*Appears in 21st National Conference on Artificial Intelligence (AAAI06), Boston, MA 2006*

# Towards a Validated Model of "Emotional Intelligence"

**Jonathan Gratch, Stacy Marsella and Wenji Mao**

University of Southern California
gratch@ict.usc.edu, Marsella@isi.edu, mao@ict.usc.edu

**Abstract**

This article summarizes recent progress in developing a validated computational account of the cognitive antecedents and consequences of emotion. We describe the potential of this work to impact a variety of AI problem domains.

## Introduction

The last decade has seen an explosion of interest in emotion in both the social and computational sciences. Within artificial intelligence, we see the development of computational models of emotion as a core research focus that will facilitate advances in the large array of intelligent systems that strive for human-level competence in dynamic, semi-predictable and social environments:

- Applications that presume the ability to interpret the beliefs, motives and intentions underlying human behavior can benefit from a model of how emotion motivates human action, distorts perception and inference, and communicates information about mental state. Some tutoring applications already incorporate emotion into user models [1]. Dialogue and collaborative planning systems could also benefit from such an approach.
- Emotions play a powerful role in social influence: emotional displays seem designed to elicit social responses from other individuals. Such responses can be difficult to suppress and the responding individual may not even be consciously aware of the manipulation. A better understanding of this phenomena would benefit applications that attempt to shape human behavior, such as psychotherapy [2], tutoring [3] and marketing.
- Modeling techniques increasingly strive to simulate emotional-evoking situations such as how crowds react in disasters [4], how military units respond to the stress of battle [5], and even large social situations as when modeling the economic impact of traumatic events such as 9/11 or modeling inter-group conflicts [6]).



More generally, an understanding of the cognitive and social function of human emotion complements traditional rational views of intelligence. Debates about the benefit of emotion span recorded history and were prominent in the early days of artificial intelligence. Several have argued that emotional influences that seem irrational on the surface have important social and cognitive functions that are lacking from the individualistic and disembodied view of cognition from which artificial intelligence stems. For example, Simon [7] argued that emotions make us more reactive by interrupting normal cognition when unattended goals require immediate servicing in the world. Frank argues social emotions such as anger reflect a mechanism that improves group utility by minimizing social conflicts, and thereby explains peoples "irrational" choices to cooperate in social games such as prison's dilemma [8]. Similarly, "emotional biases" such as wishful thinking may reflect a rational mechanism that is more accurately accounting for certain social costs, such as the cost of betrayal when a parent defends a child despite strong evidence of their guilt in a crime [9]. Finally, the exercise of accurately modeling emotion can often spur the development of new agent capabilities. For example, Mao's effort to model anger has led to a general mechanism of social credit assignment and a model of social coercion [10].

This article draws on recently published papers to summarize our recent progress in developing a validated computational account of the cognitive antecedents and consequences of emotion [10-13]. Our goal is to create working models that simulate emotional human behavior for a variety of possible applications, but with a focus on virtual reality-based training [14]. Here, we highlight recent progress in validating this model against human performance data, along the way emphasizing differences between "rational" and emotionally influenced information processing.

## Emotion Theory

Contemporary emotion research suggests emotion exacts pervasive control over cognitive processes. Emotional state can influence what information is available in working memory [15], the subjective utility of alternative choices [16], and even the style of processing [17]. For example,

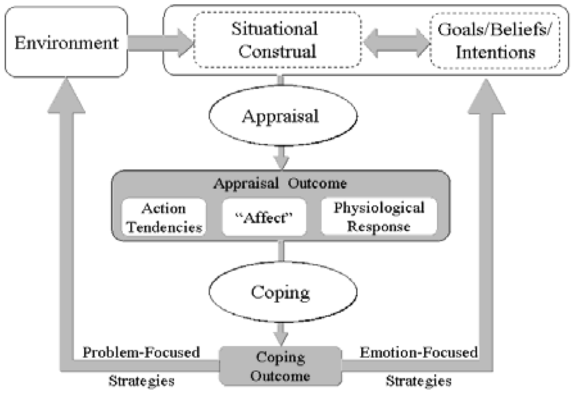


Figure 1: Schematic view of appraisal theory.

people who are angry or happy tend to perform shallower inference and are more influenced by stereotypical beliefs than sad individuals. Neuroscience evidence also underscores the close connection between emotion and centers of the brain associated with higher-level cognition. For example, damage to the connections between emotion and decision-making centers of the brain lead to maladaptive behavior in certain gambling tasks [18]. Collectively, these findings demonstrate that emotion and cognition are closely coupled and suggest emotion has a strong, pervasive and controlling influence over cognition.

There are several theoretical perspectives on the relationship between emotion and cognition. *Appraisal theory* [19] is the predominant psychological theory of emotion (Figure 1). We argue that it is the most fruitful theory of emotion for those interested in the design of symbolic AI systems as it emphasizes the connection between emotion and cognition. Emotion is argued to arise from patterns of individual judgment concerning the *person-environment relationship* (i.e., the perceived relationship between events and an individual's beliefs, desires and intentions). These judgments, formalized as *appraisal variables*, characterize aspects of the personal significance of events (e.g., was this event expected in terms of my prior beliefs? is this event congruent with my goals; do I have the power to alter the consequences of this event?). Patterns of appraisal elicit emotional behavior, but they also trigger stereotypical cognitive responses formalized as qualitatively distinct *coping strategies* (e.g., planning, procrastination or resignation).

Due to its reliance on cognitive judgments and responses, appraisal theory can be recast as a requirement specification for how to build an intelligent system – it claims a superset of the judgments and cognitive strategies considered by most AI systems must be supported in order to correctly detect, classify, and adaptively respond to significant changes to their physical and social environment.

## EMA

EMA is a computational model of the cognitive antecedents and consequences of emotion as posited by appraisal theory [11, 13]. In general terms, we characterize a computational model as processes operating on representations. In this case, the processes involve the interpretation (appraisal) and manipulation (coping) of a representation of the person-environment relationship. In realizing this abstract psychological theory, we draw extensively on common artificial intelligent methods of reasoning and representation. To this end, EMA represents the relationship between events and an agent's internal beliefs desires and intentions by building on AI planning to represent the physical relationship between events and their consequences, and BDI frameworks to represent the epistemic factors that underlie human (particularly social) activities.

Appraisal processes characterize this representation in terms of individual appraisal judgments. These extend traditional AI concerns with utility and probability:

- Desirability: what is the utility (positive or negative) of the event if it comes to pass.
- Likelihood: how probable is the outcome of the event.
- Causal attribution: who deserves credit/blame.
- Controllability: can the outcome be altered by actions under control of the agent.
- Changeability: can the outcome change on its own.

Patterns of appraisal elicit emotional displays, but they also initiate coping processes to regulate the agent's cognitive response to the appraised emotion. Coping strategies work in the reverse direction of appraisal, identifying plans, beliefs, desires or intentions to maintain or alter. These include "problem focused" strategies (e.g. planning) directed towards improving the world (the traditional concern of AI techniques) but also encompasses "emotion-focused" strategies that impact an agent's epistemic and motivational state:

- Planning: form an intention to perform some act (the planner uses intentions to drive its plan generation)
- Seek instrumental support: ask someone that is in control of an outcome for help
- Procrastination: wait for an external event to change the current circumstances
- Denial: lower the perceived likelihood of an undesirable outcome
- Mental disengagement: lower utility of desired state
- Shift blame: shift responsibility for an action toward some other agent

Strategies give input to the cognitive processes that actually execute these directives. For example, planful coping generates an intention to act, leading the planning system to generate and execute a valid plan to accomplish this act. Alternatively, coping strategies might abandon the goal, lower the goal's importance, or re-assess who is to blame.

EMA uses an explicit representation of plans, beliefs, desires and intentions to capture output and intermediate results of processes that relate the agent to its physical and social environment. This represents the agent's current view of the agent-environment relationship, which changes

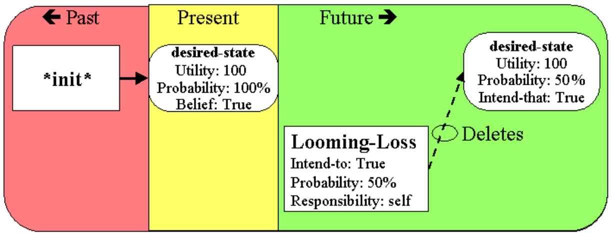


Figure 2: EMA's encoding of the SCPQ loss condition

with further observation or inference. We treat appraisal as a mapping from syntactic features of this representation to individual appraisal variables. Multiple appraisals are aggregated into an overall emotional state that influences behavior. Coping directs control signals to auxiliary reasoning modules (i.e., planning, or belief updates) to overturn or maintain features of the representation that lead to individual appraisals. For example, coping may abandon a cherished desire in response to an uncontrollable threat.

## Validation

Our recent efforts have been directed towards validating EMA's effectiveness in modeling the influence of emotion over human judgments. This involves assessing its consistency with human emotional responses (I/O validity). We are further interested in the more challenging test of whether the inferential mechanisms underlying EMA are consistent with human inference (process validity). Rather than using an abstract overall assessment, such as a subject's assessment of "believability," we directly compare the internal variables of the model to human data, assessing emotional responses, but also the value of appraisal variables, coping tendencies, and in particular, how these assessments change in response to an evolving situation.

There is little established methodology for evaluating emotion models and our group spearheaded such efforts. Our current efforts adopt the following approach: identify a corpus of emotional situations used to validate psychological theories of emotion; encode these situations in our model; contrast the predictions of the model with human responses. We have recently completed two major studies based on this approach which we summarize here.

**Appraisal and Coping Dynamics:** Although human mental processes cannot be observed directly, emotion psychologists assess this information indirectly through interactive questionnaires. The Stress and Coping Process Questionnaire (SCPQ) [20] is one such instrument used to assess coping processes. Subjects are presented stereotypical emotion-evoking episodes and their responses are queried as the episodes evolve. Episodes are constructed from a grammar that encodes prototypical causal relationships between events and goals. For example, in the *loss condition*, a subject might be told of a looming threat to an important goal (e.g. your spouse is threatening a divorce). Subjects are queried on how they would feel in this situation (*emotional response*), how they appraise certain aspects of the current situation (*appraisal variables*) and what strategies they would use to confront the situation (*coping strategies*). They are then presented updates to the situation (e.g., some time has passed and the situation has not improved) and asked how their interpretation changes.

The grammar underlying SCPQ elicits specific patterns of appraisal and coping responses. We use this characteristic to assess the validity of EMA's by comparing these patterns with those produced by the model. Rather than attempting to parse English and use the scale directly, we take advantage of the fact that all of the episodes in the scale correspond to one of four dynamic causal theories. For example, Figure 2 illustrates EMA's encoding of the loss condition. See [12] for details.

*Results:* The results show strong support for the model. SCPQ identifies nine trends that indicate normal emotional responses. EMA is consistent with eight of these trends. EMA also shows close correspondence with the temporal patterns of appraisal and emotional response across the phases of the dynamic scenarios. One departure from the human data is that people often felt they had more control over situations than are predicted by the model, suggesting people were bringing commonsense knowledge to bear that was not explicitly mentioned in the episode descriptions. Another limitation of EMA concerns its ability to reason about social emotions. This is addressed in the next study.

**Causal Attributions:** Some emotions such as guilt and anger involve social judgments of blame and responsibility. Although many intelligent systems reason about the physical causes of outcomes, traditional notions of causality are simply inadequate for explaining such social judgments. Instead, social causality, in theory and as practiced in everyday folk judgments, emphasizes multiple causal dimensions, involves epistemic variables, and distinguishes between physical cause, responsibility and blame.

We have begun to model how people form judgments of blame and responsibility, including not only causal factors, but also epistemic variables such as freedom of choice, intention and foreknowledge [10, 21]. As a result, an actor may physically cause an event, but be absolved of responsibility and blame, or conversely, blamed for what she did not physically cause.

Using a variation of our methodology, we contrasted performance of this model against human performance data on hypothetical scenarios, but used the model itself to systematically generate scenario variants that should be appraised differently. As a starting point, we adopt the well-known "company program scenario that involves two corporate executives discussing a policy that may harm the environment [22]. In our study, descriptions of the scenario are organized into separate labeled statements of evidence. We then added, deleted or altered these lines in order to change intermediate inferences made by the model. For example, if our model suggests that a particular line of evidence is necessary to infer coercion, than an obvious variation would be to eliminate that line of evidence.

*Results:* A questionnaire followed the presentation of each scenario. Each question was designed to test the belief about one judgment underlying the model (e.g., did agent A intend X). In terms of I/O validity, we measured the agreement of the model and each subject using *Kappa statistic*. The average *Kappa* agreement of the model and subjects is 0.732, indicating substantial agreement. In terms of process validity, we compared subject responses with intermediate inferences of the model. In the model, each belief is derived by a specific inference rule, so each question in the questionnaire corresponds to the firing of one rule. Currently, we have 37 dialogue and causal inference rules in the model. This survey study covers 19 of them. To assess the inference rules, we compare the conditions of each rule with the evidence people use in forming each answer. We derive an accuracy value for each rule based on a confusion matrix built from subjects responses. The results show high accuracy (in the range of 70-90% for each rule).

## Conclusion

Our empirical studies to date show strong empirical support for our approach. One concern with our current validation methodology is its reliance on self-reports of imagined situations. Although this is standard in contemporary emotion research and results are generally consistent with those obtained by other means, self-reports are rightly criticized as possibly saying more about how people think about emotion retrospectively rather then how they actually behave in emotional situations. As self-reports are the primary means for assessing appraised emotional state, this is a concern, not just for the present study, but for the field of emotion research in general. The use of virtual humans and virtual environments points to one way to address this concern. Rather than presenting subjects a fixed textual description of a situation, they could be presented with a virtual facsimile of the episode. And rather than asking subject how they might act in such a situation, they could be provided the means of actually acting out in the episode and possibly changing its evolution through their actions.

More generally, we see the modeling emotion as increasingly vital as AI matures beyond simple, static and nonsocial problem solving. Human emotion clearly exacts a controlling influence over cognition and a functional analysis of emotion's impact can contribute to the discourse on how to achieve human-level intelligence. As a theory designed to characterize emotional responses to a wide span of human situations, appraisal theory can suggesting core cognitive functions often overlooked by traditional AI.

## References


1. Conati, C. and H. MacLaren. *Evaluating A Probabilistic Model of Student Affect*. in *7th International Conference on Intelligent Tutoring Systems*. 2004. Maceio, Brazil.
2. Marsella, S., W.L. Johnson, and C. LaBore. *Interactive Pedagogical Drama*. in *Autonomous Agents*. 2001. Montreal.
3. Lester, J.C., B.A. Stone, and G.D. Stelling, *Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments.* User Modeling and User-Adapted Instruction, 1999. **9**(1-2): p. 1-44.
4. Silverman, B.G., et al. *Constructing Virtual Asymmetric Opponents from Data and Models in the Literature: Case of Crowd Rioting*. in *CGF-BR*. 2002. Orlando, FL.
5. Gratch, J. and S. Marsella, *Fight the way you train: the role and limits of emotions in training for combat.* Brown Journal of World Affairs, 2003. **X(1)**(Summer/Fall): p. 63-76.
6. Marsella, S., D. Pynadath, and S. Read. *PsychSim: Agent-based modeling of social interactions and influence*. in *International Conference on Cognitive Modeling*. 2004.
7. Simon, H.A., *Motivational and emotional controls of cognition.* Psychological Review, 1967. **74**: p. 29-39.
8. Frank, R., *Passions with reason: the strategic role of the emotions*. 1988, New York, NY: W. W. Norton.
9. Mele, A.R., *Self-Deception Unmasked*. 2001, Princeton, NJ: Princeton University Press.
10. Mao, W. and J. Gratch. *Evaluating a computational model of social causality and responsibility*. in *AAMAS* 2006.
11. Gratch, J. and S. Marsella. *Tears and Fears: Modeling Emotions and Emotional Behaviors in Synthetic Agents*. in *Autonomous Agents*. 2001. Montreal, Canada: ACM Press.
12. Gratch, J. and S. Marsella. *Evaluating the modeling and use of emotion in virtual humans*. in *AAMAS* 2004. New York.
13. Marsella, S. and J. Gratch. *Modeling coping behaviors in virtual humans: Don't worry, be happy*. in *AAMAS* 2003..
14. Swartout, W., et al., *Toward Virtual Humans.* AI Mag, 2006.
15. Bower, G.H., *Emotional mood and memory.* American Psychologist, 1991. **31**: p. 129-148.
16. Lerner, J.S. and D. Keltner, *Beyond valence: Toward a model of emotion-specific influences on judgement and choice.* Cognition and Emotion, 2000. **14**: p. 473-493.
17. Bless, H., N. Schwarz, and M. Kemmelmeier, *Mood and stereotyping: The impact of moods on the use of general knowledge structures.* European Review of Social Psychology, 1996. **7**: p. 63-93.
18. Bechara, A., et al., *Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-Making.* Journal of Neuroscience, 1999. **19**(13).



This work was sponsored by the U. S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

19. Scherer, K.R., A. Schorr, and T. Johnstone, eds. *Appraisal Processes in Emotion*. Affective Science, ed. R.J. Davidson, P. Ekman, and K.R. Scherer. 2001, Oxford University Press.
20. Perrez, M. and M. Reicherts, *Stress, Coping, and Health*. 1992, Seattle, WA: Hogrefe and Huber Publishers.
21. Mao, W. and J. Gratch. *Social Judgment in Multiagent Interactions*. in *AAMAS* 2004.
22. Knobe, J., *Intentional Action and Side-Effects in Ordinary Language.* Analysis, 2003. **63**: p. 190-193.