«Consciousness» as the Foundation for Diagnosis in a Human-like Tutoring Agent

Daniel Dubois, Roger Nkambou, Patrick Hohmeyer
GDAC Laboratory, University of Quebec in Montreal
P.O. Box 8888, Centre-ville Station, Montreal, Quebec, Canada

Abstract

Interacting with humans requires subtlety if one is to be perceived as a great and pleasant tutor. «Consciousness» mechanisms can allow the tutoring agent to filter through the various sources of information and efficiently guide learners. We describe in the present paper an architecture implementing these processes in the specific context of astronauts training on Canadarm2.

Introduction

In a simulator, free exploration may suit some, but being guided by an experienced coach has been shown to help significantly, either in speed or in quality of learning. But coaching is challenging, and it gets no easier when attempted with an artificial agent. A virtual Simulator of the International Space Station and its Robotic Manipulator System, Canadarm2, has been developed in our lab [1]. Its reactive feedbacks are being augmented by those of our "conscious" cognitive tutoring agent.

What *Consciousness* Is About

Aside from more popular notions, Block's [2] clarifies that *consciousness* encompasses four internal mechanisms that allow us to represent and make the content of the present experience available to the rest of our internal, unconscious processes. The *access consciousness* grants access to unconscious, unreachable resources; *monitoring consciousness* keeps us informed about the activities of our senses; *self-consciousness* is about being aware of existing as an entity distinct from the rest of the world; *phenomenal consciousness* describes the qualities of the experiences. We attempt to reproduce the first three in our architecture.

Baars' Global Workspace Theory

In Baars' view [1], consciousness plays fundamental roles in nine functions, among which we find: Adaptation and learning, contextualizing, Prioritizing and access control, Recruitment, Decision-making and Self-monitoring. All these functions, all brain operations, are carried by a multitude of globally distributed, unconscious, specialized processors (which we implement computationally as *codelets*). Each has a limited ability, and a limited range of knowledge processing, but is very efficient. When one of them cannot complete its operation, it tries to make this situation published, in other words have the whole system become conscious of the situation. Processes that recognize the fact and know what to do about it, or how to take over from this step, grab a copy of the information and process it (no need for any central coordination mechanism). A situation brought to consciousness for an explicit, collective processing, is described by coalitions of processors presenting various aspects of it. Many such coalitions may compete in the Global Workspace to come to consciousness.

Our Implementation of Baars Theory

Our architecture is essentially rooted in professor Franklin's IDA architecture [4], but brings some domain specific extensions, such as learner modeling. The conceptual architecture (Fig. 1) covers every major aspect of cognition,

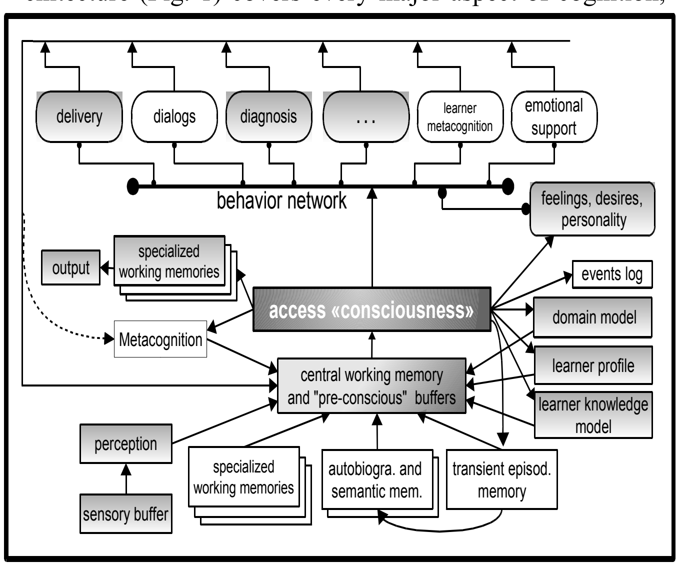

Fig. 1 Conceptual architecture. Grayed boxes indicate which functions are in the process of being implemented.

with many functional parallels to the physiology of the brain. Not readily visible on the diagram are some internal operations such as action selection, deliberation and inhibitions. At the center of all this is the *access "consciousness"* and the computational reproduction of Baars' mind processes (neuronal groups): the *codelets*. These are simple unintelligent agents that, as Baars processes, have specialized, limited processing abilities and range, but are very efficient. We now give a tour of main aspects of the architecture.

Senses and Perception. Every dynamic aspect of the "environment" appears in the messages received from the Simulator. The perceptual codelets scan the sensory buffer and activate nodes in the Perceptual Network (PN), such as the node for the rotation angle of a Canadarm2 joint, or for an *environmental consequence* (e.g. collision risk), resulting in "concepts" the agent can recognize. They also grant the activation level on a semantic basis (the importance of the information). A language has been developed to implement the communication between perceptual codelets and the Simulator.

A Distributed Learner Model. The learner model is distributed throughout Transient Episodic memory (TEM), Learner Profile (LP), and Learner Knowledge Model (LKM). The LKM is the main mechanism in establishing the causes of the learner's difficulties. It is implemented as a bayesian network coupled to codelets for its inputs and outputs. The LKM may also volunteer information, eventually priming some "*feeling"* in the agent

The Behavior Network (BN) and the Codelets. Based on an idea from Maes [5] and modified by Negatu and Franklin [6] the BN holds the repertoire of the agent's know-how in the form of *streams* of behavior nodes, and offers a way to decide which should activate. Nodes accumulate the energy that comes from the agent's "feelings" and "desires" (see Personality section) and from the "environment" until they are elected for action. The BN serves as the coordinator for the agent's actions, and generally counts on other functionalities' reactions to render a service. As an example, when all the conditions are met, the Diagnosis stream declares its wish for the probable cause, and LKM will very likely respond to this request "heard" in the broadcast. Negatu and Franklin also modified Maes model so that each behavior is realized by a collection of codelets. Codelets inherit the energy level of the information they carry, giving urgent information a better chance of being selected for broadcasting.

The Personality of the Agent. "Feelings" and "desires", implemented as high-level goal nodes, are the motivational mechanisms that feed the BN with activation and so orient action selection in line with the agent's high-level goals.

The Access "Consciousness". This mechanism selects the most important coalition of codelets in the central WM, and broadcasts its information, allowing all other systems to become aware of the situation. This mechanism is crucial for the collaboration of the parts, for instance in reaching a diagnosis.

The Architecture in Action

Let's illustrate the collaboration of the subsystems about a collision risk. 1) The Simulator sends a message about Canadarm2 configuration and about a collision risk. 2) Perceptual codelets stimulate their respective nodes in the PN. 3) Attention codelets form coalitions with various portions of the network sent to WM. 4) The coalition about the collision risk is selected and broadcast by the *access "consciousness"*. 5) All subsystems "hear" the broadcast, among which the "feeling" of the need for a diagnosis. Energy starts flowing in the Diagnosis stream of the BN. 6) Eventually, its first behavior, a primary diagnosis candidate, sends to the central WM a codelet to advertise its need for the probable cause. 7) If selected, the coalition formed with this codelet is broadcast. The LKM reacts and sends an information codelet containing its hypothesis. 8) This new codelet confirms the Primary Diagnosis coalition, and this is broadcast. 9) The "feeling" for a diagnosis cools down. Although this streamlined explanation makes the process sound very deterministic, in reality, depending on the context, many entities may respond to a broadcast. Only the most important information is selected.

Conclusion

Considering various aspects from various sources is very natural for our "conscious" architecture: every "module" is made aware of selected elements, and can contribute information to the description of the situation or participate in a decision. Then, internal deliberation offers great flexibility in the behavior adaptation. Thanks to numerous filtering and prioritizing mechanisms reproducing the various levels of consciousness, the "conscious" cognitive Tutor is able to fix its attention to the important things and take very sensible actions to support learner's efforts.

References

[1] F. Kabanza, R. Nkambou, K. Belghith, L. Hartman (2005). Path-Planning for Autonomous Training on Robot Manipulators in Space. Proceedings of IJCAI'2005, pp. 1729-1732.

[2] Block, N. 2002. Some Concepts of Consciousness. *Philosophy of Mind: Classical and Contemporary Readings*. D. Chalmers ed. Oxford University Press.

[3] Baars, B.J. 1988. *A cognitive theory of consciousness* NY: Cambridge University Press.

[4] Franklin S.; Baars, B.J.; and Ramamurthy, U.; Ventura, M. 2005. The Role of Consciousness in Memory. *Brains, Minds and Media* 1(bmm150).

[5] Maes, P. (1989). How to Do the Right Thing. *Connection Science Journal.* 1(3):291-323.

[6] Negatu, A.S. and Franklin, S. 2002. An Action Selection Mechanism for "Conscious" Software Agents. *Cognitive Science Quarterly* 2:363-386.

Acknowledgments. Special thanks to professor Franklin and to the University of Memphis, who graciously granted us access to their IDA technology. We also thank professors P. Poirier and F. Kabanza, and K. Belghith for their help.