Foundation for Emotion Capable Cognitive Architectures

Jerry Lin[1], Marc Spraragen[2], Jim Blythe[1], and Michael Zyda[2]

[1] Information Sciences Institute
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292
{jerrylin,blythe}@isi.edu
[2] University of Southern California
Department of Computer Science GamePipe Laboratory
941 W. 37th Pl., SAL 300, Los Angeles, CA 90089
{sprarage,zyda}@usc.edu

Abstract. Emotions play an important role in both human intelligence and behavior. Advances in understanding emotions can improve our understanding of human intelligence as well as allow higher fidelity modeling of human behavior. Current research in computational models of emotion is trying to address integration between emotion and traditional cognitive architecture. Different interactions between emotion and cognition have been observed in cognitive science, neuroscience, and psychology. However, only relatively small subsets of the observed phenomena can be explained or replicated by existing computational models. In this paper we identify a small but representative set of phenomena that we believe should be covered by an architectural design to integrate computational emotion and cognition. We discuss a set of computational mechanisms that is sufficient to cover the phenomena discussed and present the embodiment of the proposed design. This design is also parsimonious in the sense that omitting any of the mechanisms would remove the ability to model at least one of the phenomena. This broader level of integration should allow both advances in human intelligence and virtual human modeling.

1 Introduction

Emotions characterize the human experience. Emotions help us make good decisions and are a constant part of every behavior we exhibit. In the context of understanding both human intelligence and human behavior, it has become vital to understand emotion.

Since the early days of philosophy, emotion was regarded as only being able to cloud the mind from rational thought. Everyone is familiar with how emotions can lead us astray - we do something we later regret because we were angry, we gamble with bad odds, etc. Bechara and Damasio also showed that the human mind fails to make normal decisions in the absence of emotions in cognitive processing. These findings contradicted traditional opinions of emotion, and Bechara

and Damasio asserted that emotions contain important information to aid cognitive processes in optimizing benefit of decisions [3]. Similarly, other studies have shown that emotions provide a heuristic in avoiding slow cognitive processing while still allowing beneficial decision making when constrained for time [24, 8].

A recent spike in research activity investigating the interaction between emotion and cognition includes a few computational models. Each of these models seeks to address issues of how emotion is generated in the context of cognition and how such generated emotion, in turn, affects cognitive processes. Currently, these models only contain enough explanatory power to cover relatively small subsets of emotional-cognitive phenomena reported in psychology, neuroscience, and cognitive science literature. We define emotional-cognitive phenomena as observable emergent behavior caused by the interplay between emotion and cognition.

Further, these models have been primarily focused on problems of computational emotions in isolation and for the most part, ignore the overall context of emotion in a cognitive system. This paper seeks to address the bigger picture by studying the architectural characteristics that yield more powerful emotional models. In doing so, this paper considers two complementary sources for requirements of a successful emotional-cognitive architecture. First, we recognize that for an intelligent agent to fully gain emotional competence, the architecture must support the interplay between affect generation and cognition. The second source comes from observations of human emotional behavior in the literature; we aim for coverage of a small representative set of emotional-cognitive phenomena as a basis for measuring depth of integration between emotional and cognitive mechanisms. The key ideas in this set lie in the choice of computational components, memory structure, and emotion representation.

This paper proposes a fundamental set of architectural design decisions and mechanisms intended to allow integration of emotion and cognition at a greater depth than previous models. This set includes components and mechanisms for appraisal, arousal and decay of emotions, associations in memory, cognitive focus, expectation and alternate possibility generation, meta-control, mood based on recent emotions, and a universal emotion representation for use within any cognitive process. Our first hypothesis is that this set is sufficient to address a wider range of emotional-cognitive phenomena than any previous system. Our second hypothesis is that any subset of these mechanisms is insufficient. In developing these hypotheses, we analyze each phenomenon for the conceptual factors which produce it. We will show that each of the aforementioned architectural mechanisms satisfies an aspect of one or more phenomena, and that no previous system encompasses all these mechanisms.

1.1 Related Work

Various theories of emotion are relevant, but the most important are appraisal theories and work by Gordon Bower. Appraisal theories assert that people continuously evaluate their environment, and the results from these evaluations result in emotions. There are many different appraisal theories, notably those of OCC

[21], Frijda [9], Smith and Lazarus [25], and Scherer [22]. Each theory differs in its appraisal variables and the manner in which appraisals are generated (e.g. simultaneously vs. specific order) but they all have some notion of arousal and valence. Bower's key work was about the interplay between emotion and memory [5] and provides some insight into how emotion integrates with cognition as a whole.

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 [1] and Soar [12]. See [13] for discussion on this topic.

Some researchers have theoretically integrated emotion and cognition [23, 6] but leave out many details about the processes and the data that underlie this integration. Several computational models have been developed in attempt to flesh out some of these details.

A few of these key systems and theories which have demonstrated some explanatory power and are the focus of our discussion section are EMA [18], MAMID [10], Soar-Emote [17], and WASABI [4].

1.2 Overview

We first identify a relatively wide range of emotional-cognitive phenomena. We follow by discussing the components and mechanisms sufficient in an architecture to be able to cover said phenomena. Discussion follows, about the necessity of the proposed components and mechanisms to address the emotional-cognitive phenomena set forth. Finally we present our current status and identify future work.

2 Emotional-Cognitive Phenomena

To measure the strength of integration between emotion and cognition in our system, we propose that better integrated models can explain a wider set of phenomena caused by interplay between emotion and cognition. These phenomena are gleaned from literature in cognitive science, neuroscience, and psychology. The phenomena themselves are interesting to address because some may allow us to make progress towards modeling human level intelligence, help us to understand human cognition, and allow us to reproduce human behavior with higher fidelity.

Currently, explanatory power of current computational models of emotion is limited to small subsets of known emotional-cognitive phenomena. We have carefully hand-picked the following small set of phenomena which we believe covers a spectrum of mechanisms wide enough to require strong emotional-cognitive integration. The phenomena are described below, and the integrated mechanisms which produce them are covered in the following section.

2.1 Incidental Emotions

Incidental emotion is defined as emotion which is unrelated to the task or decision at hand. Research has shown that this type of emotion strongly influences the way humans think and behave [15]. This phenomenon is particularly important in understanding human decision making, and thus aids in areas such as behavioral modeling, game theoretic models, and believable agents.

In an example inspired by some of Lerner's work, a retail store playing a song which invokes feelings of sadness while a customer shops results in the customer spending more money and being less discriminating with purchases [14]. In another example of incidental emotion, someone is sitting on a hard chair during an interview and feels more confident throughout the interview (affective priming).

2.2 Prejudiced Like/Dislike

Humans exhibit prejudice towards a number of things; for instance, we tend to like babies with big eyes and to dislike animals baring their teeth. Prejudice also extends to simple decisions like choosing a restaurant from a vast sea of options.

In the Iowa Gambling task, people experienced emotional preference for certain decks before they could explain why they liked or disliked them. People who lacked capacity for emotions due to brain lesions also lacked this ability to favor beneficial decks, though they could eventually explain the optimal strategy [2].

Separate studies by Bechara and Damasio revealed that the same people who lost the capacity for emotions often had a hard time making simple decisions about where to eat. It is as if they were trying to weigh all the benefits and consequences of all possibilities. Any computer scientist would recognize this approach and instantly realize this leads to an explosion in the solution space and is infeasible to completely search in any reasonable amount of time. This suggests humans make intelligent decisions by using emotions as a bias, so we don't need to spend as much time deliberating.

2.3 Panic, Regret, and Greedy Exuberance

Most people are familiar with feelings of panic and regret. Greedy exuberance is usually described as a period of exciting desire to acquire or possess (sometimes caused by panic and regret). These are all similar in that they each tie in closely with future and alternate projections we make. Understanding why we sometimes experience panic, regret, and greed can inform us of how humans make predictions about the future (or analyze the past), which has broad applications in many areas of AI.

The fact that history repeats itself leads us to believe that there is something inherent to human nature that drives these repetitions. The dot com boom, the Dow of 1929, the S&P 500 of 1987, and the Nasdaq of 1999 all exemplify this. In a 2007 experiment, subjects were given money to invest unknowingly in simulations of the three aforementioned events while having their brains scanned

with an fMRI machine. Results showed that the frenzy experienced during these simulations were caused by emotional regions of the brain [16]. People experienced greed as they foresaw continuing market raises as an opportunity to make money. Subjects experienced regret and panic as they analyzed missed opportunities of not investing everything in the raising market. This, as one might imagine, ended badly for many of the subjects.

2.4 Mood Congruent Cognition

Bower pioneered the idea of mood-congruent cognition, which is the concept that various cognitive processes behave in correspondence to the mood state (e.g. a person is more likely to recall happy memories while in a happy mood). He demonstrated mood-congruent memory recall [5] and other mood-congruent cognition such as learning, judgment, expectation generation, imagination, etc [6, 7]. Understanding this phenomenon certainly will allow us to better understand human behavior, but more interestingly it may help us pioneer areas such as imagination and creativity in artificial agents.

Many other researchers have confirmed this phenomenon in different experiments, including but not limited to work by Lerner, Loewenstein (see above), Mayer [20], and Isen [11].

Subjects of Bower's studies were induced into happy or sad moods, and he found that the emotion powerfully induced cognitive processes such as free association, imagination, and snap judgments to be congruent with the subject's current mood [5].

3 Components and Mechanisms

We advocate a small set of components for computational emotions that we argue are sufficient to cover these phenomena as part of a general architecture for intelligence. In doing so, we also follow principles for general architectures, including universal representability. Emotion should be capable, at least in principle, of affecting any cognitive facility modeled in the architecture and, likewise, the results of cognitive processing should provide for changes to emotions. Therefore we require a universal representation for emotion that is independent of any module within the cognitive architecture and we seek a general mechanism for feedback from cognition as input to emotion processing.

We argue that to be able to explain the above phenomena, the following theoretical concepts are sufficient, and the lack of any single concept will result in the remaining subset's inability to completely cover the presented phenomena:

- Appraisal process which specifies preference (like/dislike) based on past experience
- Arousal and decay of emotions
- Associations in memory
- Emotional memory

– Expectation and alternate possibility generation
– Mood based on various arousal and decay of recent emotions

Theoretical integration and implementations for these concepts are suggested in section 4.

3.1 Appraisal

Appraisal theory is currently the dominant approach in computational emotions approach and has shown the most promise in existing systems today (see [19] for discussion). Other approaches include dimensional, anatomic, rational, and communicative theories.

Most appraisal theories at the very least embody some notion of like/dislike for specific stimuli. The annotation of like/dislike is the base mechanism we mean should be included in an emotionally competent cognitive architecture. It is intuitive to understand why it is difficult to have basic emotions without notion of like/dislike.

Simpler mechanisms, such as those which just assign a discrete emotion to the agent state lack enough expressive power to address incidental emotion and very context-sensitive feelings of panic, regret, and greed. Without the use of appraisal or some other emotion theory that embodies like/dislike of individual stimuli, none of the phenomena would be completely addressed.

3.2 Arousal and Decay

Arousal is similar to excitement of a specific emotion. Decay occurs when the level of excitement atrophies. Human experience tells us that emotions come and go and some mechanisms should exist to address this sort of emotional behavior. We propose inclusion of annotation of arousal over each specific stimulus alongside like/dislike (or valence) and algorithms for controlling arousal and decay of various appraisals.

The limited size of working memory can also explain the decay of emotions in some cases. As new stimuli enter working memory, older concepts and their emotional information are pushed out, simulating loss of those emotional signals over time. However this approach may have the undesirable side effect of dictating the size of working memory based on the rate of emotion decay.

3.3 Associations in Memory

Associative memory is a network of associative links made between various semantic concepts and schemata. Links are used to create clusters describing an event or episode, or used to bind concepts' relevant objects together. Early arguments for associative memory for emotions were notably made in the associative network theory of memory and emotion [5]. This paper proposes associative networks as the best option as a substrate for both working memory and long term memory.

3.4 Emotional Memory

The concept of emotional memory is that emotions should be recoverable from long term memory as well as short term memory. To achieve this, we propose that the appraisals over various nodes should be imprinted in memory along with concepts and associations. This could be easily achieved by saving average valence and arousal values with each node.

3.5 Expectations and Alternate Possibilities

The generation of expectations is a projection of the future, and alternate possibilities are similar projections except they can also be generated for past events (after knowing actual outcomes). This mechanism is likely to be closely tied to any kind of planning processes an agent possesses, but in the case of expectations it should contain some enumeration of possible result states and perceived probabilities. Alternate possibilities can most easily be represented as a state-action path within the plan, different from the one actually taken, at some branching of events or results.

3.6 Mood

Mood is typically defined in the area of computational emotions as a long lasting, stable, overall emotional state. This is opposed to immediate emotions which may be attributed to an individual stimulus. We propose that mood is a function of all emotional values in working memory, such as a simple average. Taking an average will have a damping effect on the fluctuations of emotions as individual appraisals occur and decay.

4 Discussion

In this discussion, we will reason about how each presented phenomenon can be generated by a combination of the proposed components and mechanisms. In this discussion, we will reason about how each presented phenomenon can be generated by a combination of the proposed components and mechanisms. It is trivial to understand from our descriptions how the absence of any one of these mechanisms then results in inability to address said phenomenon.

The following table summarizes the correspondence between the sets of phenomena and components.

4.1 Incidental Emotion

Incidental emotion can be covered by appraisal, associative memory, arousal and decay, emotional memory, and mood. Following the example, as an agent perceives an object such as the song being played, it is placed in working memory.

Table 1. Phenomena and Components Sufficient to Address Them

	Appraisal	**Arousal / Decay**	**Association**	**Emotional Memory**	**Expectations / Alternates**	**Mood**
Incidental Emotion	✓	✓	✓			
Prejudice	✓		✓	✓		
Panic, Regret	✓			✓	✓	
Mood-Congruent Cognition		✓	✓			✓

The association management module associates the song with long term memories, such as music with similar sound, and the appraisal module makes an appraisal based on these associations (may include lookup of remembered appraisals). Suppose that the appraisal is of negative valence and strong arousal. Since we have assumed mood is based on all emotional appraisals in working memory, mood is shifted negatively. Planning and other cognitive processes that use this mood, modify their behavior to bias purchase of new items more heavily, resulting in the behavior observed by Lerner.

4.2 Prejudiced Like/Dislike

The key component needed is emotional memory, but appraisal and associative memory also have a role in explaining prejudiced like and dislike. As stimuli enter working memory, associations are built to long term memories, based on traits or features. These associations we hold influence appraisal of the original object based on strength of association, leading to prejudice. Since appraisal is generally assumed to be automatic and independent of most cognitive processes, this not only explains early bias but also the capacity to place heuristics on concepts before full cognitive evaluation takes place. Since all concepts are proposed to be represented in a universal associative substrate, plan steps will also have appraisals that may act as a heuristic to short circuit decisions in time constrained, open, and dynamic environments.

4.3 Panic, Regret, and Greedy Exuberance

To replicate behavior observed in the various stock crashes, aspects of appraisal, expectation generation, and emotional memory come into play. The agent may initially have a decision to invest its money in multiple places (based on statistically good practice), but when it observes the bubble's increase in value, it infers an alternate possibility: if it had invested all its money in that specific market, it would have made much more money. The appraisal for the missed opportunity is much better than the actual current situation, resulting in the feeling of regret. The regretted decision has a negative appraisal attributed to it, so in the future, when a similar situation arises, it wants to avoid the previously

experienced regret and invests more and more into the bubble. Some agents may have better meta-management rules that allow the agent to avoid regret affecting future decisions.

4.4 Mood-Congruent Cognition

Mood, arousal and decay, and the associative network can be used to explain mood-congruent cognition. The need for mood is straightforward, but arousal and decay are very closely tied to the temporal dimension of emotions, including mood. The associative network in our proposed case is represented similarly to all cognitive processes, so as long as a cognitive process is able to take nodes from the network as input, it can participate in mood-congruent cognition. Many processes have been shown to work with such a data structure in ACT-R. The mood itself is a value in the working memory that can be accessed in conjunction with all nodes. From there, if we represent the mood as a valence and arousal then a simple function calculating difference in arousal values can quantify the level of congruency between appraisals on nodes and mood.

5 Conclusion

We presented a small but broad set of emotional-cognitive phenomena and argued that they can be addressed by the integration of a small number of components into a cognitive architecture. Exclusion of any component will result in inability to address some phenomenon. For example, our approach to incidental emotion depends on associative memory, while arousal and decay are required for mood-congruent cognition. It is in this way that this set of components is parsimonious for an emotionally competent architecture.

Existing systems such as EMA, MAMID, Soar-Emote, and WASABI have some notion of appraisal, mood, and expectations, but none of the other mechanisms. From the fact that all stated phenomena leverage some aspect of associative memory, we believe associative memory is the most vital component in the problem of fully integrating emotion and cognition.

With these arguments set forth, we have laid the groundwork for our two initial hypothesis we set forth in Section 1; the set of components and mechanisms described in this paper are sufficient to address a wider range of emotional-cognitive phenomena than any previous system and any strict subset of these mechanisms is insufficient.

With emotions playing a strong role in both human intelligence and behavioral modeling, using these fundamental design decisions and mechanisms can facilitate unification of a wide range of research and provide a framework for newer, more realistic agents.

References

1. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An Integrated Theory of the Mind. Psychological review 111(4), 1036–60 (Oct 2004)

2. Bechara, A., Damasio, H.: Deciding advantageously before knowing the advantageous strategy. Science (1997)
3. Bechara, A., Damasio, H., Damasio, A.: Emotion, decision making and the orbitofrontal cortex. Cerebral cortex (2000)
4. Becker-Asano, C., Wachsmuth, I.: Affective computing with primary and secondary emotions in a virtual human. Autonomous Agents and Multi-Agent Systems 20(1), 32–49 (May 2009)
5. Bower, G.: Mood and Memory. American psychologist (1981)
6. Bower, G.: Affect and Cognition. Philosophical Transactions of the Royal Society of London B(302), 387–402 (1983)
7. Bower, G.: How might emotions affect learning (1992)
8. Finucane, M.L., Alhakami, A., Slovic, P., Johnson, S.M.: The affect heuristic in judgments of risks and benefits. Journal of Behavioral Decision Making 13(1), 1–17 (Jan 2000)
9. Frijda, N.: Emotion, cognitive structure, and action tendency. Cognition & Emotion 1(2), 115–143 (1987)
10. Hudlicka, E.: Reasons for Emotions : Modeling Emotions in Integrated Cognitive Systems, pp. 1–37 (2007)
11. Isen, A.M.: Positive Affect and Decision Making. Guilford Press (2000)
12. Laird, J.: Extending the Soar Cognitive Architecture. In: Proceeding of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference. pp. 224–235. Amsterdam (2008)
13. Langley, P., Laird, J., Rogers, S.: Cognitive architectures: Research issues and challenges. Cognitive Systems Research 10(2), 141–160 (2009)
14. Lerner, J.S., Small, D.A., Loewenstein, G.: Heart Strings and Purse Strings: Carryover Effects of Emotions on Economic Decisions. Psychological Science 15(5), 337–341 (May 2004), http://www.ncbi.nlm.nih.gov/pubmed/15102144
15. Loewenstein, G., Lerner, J.S.: The role of affect in decision making. Handbook of affective science pp. 619–642 (2003)
16. Lohrenz, T., McCabe, K., Camerer, C.F., Montague, P.R.: Neural signature of fictive learning signals in a sequential investment task. Proceedings of the National Academy of Sciences of the United States of America 104(22), 9493–8 (May 2007)
17. Marinier, R., Laird, J., Lewis, R.: A computational unification of cognitive behavior and emotion. Cognitive Systems Research (2009)
18. Marsella, S., Gratch, J.: EMA: A process model of appraisal dynamics. Cognitive Systems Research (2009)
19. Marsella, S.C., Gratch, J.: Computational Models of Emotion. Oxford University Press (2010)
20. Mayer, J., Gaschke, Y.: Mood-congruent judgment is a general effect. Journal of Personality and Social Psychology 63(1), 119–132 (1992)
21. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1988)
22. Scherer, K.R.: Appraisal Considered as a Process of Multilevel Sequential Checking, pp. 92–120. Oxford University Press (2001)
23. Schorr, A.: Appraisal: The evolution of an idea., pp. 20–33 (2001)
24. Slovic, P., Finucane, M., Peters, E., Macgregor, D.G.: The Affect Heuristic, pp. 397–420. Cambridge University Press (2007)
25. Smith, C., Lazarus, R.: Emotion and adaptation. Handbook of personality: Theory and research (1990)

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 Yerkes-Dodson law (Yerkes 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 in 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).

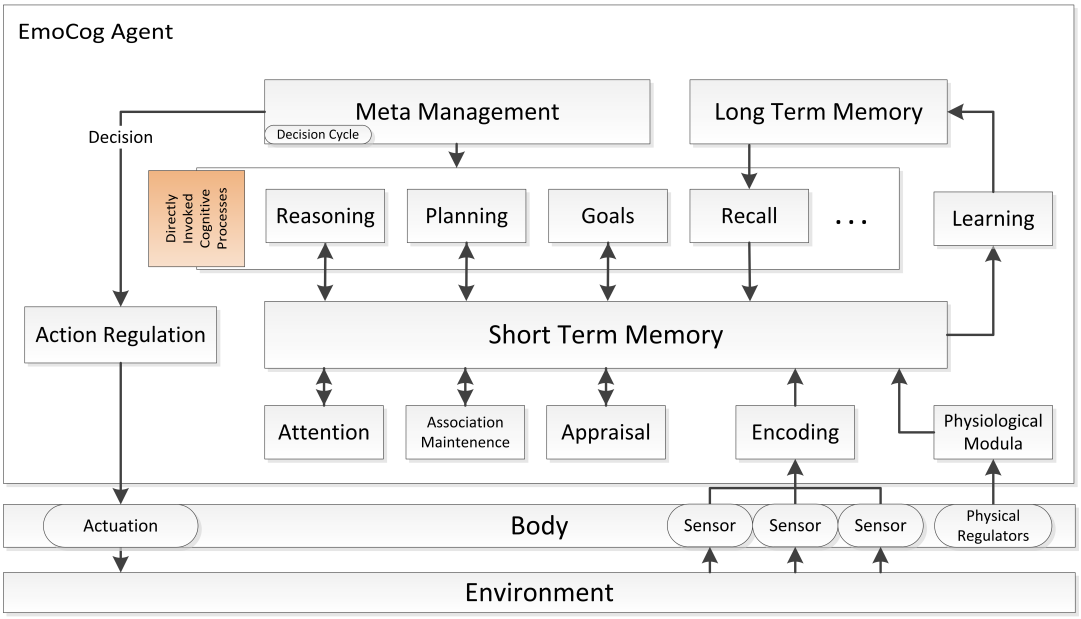

Figure 1: EmoCog cognitive architecture

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 which 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 metamanagement. 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 runs[1].
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

[1]Processes 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 metamanagement 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

Anderson, J.; Bothell, D.; Byrne, M.; Douglass, S.; Lebiere, C.; and Qin, Y. 2004. An integrated theory of the mind. *Psychological Review* 111(4):1036–1060.

Anderson, J. R. 1983. A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior* 22(0-00).

Bechara, A.; Damasio, H.; and Damasio, A. 2000. Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*.

Becker-Asano, C., and Wachsmuth, I. 2009. Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems* 20(1):32–49.

Bower, G. 1981. Mood and Memory. *American psychologist*.

Bower, G. 1983. Affect and Cognition. *Philosophical Transactions of the Royal Society of London* B(302):387–402.

Bower, G. 1992. *How might emotions affect learning*.

Cochran, R.; Lee, F.; and Chown, E. 2006. Modeling Emotion: Arousals Impact on memory. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 1133–1138. Citeseer.

Damasio, A. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam Adult.

De Martino, B.; Kumaran, D.; Seymour, B.; and Dolan, R. 2006. Frames, biases, and rational decision-making in the human brain. *Science* 313(5787):684.

Easterbrook, J. 1959. The Effect of Emotion on Cue Utilization and the Organization of Behavior. *Psychological Review* 66(3):183–201.

Finucane, M. L.; Alhakami, A.; Slovic, P.; and Johnson, S. M. 2000. The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making* 13(1):1–17.

Frijda, N. 1987. Emotion, cognitive structure, and action tendency. *Cognition & Emotion* 1(2):115–143.

Fum, D., and Stocco, A. 2004. Memory, emotion, and rationality: An ACT-R interpretation for Gambling Task results. In *Proceedings of the Sixth International Conference on Cognitive Modelling. Mahwah, NJ: Lawrence Erlbaum*. Citeseer.

Gmytrasiewicz, P., and Lisetti, C. 2000. Using decision theory to formalize emotions in multi-agent systems. *Proceedings Fourth International Conference on MultiAgent Systems* 391–392.

Ingrand, F.; Georgeff, M.; and Rao, A. 1992. An architecture for real-time reasoning and system control. *IEEE EXPERT INTELLIGENT SYSTEMS and THEIR APPLICATIONS* 34–44.

Kaufman, B. E. 1999. Emotional arousal as a source of bounded rationality. *Journal of Economic Behavior & Organization* 38:135–144.

Laird, J. 2008. Extending the Soar cognitive architecture. In *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*.

Langley, P.; Laird, J.; and Rogers, S. 2009. Cognitive architectures: Research issues and challenges. *Cognitive Systems Research* 10(2):141–160.

Marinier, R., and Laird, J. 2008. Emotion-Driven Reinforcement Learning. *Cognitive Science*.

Marinier, R.; Laird, J.; and Lewis, R. 2009. A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*.

Marsella, S., and Gratch, J. 2009. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*.

Ohman, A.; Flykt, A.; and Esteves, F. 2001. Emotion Drives Attention : Detecting the Snake in the Grass. *Emotion* 130(3):466–478.

Ortony, A.; Clore, G.; and Collins, A. 1988. *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

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

Schorr, A. 2001. *Appraisal: The evolution of an idea.* 20–33.

Simon, H. 1967. Motivational and Emotional Controls of Cognition. *Psychological Review*.

Sloman, A. 2001. Varieties of affect and the cogaff architecture schema. *Proceedings Symposium on Emotion, Cognition, and ...*.

Slovic, P.; Finucane, M.; Peters, E.; and Macgregor, D. G. 2007. *The Affect Heuristic*. Cambridge University Press. 397–420.

Smith, C., and Lazarus, R. 1990. Emotion and adaptation. *Handbook of personality: Theory and research.*

Sun, R. 2006. The CLARION cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction: From cognitive ...*.

Yerkes, J. D., and Dodson, R. M. 1908. The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology* 18:459–482.