To appear in the *Journal of Autonomous Agents and Multiagent Systems (special issue on the best of AAMAS 2004)*
*** DRAFT ***

Evaluating a computational model of emotion

Jonathan Gratch
University of Southern California, Institute for Creative Technology
13274 Fiji Way, Marina del Rey, CA 90292
gratch@ict.usc.edu

Stacy Marsella
University of Southern California, Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA 90292
marsella@isi.edu

Abstract

Spurred by a range of potential applications, there has been a growing body of research in computational models of human emotion. To advance the development of these models, it is critical that we evaluate them against the phenomena they purport to model. In this paper, we present one method to evaluate an emotion model that compares the behavior of the model against human behavior using a standard clinical instrument for assessing human emotion and coping. We use this method to evaluate the EMA model of emotion [1-3]. The evaluation highlights strengths of the approach and identifies where the model needs further development.

1. Introduction

The interest in general computational models of emotion and emotional behavior has been steadily growing in the agent research community. Although such models can ideally inform our understanding of human behavior, we see the development of computational models of emotion as a core research focus that will facilitate advances in the large array of computational systems that model, interpret or influence human behavior:

- Many applications presume the ability to correctly interpret the beliefs, motives and intentions underlying human behavior and could benefit from a model of how emotion motivates action, distorts perception and inference, and communicates information about mental state. Indeed, some tutoring applications have explored this potential to inform user models [4, 5] and dialogue systems, mixed-initiative planning systems, or systems that learn from observation could also benefit from such an approach.

- Emotions play a powerful role in social influence: certain emotional displays seem designed to elicit particular social responses from other individuals, and arguably, 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 applications [6, 7], tutoring systems [8-10], and marketing applications [11, 12].

- Modeling applications must account for how people behave when experiencing intense emotion including disaster preparedness (e.g., when modeling how crowds react in a disaster [13]), training (e.g., when modeling how military units respond in a battle [14]), and even large scale social simulations (e.g., when modeling the economic impact of traumatic events such as 9/11 or modeling inter-group conflicts [15]).

- Models of emotion may give insight into building models of intelligent behavior *in general*. Several authors have argued that emotional influences that seem irrational on the surface have important social and cognitive functions that would be required by any intelligent system [16-21]. For example, social emotions such as anger and guilt may reflect a mechanism that improves group utility by minimizing social conflicts, and thereby explains peoples "irrational" choices in social games such as prison's dilemma [22]. Similarly, "delusional" coping strategies such as wishful thinking may reflect a rational mechanism that is more accurately accounting for certain social costs [23]. Finally, the exercise of accurately modeling emotion can often spur the development of new mechanisms that may be of general use to agent systems (e.g., Mao's effort to model anger led to a general mechanism of social credit assignment and a model of social coercion [24]).

Figure 1: The first author interacting with a virtual character in the Mission Rehearsal Exercise, which allows trainees to speak with life-sized characters for task-oriented training. Characters incorporate the EMA emotional model to inform decision making, perceptual attention and nonverbal behavior.

Our work is particularly influenced by a growing body of work in the design of virtual humans, software artifacts that act like people but exist in virtual worlds, interacting with immersed humans and other virtual humans (see [25] for an overview of developments in this area). Virtual human technology is being applied to training applications [26], health interventions [27], marketing [11] and entertainment [28]. Emotion models have also been proposed as a critical component of more effective human computer interaction that factors in the emotional state of the user [29, 30].

Emotion models can play a critical role in advancing the capabilities of virtual humans. Virtual humans are designed to behave like people and emotions impact human behavior in many ways: emotion impacts decision making, action selection, memory, attention, voluntary muscles, social interactions, etc., all of which may subsequently impact emotional state (e.g., see [2]). Further, emotions are an important cue to a person's mental state and are frequently attributed to humans in the absence of any visible signal (e.g., he is angry but suppressing it) so failure to model and express emotions in virtual humans leads users to misinterpret the virtual human behavior. Virtual humans that model and express emotions also provide more engaging experiences for the immersed human users [31] .

Whereas emotion models can aid virtual human design, virtual humans can play a complementary role in advancing the state of emotion models. Incorporating an emotion model into a virtual human provides the opportunity to address a broad range of mental and physical behaviors and highlights the narrowness of existing computational models. For example, although there are now several models of the sources of emotion (i.e., appraisal), there has been far less computational work in modeling the wide-ranging impact human emotions have on the cognitive and behavioral mechanisms that virtual humans provide. The broad, interactive nature of virtual human systems begs the possibility to explore these richer emotional influences.

In our research, we have been developing a general computational model of human emotion, EMA (**Em**otion and **A**daptation) [1-3], that attempts to account for both the factors that give rise to emotions as well as the wide-ranging impact emotions have on cognitive and behavioral responses, particularly coping responses. The model has been implemented and used to create a significant application where people can interact with the virtual humans through natural language in high-stress social settings (Figure 1) [26, 32].

Given the broad influence emotions have over behavior, evaluating the effectiveness of such a general architecture presents some unique challenges. Emotional influences are manifested across a variety of levels and modalities. For instance, there are telltale physical signals: facial expressions, body language, and certain acoustic features of speech. There are also influences on cognitive processes, including coping behaviors such as wishful thinking, resignation, or blame-shifting. Unlike many phenomena studied by cognitive science, emotional responses are also highly variable, differing widely both within and across individuals depending on non-observable factors like goals, beliefs, cultural norms, etc. And unlike work in rational decision making, there is no accepted, idealized model of emotional responses or their dynamics that we can use as a gold standard for evaluating techniques.

In the virtual human research community, the current state-of-the-art in evaluation has relied largely on the concept of "believability" in demonstrating the effectiveness of a technique: A human subject is allowed to interact with a system or see the result of some system trace, and is asked how believable the behaviors appear; it is typically left to the subject to interpret what is meant by the term. One obvious limitation with this approach is that there seems to be no generally agreed definition of what "believability" means, how it relates to other similar concepts such as realism (or example, in a health-intervention application developed by one of the authors, stylized cartoon animation was judged to be highly believable even though it was explicitly designed to be unrealistic along several dimensions [6]).

In our view, research into emotion models, and more generally virtual human technologies, would benefit from evaluation methodologies that go beyond abstract overall assessments such as self-reports of believability. In particular, we seek evaluations that address more specific questions about the functional and dynamic behavior of our models and how that behavior compares to human behavior. In cases where relevant data on human behavior is available, then model behavior can be contrasted with such behavior. Often it is not available, in which cases we may need to collect corresponding human data. Comparisons between model and human data can then be done with respect to the input and behavior of the model. More ambitiously, even finer grain comparisons can be done between the internal variables of the model that mediate its behavior and corresponding mediating variables in the human data, when such variables are available in the human data.

Consistent with this view, the study described here seeks to evaluate the "process validity" of an emotion model: does the EMA model generate cognitive influences that are consistent with human data on the influences of emotion, specifically with regard to how emotion shapes perceptions and coping strategies, and how emotion and coping unfold over time. In other words, does the EMA model of emotion create the right cognitive dynamics? To assess this question, we directly compare the internal and external variables of the model to human data and how these variables change in response to an evolving situation.

2. Appraisal Theory (a review)

Motivated by the need to inform the design of symbolic systems, our work is based on appraisal theories of emotion that emphasize the cognitive and symbolic influences of emotion and the underlying processes that lead to this influence [33], in contrast to models that emphasize lower-level processes such as drives and physiological effects [34]. In particular, our work is informed by Smith and Lazarus' cognitive-motivational-emotive theory [35].

Appraisal theories argue that emotion arises from two basic processes: appraisal and coping. Appraisal is the process by which a person assesses their overall relationship with its environment, including not only their current condition but past events that led to this state as well as future prospects. Appraisal theories argue that appraisal, although not a deliberative process in of itself, is informed by cognitive processes and, in particular, those processes involved in understanding and interacting with the environment (e.g., planning, explanation, perception, memory, linguistic processes). Appraisal maps characteristics of these disparate processes into a common set of terms called *appraisal variables*. These variables serve as an intermediate description of the person-environment relationship – a common language of sorts – and mediate between stimuli and response (e.g. different responses are organized around how a situation is appraised). Appraisal variables characterize the significance of events from the individual's perspective. Events do not have significance in of themselves, but only by virtue of their interpretation in the context of an individual's beliefs, desires and intention, and past events.

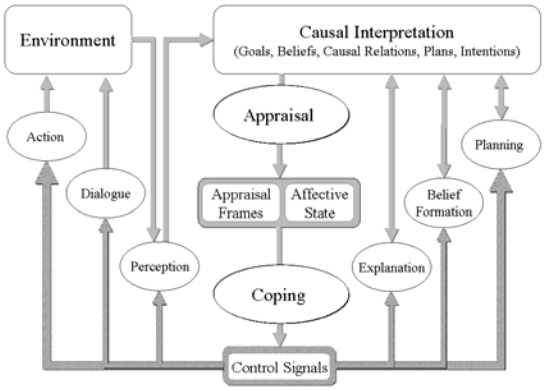

Figure 2: A computational view of Smith and Lazarus

Coping determines how one responds to the appraised significance of events. People are motivated to respond to events differently depending on how they are appraised [36]. For example, events appraised as undesirable but controllable motivate people to develop and execute plans to reverse these circumstances. On the other hand, events appraised as uncontrollable lead people towards denial or resignation. Psychological theories often characterize the wide range of human coping responses into two broad classes: *problem-focused coping* strategies attempt to change the environment; *emotion-focused coping* [33] involves inner-directed strategies for dealing with emotions, for example, by discounting a potential threat or abandoning a cherished goal. The ultimate effect of these strategies is a change in the person's interpretation of their relationship with the environment, which can lead to new (re-) appraisals. Thus, coping, cognition and appraisal are tightly coupled, interacting and unfolding over time [33]: an agent may "feel" distress for an event (appraisal), which motivates the shifting of blame (coping), which leads to anger (re-appraisal). A key challenge for a computational model is to capture this dynamics.

3. A Computational Model

EMA is a computational model based on appraisal theory and described in detail elsewhere [1-3]. Here we sketch the basic outlines. A central tenant in cognitive appraisal theories in general, and Smith and Lazarus' work in particular, is that appraisal and coping center around a person's *interpretation* of their relationship with the environment. This interpretation is constructed by cognitive processes, summarized by appraisal variables and altered by coping responses. To capture this interpretative process in computational terms, we have found it most natural to build on the causal representations developed for decision-theoretic planning (e.g., [37]) and augment them with methods that explicitly model commitments to beliefs and intentions [38]. Plan representations provide a concise representation of the causal relationship between events and states, key for assessing the relevance of events to an agent's goals and for assessing causal attributions. Plan representations also lie at the heart of many autonomous agent reasoning techniques (e.g., planning, explanation, natural language processing). The decision-theoretic concepts of utility and probability are key for modeling appraisal variables of desirability and likelihood. Explicit representations of intentions and beliefs are critical for properly reasoning about causal attributions, as these involve reasoning if the causal agent intended or foresaw the consequences of their actions [39]. As we will see, commitments to beliefs and intentions also play a role in modeling coping strategies.

In EMA, the agent's interpretation of its "agent-environment relationship" is reified in an explicit representation of beliefs, desires, intentions, plans and probabilities, which we refer to as the *causal interpretation* to emphasize the importance of causal reasoning as well as the interpretative (subjective) character of the appraisal process. Following a blackboard-style model, the causal interpretation (corresponding to the agent's working memory) encodes the input, intermediate results and output of reasoning processes that mediate between the agent's goals and its physical and social environment (e.g., perception, planning, explanation, and natural language processing). At any point in time, the causal interpretation represents the agent's current view of the agent-environment relationship, which changes with further observation or inference. We treat appraisal as a set of feature detectors that map features of the causal interpretation into appraisal variables. For example, an effect that threatens a desired goal is assessed as a potential undesirable event. Coping is treated as a control mechanism that identifies a particular intense emotional response to overturn (in the case of negative emotions) or support (in the case of positive ones) and directs control signals to auxiliary reasoning modules to influence their processing (i.e., planning, belief updates, etc.). Coping aims to overturn or maintain those features that yielded the appraisals. For example, coping may attempt to resign the agent to a threat by suggesting the planner abandon a goal. Figure 2 illustrates a reinterpretation of Smith and Lazarus' cognitive-motivational-emotive system consistent with this view.

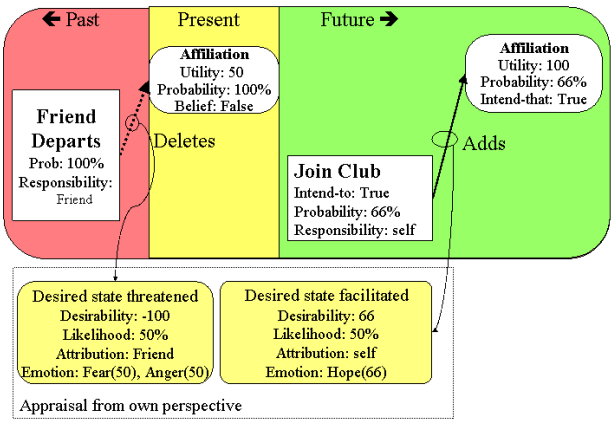

Figure 3: A causal interpretation (aversive condition)

Figure 3 illustrates a causal interpretation and associated appraisal frames. The interpretation is temporally divided into three sections: past, present and future. Rectangles represent actions, ovals represent states and arrows between actions and states represent causal relationships. States are annotated with information about their utility for the agent, their current truth value, any intentions toward the state (intended states are viewed as goals) and probability (indicating a measure of belief for present states and a derived likelihood of goal attainment in the case of future states). Actions are annotated with the agent responsible for executing the action, a derived probability that the action will be executed, and any intentions associated with the action. In the figure, an agent has a single goal (affiliation) that was defeated by the recent departure of a friend (the past "friend departs" action has one effect that deletes the "affiliation" state). This goal might be re-achieved if the agent joins a club. Appraisal assesses each case where an act facilitates or inhibits a state in the causal interpretation and the output of appraisal is represented by explicit frames: the yellow rectangles below the causal interpretation. In the figure, the interpretation encodes two "events," the threat to the currently satisfied goal of affiliation, and the potential re-establishment of affiliation in the future.

Each event is appraised along several appraisal variables by domain-independent functions that examine the syntactic structure of the causal interpretation:

- Perspective: from whose viewpoint is the event judged
- Desirability: what is the utility (positive or negative) of the event if it comes to pass, from the perspective taken (e.g., does it causally advance or inhibit a state of some utility). The utility of a state may be intrinsic (agent X attributes utility Y to state Z) or derived (state Z is a precondition of a plan that, with some likelihood, will achieve an end with intrinsic utility).
- Likelihood: how probable is the outcome of the event. This is derived from the decision-theoretic plan.
- Causal attribution: who deserves credit or blame. This depends on what agent was responsible for executing the action, but may also involve considerations of intention, foreknowledge and coercion (see [24]).
- Temporal status: is this past, present, or future
- Controllability: can the outcome be altered by actions under control of the agent whose perspective is taken. This is derived by looking for actions in the causal interpretation that could establish or block some effect, and that are under control of the agent who's perspective is being judged (i.e, agent X could execute the action).
- Changeability: can the outcome be altered by some other causal agent.

Each appraised event is mapped into an emotion instance of some type and intensity, following the scheme proposed by Ortony et al. [40]. A simple activation-based focus of attention model computes a current emotional state based on most-recently accessed emotion instances.

Coping determines how one responds to the appraised significance of events. Coping strategies are proposed to maintain desirable or overturn undesirable in-focus emotion instances. Coping strategies essentially work in the reverse direction of appraisal, identifying the precursors of emotion in the causal interpretation that should be maintained or altered (e.g., beliefs, desires, intentions and expectations).

Strategies include:

- Action: select an action for execution
- Planning: form an intention to perform some act (the planner uses intentions to drive its plan generation)
- Seek instrumental support: ask for help from someone that has control over an outcome
- Procrastination: wait for an external event to change the current circumstances
- Positive reinterpretation: increase utility of positive side-effect of an act with a negative outcome
- Acceptance: drop a threatened intention
- Denial: lower the probability of a pending undesirable outcome
- Mental disengagement: lower utility of desired state
- Shift blame: shift responsibility for an action toward some other agent
- Seek/suppress information: form a positive or negative intention to monitor some pending or unknown state

Strategies give input to the cognitive processes that actually execute these directives. For example, "planning" will generate an intention to perform the "join club" action, which in turn leads to 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.

Not every strategy applies to a given stressor (e.g., an agent cannot engage in problem directed coping if it is unaware of an action that impacts the situation), however multiple strategies can apply. EMA proposes these in parallel but adopts strategies sequentially. EMA adopts a small set of search control rules to resolve ties. In particular, EMA prefers problem-directed strategies if control is appraised as high (take action, plan, seek information), procrastination if changeability is high, and emotion-focus strategies if control and changeability are low.

In developing a computational model of coping, we have moved away from the broad distinctions of problem-focused and emotion-focused strategies. Formally representing coping requires a certain crispness lacking from the problem-focused/emotion-focused distinction. In particular, much of what counts as problem-focused coping in the clinical literature is really inner-directed in a emotion-focused sense. For example, one might form an intention to achieve a desired state – and feel better as a consequence – without ever acting on the intention. Thus, by performing cognitive acts like planning, one can improve ones interpretation of circumstances without actually changing the physical environment.

4. Related Work

Computational work on emotion can be roughly divided into "communication-driven" approaches that focus on surface manifestation of emotion and its potential for influencing human-computer interaction, and "simulation-driven" approaches that attempt to model the cognitive mechanisms underlying emotion and its potential for influencing cognitive processes (see [25, 41]). Although they are not exclusive, computational models tend to focus exclusively on one of these perspectives. EMA is primarily a simulation-based approach.

In communication-driven approaches, the system chooses emotional behaviors on the basis of its desired impact on the user. For example, Catherine Pelachaud and her colleagues use facial expressions to convey the performative of a speech act [42]. Klesen models the communicative function of emotion, using stylized animations of body language and facial expression to convey a character's emotions and intentions with the goal of helping students understand and reflect on the role these constructs play in improvisational theater [43]. Nakanishi et al. [44] and Cowell and Stanney [45] each evaluated how certain non-verbal behaviors could communicate a character's trustworthiness for training and marketing applications, respectively. Several applications have also tried to manipulate a student's motivations through emotional behaviors: Lester utilized praising and sympathetic emotional displays to provide feedback and increase student motivation in a tutoring application [46]; The VICTEC system (www.vitec.org) exploits general framing effects to promote student empathy with animated characters with the goal of bullying prevention in schools; Biswas et al. [47] also use human-like traits to promote empathy and intrinsic motivation in a learning-by-teaching system.

Simulation-driven approaches aim at simulating the cognitive process underlying emotional behavior, especially the presumed function these processes have in guiding an organism toward adaptive responses to its environment. Simulation-driven approaches are almost exclusively based on appraisal theory as it is the dominant mechanistic account of emotion in contemporary psychology (although this may change with the rise of neuroscience). Simulation-driven models vary considerably from simplistic approaches that require events to be hand-annotated with the appraisals they would produce, to full mechanistic accounts that attempt a deep theory of appraisal-related mechanisms.

EMA relates to a number of past appraisal models of emotion. Although we are perhaps the first to provide an integrated account of coping, computational accounts of appraisal have advanced considerably over the years. In terms of these models, our work contributes primarily to the problem of developing general and domain-independent algorithms to support appraisal, and by extending the range of appraisal variables amenable to a computational treatment. Early ap-

Phase 1: You are unable to find an important document which details your professional qualifications. You need it urgently.
Phase 2: After a few days, the missing document still hasn't appeared. It is highly important that you should always have this document at your disposal
Phase 3: You could not find the certificate in time. You had to show your credentials in another, less satisfactory way

Figure 4: An episode from the SCPQ

praisal models focused on the mapping between appraisal variables and behavior and largely ignored how these variables might be derived, focusing on domain-specific schemes to derive their value variables. For example, Elliott's [48] Affective Reasoner, based on the theory of Ortony, Clore and Collins (the "OCC model") [40], required a number of domain specific rules to appraise events. A typical rule would be that a goal at a football match is desirable if the agent favors the team that scored. More recent approaches have moved toward more abstract reasoning frameworks, largely building on traditional artificial intelligence techniques. For example, El Nasr and collogues [49] use markov-decision processes (MDP) to provide a very general framework for characterizing the desirability of actions and events. This method can represent indirect consequences of actions by examining their impact on future reward (as encoded in the MDP), but it retains the key limitations of such models: they can only represent a relatively small number of state transitions and assume fixed goals. The closest computational approach to what we propose here is WILL [50] that ties appraisal variables to an explicit model of plans (which capture the causal relationships between actions and effects), although WILL does not address the issue of blame/credit attributions, or how coping might alter this interpretation. EMA builds on these prior models, extending them to provide a better characterization of causality and the subjective nature of appraisal that facilitates coping. Prior computational work on the motivational function of emotions has largely focused on using emotion or appraisal to guide action selection. EMA appears to be the first attempt to model the wider range of human coping strategies such as positive reinterpretation, denial, acceptance, shift blame, etc that alter beliefs, goals, etc.

Some simulation-based models are inspired by models other than appraisal theory and, with a few exceptions, focus on low-level cognitive functions. Some have tried to faithfully model what is known about the neuroscience of emotion to give better insight into these processes. For example, LeDoux and colleagues have build a model of the fear circuit identified in his research [51]. Several robotics researchers have been influenced by ethology-inspired drive models to help inform robotic control systems [52, 53]. A few researchers have explored non-appraisal models of the influence of emotion on higher-level cognition, typically by extending classical decision models. For example, Busemeyer's [54] Decision Field Theory attempts to integrate a notion of drives into classical decision theory to explain the influence of emotions on decision making, and Gmytrasiewicz and.Lisetti cast emotion in terms of short-cuts in expected utility calculations [55].

Few computational models of emotion have been formally evaluated and most evaluations have focused on external behaviors driven by the model rather than directly assessing aspects of the emotion process. For example, most evaluations consider the interpretation of external behavior (e.g., are the behaviors believable?). More sophisticated work in this vein has tested more specific effects. For example, Prendenger [56] considered the impact of emotional displays on user stress and confidence and Lester [46] evaluated the impact of emotional feedback on student learning. Additionally, there is now a sizable body of work on the impact of virtual human non-verbal behavior *in general* on human observers (e.g., [57]). A small number of studies have tried to evaluate internal characteristics of an emotion process model. For example, Scheutz [58] illustrated that the inclusion of an emotion process led artificial agents to make more adaptive decisions in a biologically inspired foraging task. We are unaware of any work, other than the work presented here, that has directly compared the dynamic processes of an emotion model against human data.

5. Assessing Cognitive Dynamics

A key question for our model concerns its "process validity": does the model capture the unfolding dynamics of appraisal and coping. Rather than using an abstract overall assessment, such as a subjects 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.

Although human mental processes cannot be observed directly, several clinical instruments have been developed to assess this information indirectly through interactive questionnaires. For example, the Stress and Coping Process Questionnaire (SCPQ) [59] is a clinical instrument used to assess a human subject's coping process against an empirical model of normal, healthy adult behavior. A subject is presented a stereotypical episode and their responses are measured several

times as the episode evolves. For example, they are told to imagine themselves in an argument with their boss and are queried on how they would feel (*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., they are told some time has passed and the situation has not improved) and asked how their emotions/coping would dynamically unfold in light of these manipulations. The episodes are evolved systematically to alter expectations and perceived sense of control. Based on their evolving pattern of responses, subjects are scored as to how closely their reactions correspond to a validated profile on how normal healthy adults respond.

Using such a scale has the advantage that it provides an independently derived corpus of evolving situations and a ready source of human data, though it does not provide data on individual differences. Ideally, we would like to show that EMA captures how an arbitrary individual appraises a situation given knowledge of their initial beliefs and preferences, or at least models the most common response. As a start however, and given the practical difficulties in obtaining individual information, we compare EMA against aggregate data from the SCPQ. This instrument averages observations across multiple subjects and attempts to characterize "typical" human responses. Given the variability of human emotional behavior, we believe it is important to start by comparing against such normalized responses.

Figure 4 illustrates one of the episodes from the SCPQ. The scale consists of several distinct episodes but all are generated from a grammar that encodes two prototypical stressful episodes. Episodes evolve over three discrete phases: an initial state (phase 1), a state where some time passes without change (phase 2), and one of two possible ending phases which can either result in a good or bad conclusion. The *loss condition* prototype presents an episode where some loss is looming in the future (i.e., a threat to an important goal), the loss continues to loom for some time, and then the loss either occurs or is averted. In the *aversive condition* prototype, some bad outcome (i.e., a goal is violated) has occurred but there is some potential to reverse it. After some time, the undesirable outcome is either reversed or the attempt to reverse it fails. In all, there are four canonical situations (loss/good outcome, loss/bad outcome, aversive/good outcome and aversive/bad outcome). The scale contains multiple variants of each of these canonical situations, each with the same underlying causal structure and dynamics but different surface text. The aversive condition is intended to convey a greater sense of control/changeability than the loss condition, and the surface description of these episodes is selected and empirically validated to produce this effect. Figure 4 illustrates one of the "loss/bad outcome" episodes.

When used as a diagnostic tool, a patient would be presented each phase of each episode, would be asked to imagine themselves in that situation, and would then be asked to answer a series of questions to assess their mental state. One set of questions assesses the subject's imagined emotional state. For example, they would be asked to note on a continuum the extent to which the situation would make them feel anxious/nervous versus composed/calm. A second set of questions asks people to characterize the event along standard appraisal dimensions. For example, they would be asked to assess the extent to which the situation would improve of its own accord (corresponding to the appraisal variable of changeability). Finally, a third set of questions queries their propensity for adopting particular coping strategies. For example, they would be asked to indicate on a numeric scale their tendency to behave passively and wait for something to happen. After these questions, the subject is presented the next phase of the episode. After presenting the three phases of an episode and their corresponding questions, a new episode is presented. Note the situations evolve in a fixed way, regardless of how subjects indicate they would act in the situation (an issue we will revisit later in the article).

The scale combines answers across canonical situations to create a profile of how the subject tends to respond emotional, tends to appraise and tends to cope with these situations. These are scored with respect to how closely they follow the trends exhibited by healthy adults.

Trends include:

1.1 Aversive condition should yield appraisals of higher controllability and changeability than the loss condition (this follows from the design of the stimuli)
1.2 Appraisal of controllability and changeability decreases over phases (as likelihood of change drops)
1.3 Valence, a measure of how positive the situation feels, should decrease over phases and there should be a strong difference in valence on negative vs. positive outcomes
1.4 Aversive condition should lead to more anger and less sadness (the developers of the scale claim that this follows from the lack of appraised control in the loss condition)
2.1 Less appraised control should lead to less problem-directed coping
2.2 Less appraised control may produce more passivity
3.1 Lower ambiguity should produce a more limited search for information
3.2 Lower ambiguity should yield more suppression of information about stressor
4 Less appraised control should produce more emotion-focused coping[1]

[1] SCPQ treats this as two distinct sub-trends, distinguishing between two types of emotion-directed strategies. As Smith and Lazarus do not make this distinction, we collapse them.

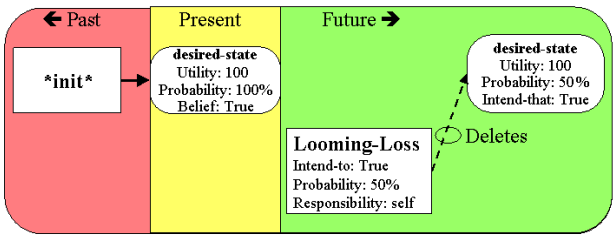

Figure 5: The initial loss condition

We use the scale as a diagnostic instrument to ascertain if the judgments made by our model fall within the expected range of responses of normal healthy adults. 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 canonical causal structures. Thus, we encode the causal structure of these four episodes into EMA.

Method

The four canonical episodes in the SCPQ are encoded as dynamic causal theories. We then compare the model's appraisals and coping strategies to the trends indicated by the scale. Consistent with subjects' inability to actually act on the SCPQ scenarios, we allow EMA to propose coping strategies, but these proposals do not influence subsequent phases (the model proposes strategies but their effects are preempted). As in the SCPQ, the evolution of each episode is determined in advance and occurs in three discrete phases.

EMA requires an encoding of a) the causal structure of the scenario, b) the beliefs and intentions of the agent, and c) the probability and utility of states and d) the probability of events. The first two factors follow straightforwardly from underlying grammar used to generate SCPQ episodes (e.g., there is a goal that is threatened by a possible future action). A remaining issue is to map the qualitative text descriptions into EMA's underlying numeric representation of probabilities and utilities. The SCPQ authors provide some basic constraints to inform this translation: All scenarios involve an important goal; the likelihood of goal attainment should drop in Phase 2, reaching zero (one) in the bad (good) outcome; finally, the aversive condition should be perceived as more controllable than the loss condition. In each condition, we assign the maximum possible utility (100) to represent the importance of the goal. We represent the drop in likelihood of goal attainment (the probability that the looming loss occurs for the loss condition; the probability that the violated goal gets reestablished in the aversive condition) by a function that drops linearly across phases but has a flatter slope (lower initial probability) for the loss condition.

Figure 3 illustrates the initial phase of the domain used for the aversive condition: an action executed by some other agent in the past (friend leaving) makes false some desired state (friendship), but there is some potential action under the control of the agent with no preconditions and one effect that could lead to the desired outcome (join a club). (Labels on states and actions do not impact the model.) In subsequent phases, we alter the subjective probability that the future action will succeed/fail. In the aversive condition, the future action has 66% chance of succeeding, this drops to 33% in phase two, and in phase three is set to either zero or 100%, depending on if the bad or good outcome is modeled. The violated goal has high positive utility (100).

Figure 5 illustrates the initial phase of the domain for the loss condition: a desired state is initially true and a future action potentially executed by another agent may make this state false. Again, probability across phases is adjusted. The chance of the loss succeeding is initially 50%, raises to 75% in phase two, and then is set to either 100% or 0%, depending on if the bad or good outcome is modeled. The desired state has high positive utility (100).

Some terms used in the SCPQ do not map directly to representational primitives in EMA and had to be reinterpreted. EMA does not currently model ambiguity as an explicit appraisal variable. Since the only ambiguity in the SCPQ scenarios relates to the success of pending outcomes, we equate ambiguity with changeability for the purposes of this evaluation. As EMA incorporates the OCC mapping of appraisal variables to emotion types [40], our model also does not directly appraise "sadness" but rather derives "distress" (an undesired outcome has occurred). For this evaluation we equate "sadness" with "distress." Finally, trend 1.3 depends on an overall measure of "valence" that our model does not support. Given that we appraise individual events and an event may have good and bad aspects, for the purpose of this evaluation we derive an aggregate valence measure that sums the intensities of undesirable appraisals and subtracts from the intensities of positive appraisals. We revisit some of these decisions in the discussion.

Table 1: Summary of main results		
	Predicted Trend	EMA
1.1	Aversive more controllable	Yes
	Aversive more changeable	Yes
1.2	Controllability decreases	No
	Changeability decreases phases	Yes
1.3	Negative valence increases	Yes
	Good outcome strongly positive	Yes
1.4	Aversive yields more anger	Yes
	Loss yields more sadness	No
2.1	Low control → low problem-focused	Yes
2.2	Low control → passivity	Yes
3.1	Low ambiguity → low seek info	Yes
3.2	Low ambiguity → suppression	Yes
4	Low control → emotion-focused	Yes

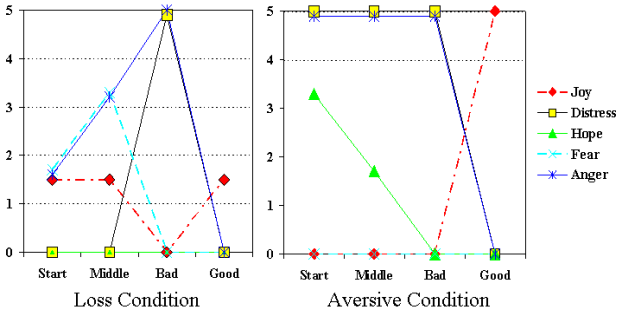

Figure 6: Emotional response of the EMA model across each phase of the two conditions.

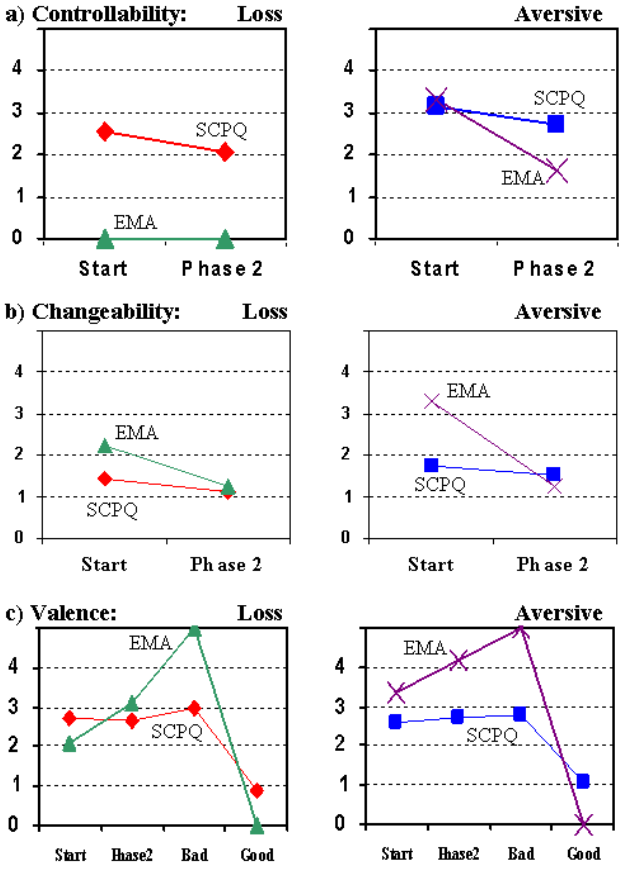

Figure 7: These charts contrast EMA's determination of key appraisal variables with the human SCPQ data reported by Perrez and Reicherts.

Results

Key results are summarized in Table 1 and the raw findings are illustrated in Figures 6 and 7 and Table 2. Trend 1.1 is fully supported by the model: the aversive condition is appraised as more controllable and changeable (Figure 7a and 7b). Trend 1.2 is fully supported for the aversive condition but only partially supported in the loss condition: EMA correctly deduces that the situation is less likely to change across phases, but it determines that the agent has no control over the loss, even in phase 1. This suggests our encoding of the loss condition is overly simplistic, as we will discuss below. Trend 1.3 is fully supported: negative valence increases across phases in both conditions (Figure 7c). Trend 1.4 is also partially supported: there is more anger in the aversive condition, however these is also more sadness, contrary to the prediction (Figure 6). Rather than yielding higher sadness, EMA appraised only fear in the initial phases of the loss condition. Sadness arises only in the bad outcome, when the looming loss becomes certain.

Trends 2.1 and 2.2 are both supported (Table 2). In the aversive condition, the model forms an intention to restore the loss only when its probability of success is high (phase 1). In the loss condition, no known action can influence the pending loss so control is low and no problem-directed strategies are selected. When changeability is high (phase 1 of both conditions), the model suggests a wait-and-see strategy, which is rejected in later phases.

Trends 3.1 and 3.2 are fully supported (Table 2). When the model finds the situation likely to improve on its own (high changeability), it proposes monitoring the truth-value of the state predicate that has high probability of changing. As changeability drops, the model proposes strategies that suppress the monitoring of these states.

Trend 4 is supported (Table 2). As the control drops, proposed strategies tend towards emotion-focused (see Table 1). In the aversive condition, for example, EMA initially forms an intention to execute the "join a club" action (take action) and forms an intention to monitor the truth value of the desired state (seek information). As the likelihood that the action will succeed diminishes, the agent forms an intention to avoid monitoring the status of the desired state (suppress informa-

tion) and begins to lower its attachment to the goal by lowing its utility (mental disengagement). This trend is reinforced in the bad outcome, but is reversed if the action succeeds (good outcome).

Discussion

The model supports most of the trends predicted by SCPQ. Two departures deserve further mention. The loss condition should have produced more sadness than the aversive condition but the opposite occurred. This may indicate that the OCC model's definition of distress, which we have adopted in the current version of EMA, is inappropriate for modeling sadness. OCC appraises distress whenever an undesirable event has occurred, however, many theories argue that the attribution of sadness is also related to the perceived sense of control over the situation (e.g., [33]). This alternative definition could be straightforwardly added to EMA.

A second departure from the human data is that the model appraises zero control in the loss condition across all phases. This is due to the fact that, in our encoding, another agent is represented as the actor for the "looming loss" action, meaning the agent has no direct control and, as this action has no preconditions that could be confronted, there is no indirect control as well. This is clearly too strong and probably does not reflect the causal structure that people recover when they read the SCPQ episodes. This assumption could be relaxed by adding some other action to the domain model executable by the agent that could influence the likelihood of the loss, or to incorporate a notion of shared responsibility.

There are pros and cons to our current methodology from the standpoint of evaluation. On the plus side, the situations in the instrument were constructed by someone outside our research group, and thus constitute a fairer test of the approach's generality than what is often performed (though we are clearly subject to bias in our selection of a particular instrument). Further, by formalizing an evolving situation, this instrument directly assesses the question of emotional dynamics, rather than single situation-response pairs typically considered in evaluations. On the negative side, the scenarios were described abstractly and we had some freedom in how we encoded the situations into a causal model, potentially biasing our results, though this could be mitigated in future experiments. For example, we could employ multiple independent coders.

A more general concern is the use of aggregate measures of human emotional behavior. People show considerable individual difference in their appraisal and coping strategy. In this evaluation, however, we compare the model to aggregate trends that may not well-approximate any given individual. This concern is somewhat mitigated by the fact that the SCPQ scale is intended to characterize individuals in terms of the "normalcy" of their emotional behavior and has been validated for this use. However, a more rigorous test would be to fit to individual reports based on their perceived utility and expectations about certain outcomes.

A final concern is the accuracy of self-reports against which we are contrasting our model's behavior. Self-reports may well say 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. To contrast subject performance against model performance, we could replace the subject with the model and thereby collect data on model performance in the same virtual environment. Such an approach has the potential to not only address the methodological concerns of the present study, but to make a contribution to the field of emotion research in general.

Moving forward in this work, we see closely linked experiments on both the model and human subjects as a particularly effective way to extend our evaluation work. In addition to addressing the self-report issue, collecting human data in concert with experiments on the model will allow us to mitigate several other concerns mentioned above. To mitigate potential encoding biases, we can use the same formal representation of the scenario for both human and model experiments. The specification would be used directly in the model experiments while an automated model to text conversion schema would be used to generate the scenarios for the human experiments. To mitigate the concern about contrasting model performance against aggregate data, we could assess and group subjects based on their dispositional tendencies towards perceived utility and expectations about certain outcomes. The related manipulations would also be performed on the model. Thus, by performing our own human subject experiments that are closely correlated with the model experiments, we can refine the evaluation of the model, as well as potentially refine our understanding of human emotional processes.

6. Summary

Spurred by a range of potential applications, there has been a growing body of research in computational models of human emotion. To advance the development of these models, it is critical that we begin to contrast them against the phenomena they purport to model.

In this article, we presented one method to evaluate an emotion model. We compared the behavior of the computational model against normative behavior, using a standard clinical instrument. Remarkably, the model did quite well. And, as expected, the comparison helped identify where the model needs further development.

As with any new discipline, evaluation of affective systems has lagged far behind advances in computation models. This situation is slowly changing as a number of groups move beyond simple metrics and move toward more differentiated notions of the form and function of expressed behavior (e.g. [45, 56]). This article contributes to this evolution.

Acknowledgements

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.

References

[1] J. Gratch and S. Marsella, "Tears and Fears: Modeling Emotions and Emotional Behaviors in Synthetic Agents," presented at Fifth International Conference on Autonomous Agents, Montreal, Canada, 2001.

[2] S. Marsella and J. Gratch, "Modeling coping behaviors in virtual humans: Don't worry, be happy," presented at Second International Joint Conference on Autonomous Agents and Multi-agent Systems, Melbourne, Australia, 2003.

[3] J. Gratch and S. Marsella, "A domain independent framework for modeling emotion," *Journal of Cognitive Systems Research*, vol. 5, pp. 269-306, 2004.

[4] C. Conati, "Probabilistic Assessment of User's Emotions in Educational Games," *Journal of Applied Artificial Intelligence, special issue on "Merging Cognition and Affect in HCI"*, vol. 16, pp. 555-575, 2002.

[5] C. Conati and H. MacLaren, "Evaluating A Probabilistic Model of Student Affect," presented at 7th International Conference on Intelligent Tutoring Systems, Maceio, Brazil, 2004.

[6] S. Marsella, W. L. Johnson, and C. LaBore, "Interactive Pedagogical Drama," presented at Fourth International Conference on Autonomous Agents, Montreal, Canada, 2000.

[7] B. O. Rothbaum, L. F. Hodges, R. Alarcon, D. Ready, F. Shahar, K. Graap, J. Pair, P. Hebert, B. Wills, and D. Baltzell, "Virtual Environment Exposure Therapy for PTSD Vietnam Veterans: A Case Study," *Journal of Traumatic Stress*, vol. 12, pp. 263-272, 1999.

[8] J. C. Lester, 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*, vol. 9, pp. 1-44, 1999.

[9] K. Ryokai, C. Vaucelle, and J. Cassell, "Virtual Peers as Partners in Storytelling and Literacy Learning," *Journal of Computer Assisted Learning*, in press.

[10] E. Shaw, W. L. Johnson, and R. Ganeshan, "Pedagogical Agents on the Web," presented at Proceedings of the Third International Conference on Autonomous Agents, Seattle, WA, 1999.

[11] E. André, T. Rist, S. v. Mulken, and M. Klesen, "The Automated Design of Believable Dialogues for Animated Presentation Teams," in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, Eds. Cambridge, MA: MIT Press, 2000, pp. 220-255,.

[12] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan, "Human conversation as a system framework: Designing embodied conversational agents," in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, Eds. Boston: MIT Press, 2000, pp. 29-63.

[13] B. G. Silverman, "Human Behavior Models for Game-Theoretic Agents: Case of Crowd Tipping," *CogSci Quarterly*, vol. Fall, 2002.

[14] J. Gratch and S. Marsella, "Fight the way you train: the role and limits of emotions in training for combat," *Brown Journal of World Affairs*, vol. X(1), 2003.

[15] S. Marsella, D. Pynadath, and S. Read, "PsychSim: Agent-based modeling of social interactions and influence," presented at International Conference on Cognitive Modeling, 2004.

[16] M. Minsky, *The Society of Mind*. New York: Simon and Schuster, 1986.

[17] H. A. Simon, "Motivational and emotional controls of cognition," *Psychological Review*, vol. 74, pp. 29-39, 1967.

[18] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Avon Books, 1994.
[19] K. Oatley and P. N. Johnson-Laird, "Cognitive Theory of Emotions," *Cognition and Emotion*, vol. 1, 1987.
[20] A. Sloman and M. Croucher, "Why robots will have emotions," presented at International Joint Conference on Artificial Intelligence, Vancouver, Canada, 1981.
[21] C. Lisetti and P. Gmytrasiewicz, "Can a rational agent afford to be affectless? A formal approach," *Applied Artificial Intelligence*, vol. 16, pp. 577-609, 2002.
[22] R. Frank, *Passions with reason: the strategic role of the emotions*. New York, NY: W. W. Norton, 1988.
[23] A. R. Mele, *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press, 2001.
[24] W. Mao and J. Gratch, "Social Judgment in Multiagent Interactions," presented at Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004.
[25] J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan, and N. Badler, "Creating Interactive Virtual Humans: Some Assembly Required," in *IEEE Intelligent Systems*, vol. July/August, 2002, pp. 54-61.
[26] J. Rickel, S. Marsella, J. Gratch, R. Hill, D. Traum, and W. Swartout, "Toward a New Generation of Virtual Humans for Interactive Experiences," in *IEEE Intelligent Systems*, vol. July/August, 2002, pp. 32-38.
[27] S. Marsella, W. L. Johnson, and C. LaBore, "Interactive pedagogical drama for health interventions," presented at Conference on Artificial Intelligence in Education, Sydney, Australia, 2003.
[28] M. Cavazza, F. Charles, and S. J. Mead, "Agents' Interaction in Virtual Storytelling," presented at Third International Workshop on Intelligent Virtual Agents, 2001.
[29] C. L. Lisetti and D. Schiano, "Facial Expression Recognition: Where Human-Computer Interaction, Artificial Intelligence, and Cognitive Science Intersect," *Pragmatics and Cognition*, vol. 8, pp. 185-235, 2000.
[30] R. W. Picard, *Affective Computing*. Cambridge, MA: MIT Press, 1997.
[31] S. van Mulken, E. André, and J. Muller, "The Persona Effect: How Substantial Is It," presented at Human Computer Interaction Conference, Berlin, 1998.
[32] W. Swartout, R. Hill, J. Gratch, W. L. Johnson, C. Kyriakakis, C. LaBore, R. Lindheim, S. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. Thiebaux, L. Tuch, R. Whitney, and J. Douglas, "Toward the Holodeck: Integrating graphics, sound, character and story," presented at Fifth International Conference on Autonomous Agents, Montreal, Canada, 2001.
[33] R. Lazarus, *Emotion and Adaptation*. NY: Oxford University Press, 1991.
[34] J. Velásquez, "When robots weep: emotional memories and decision-making.," presented at Fifteenth National Conference on Artificial Intelligence, Madison, WI, 1998.
[35] C. A. Smith and R. Lazarus, "Emotion and Adaptation," in *Handbook of Personality: theory & research*, L. A. Pervin, Ed. NY: Guilford Press, 1990, pp. 609-637.
[36] E. Peacock and P. Wong, "The stress appraisal measure (SAM): A multidimensional approach to cognitive appraisal," *Stress Medicine*, vol. 6, pp. 227-236, 1990.
[37] J. Blythe, "Decision Theoretic Planning," in *AI Magazine*, vol. 20(2), 1999, pp. 37-54.
[38] B. Grosz and S. Kraus, "Collaborative Plans for Complex Group Action," *Artificial Intelligence*, vol. 86, 1996.
[39] K. G. Shaver, *The attribution of blame: Causality, responsibility, and blameworthiness*. NY: Springer-Verlag, 1985.
[40] A. Ortony, G. Clore, and A. Collins, *The Cognitive Structure of Emotions*: Cambridge University Press., 1988.
[41] J. Gratch and S. Marsella, "Lessons from Emotion Psychology for the Design of Lifelike Characters," in *Applied Artificial Intelligence*, to appear.
[42] C. Pelachaud, V. Carofiglio, B. D. Carolis, F. d. Rosis, and I. Poggi, "First International Joint Conference on Autonomous Agents and Multiagent Systems," presented at Embodied Contextual Agent in Information Delivering Application, Bologna, Italy, 2002.
[43] M. Klesen, "Using Theatrical Concepts for Role-Plays with Educational Agents," *Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors"*, 2005.
[44] Nakanishi, Shimuzu, and K. Isbister, "Social Agents for Virtual Training," *Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors"*, 2005.
[45] A. Cowell and K. M. Stanney, "Embodiement and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents," presented at Intelligent Virtual Agents, Kloster Irsee, Germany, 2003.
[46] J. C. Lester, S. G. Towns, C. B. Callaway, J. L. Voerman, and P. J. FitzGerald, "Deictic and Emotive Communication in Animated Pedagogical Agents," in *Embodied Conversational Agents*, J. Cassell, S. Prevost, J. Sullivan, and E. Churchill, Eds. Cambridge: MIT Press, 2000, pp. 123-154.
[47] G. Biswas, D. Schwartz, K. Leelawong, N. Vye, and TAG-V, "Learning by Teaching. A New Agent Paradigm for Educational Software," *Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors"*, vol. 19, 2005.

[48] C. Elliott, "The affective reasoner: A process model of emotions in a multi-agent system," Northwestern University Institute for the Learning Sciences, Northwestern, IL, Ph.D Dissertation 32, 1992.
[49] M. S. El Nasr, J. Yen, and T. Ioerger, "FLAME: Fuzzy Logic Adaptive Model of Emotions," *Autonomous Agents and Multi-Agent Systems*, vol. 3, pp. 219-257, 2000.
[50] D. Moffat and N. Frijda, "Where there's a Will there's an agent," presented at Workshop on Agent Theories, Architectures and Languages, 1995.
[51] J. L. Armony, D. Servan-Schreiber, J. D. Cohen, and J. E. LeDoux, "Computational modeling of emotion: Explorations through the anatomy and physiology of fear conditioning," *Trends in Cognitive Science*, vol. 1, pp. 28-34, 1997.
[52] D. Cañamero, "Modeling motivations and emotions as a basis for intelligent behavior," presented at International Conference on Autonomous Agents, Marina del Rey, CA, 1997.
[53] R. C. Arkin, "Moving Up the Food Chain: Motivation and Emotion in Behavior-based Robots," in *Who Needs Emotions: The Brain Meets the Robot*, J. Fellous and M. Arbib, Eds.: Oxford University Press, 2005.
[54] J. R. Busemeyer, J. T. Townsend, and J. C. Stout, "Motivational Underpinnings of Utility in Decision Making: Decision Field Theory Analysis of Deprivation and Satiation," in *Emotional Cognition*, S. Moore, Ed. Amsterdam: John Benjamins, 2002.
[55] P. Gmytrasiewicz and C. Lisetti, "Using Decision Theory to Formalize Emotions for Multi-Agent Systems," presented at Second ICMAS-2000 Workshop on Game Theoretic and Decision Theoretic Agents, Boston, 2000.
[56] H. Prendinger, S. Mayer, J. Mori, and M. Ishizuka, "Persona Effect Revisited," presented at Intelligent Virtual Agents, Kloster Irsee, Germany, 2003.
[57] N. C. Kramer, B. Tietz, and G. Bente, "Effects of embodied interface agents and their gestural activity," presented at Intelligent Virtual Agents, Kloster Irsee, Germany, 2003.
[58] M. Scheutz and A. Sloman, "Affect and agent control: experiments with simple affective states," presented at IAT, 2001.
[59] M. Perrez and M. Reicherts, *Stress, Coping, and Health*. Seattle, WA: Hogrefe and Huber Publishers, 1992.