A New Emotional Architecture for Cognitive Tutoring Agents

Usef Faghihi, Pierre Poirier, Daniel Dubois, Mohamed Gaha

Université du Québec à Montréal, P.O. Box 8888, Centre-ville Station, Montreal, Quebec, Canada

Abstract

In this paper we describe the reproduction of emotional mechanisms in the cognitive agent CTS (Conscious Tutoring System). The model replicates the amygdala's emotional influence over different parts of the system as a means of preparing the system for action, biasing action selection and improving the agent's self-resilience in an unknown or changing environment.

Introduction

Most cognitive scientists now recognize that emotions play an important part in cognition, and cognitive modelers accordingly have sought to include emotional mechanisms into their agents' architecture. In this paper, we present first steps to insert biologically plausible emotional mechanisms in an artificial agent. Research in neurobiology suggests that there are two routes from perception to the amygdala (Rolls 1999). The first route is short and direct: information flows from the sensory thalamus directly to the amygdala. Motor reaction is then rapid (Squire & Kandel 1998) as the received information is not interpreted by other brain structures (Rolls 1999). In the second route, information from the external environment is analyzed by various cortical areas (primary sensory cortex, unimodal associative cortex, polymodal associative cortex). It is then sent to the hippocampus, presumably for memory retrieval and temporary storage. All this processing serves to give meaning to the external stimulus and link it to other events (the hippocampus' episodic memory), before it goes again to the amygdala for emotional appraisal and response.

CTS's Emotional Architecture

Our tutoring agent is an extension to the generic cognitive architecture underlying it, which relies on functional "*consciousness*" mechanisms for much of its operations (described in Dubois 2007). Its functional architecture is inspired by LIDA's (Franklin and Patterson 2006) which is essentially an implementation of Baars' Global Workspace theory (Baars 1997). Modules (denoted with capitals to segregate them from their biological counterpart) seldom communicate with one another and mostly contribute in-

formation to Working Memory (WM) through information *codelets*[1]. These travel back to modules and other unconscious resources through cycles of "*conscious publications*" that broadcast only the most important, urgent, or relevant information (as suggested by Baars). But, like LIDA, CTS' original architecture does not distinguish the two routes of emotions. We believe that if we wish to create a human-like agent, an important step is considering and implementing this essential feature of human cognition. To address this issue, we propose a modification to the architecture of CTS, one which is both based on Baars' Global Workspace theory and inspired by recent advances in neurobiology concerning the brain's functional wiring and the amygdala's functionalities (see Ledoux 2000).

The Two Routes to Emotions in CTS

In CTS, the percept constituted by the active semantic nodes of the Perceptual Network (PN) enters Working Memory as a single network of codelets (step 1 in Figure 1), but is looked upon by the Coalition Manager as multiple possible *coalitions* of codelets, each describing various aspects of the situation. The emotional codelets inspect each coalition's informational content, and infuse it with a level of activation proportional to its emotional valuation. This increases the likeliness of some coalitions to draw Attention upon themselves. This emotional intervention in WM is how the Amygdala gets involved in CTS' long route. In the short route (Figure 1, ESR rectangles), emotional involvement happens sooner, but based on a rough evaluation of the situation. The perception codelets connect in parallel both to the PN and to the emotional codelets. Before their collective interpretation create a rich and complete interpretation of the stimulus, their individual activation is sent directly to emotional codelets, the first leg of the short route. This direct stimulation finds its application mostly in dangerous situations. When low-level basic information coming from the perception codelets recognize aspects of the situation as highly dangerous, there is no time to think and the emotional codelets will force an action to fire in the Behavior Network. This makes CTS act before it had time to become "conscious" of the situation and consciously plan a course of action.

---

[1] Based on Hofstadter *et al.'s* idea, a *codelet* is a very simple agent, "a small piece of code that is specialized for some comparatively simple task". Implementing Baars theory's *simple processors*, codelets do much of the processing in the architecture.
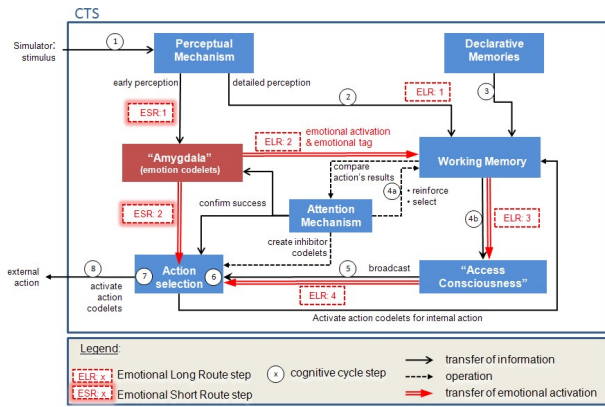
Figure 1 The long and short routes in CTS.

The firing of a behavior node releases one or more expectation codelets. Those are processes that watch the arrival in WM of some specific information, expecting to see within a given time frame some specific result(s) for the action taken by CTS. If they see information coming in WM that confirms the Behavior's expected result, they tacitly send reinforcement activation to the Behavior node that created them. In case of failure, however, relevant resources need to be recruited, both to allow them to analyze the cause, and for a deliberation to take place concerning supplementary and/or alternative actions; the expectation codelets thus work to have discrepancies brought to the attention of the whole system in an eventual conscious broadcast of the noted discrepancy. However, the instantaneous, emotional reflex initiated by the "short route" must be evaluated following the more thorough analysis of the situation that concludes later, through the long route. CTS does this because instinctive reactions do not block the longer analytical process; they merely step in first. Eventually, however, a better idea of the situation is worked out and comes to consciousness (long route), allowing normal action selection to take place. When the action thereby proposed comes into WM, the expectation codelets compare it to the reflex action that has been prompted. If roughly in correspondence, they put into WM a confirmation to that effect, which will serve, when broadcast, as a reinforcer to the emotional codelet(s) that were instrumental in setting off the reflex. In effect, this will have CTS' Amygdala reinforce its relevant rule. However, when the initial reaction diverges from the behavior proposed by the more methodical analysis, some mechanism must be in place to control the incorrect reaction. From a neurological point of view, control over actions is the role of cortical areas. We implement the cortical controls here with *inhibition codelets,* which are spun by the expectation codelet that spotted the discrepancy. The inhibition codelets attach themselves to a Behavior node in the action selection mechanism (BN) and constantly subtract some activation from it. This way of implementing the control seems in accordance with the fact that the amygdala never unlearns a "rule", always reacts to the same stimulus (Rolls

1999) and needs cortical interventions to temper it (Morén 2002). It accounts for our ability to willfully prevent a spontaneous urge to act (one can become conscious of being on the verge of taking action). This description highlights the fact that CTS' *emotional mechanism,* which responds implicitly to the events, reacts faster than the conscious process, but may react in ways that are different from what conscious planning would decide. As a concrete example, during Canadarm2 manipulations in the virtual world, information coming from the simulator describes an imminent collision. This is a situation to which the tutor must react at once and immobilize Canadarm2. This *implicit* reaction corresponds to the process described by Squire and Kandel (1998).

Conclusion

Our main objective in this paper was to suggest a new way to implement "emotions" in cognitive agents. The resulting architecture is quite plausible neurologically, integrating a recent view of the amygdala's double role with respect to emotions: to get the agent to react swiftly in emotionally-laden situations, and provide an emotional valuation to all stimuli coming into working memory. CTS' processing is now closer to human phenomenology. Its emotional mechanism is kept aware of the ongoing situation and can, in real-time, contribute emotional valance to the description of the situation. When information becomes "conscious", it may then participate in a richer way to decisions made by the system. The new emotional mechanism thus offers great suppleness in behavior adaptation.

References

Baars, B. J. 1997. *In the Theater of Consciousness*. OUP.

Dubois, D. 2007. *Constructing an agent equipped with an artificial consciousness: Application to an intelligent tutoring system*. PhD Thesis, Université du Québec à Montréal.

Franklin, S., Kelemen, A. and McCauley, L. (1998). IDA: A Cognitive Agent Architecture. In *IEEE Conf on Systems, Man and Cybernetics*. IEEE Press.

Franklin, S. and Patterson, F. G. Jr. (2006). The LIDA Architecture: Adding New Modes of Learning to an Intelligent, Autonomous, Software Agent. In *Proceedings of Integrated Design and Process Technology. IDPT-2006*, San Diego, CA, Society for Design and Process Science.

Hofstadter, D. R., and Mitchell, M. 1994. The Copycat Project: A model of mental fluidity and analogy-making. In *Advances in connectionist and neural computation theory, Vol.2:logical connections*, Holyoak, and Barnden eds.. Norwood N.J.: Ablex.

Ledoux, J. E. 2000. Emotion circuits in the brain. *Annual Review of Neuroscience*, 23, 155-184.

Morén, J. 2002. *Emotional and learning: A Computational Model of the Amygdala.* Doctoral Thesis.

Rolls, E.T. 1999. *The Brain and Emotion*. Oxford: OUP.

Squire, R. L. and Kandel, E. R. 1998. *Memory: From mind to molecules*. New York: Owl Books.