.

Halim Djerroud and Arab Ali Cherif are with the LIASD Laboratoire d'Informatique Avancée de Saint-Denis (Saint-Denis Advanced Computer Science Laboratory) Université Paris 8. 2 Rue de la Liberté, 93526 Saint-Denis, France (e-mail: hdd@ai.univ-paris8.fr, aa@ai.univ-paris8.fr).

II. State of the Art

In recent years, several researchers have tried to model consciousness. As is clear from the description of their models these are all grounded in almost the same definition, we can even say that there is an emerging convergence towards a "standard model" of consciousness, at least for the set of ideas proposed since 1950-1960 (Stanislas Dehaen).

- A system of centralised supervision
- Limited capacity
- Necessitation for slow loops, re-entry and descending from the "top-down" as opposed to the process rising from the "bottom-up" rapidly and non-consciously (Edelman).
- An internal space of synthesis, retention and data sharing: "theatre[1]", "the blackboard", "the global workspace"

A. The Sketchboard

For Gérard Sabah consciousness is modelled as a controlled process that manages multiple sub-processes, the data from which is stored in a blackboard. The sketchboard is an extension of the blackboard for the automatic establishment of feedback loops from higher levels towards the lower levels. The modules construct on the one hand their own result (a sketch), and on the other hand, refer to the modules which they use the result is a response that reflects their satisfaction to what they have built. This leads the first modules to modify their sketches in order to optimise the response. These relationships are widespread across all modules used in problem solving in order to enable the construction of increasingly accurate sketches, using knowledge from the whole of the system involved. This model was implemented in Smalltalk under the name of CARAMEL (in french: Compréhension Automatique de Récits, Apprentissage et Modélisation des Échanges Langagiers) [9]-[11].

B. The CogAff Project

The Cognition and Affect Project (CogAff) focuses on emotions, with the intention of modelling and explaining them

[1]The Theatre of consciousness metaphor was proposed by Taine in 1870: "We can compare the mind of a man to a theatre of undefined depth, in which the footlights are very narrow, but with a scene opening up from the footlights. Right in front of these bright footlights, there is hardly room for one actor... but beyond them, on the various planes of the scene, there are other groups that are rather less distinct, that are further away from the footlights.".

in relation to living beings and providing artificial systems with the means of understanding and reproducing emotional behaviours [14]. The project seeks to define how to develop a comprehensive architecture of intelligent agents, whether artificial or natural [13].

The architecture most successful in this project, labelled H-CogAff, consists of three levels:

- The reagent level that instinctively reacts to modifications in the environment by actions exerting in their turn an influence on the environment
- The deliberative level that overlaps with the previous one and which deals with planning, decision-making, etc. that is to say the selection of action
- The meta-management level that permits reflection on ones own behaviour, to evaluate or even modify it. We have reproduced in Fig. 1 the graphic representation as proposed by [14].

C. The Global Workspace

In this section, we will present Baars' consciousness model which he entitled the "the global workspace" [3]. Our work is primarily based on this model, for this reason we will study its operation and will review the hypotheses that led Bernard Baars to develop the model. Prior to the presentation of the model, Baars first presented the hypotheses on which the theory rests, thus providing clarification of the model he made for consciousness.

1) The Activation Hypothesis: (An element is conscious when its activation exceeds a certain threshold) Consciousness involves a form of activation of psychological factors so that they access the memory, either by their intensity or by certain associations. An activation threshold exists beyond which the elements become aware.
2) The Novelty Hypothesis: Consciousness tends to focus on new facts, it considers that one only becomes conscious of elements that bring new information, however, an experience cannot be totally new otherwise it would not be interpretable, it is necessary therefore to have a partial correspondence with past experiences.
3) The tip of the iceberg hypothesis: That which is conscious is that which emerges from a set of unconscious experiences. It also implies that consciousness is very limited and that if the tip of the iceberg is consciousness, the rest is unconscious.
4) The Theatre Hypothesis [1]: Assumes that there exists a place in the brain where information is collected to be rendered conscious. Consciousness is seen here as the place of presentation of the results produced by the treatment of our senses (the Cartesian theatre).

Based on these assumptions, we can sketch consciousness such as it is represented by Baars's theory: Consciousness is a space to which access of the components follow from an activation resulting from new, dissonant or unusual information. These elements are related to other elements that remain unconscious. In Baars's theory consciousness is considered as being a distributed cognitive system in which the modules specialise in their own skills. These modules can work co-operatively (in series) or in competition (in parallel) in order to access a memory zone called the "global Workspace" whose content is broadcast to the entire system. Baars considers that consciousness involves such a space that enables various specialised modules to interact through the exchange of centralised information. This system only shares the information present in the global workspace, in some ways this system imposes a minimum of one bottleneck (Fig. 2), thus forcing the individual modules either to cooperate or to compete (collaborative contexts). Contexts are processors or the coalition of processors fixed to the system. They are able to assimilate reflexes, automation and intuitions. They thus restrict the "framework" of processors, limiting recourse to consciousness. Baars considers that the effective realisation of predictions about our feelings is what gives stability to the outside world, which allows it to define an internal context first: A system that shapes our conscious experiences without being itself even conscious at the time.

Different types of contexts are considered: The context of perception (in certain ways the presuppositions that make a perception ambiguous are interpreted preferentially), the context of conceptual thinking (the assumptions related to our beliefs that are considered certainties [alternative solutions are forgotten]), the context of intentions (the hierarchy of objectives inherent for the individual or specific to the current location), and the context of communication (what is believed to be common at the moment of communication with another). These different contexts interact. A context is considered from two perspectives: either as a goal to achieve, which guides the interpretation of perceptions, or as an association of specialised processors that dominate the global workspace, contexts can cooperate (interlink) or be competition (parallel).

Fig. 2 shows the three essential components of this model:

- The contexts representing the general framework in which the objects are situated, as expectations, intentions, perceptions of the environment, etc.
- Consciousness, or the global work space, in which concepts such as attention, working memory, process control, etc. take place,
- The specialised unconscious processors that represent all of the functions of the nervous system, such as long-term memory, the reflexes, physical and intellectual faculties, parallel distributed processors etc. A specialised unconscious processor receiving information omitted by consciousness can for example react by trying in its turn to access consciousness or to rejoin again with it or to form a coalition, represented by a feedback loop in the diagram above. This feedback loop may also have a purely informative nature, to inform the conscious processes of their own finality.

In the model proposed by Baars, learning corresponds to the mechanism that allows the creation of contexts (for new information) and then the passage to conscious information and finally the detection of redundancy (the cycle of adaptation); learning is therefore a reduction of possibilities within a domain (items processed unconsciously become

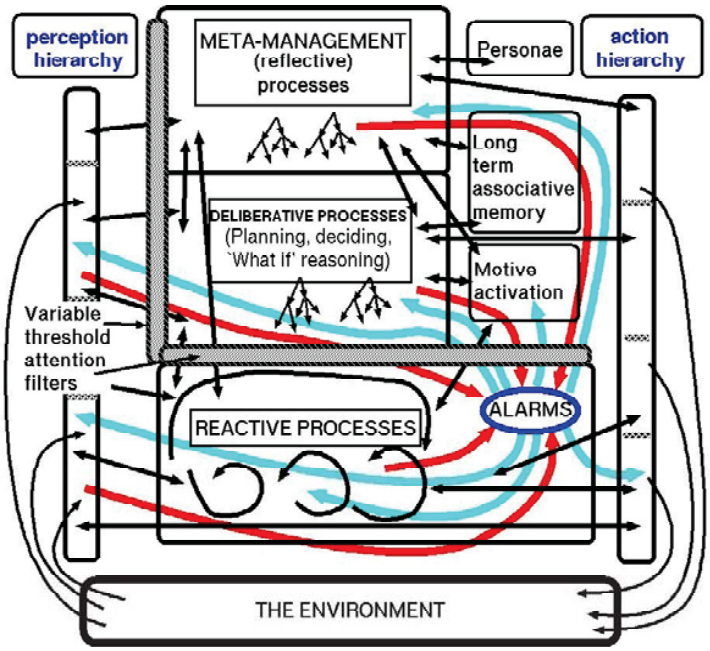

Fig. 1 The H-CogAff architecture as shown in [14]

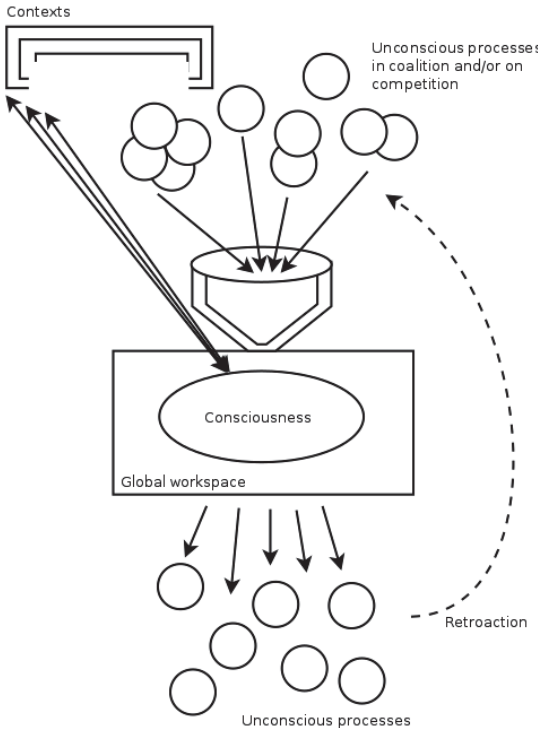

Fig. 2 The global workspace

completely predictable – the loss of consciousness is a sign of successful learning). This phenomenon exists at the perceptual, conceptual and intentional level but also at unconscious levels. In other words, it adapts to the information on a multitude of levels (and this adjustment corresponds to a reduction in the amount of input information); symmetrically, equally we actively search (by raising the level of consciousness) for information on very different levels (to increase the amount of input information). Inspired by the Global Workspace, Stan Franklin and his team (CCRG - Cognitive Computing Research Group) provide a framework which implements Bernard Baars' theory within a cognitive architecture called LIDA (Learning Intelligent Distribution Agent) [2]. In Fig. 3, we have reproduced the graphic representation as proposed by [2].

The cognitive cycle that LIDA proposes is divided into three phases, comprehension, attention (conscious) and action and learning selection. These phases repeat themselves indefinitely. The understanding phase is based on the CopyCat system [12] in order to create some association with the objects of perception with objects that were previously perceived in memory. The second phase permits the selection of the most visible or urgent? processes, in order to disseminate their information (result) to the other processes. The last phase permits the selection of the effectual action from amongst a set of choices found in the procedural memory.

III. Global Abstraction Workspace

The model that Baars proposes is a computational model. That is to say it is possible to make a computer implementation which presents no algorithmic limitations. But the mechanism of process selection that will diffuse the information, such as (to access consciousness) the results deem relevant to the context, and by the effect of general diffusion, entitles that which is conscious to influence whatever else (isotropy — the relevance is then seen as a rather "magic" phenomenon comparable to the not always well justified heuristics of artificial intelligence). From this point of view, the Baars'

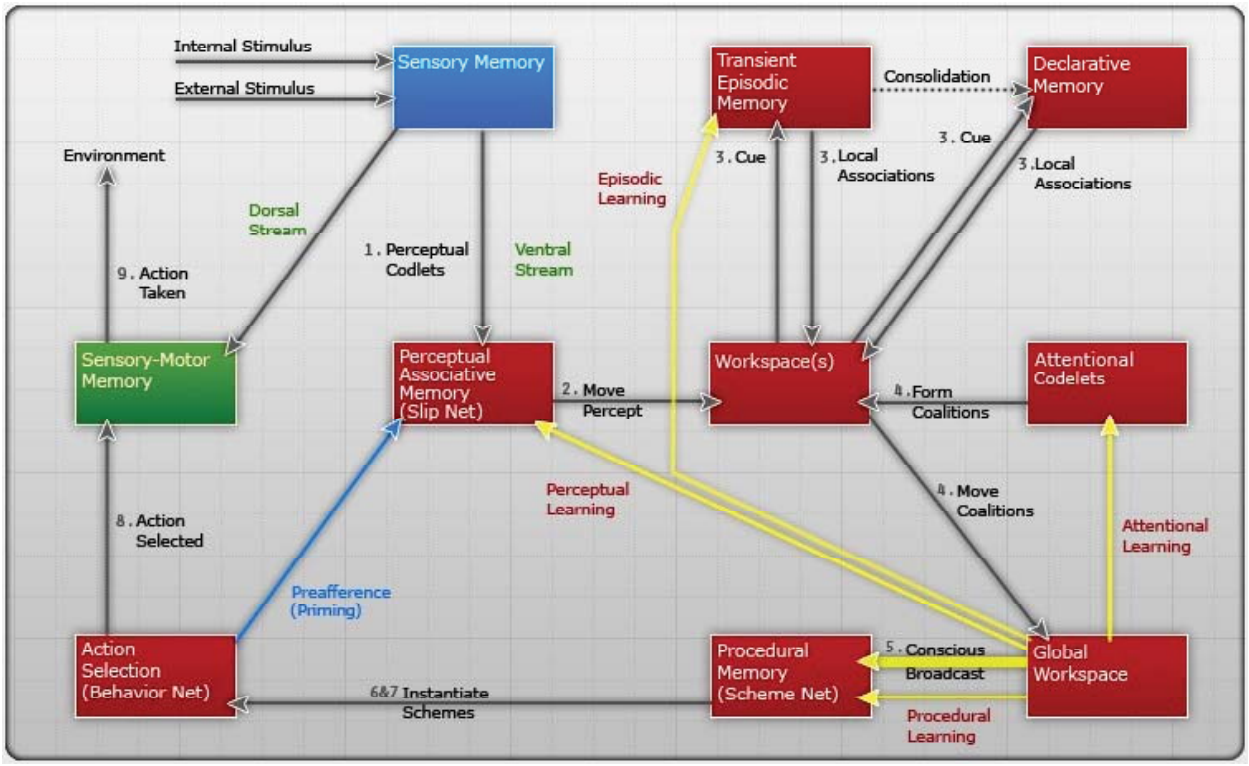

Fig. 3 LIDA in [2]

model has been heavily criticised by the philosopher Daniel Dennett [8].

The second practical problem posed by Baars' model is the energy consumption (high energy), where this model mobilises all the unconscious processes to finally select one and reject all the other results. Except, as we are well aware, nature favours a principle of economy (combinatory explosion). In the rest of this section we will outline our model we call the global abstraction workspace which draws heavily on ideas strongly inspired by Baars and which respond well to the definition of consciousness proposed by Baars (voir II-C).

A. Hypothesis

We assume that consciousness corresponds to a mental image (memory space) that a mental process updates throughout its life cycle. This mental image corresponds to the different knowledge that the subject has regarding the problem that is in the course of treatment (context). We call this memory space "the global abstraction workspace".

Our model of consciousness is seen as a cognitive architecture if we assume its existence as a hypothesis:

1) **Global Abstraction Workspace:** which is a volatile memory that reflects the state of the environment observed at any given moment to add to the data issued from the memory (we will see later in this article how data is selected). The space will serve as a mental image in which several processes come to perform experiments in order to make predictions, extract information and produce contexts.
2) **Automatic Processes:** There are automatic and unconscious processes that act on the information within the global abstraction workspace. These processes produce information which is itself registered in this global abstraction workspace. These processes are selected (the selection mechanism will be explained in the next section) by the background process. These processes can work either in coalition (standard) or in competition (in parallel).
3) **Working Process:** A number of processes permanently reside in the global abstraction workspace; their role is to regulate the space.

B. Functioning

In our architecture we propose that human cognition consists of cascading cycles of recurring events in the brain. Each cognitive cycle detects the current situation; it interprets it with reference to most relevant objectives then selects an internal or external action in response (Fig. 4). The cognitive cycle comprises of the following elements:

1) Perception: The perception phase strives to accomplish two functions, the first is responsible for finding representations in memory in the perceived environment, then subsequently to agentify the entity perceived in the global abstraction workspace, for example this phase could be implemented in relation to the *copycat model* by [12]. The second function will enable the location of the skill or skills (the process or processes) which are apt to process the data

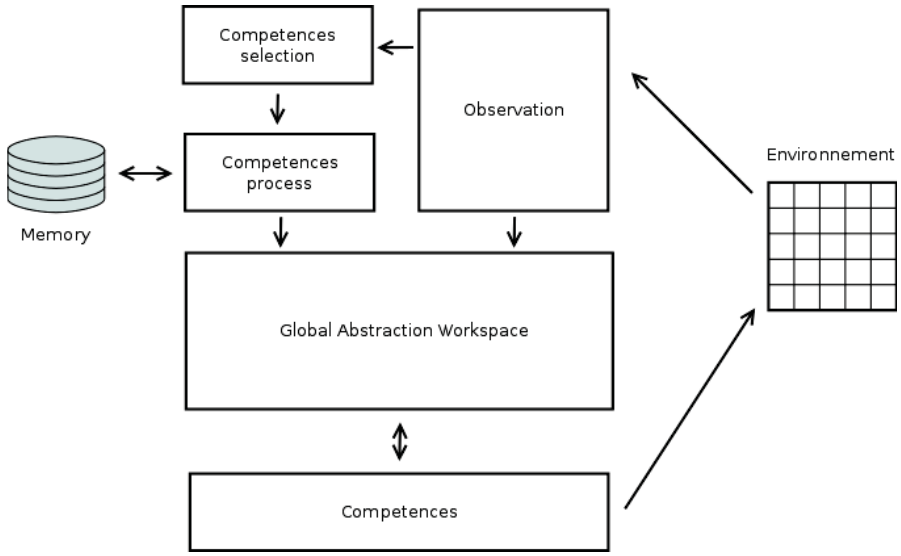

Fig. 4 Gloabal Abstraction Workspace

collected in this process, this phase could well be carried out by the *Contract Net* [15].

2) Competences: A set of dormant agents with well defined skills, which can be elected to act on the global abstraction workspace. The fundamental hypothesis of our model is that consciousness is an automatic phenomenon realised by a set of agents with diverse skills, these agents are activated only if their competences match the existing data in the global abstraction workspace.

3) Consciousness and Global Abstraction Workspace: The global abstraction workspace is an internal model of the environment perceived as a multi-agent system, this system is capable of evolving and generating new data or new processes.

Fig. 4 gives an overview of the system and how the modules are interconnected.

C. Contexts and Their Emergence

The contexts determine the scope/framework within which the system is located, and directly influences the behaviour of the unconscious processors. Contexts are unconscious, the effective implementation of the predictions generated by the system reinforces the persistence of the context, and increases the life of the agents elected in the global abstraction workspace and produces a new specialised process.

How do we decide the value of information? The essential mechanism is the feedback produced by specialised processes that indicate their value, more or less equivalent to their adaptation to the global conscious message. By a stack mechanism, a conscious event that becomes redundant gives way (in consciousness) to the next most informative message (conscious elements are treated as objects, non-conscious elements as contexts).

In this context, the major function of consciousness is then revealed, to modify or create new contexts which in their turn will influence conscious experiences in the future.

The contexts are processes or coalitions of processes fixed in the system. They can assimilate reflexes, automation and intuitions. They thus restrict the ”working framework” of the process.

A specialised unconscious processor receiving information issued by consciousness can for example react by trying in its turn to access the conscience or furthermore, to join or form a coalition, represented by its feedback mechanism. This feedback loop may also be purely informative, informing the conscious process on its own finality.

If in a context considered stable, the finality of the conscious process is still the same, that is to say that the choice of a solution from among a finite set of alternatives is still identical, then this process will become less and less conscious to form either a new context or a new specialised unconscious processor, depending on the scope of the process. It is the capacity for adoption even a form of learning. The formation of a context consists of automatising the formation of a coalition of agents, that is to say to fix a coalition to make a reflex or any other form of automatism.

IV. Discussion

When we observe nature and the way in which living organisms address the problems that they encounter, it becomes possible to contemplate the conception of generic machines able to face new situations and adapt. Furthermore, in terms of animal species, the difference between species is minimal. Take the hand or the heart for example, these organs are present in most animals and carry out the same functions; catching things in the case of the hand and pumping blood in the case of the heart. These form a kind of *design pattern*, a common generic solution for most species. It seems obvious that the brain, also present in most species, has the same purpose and the same content: to solve problems in a similar fashion. With this in mind, we will consider consciousness as a solution adopted by nature to run these generic machines.

The research we present in this text aims at producing an implementable model of consciousness and in proving its effectiveness through its implementation in robots.

In effect, we adhere to the model of consciousness provided by the field of psychology: [3], [6]-[8], and try to reproduce their conclusions regarding the function of consciousness within the domain of computing as a cognitive architecture.

In terms of the cognitive architecture that we are in the process of constructing now in the laboratory, we can assimilate it into a program that produces specific abstract machines for each problem posed in a given context.

The architecture - LIDA [16] - that we describe in (Section II-C) is a partial implementation of the Baars model [3] realised by [5]. We believe this implementation remains partial and has never completely described the model proposed by Baars, because of the interpretation of the psychological processes that Baars describes and the computer processes applied by [5] which are not suitable for describing these representations. To do this we have proposed a model close to that which Baars has proposed and conform to the definition (Section II-C), but we explain these phenomena using computer processes that are already well known.

In our next articles we will publish our results and compare the two architectures in order to remove any doubt regarding misinterpretations of the processes.

V. Conclusion

The global abstraction space suggests that consciousness enables the creation of an internal representation of a problem or else a modelling of the problem in the form of a multi-agent system which allows multiple processes to cooperate and participate in solving problems.

The conscious content is able to correspond to the organisation of agents (multi-agents system) that evolve and generate new information that most likely corresponds to the state of the environment in the near future (prediction).

References

[1] Bernard J Baars. *In the theater of consciousness: The workspace of the mind*. Oxford University Press, 1997.

[2] Bernard J Baars and Stan Franklin. Consciousness is computational: The lida model of global workspace theory. *International Journal of Machine Consciousness*, 1(01):23–32, 2009.

[3] BJ Baars. A cognitive theory of consciousness. cambridge universitypress.[anc, bjb, rnc](1993) how does a serial, integrated and very limited stream of consciousness emerge from a nervous system that is mostly unconscious, distributed, parallel and of enormous capacity? theoretical and experimental studies of consciousness. In *Ciba Foundation Symposium*, volume 174, page 282303, 1988.

[4] Ned Block. How many concepts of consciousness? *Behavioral and brain sciences*, 18(02):272–287, 1995.

[5] CCRG. Cognitive computing research group. http://ccrg.cs.memphis.edu/, 1994-2016.

[6] Antonio R Damasio. *Sentiment même de soi (Le): Corps, émotions, conscience*. Odile Jacob, 1999.

[7] Stanislas Dehaene. *Le code de la conscience*. Sciences. O. Jacob, Paris, 2014.

[8] Daniel C Dennett. *Consciousness explained*. Penguin UK, 1993.

[9] Sabah Gérard. *CARAMEL : A computational model of natural language understanding using a parallel implementation.* Actes ECAI, Stockholm, p. 563-565, 1993.

[10] Sabah Gérard. *CARAMEL : A flexible model for interaction between the cognitive processes underlying natural language understanding.* Actes Coling, Helsinki., 1993.

[11] Sabah Gérard. *CARAMEL : Un système multi-experts pour le traitement automatique des langues, Modèles linguistiques.* Fasc 1, p. 95-118., 1993.

[12] Douglas R Hofstadter, Melanie Mitchell, et al. The copycat project: A model of mental fluidity and analogy-making. *Advances in connectionist and neural computation theory*, 2(31-112):29–30, 1994.

[13] Ingo Lütkebohle. The cognition and affect project. http://www.cs.bham.ac.uk/~axs/cogaff.html/, 2004.

[14] Aaron Sloman. Varieties of affect and the cogaff architecture schema. In *Proceedings of the AISB'01 symposium on emotions, cognition, and affective computing. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour*, 2001.

[15] Reid G Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on computers*, (12):1104–1113, 1980.

[16] Javier Snaider, Ryan McCall, and Stan Franklin. The lida framework as a general tool for agi. In *Artificial General Intelligence*, pages 133–142. Springer, 2011.