Affective Agent Architecture: A Preliminary Research Report

Kevin Raison[1] and Steve Lytinen[2]

[1] Chatsubo Labs, Seattle WA 98117, USA
[2] DePaul University, Chicago IL 60604, USA

Keywords. Affect theory, affective agents, affective computing, planning, artificial life, multi-agent systems.

Our work with affective agents is motivated by a desire to develop a biologically-inspired model of the human system that includes sensation, feeling, and affect integrated with the logical and cognitive models traditionally used in AI. Our work can be situated alongside Breazeal's robot Kismet (1998), Elliott's Affective Reasoner (1992), as well as Baillie's architecture (2002) for intuitive reasoning in affective agents; we too seek to model the motivations and reasoning of agents with an affective foundation. Our approach differs in how we model affect and the transition of biological impulses from sensation through affect into action. We base our work on the Affect Theory of Silvan Tomkins; Thomkins' (2008) theory comprises both a biologically hardwired notion of affect, similar to Ekman's basic emotions and Izard's DET, as well as a more culturally and biographically situated theory of scripts for describing how affect moves from feeling into action (Nathanson, 1996). As a proof of concept, we have consciously chosen to work with a simplified implementation of Tomkins' idea of affect, and explore how both internal (drives, organ sensations) and external stimuli move through the agents via the affect system and transition into action.

As pointed out by Scarantino (2012), the field of emotion research is in need of a properly defined set of "natural kinds" of affect as a basis for research. Tomkins provides such a set with his nine basic affects, which he considers biological and cross-cultural. They are positioned as an evolutionary response to the problem of the limited channel of consciousness (Vernon, 2009), insofar as the affect system provides not just distinct feelings, but thereby filters stimuli and focuses attention in an advantageous way. Because the negative (inherently punishing) affects of fear/terror, anger/rage, etc. trump the positive (inherently rewarding) interest/excitement and enjoyment/joy, the affect system favors responses that evade or directly confront threats to survival. Such a goal-oriented filtering of stimulus provides a clear foundation for a computational model of affect dynamics. With Sheutz's (2002) caution against prematurely assigning emotional labels to the states of artificial agents in mind, we have chosen to bracket out problems of qualia for now, and focus on developing our agent architecture with this filtering system as its core (see figure 1).

For this proof of concept, we implemented the basic structure of Tomkins' model with a limited number of drives, senses, affects, deliberative faculties

B. Goertzel et al. (Eds.): AGI 2014, LNAI 8598, pp. 250–253, 2014.
© Springer International Publishing Switzerland 2014

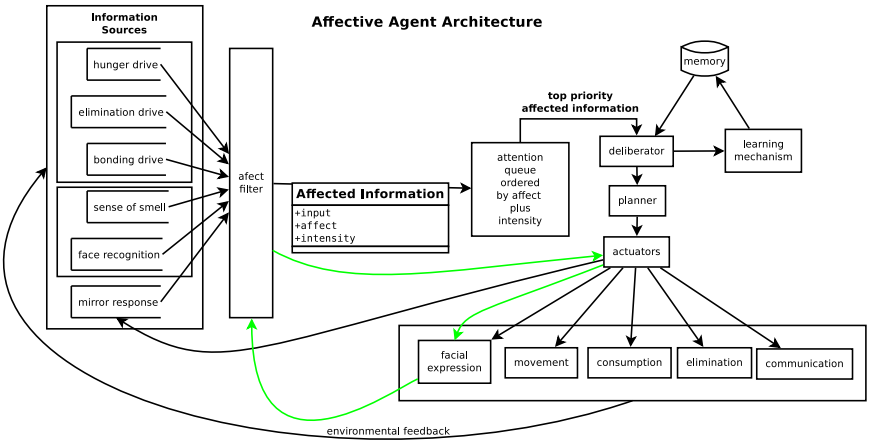

Fig. 1.

and actuators. We built our agents in Common Lisp, on the foundation of a blackboard system called GBBopen (Corkill, 2003). In order to test the agents' responses to their environment, we constructed a multi-agent game grid based on the Wumpus world (Russel & Norvig, 2003). We replaced the search for gold with a biologically-inspired search of food, but kept the Wumpus as a predator and pits as obstacles. We also built in the ability of individual agents to perceive one another's expressions as a starting point for experimenting with social behavior; this component also includes a simplified model of empathy based on Rizzolatti's (2004) mirror neurons.

Each agent is composed of a set of blackboard objects representing instances of various drives, a limited number of sensors (smell and facial feature recognition), an affective filter that is capable of combining with any type of information produced by other components of the system, an ordered queue of information and associated affect (called affected information), a planning / goal seeking component that is capable of processing one piece of affected information at a time, a memory store, and a set of actuators capable of operating in the game grid.

We use the notion of drives as a simplification of the complexities of organ sensations associated with biological needs such as hunger and elimination. The drive system operates largely in the background, independent of the other components of the system, except when a drive needs satisfaction. In that case, the drive will activate and send a signal into the affective filter. The signal will then be interpreted based on its intensity and an affect will be assigned. This affected information will be placed on the attention queue for processing by the planning system. If the affected information is on the top of the queue, the deliberative component will plan accordingly and attempt to satisfy it, until such time as a more urgent piece of affected information is placed on the queue. The information provided by the drives is in no way static; it can change as the intensity of the drive changes, in turn intensifying the associated affect, or perhaps even changing the affect from one type to another. This will cause a reordering of the

queue. E.g., hunger manifests as interest/excitement; if the hunger drive is not satisfied presently, it will escalate until it transforms into distress, thus raising the likelihood that the affected information will end up on the top of the queue. Other drives follow similar patterns appropriate to their function.

A similar model is used for filtering sensory information; if the smell of a Wumpus is detected, this information is filtered and assigned an affect of fear/terror appropriate to the proximity of the threat. The same is true of the breeze associated with a pit. Again, as these pieces of affected information make their way to the top of the queue, the deliberative mechanism plans accordingly so as to avoid the threat. The one exception to this flow of sensory information is in the case of agents coming into contact with each other. Depending on each agents' current affective state, the presence of the other agent may become associated with the bonding drive and then filtered via the affect system. If no higher-intensity affects are being processed, the agents will read each other's expressions and internalize them, thus activating their affective system a second time and reinterpreting the perceived state of the other agent as their own (an approximation of empathy). If the response is of a high enough threshold, the agents will then activate their memories of their own previous affective states and attempt to guess the reason for the other agent's expression. This internalizing of the other's affect will prompt the same action as the same affect from a different source.

The results of our experiments with affective agents in the grid world were very satisfying, though limited in this preliminary study. Agents were generally able to navigate the grid world, successfully evading threats and satisfying their drives based on the information provided to them by their affective filters. Competing drives and sensory inputs would often lead to oscillations between which information was on the top of the attention queue, but the flexibility of the planning system allowed for smooth transitions between competing goals. E.g., if an extremely distressed and hungry agent wandered close to a Wumpus, the hunger-satisfaction goal would be temporarily overridden by the fear-induced need to evade a threat, and once the threat was removed, the distress-laden hunger goal would reassert itself and the agent would go back to seeking food. There were a number of instances where an agent was dealing with extreme hunger and would discover that there was food right next to a Wumpus; in these cases, if the affected hunger was stronger than the fear of the Wumpus, the agent would risk getting close to the Wumpus just to eat. As soon as the hunger drive was satisfied, the agent's fear of the Wumpus would dominate and the agent would engage in evasive behavior. This sort of emergent behavior was not expected, but certainly validates the usefulness of the model. Results of the social experiments will be discussed in a forthcoming paper.

Overall, we would call these experiments a success; we were able to develop a model of affect-motivated deliberation and goal-striving in artificial agents. While at this stage, the agents are rather simple, enough has been accomplished to justify further exploration. A few things in particular stand out as needing improvement; first is adding a communication protocol for inter-agent message

passing beyond simple facial expression recognition. Second would be to flesh out the affect filtering system so that it operates more in line with Tomkins' notion of the density of neural firing; in his view, a particular affect is not caused by a symbolically recognized entity (as in our simplified model), but rather is the result of a particular pattern of neural activity (see Tomkins, 2008, p. 139). Layered neural networks might act as a better approximation of Tomkins' idea than the current symbolic architecture. Our model of sociability should also be greatly expanded; it currently only encompasses one aspect of the incredibly complex set of social motivations and behaviors. A final, longer-range goal would be to explore in more detail the complex interaction of affect, memory and imagery that Tomkins described in his script theory. Implementing this idea as a computer model could contribute positively to the looming possibility of the emergence of an artificial general intelligence; if such an entity were to emerge, it would be important that it resemble its human creators on the level of emotional experience so that we have some chance of actually relating to one another.

References

Baillie, P.: The Synthesis of Emotions in Artificial Intelligences: An Affective Architecture for Intuitive Reasoning in Artificial Intelligence. The University of Southern Queensland (2002)

Breazeal(Ferrell), C.: Early Experiments using Motivations to Regulate Human-Robot Interaction. In: Proceedings of 1998 AAAI Fall Symposium: Emotional and Intelligent, The Tangled Knot of Cognition (1998)

Corkill, D.: Collaborating Software: Blackboard and Multi-Agent Systems & the Future. In: International Lisp Conference 2003, New York (2003)

Elliott, C.: The affective reasoner: a process model of emotions in a multi-agent system. Northwestern University (1992)

Nathanson, D.: What's a Script. Bulletin of the Tomkins Institute 3, 1–4 (1996)

Rizzolatti, G., Craighero, L.: The Mirror-Neuron System. Annual Review of Neuroscience 27, 169–192 (2004)

Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River (2003)

Scarantino, A.: How to Define Emotions Scientifically. Emotion Review 4, 358–368 (2012)

Sheutz, M.: Agents with or without Emotions? AAAI, Palo Alto (2002)

Tomkins, S.: Affect Imagery Consciousness. Springer, New York (2008)

Vernon, K.: A Primer of Affect Psychology. Tomkins Instutite, Lewisburg (2009)