

# Modeling the Cognitive Antecedents and Consequences of Emotion

Jonathan Gratch

Institute for Creative Technologies, University of Southern California

13274 Fiji Way, Marina del Rey, CA 90292

[gratch@ict.usc.edu](mailto:gratch@ict.usc.edu)

Stacy Marsella

Information Sciences Institute, University of Southern California

4676 Admiralty Way, Marina del Rey, CA 90292

[marsella@isi.edu](mailto:marsella@isi.edu)

Paolo Petta

Center for Brain Research, Medical University of Vienna

Freyung 6/2, A 1010 Vienna, Austria

[paolo.petta@meduniwien.ac.at](mailto:paolo.petta@meduniwien.ac.at)

## Introduction

The last decade has seen an explosion of interest in emotion in both the social and computational sciences. Emotions arise from interactions with both people and technology. They color human perception and decision making and shape a person's moment-to-moment responses to their social and physical environment. Emotions are expressed through changes in speech, facial expression, posture and physiological processes, and these changes provide essential clues to a person's beliefs, desires, intentions and likely future behavior. Recognizing and exploiting such influences can have broad impact across a variety of disciplines: Incorporating the influence of emotion increases explanatory power of models of human decision making (Loewenstein & Lerner, 2003 and the articles by Parkinson and by Clore and Palmer in this special issue); Responding to a student's emotions can enhance the effectiveness of human or computer tutors (Conati & MacLaren, 2004; Graesser et al., 2008; Lepper, 1988); And modeling emotional influences can enhance the fidelity of social simulations, including how crowds react in disasters (Lyell, Flo, & Mejia-Tellez, 2006; Silverman, Johns, O'Brien, Weaver, & Cornwell, 2002), how military units respond to the stress of battle (Gratch & Marsella, 2003), and even large social situations as when modeling the economic impact of traumatic events such as 9/11 or modeling inter-group conflicts (Marsella, Pynadath, & Read, 2004).

More generally, an understanding of the cognitive and social function of human emotion complements the rational, individualistic and disembodied view of cognition that underlies most artificial intelligence and cognitive system research. Emotional influences that seem irrational on the surface may have important social and cognitive functions that would be required by any intelligent system. For example, Herb Simon (1967) theorized that emotions serve to interrupt normal cognition when unattended goals require servicing. Robert Frank argues that social emotions such as anger and guilt reflect a mechanism that improves group utility by minimizing social conflicts, and thereby explains people's "irrational" choices to cooperate in social games such as the prisoner's

dilemma (Frank, 1988). Similarly, Alfred Mele (2001) claims that "emotional biases" such as wishful thinking reflect a rational mechanism that more accurately accounts for social costs, such as the cost of betrayal when a parent defends a child despite strong evidence of their guilt in a crime (see also Ito, Pynadath, & Marsella, 2008). At the same time, findings on non-conscious judgments (e.g., Barrett, Ochsner, & Gross, 2007; Moors, De Houwer, Hermans, & Eelen, 2005) have enriched our understanding of how cognitive style is shaped by the socio-emotional context, often in adaptive ways. More broadly, appraisal theorists such as Lazarus (1991), Frijda (1987) and Scherer (2001) have argued that emotions are intimately connected with how organisms sense events, relate them to internal needs (e.g., is this an opportunity or a threat?), characterize appropriate responses (e.g., fight, flight or plan) and recruit the cognitive, physical and social resources needed to adaptively respond. Thus, an understanding of emotion's function can inform the design of cognitive systems that must survive in a dynamic, semi-predictable and social world.

This special issue of the Journal of Cognitive Systems Research gives a cross-section of contemporary psychological and computational research on the interplay of cognition and emotion. The articles arise from a recent interdisciplinary symposium on Modeling the Cognitive Antecedents and Consequences of Emotion that brought together leaders in psychological and computational approaches to emotion for three days of intense discussion. The articles represent the current state of an ongoing discussing to bridge the divide between computational and psychological perspectives on emotion, illustrating both that theories on the function of emotion in human cognition can yield key insights into the design and control of intelligent entities in general, and that computational models of human mental processes can inform psychological theories through the exercise of concretizing them into working and testable systems.

## **Background**

Emotion research spans an enormous body of work across a wide range of scientific disciplines and contains within it a diversity of competing theoretical perspectives. This special issue emphasizes appraisal theories of emotion (Ellsworth & Scherer, 2003), the dominant psychological theory over the last twenty years and, with its emphasis on cognitive processes, the most congenial to cognitive systems research. Before describing the contributions to the special issue we first review differing theoretical perspectives and recent progress on computational models of emotional processes.

Although there is no consensus on the definition of the term *emotion*, the following proposal by Scherer is illustrative: emotions are episodes of synchronized recruitment of mental and somatic resources allowing an organism to adapt to or cope with a stimulus event subjectively appraised as being highly pertinent to its needs, goals, and values (Scherer, 2004). This definition emphasizes the alignment of several distinct components including cognitive processes (e.g., appraisal), physiological processes (e.g., ANS arousal), behavioral tendencies and responses (e.g., facial expressions). Other theorists also emphasize the importance of awareness (e.g., "I feel mad!"). Most research treats emotion as short-term and changeable in contrast with longer term moods or dispositional tendencies (e.g., personality). Some theorists also emphasize that emotions are intentional

in the sense that they make reference to a specific entity or situation. For example, one may be in an angry mood (a non-intentional state) but to experience the emotion of anger, one must be angry at something.

Theories differ in which components are intrinsic to an emotion (e.g., cognitions, somatic processes, behavioral tendencies and responses), the relationship between components (e.g. do cognitions precede or follow somatic processes), and representational distinctions (e.g. is anger a prototype or a natural kind). For example, *discrete emotion* theories argue that emotions are best viewed as a set of discrete sensory-motor programs (Ekman, 1992; LeDoux, 1996; Öhman & Wiens, 2004). Each of these programs consists of a coherent brain circuit that links eliciting cognitions and somatic responses into a single neural system. At the other extreme, *dimensional* theories (e.g., Russell, 2003) argue emotions are simply cognitive labels we apply retrospectively to sensed physiological activation, which, rather than consisting of discrete motor programs, is characterized in terms of broad bipolar dimensions such as valence and arousal (e.g. I feel negative arousal in a context where I've been wronged, therefore I must be angry). Each of these perspectives has merit and its own body of empirical support and it remains an open challenge to reach an overall synthesis (see Parkinson's article in this special issue).

Since the 1980s, appraisal theories have become a major theoretical perspective in the study of emotion and the dominant contemporary theory underlying computational models of emotional processes. Appraisal theories emphasize the cognitive antecedents of emotional experience. The central tenant of appraisal theories is that the organism's evaluation of its circumstances plays the primary role in eliciting and differentiating emotional responses. Appraisal theories posit a set of discrete judgments, called appraisal variables, which characterize the impact of events (real or imagined) on the organism's beliefs and desires. Some of these proposed variables include pleasantness, expectedness and coping potential. According to appraisal theories, these judgments largely determine the organism's emotions and behavioral responses. For example, an unexpected negative event may provoke fear and a tendency to freeze or run away. In this sense, appraisal theories resemble the discrete-emotion perspective in proposing a coherent linkage between elicitation (in terms of appraisal) and somatic response, but they differ in claiming a far richer and more flexible mapping between elicitation and response that better captures the subtlety and richness of human emotion.

Although individual appraisal theories differ in terms of their posited appraisal dimensions and their process assumptions, computational models of emotion have been most influenced by the appraisal theory of Ortony, Clore and Collins (1988), chiefly as it is described with a clarity that can be readily translated into a computer program. Clark Elliott's Affective Reasoner was the first attempt to realize this theory (Elliott, 1992). Most subsequent computational approaches have focused on the appraisal component of emotion, proposing more general and comprehensive techniques for deriving appraisal variables from a representation of perceptions, knowledge and goals. For example, whereas Affective Reasoner used hand-crafted rules (e.g., during a football match, a goal scored by my opponent is *undesirable*), the subsequent EM system (Neal Reilly, 1996)

divided appraisal into general reasoning mechanisms that operated over domain-specific knowledge structures.

One unfortunate consequence of this reliance on Ortony, Clore and Collins' model is that it focused most concretely on the cognitive *structure* of emotions (i.e., appraisal dimensions) but not the overall emotion process, and the resulting computational models reflect this narrowness. Other appraisal theorists, such as Richard Lazarus (1991), Nico Frijda (1987) and Klaus Scherer (2001), proposed more comprehensive theories that not only encompassed a wider range of emotional components (e.g., cognitions, somatic processes, behavioral tendencies and responses) but also articulated basic process assumptions whereby emotions continuously influence and are influenced by cognition. A "second wave" of computational models has begun to address these more comprehensive theories (Dias & Paiva, 2005; Gratch & Marsella, 2004; Hudlicka, 2006; Louchart, Aylett, & Dias, 2007; Marinier & Laird, 2004; Rank & Petta, 2005), two of which are presented in this special issue.

These more comprehensive appraisal theories and corresponding computational models are not without their limitations, and as appraisal theories have developed and become established, their limits have also become increasingly clear. For example, whereas appraisal theories focus on how goal-relevant aspects of a situation impact emotion; it is now clear that irrelevant factors (such as sad music or the rainy weather) can profoundly impact one's emotional state and subsequent action tendencies. Such effects hint at a more fundamental criticism of appraisal theories: namely that they reflect an outdated "Cartesian" view of the mind as a disembodied symbol system. Recent work on embodied cognition emphasizes that emotion develops moment-by-moment as the person adjusts to the changing demands of their environment and is perhaps best seen as a dynamic, embodied and situated process rather than an appraisal of cognitive representations (see Niedenthal, 2007 and the articles by Parkinson and by Clore and Palmer in this special issue).

### **Preview of the Articles**

The special issue brings together five papers addressing recent developments on psychological and computational appraisal theory. These papers summarize the current state-of-the-art, identify recent criticisms of appraisal models and propose potential solutions to these criticisms.

In the first article, Rainer Reisenzein provides an elegant framework for conceptualizing appraisal theories of emotion: that emotions arise from an appraisal of the state of an agent's beliefs and desires, that appraisals summarize the state of this belief-desire system, and that this mechanism focuses cognitive resources on important changes. Specifically, he proposes that core emotions arise from two appraisal mechanisms: a belief-belief comparator (e.g., surprise arises from a discrepancy from a previously held belief that differs from current perceptions) and a belief-desire comparator (e.g., unhappiness arises from the belief that a desired proposition fails to hold). According to the model, the output of appraisal is non-symbolic, much like sensory transducers, but provides essential feedback to focus a belief-desire system towards maintaining desired

and avoiding undesired situations. Reisenzein claims the theory has several important advantages. It provides a principled definition of emotions in terms of these two comparison processes. It addresses a recent controversy within appraisal theory research concerning the object-directedness (i.e., intentionality) of emotions, a point that will be emphasized in a later chapter by Clore. Finally, it contributes to the debate within cognitive science on the relationship between emotion and cognition.

In the next article, Clore and Palmer critique appraisal theory by illustrating how emotion can arise from factors outside of appraisal judgments. Appraisal theories, with their emphasis on a cognitive assessment of the situation, imply a direct and cognitively penetrable linkage between appraisal and the resulting emotion, and though some appraisal theorists are careful not to form this implication (see Ellsworth & Scherer, 2003), many computational models assume 1) that emotions arise if and only if specific appraisal patterns occur, and 2) that these appraisals uniquely determine the subsequent cognitive response. Clore and Palmer discuss several empirical findings that undermine this notion of a direct linkage: on the one hand, seemingly unrelated factors (such as the weather) can influence emotional responses; on the other hand, people can generate incorrect self-explanations of why they feel an emotion, suggesting that appraisal might be a retrospective rather than a prospective process. This has strong implications for computational models of appraisal theory.

Brian Parkinson's article reviews the evidence for appraisal theories and raises useful criticisms and limits to their explanatory power. He brings up several alternative perspectives on emotion including transactional, attributional and communicative approaches and proposes a synthesis based on the idea of emotion as a dynamic and continuously readjusted "relational alignment" to the social and physical environment. The article goes into considerable detail on the empirical evidence and criticisms of these approaches. For a computer modeling perspective, one of the greatest benefits of the article is that it identifies how the alternative theoretical perspectives conceptualize emotion processes differently and even define emotion differently. First, it expands a modeler's intellectual horizons beyond appraisal theory that has largely dominated the computational community. It also identifies that there is a real debate out there as to how emotion processes work and even what emotion is. This in turn suggests that a modeler may want to consider these alternatives. To that end, it touches on the strengths and weaknesses of the approaches and hints at ways to combining them.

The remaining two articles present detailed implemented models inspired by psychological theories but illustrating how the exercise of translating theory into a computational system can call assumptions of these theories into question and suggest interesting reformulations. Illustrating the more comprehensive approach to emotion seen in current computational methods, these models go beyond a focus on the elicitation of emotion (i.e., appraisal) seen in many earlier approaches and towards a more comprehensive treatment of the emotion's cognitive antecedents and consequences. They also illustrate how a more comprehensive perspective on appraisal theory can suggest solutions to the criticisms raised by Parkinson, Clore and Palmer.

Mariner, Laird and Lewis aim for a synthesis of cognitive and emotional models by integrating Scherer's (2001) Sequential Checking Theory of appraisal with Newell's (1990) PEACTION theory of cognitive control. This work aspires to advance the state of cognitive systems by using appraisal theory to enable traditional cognitive models (in this case the Soar cognitive architecture) to predict the emotions and behaviors of other entities. It also seeks to advance appraisal theory by relating emotion-eliciting appraisals to specific processing stages within a well-specified cognitive theory. The authors emphasize how this integration can begin to address the dynamic nature of emotions highlighted by Parkinson's article by combining the sequential patterning of appraisal suggested by Scherer with Newell's view of a continuous cycle of perception, encoding, comprehending and acting.

Marsella and Gratch also tackle the question of emotion as a moment-by-moment adjustment to the changing demands of a responsive environment. They use their computational appraisal model, EMA, to carefully analyze a naturalistic emotional episode. They illustrate how emotions in their model unfold over time as a function of dynamics in the world and the unfolding time-course of cognitive processes. This approach achieves a form of sequential patterning similar to Scherer's (2001) Sequential Checking Theory, but in a very different manner than realized by Mariner et. al. By emphasizing a sharp distinction between the construction of an internal representation of the situation (which may be slow and sequential) and appraisal (which is rapid, parallel and continuously shapes this unfolding construction), they argue that there really is no necessary sequence to appraisal. Rather, this order "emerges" from the agent's task demands, the underlying dynamics the environment and the sequential nature of some cognitive processes.

### **Summary**

As cognitive systems research moves beyond simple, static, and nonsocial problem solving, researchers must increasingly confront the challenge of how to allocate and focus mental resources in the face of other (potentially adversarial) social actors, conflicting goals, and events that unfold with uncertainty across a variety of timescales. This leads us naturally into the domain of emotion. Emotions arise from social interaction. They arise from the dissonance we feel between competing goals and conflicting interpretations of the world around us. They arise from the need to make moment-to-moment decisions in the face of a dynamic and uncertain world where we have limited control over direction and time-course of future events. Emotion researchers have long argued that emotions have evolved to help us successfully navigate an uncertain, social and dynamic world. This special issue illustrates how emotion research can spur the development of cognitive systems with this, until now, uniquely human ability.

### **Acknowledgements**

Several individuals contributed to the engaging multidisciplinary workshop and subsequent discussions that led to the creation of this special issue including Jerry Clore, Susanne Kaiser, Agnes Moors, Scott Neal Reilly and Ira Roseman. We gratefully acknowledge the efforts of Robert Trappi to support the initial meeting in Vienna. This

work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM) and the Air Force Office of Scientific Research under the grant #FA9550-06-1-0206. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

- Barrett, L. F., Ochsner, K. N., & Gross, J. J. (2007). On the automaticity of emotion. In J. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 173-217). New York: Psychology Press.
- Conati, C., & MacLaren, H. (2004). *Evaluating a probabilistic model of student affect*. Paper presented at the 7th International Conference on Intelligent Tutoring Systems, Maceio, Brazil.
- Dias, J., & Paiva, A. (2005). *Feeling and Reasoning: a Computational Model for Emotional Agents*. Paper presented at the Proceedings of 12th Portuguese Conference on Artificial Intelligence, EPIA 2005.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169-2000.
- Elliott, C. (1992). *The affective reasoner: A process model of emotions in a multi-agent system* (Ph.D Dissertation No. 32). Northwestern, IL: Northwestern University Institute for the Learning Sciences. (Document Number)
- Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, H. H. Goldsmith & K. R. Scherer (Eds.), *Handbook of the affective sciences* (pp. 572-595). New York: Oxford University Press.
- Frank, R. (1988). *Passions with reason: the strategic role of the emotions*. New York, NY: W. W. Norton.
- Frijda, N. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, 1, 115-143.
- Graesser, A. C., D'Mello, S. K., Craig, S. D., Witherspoon, A., Sullins, J., McDaniel, B., et al. (2008). The relationship between affect states and dialogue patterns during interactions with AutoTutor. *Journal of Interactive Learning Research*, 19, 293-312.
- Gratch, J., & Marsella, S. (2003). Fight the way you train: the role and limits of emotions in training for combat. *Brown Journal of World Affairs*, X(1)(Summer/Fall), 63-76.
- Gratch, J., & Marsella, S. (2004). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4), 269-306.
- Hudlicka, E. (2006). *Depth of Feelings: Alternatives for Modeling Affect in User Models*. Paper presented at the Text, Speech and Dialogue, Brno, Czech Republic.
- Ito, J., Pynadath, D., & Marsella, S. (2008). *Modeling Self-Deception within a Decision-Theoretic Framework* Paper presented at the 8th International Conference on Intelligent Virtual Agents, Tokyo.
- Lazarus, R. (1991). *Emotion and Adaptation*. NY: Oxford University Press.
- LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York, NY: Simon & Schuster.
- Lepper, M. R. (1988). Motivational Considerations in the Study of Instruction. *Cognition and Instruction*, 5(4), 289-309.
- Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In R. Davidson, H. H. Goldsmith & K. Scherer (Eds.), *Handbook of Affective Science* (pp. 619-642): Oxford: Oxford University Press.
- Louchart, S., Aylett, R., & Dias, J. (2007). *Double appraisal for synthetic characters*. Paper presented at the 7th International Conference on Intelligent Virtual Agents, Paris, France.

- Lyell, M., Flo, R., & Mejia-Tellez, M. (2006). *Agent-based simulation framework for studies of cognitive pedestrian agent behavior in an urban environment with crisis*. Paper presented at the 6th International Conference on Complex Systems.
- Marinier, R. P., & Laird, J. E. (2004). *Towards a comprehensive computational model of emotions and feelings*. Paper presented at the 6th International Conference on Cognitive Modeling, Pittsburgh, PA.
- Marsella, S., Pynadath, D., & Read, S. (2004). *PsychSim: Agent-based modeling of social interactions and influence*. Paper presented at the International Conference on Cognitive Modeling.
- Mele, A. R. (2001). *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press.
- Moors, A., De Houwer, J., Hermans, D., & Eelen, P. (2005). Unintentional processing of motivational valence. *The Quarterly Journal of Experimental Psychology*.
- Neal Reilly, W. S. (1996). *Believable Social and Emotional Agents* (Ph.D Thesis No. CMU-CS-96-138). Pittsburgh, PA: Carnegie Mellon University. Document Number)
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Niedenthal, P. M. (2007). Embodying Emotion. *Science*, 316, 1002-1005.
- Öhman, A., & Wiens, S. (2004). The concept of an evolved fear module and cognitive theories of anxiety. In A. Manstead, N. Frijda & A. H. Fischer (Eds.), *Feelings and Emotions* (pp. 58-80). Cambridge: Cambridge University Press.
- Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive Structure of Emotions*: Cambridge University Press.
- Rank, S., & Petta, P. (2005). *Appraisal for a Character-based Story-World*. Paper presented at the 5th International Working Conference on Intelligent Virtual Agents, Kos, Greece.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110, 145-172.
- Scherer. (2004). Preliminary plans for exemplars: Theory. Report for EC funded IST Network of Excellence 507422 (HUMAINE) [Electronic Version]. Retrieved 22nd March 2005, from <http://emotion-research.net/deliverables/D3c.pdf>
- Scherer, K. R. (2001). Appraisal Considered as a Process of Multilevel Sequential Checking. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal Processes in Emotion: Theory, Methods, Research* (pp. 92-120): Oxford University Press.
- Silverman, B. G., Johns, M., O'Brien, K., Weaver, R., & Cornwell, J. (2002). *Constructing Virtual Asymmetric Opponents from Data and Models in the Literature: Case of Crowd Rioting*. Paper presented at the 11th Conference on Computer-Generated Forces and Behavior Representation, Orlando, FL.
- Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review*, 74, 29-39.