EmoCog: Computational Integration of Emotion and Cognitive Architecture

Jerry Lin and Marc Spraragen and Jim Blythe and Michael Zyda
Information Sciences Institute
University of Southern California
4676 Admiralty Way, Suite 1001, Marina del Rey, CA 90292
jerrylin@usc.edu, sprarage@usc.edu, blythe@isi.edu, zyda@usc.edu

Abstract

Since the reinvigoration of emotions research, many computational models of emotion have been developed. None of these models, however, fully address the integration of emotion generation and emotional effect in the context of cognitive processes. This paper seeks to unify various models of computational emotions while fully integrating with work done in cognitive architectures. We propose a perspective on how this integration would occur and EmoCog, a cognitive architecture with mechanisms for emotion generation and effects.

Introduction

Research on the interaction between emotion and cognition has become particularly active in the last twenty-five years. Notably, the work by Bechara and Damasio (Bechara, Damasio, and Damasio 2000) showed the necessity of emotion for decision making: loss of emotion likely leads to indecision or disadvantageous life decisions. This result challenged and largely overthrew the classical view that emotions could only cloud rationality, though that effect has also been documented (Gmytrasiewicz and Lisetti 2000).

Also motivating research on emotion is the characterization of emotion as an interrupt alarm signal to cognition (Simon 1967; Bower 1992). The signal is particularly responsible for heightening the importance of concepts associated with the emotional episode, and for refocusing attention (Ohman, Flykt, and Esteves 2001; Bower 1992) (causing distraction from a non-emotionally relevant task at hand when an emotional episode occurs). Damasio also asserted that emotion facilitates special recall of concepts when high emotional arousal occurs (Damasio 1994).

We believe that seemingly disparate emotional theories and experimental results can be integrated smoothly into a single computational model of human cognition. As part of the rise of emotion research in the AI and cognitive science communities, researchers have created several computational models of cognition and emotion, based on psychological theories and experimentation. A typical implementation of emotion generation is bound to a single theory, which usually conflicts with other theories on which factors generate emotion and how. Computational models of emotional effects tend to focus on a single effect of emotion on cognition or behavior. These research practices have led to incomplete, competing models which leave aside the question of a complete integration of emotion and cognition. We set forth proposals for a deeper integration than previous cognitive-emotional architectures, and present the design of a cognitive architecture, EmoCog, which embodies these ideas.

Background

Our approach is fundamentally built on theoretical and experimental work in psychology, cognitive systems, and neuroscience. For purposes of modeling emotion generation, we have particularly studied appraisal theories, which are the dominant basis for that type of computational model. Appraisal theory generally argues that people are constantly evaluating their environment, and that evaluations result in emotions such as fear or anger. Traditional game playing programs which evaluate their environment and/or self are not emotional, since they do not produce the necessary appraisal data for emotion and affect. There are many different appraisal theories, notably those of OCC (Ortony, Clore, and Collins 1988), Frijda (Frijda 1987), Smith and Lazarus (Smith and Lazarus 1990), and Scherer (Scherer 2001). Each theory differs in its appraisal variables and the manner in which appraisals are generated (e.g. simultaneously vs. specific order).

Several theories inform our work on emotional cognitive effects. The Somatic Marker Theory predicts that emotionally enhanced memory is useful for decision-making, as shown in the Iowa Gambling Task (Bechara, Damasio, and Damasio 2000). According to similar experiments, “gut feelings” during emotionally stressful moments are a heuristic to making a decision quickly, bypassing cognitive evaluation (Slovic et al. 2007; Finucane et al. 2000). The related mood congruence theory (Bower 1983; 1992) hypothesizes that facts or concepts learned during a positive or negative mood are thereafter easier to remember when in a similar mood. Conversely, the Yorkes-Dodson law (Yorkes and Dodson 1908) predicted that high levels of emotional arousal creates distraction from non-emotionally relevant tasks at hand (Kaufman 1999). The cue utilization theory (Easterbrook 1959) elaborates this effect: under high levels of arousal, environmental or internal cues not central to the arousing agent or situation will be increasingly ig-
nored.

Simon’s emotion-as-interrupt theory (Simon 1967) highlights autonomic arousal as a factor of emotion. Many of the widely cited emotion generation theories use arousal as a factor and can be applied to our model. Emotion generation theories usually also incorporate valence (degree of pleasantness or unpleasantness), which we can use to model further emotional effects on cognition, such as mood-dependent retrieval.

We also draw from a long tradition of work in computational cognitive architectures. Such systems usually try to address cognition as a whole. Our work has been most directly influenced by ACT-R (Anderson et al. 2004), CLARION (Sun 2006), PRS (Ingrand, Georgeff, and Rao 1992), and Soar (Laird 2008). See (Langley, Laird, and Rogers 2009) for discussion on this topic.

Related Work
Integration of emotions into cognitive architecture can be broken down into two separate parts:

1. Emotion generation - how cognitive processes play in the generation and decay of emotions
2. Emotional effects - how emotional signals, once generated, affect cognitive processes such as learning or planning

Some researchers have theoretically integrated emotion and cognition (Schorr 2001; Bower 1992) but leave out many details about the processes and the data that underlie them. Several computational models have been developed in attempt to flesh out some of these details.

The prominent systems that address emotion generation in a cognitive architecture include Soar-Emote (Marinier, Laird, and Lewis 2009), EMA (Marsella and Gratch 2009), and WASABI (Becker-Asano and Wachsmuth 2009). The Soar-Emote work discusses how appraisal would occur in Soar, using Newell’s theory of cognitive control. It is bound to a number of theoretical assumptions that stem from a single theory of emotion. EMA addresses the process of appraisal over a previously generated plan. Neither Soar-Emote nor EMA address how various cognitive processes would influence appraisal. WASABI is closest to our work on emotion generation. It presents primary and secondary emotions, where secondary emotions depend on past experiences and learned expectations and map to three discrete emotions (hope, fear, relief). The scope of this work lacks interaction with most cognitive processes and is limited to explaining few emotions.

Prominent systems that address effects of emotion on cognition include Soar-Emote, EMA, ACT-R (Cochran, Lee, and Chown 2006; Fum and Stocco 2004), and MAMID. Soar-Emote has work limited to how emotion may be input to reinforcement learning (Marinier and Laird 2008). EMA models generation of coping strategies following an emotional episode (e.g. change own beliefs). What are still missing are mechanisms for how the cognitive processes can be affected. Cochran’s work is limited to how emotional arousal may affect memory and Fum’s work is similarly limited to how emotional memory would affect recall and subsequently decision making. MAMID’s emotional effects on cognition are limited to altering the speed and parameters of a prescribed perception-action cognitive cycle. No previous computational model has attempted to integrate all of this work and other emotional effects on the function of cognitive processes in a single cognitive architecture or under a single theoretical perspective.

Overview
The remainder of this paper presents our propositions. This can be broken down into three sections:

1. EmoCog Architecture - modules, interactions, and data structures required
2. Mechanisms - processes that operate within EmoCog in context of emotion generation and cognitive effects
3. Discussion and Examples - rational and alternate perspectives on EmoCog’s design, and some examples to illustrate ability to model observed phenomenon

We finish by outlining our intended future work.

Approach
The primary theoretical proposals for our computational model of emotion and cognition require certain programmatical groundwork to implement in a cognitive architecture. We outline the key design decisions of EmoCog, but leave detailed discussion of implementation to a future paper. The novel features of EmoCog are the interactions between emotion and “rational” cognitive processes. In this particular version of our proposed architecture, we focus on emotion generation and emotional effect on memory, attention, and planning.

Architecture
The architecture diagram is shown in figure 1. At a high level, the architecture bears much resemblance to existing cognitive architectures such as Soar, CLARION, EPIC, and ACT-R. The potential cognitive modules are not limited to those shown.

EmoCog’s short-term memory is based on ideas outlined by Bower in his associative network theory of emotions (Bower 1981), and the spreading activation theory of memory by Anderson (Anderson 1983), similar to that which has been implemented in ACT-R (Anderson et al. 2004). EmoCog’s model of memory (both long-term and short-term) is a graph made up of generic nodes and links, and will function as an associative and semantic network.

There are several types of links between nodes, each with a label, a value, start node, end node, and optional direction. All nodes that are connected have an association link, which carries an association strength value. Associative link creation, reinforcement, and decay are all managed by the association management module (see below). In addition to associative links, there can be semantic links between nodes (e.g., causality), which can also carry values. These semantic links are maintained by cognitive processes (e.g. causal inference placing a causal link).
Each node can represent, but is not necessarily limited to, an episode, object, deadline, utility, concept, plan step, or procedure. The following node features are used by the appraisal system:

1. **Current arousal** (range 0 to 1): emotional arousal at current time
2. **Remembered arousal** (range 0 to 1): average arousal over time
3. **Current valence** (range -1 to 1): degree of like/dislike
4. **Remembered valence** (range -1 to 1): average valence over time

Other node features, such as recency of recall and how many times the node has been brought into working memory before, are not used by appraisal.

The current arousal and valence values are generated by the appraisal module. That process is presented in the following section. Remembered arousal and valence are averages of the current arousal and valence over time, which can span many episodes of the agent’s experience. The remembered arousal feature allows modeling the recall facility of nodes associated with strong emotions (Damasio 1994). The inclusion of remembered valence allows modeling mood-state dependent retrieval (Bower 1992).

**Mechanisms**

The mechanism set of EmoCog may be broken down into three key process groups: directly controlled cognitive processes, automatic cognitive processes, and meta-management. Figure 1 identifies the cognitive modules we propose to be directly controlled through meta-management. All other processes are assumed to be automatic and run in parallel.

For purposes of this paper, the details of the majority of these modules are abstracted, as we defer discussion of these to other papers. The sensory and encoding module handles the addition of new nodes into short-term memory from perception. Action regulation can be seen as the cognitive architecture’s interface (mainly output) to the body.

The attention module is responsible for selecting an associated cluster of nodes for cognitive elaboration. Selection determines the node cluster to use in cognitive elaboration by finding a single node with the greatest weighted sum of current arousal, associated utility, and associated urgency. All nodes directly and indirectly associated with the core node are also selected, using a breadth-first search until a threshold is met to form the cluster. The shifting of attention via emotional processes (Simon 1967; Bower 1992) has been marginally addressed in architectures such as CogAff (Sloman 2001). Meta-management is able to exercise limited executive control over attention by setting the weight of each of these parameters.

The association maintenance module performs spreading activation to create association links, and to reinforce current associations in working memory. With time and neglect, associations between nodes decay in long term memory. For example, when an object is perceived, a node is created for the instance of object perceived. If the object is in long term memory, an association must be made to the symbol representing that object in long term memory. If the object is previously unknown, associations can be made through various methods (e.g. matching by analogy or temporal relation).

The appraisal module adjusts the current arousal and valence values of nodes in short-term memory. When a node enters short-term memory, association maintenance occurs, and then the node is subjected to immediate first level appraisal. This appraisal is based on remembered arousal...
and valence and innate feeling (e.g., evolutionary dislike of predator or a negative utility node) if remembered arousal and valence is unavailable. The innate feeling is typically grounded in the body (e.g., pain is bad, and intensity of pain dictates arousal).

The node will be subject to reappraisal for as long as it remains in short-term memory. This may be best characterized as the influence of associated nodes on how an agent feels about the focal node. A graphical walk takes place on associated nodes, propagating the current arousal and valence values (these values are scaled down based on association strength). The traversal is terminated, if not earlier, when all nodes in short-term memory have contributed. Four values are produced by this process: sum of arousal of negative valence associations, sum of arousal of positive valence associations, average negative valence, and average positive valence. The valence with a higher summed arousal will dominate and inhibit opposing valence. The appraisal module then incorporates the average arousal and valence into the node's current arousal and valence. When a new node enters a cluster and is appraised using first level appraisal, it will similarly influence neighboring nodes in an outward fashion.

Overall mood of the agent will also be maintained by the appraisal process. The current intention is to compute mood as an average of all current arousal and valence of nodes across working memory. A single node's appraisal can still influence our mood over a long period of time, given that the node remains in working memory. This needs elaboration, however, as mood is not only an overall emotional state based on working memory, but may persist, decay, or change independent of the changing emotionally charged nodes in working memory.

Physiological signals will relate the needs of the body to the cognitive architecture. In the human body, these signals might be of hunger, thirst, or fatigue. The physiological modula interprets a body signal and maintains a node in short-term memory as well as associated urgency and utility. The strength of the signal is directly translated to an interpretation of urgency, while utility is innate.

The meta-management module is where metacognition and cognitive control will take place. The vital components of this module are the metacognitive rules, decision cycle, and list of directly invoked cognitive processes. In practice, the metacognitive rules and the rules describing cognitive processes are represented and applied within the same reasoning platform. Actions, in addition to existing as nodes in the associative memory, are also reasoned about and decomposed within the same platform. This approach gives EmoCog an unprecedented ability to represent interactions between emotional and physiological processes and cognitive processes such as planning and inference.

The decision cycle is the driving force of the meta-management. It typically progresses as follows:

1. Perception - Short-term memory is updated with information from perception.
2. Attention - Metacognitive rules determine weights of attention parameters. Attention module is invoked.
3. Elaboration - The node(s) which gain attentional focus are given limited cognitive processing. Rules of the metacognitive module choose which cognitive process runs1.
4. Decision evaluation - Metacognitive rules determine if enough elaboration has been performed.
5. Action selection - If elaboration has produced a set of candidate actions, one is selected based on metacognitive rules that weigh utility and emotional bias.
6. Action execution - If there is a selected action, it is initiated. The decision cycle is then repeated. Note that subsequent decisions, or exogenous events, may interrupt the execution of the action.

During the elaboration phase, individual cognitive processes are invoked through metacognition, although they share the same rule space. All cognitive processes execute in an anytime fashion, with a limited amount of available computation before the elaboration process repeats, possibly switching attention. Cognitive processes are only able to use the cluster of nodes under attention focus.

Discussion and Examples

We view EmoCog as an embodiment of principles needed for full integration of emotion and cognitive architecture and it will be particularly apt for modeling affective behavior as described in psychology and neuroscience literature.

One particular phenomenon we address is that of emotions both as interference and heuristic. It was observed that emotional signals can disrupt normal cognitive function, particularly when not relevant to the processing at hand.

For example, an agent is assigned a cognitive task to recall and output a list of words in order from long-term memory, under a deadline. Attention is focused on the first word and the node in associative memory representing this word. The metamanagement invokes the recall process to find the most strongly associated node. After some iteration, several nodes are recalled into short-term memory via this cycle. At some point, the word “tiger” is retrieved and following the next recall cycle, the most strongly associated node of a traumatic “tiger attack” experience is recalled. That node has high activation strength due to high remembered arousal and extreme negative valence. When the “tiger attack” node is brought into working memory, an appraisal based on the remembered arousal and valence is assigned to the node’s current arousal and valence. This causes the attention focus to be drawn away from the task to dwell on the tiger attack. Meanwhile, other nodes which do not hold attention focus have their arousal levels decay, allowing the dominance of the aroused thought.

Metamanagement, referencing the agent’s goals, attempts to refocus attention to the task by raising the weight of utility. The emotional episode, however, is so strong, that the thought of a tiger attack continues to hold the agent’s attention. The attempt to return attention to the task succeeds only when the urgency of the task also increases, due to impending deadline. These rules in metamanagement are used.

---

1Processes like learning are automatic and are not among those selected.
to reason over the various cognitive tasks. Soar’s metacognition is similar in this regard.

Our model of metagemanagement stresses the importance of metacognition when our emotions can lead us astray. A person could have been taught to ignore emotionally compelling issues to focus on his work, so he may try to do so, but emotions are very difficult to fully ignore. Sufficient emotional arousal will wrestle cognitive attention away from a rational train of thought. “People who are more rational don’t perceive emotion less, they just regulate it better” (De Martino et al. 2006).

If a person focuses on a certain task, usually irrelevant emotions fade, but it is not necessary that he has completely forgotten about the invoking fact, it’s that it has been tuned out. Neuron signal strength typically decays over time, so under the impression that emotional signals occur in the human brain as simple neurological pulses, we model current arousal of unattended nodes to decay similarly, allowing concentration on a task. That is, unless something particularly compelling draws attention away. There are also well studied mechanisms of signal inhibition and winner take all from neuroscience literature, which we leverage by having the appraisal process inhibit and suppress nodes excluded from the attention cluster.

When relevant to a task, emotion can serve as a heuristic for various types of cognitive processes. Emotion acting on the recall process can model the emotionally-enhanced recall demonstrated in the Iowa Gambling Task, and also model Bower’s mood-congruent retrieval effect. For instance, an agent wins a lottery by picking the number 7. The agent creates an association link between a node containing the number 7 and a node containing the experience of winning. The appraisal process confers higher arousal and positive valence to the number 7 via its association with winning. When the prospect of picking a number to win another lottery becomes the agent’s goal, 7 is more likely to be recalled than other numbers, as it is positively associated with winning (“lucky 7!”). The agent’s mood will also influence the choice. An agent in a positive-valence mood will be more likely to recall 7, as that number has the highest valence among the choices in long-term memory.

Since all cognitive processes work with the associative network and emotional data is embedded within all the nodes, any process can use emotion data to model emotional affect. For example, arousal and congruence may influence the action and goal choices an agent makes when it constructs a plan, and also the fidelity with which it executes a plan. The agent may omit or curtail steps whose actions or objects have lower arousal, even though they are logically necessary to the plan.

EmoCog is designed to be flexible, so that further dimensions and alternate views of emotion can be incorporated into both the associative network and mechanisms. For example, different appraisal theories can be modeled for emotion generation, as many postulate some form of arousal and valence. Other appraisal variables such as surprise can be viewed as a combination of our current appraisal and violation of expectation (generated by planner or expectation process), or the appraisal variable “causal agent” as causal inference followed by association and appraisal.

To illustrate this, consider an agent looking at a table with several objects on it. You may ask the agent how it feels about each object on the table, and it may answer very differently for each object, and why, by following the associations in working memory with each object. The emotions experienced may also depend on the co-existence of objects (e.g., a kitchen knife alone vs. a kitchen knife next to a puddle of blood). The only system with a similar capability is Soar-Emote, but its agent would only feel one momentary emotion for each object individually as it perceives it, and is limited on expressiveness in introspection.

Finally, much of our initial design subsumes previous work in computational emotion with some modification. Soar-Emote’s appraisal in PEACTIDM can be seen as appraisal during our decision cycle. EMA’s appraisal over a plan can be seen as having a series of plan steps associated in some cluster. WASABI’s primary and secondary appraisals also have equivalents in EmoCog, but the proposed system of secondary appraisal in EmoCog is more flexible, as outlined above.

**Conclusion and Future Work**

The core proposals which allow deep integration of emotions in a cognitive architecture are in associative network memory, cognitive attention, and appraisal following cognition. The associative network allows for concepts to influence each other emotionally, as well as hold emotional information for general consumption by cognitive processes, allowing effects on these processes and further emotion generation. The cognitive attention model allows for controlled elaboration and emotional rise and decay. And finally, the ideas of how appraisal and association management follow cognition in the associative network, really allows the cognition to influence emotional generation.

A majority of these ideas are not novel, but we believe the perspective on their integration has great potential. It provides a general framework to reconcile and unify existing computational models. The framework should also have greater explanatory power for emotion-related phenomenon and provide a test bed for understanding the role of emotions in a fully cognitive being.

The scope of this project is broad, encompassing aspects of cognitive architecture, emotion generation, and emotional effect. We have started to implement EmoCog, and are working to complete an initial version. After this we plan to incorporate lessons learned from its deployment in a number of settings, including behavioral simulations and computer games.

We also intend to elaborate on much of the underlying groundwork we have presented here in subsequent publications, including the topics of attention, physiological mechanisms, learning, semantic/associative networks, metacognition, and knowledge representation and the relevant algorithms, equations, and data structures.

There are also plans to demonstrate various well studied emotion-related behavioral phenomena. As we have argued here, we will be able to reproduce human behavior with greater fidelity considering both when emotions
can aid us in decision making and when emotions can lead us astray. Some of the more beneficial effects include the emotion-enhanced judgment demonstrated in the Iowa Gambling Task, and the affect heuristic used in resource-bounded decision making. Examples of negative effects are short-sighted exhilaration over a stock bubble, or extreme emotional trauma states such as PTSD.

Acknowledgements
We are very grateful to the members of the GI Lab who have contributed to our progress in developing ideas. Special thanks to Professor PR for commenting on drafts of this paper. We’d also like to thank the Office of Naval Research (ONR) for funding this research.

References
Scherer, K. 2001. Appraisal considered as a process of multilevel sequential checking.