Evaluating the modeling and use of emotion in virtual humans

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 begin to evaluate them against the phenomena they purport to model. In this paper, we present one methodology to evaluate an emotion model. The methodology is based on comparing the behavior of the computational model against human behavior, using a standard clinical instrument for assessing human emotion and coping. We use this methodology to evaluate the EMA model of emotion [1, 2]. The model did quite well. And, as expected, the comparison helped identify where the model needs further development.

Introduction

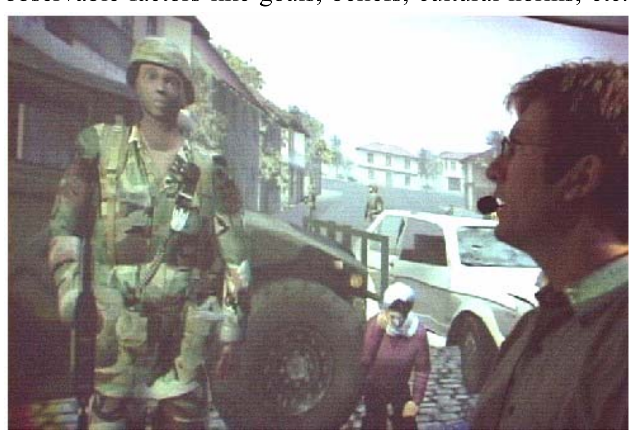
The interest in general computational models of emotion and emotional behavior has been steadily growing in the agent research community. Although the creation of general computational models of emotion is of potential interest in understanding human behavior, much of the interest in the agent community has been fueled by the application areas for such models. In particular, there has been 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. Virtual human technology is being applied to training applications [3], health interventions [4], marketing [5] and entertainment [6]. 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 [7, 8].

The critical role of emotion models in virtual human technology stems from the critical role emotions play in human experience. Virtual humans are designed to behave like people and emotions impact human behavior in many ways. It impacts their decision making, actions, memory, attention, voluntary muscles, social interactions, etc., all of which may subsequently impact their emotional state (e.g., see [2]). Further, emotions 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 a more engaging experiences for the immersed human users [9] .

Work on computational models of emotion has been significantly bolstered by a range of well-developed psychological models on the causes of emotions. In particular, cognitive appraisal theories of emotion have lead to several computational models of the causes of emotion. In contrast, there has been far less computational work in modeling the wide-ranging impact human emotions have on cognitive and behavioral responses. In part, this may stem from the simple fact that the psychological models on the impact of emotions are not as crisply defined and therefore are more difficult to leverage in creating a computational model.

In our research, we have been developing a general computational model of human emotion [1, 2]. The model 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. The model of emotion we have developed accounts for a range of such phenomena. 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 (see Figure this page) [3, 10].

Given the broad and subtle 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 [11]).

We are attempting to move beyond the concept of believability and instead evaluate more specific functional questions. The study described here addresses the question of mental models and cognitive dynamics: does the 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 a computational model of emotion create the right cognitive dynamics?

Appraisal Theory (a review)

Motivated by the need to inform the design of symbolic systems, our work is based on cognitive appraisal theory which emphasizes the cognitive and symbolic influences of emotion and the underlying processes that lead to this influence [12] in contrast to models that emphasize lower-level processes such as drives and physiological effects [13]. In particular, our work is informed by Smith and Lazarus’ cognitive-motivational-emotive theory.

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 process 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.

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 [14]. 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 have characterized the wide range of human coping responses into two broad classes. *Problem-focused coping* strategies attempt to change the environment. *Emotion-focused coping* [12] involves inner-directed strategies for dealing with emotions. Emotion-focused coping alters one’s interpretation of circumstances, 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 [12]: 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.

A Computational Model

EMA is a computational model based on appraisal theory and described in detail elsewhere [1, 2]. 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., [15]) and augment them with methods that explicitly model commitments to beliefs and intentions [16]. 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 essential 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 [17]. As we will see, commitments to beliefs and intentions also play a key role in modeling coping strategies.

In EMA, the agent's interpretation of its "agent-environment relationship" is reified by an explicit representation of beliefs, desires, intentions, plans and probabilities. Following a blackboard-style model, this representation (corresponding to the agent's working memory) encodes the input, intermediate results and output of reasoning process that mediate between the agent's goals and its physical and social environment (e.g., perception, planning, explanation, and natural language processing). We use the term *causal interpretation* to refer to this collection of data structures to emphasize the importance of causal reasoning as well as the interpretative (subjective) character of the appraisal process. At any point in time, the causal interpretation represents the agent's current view of the agent-environment relationship, which may subsequently change with further observation or inference. We treat appraisal as a set of feature detectors that map features of this representation into appraisal variables. For example, an effect that threatens a desired goal would be assessed as a potential undesirable event. Coping sends control signals to auxiliary reasoning modules (i.e., planning, action selection, belief updates, etc.) to overturn or maintain those features that yielded individual appraisals. For example, coping may resign the agent to the threat by abandoning the desired goal. Figure 2 illustrates a reinterpretation of Smith and Lazarus' cognitive-motivational-emotive system consistent with this view. The causal interpretation could be viewed as a representation of working memory (for those familiar with psychological theories) or as a blackboard.

Figure 3 illustrates a causal interpretation. In the figure, an agent has a single goal (affiliation) that is threatened 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 fluent in 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 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)
- Likelihood: how probable is the outcome of the event

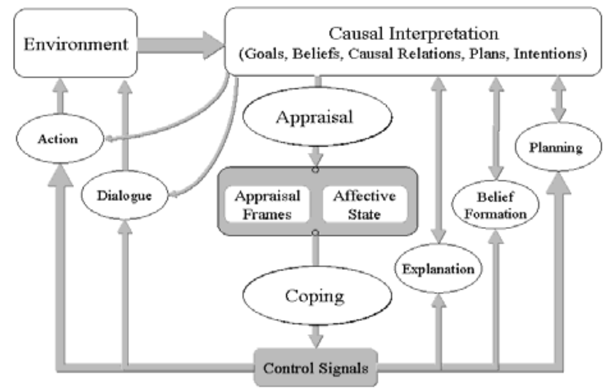


Figure 2: A reinterpretation of Smith and Lazarus

- Causal attribution: who deserves credit or blame
- 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
- 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 [18]. 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 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, 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 someone that is in control of an outcome for help
- 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, planful coping will generate in 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.

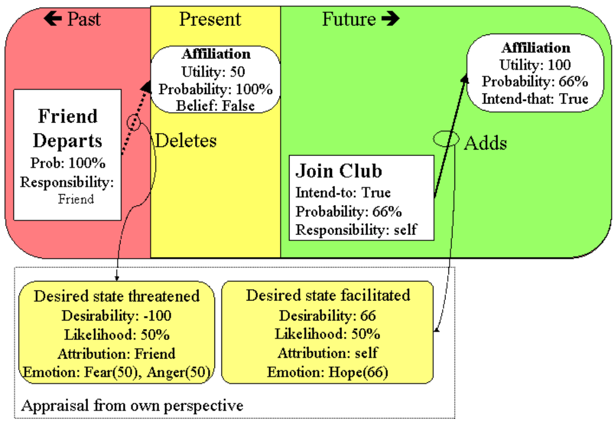


Figure 3: A causal interpretation (aversive condition)

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 is 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.

Related Work

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 appraisal 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 [19] Affective Reasoner, based on the OCC model [18], 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 [20] 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 [21] 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.

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 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 [22] considered the impact of emotional displays on user stress and confidence and Lester 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., [23]).

A small number of studies have tried to evaluate internal characteristics of an emotion process model. For example, Scheutz [REF] 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.

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 observer self-reports of “believability,” we would like to 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) [24] 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 situation and their responses are measured several times in the course of the episode. 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 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 situations are evolved systematic 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 evolving situations from the SCPQ. The scale consists of several distinct episodes but all are generated from a grammar that encodes two prototypical stressful episodes. Episodes evolves over three discrete phases: an initial state, a state where some time passes without change, and an ending phase which can either result in a good or bad conclusion. The *loss condition* presents a situation where some loss is looming in the future, the loss continues to loom for some time, and then the loss either occurs or is averted. In the *aversive condition*, some bad outcome 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, loss-bad, aversive-good and aversive-bad) each of which are represented by multiple variants in the scale. The aversive condition is designed to convey a greater sense of control/changeability, and the vocabulary is selected and empirically validated to produce this effect. Figure 4 illustrates a loss condition that ends with a bad outcome.

> **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 tat 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: A situation from the SCPQ

When used as a diagnostic tool, a patient would fill out their interpretation of the set of evolving situations. These are scored with respect to how closely they follow the trends exhibited by healthy adults. These 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 decrease over phases (as likelihood of change drops).
1.3 Negative valence should increase 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.

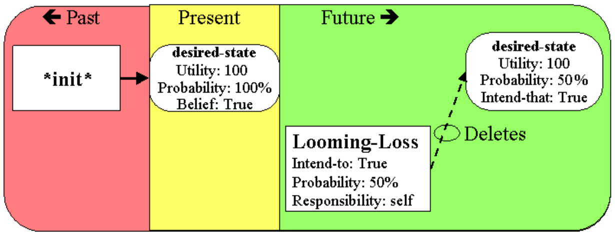


Figure 5: The initial loss condition

Our intention is to use the scale as a diagnostic instrument to ascertain if the judgments made by our model fall with 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 the four canonical scenarios. Thus, we encode the causal structure of these four episodes into EMA.

Methodology

We encode the four canonical episodes in the SCPQ as evolving causal theories and compare the model's appraisals and proposed coping strategies to the trends indicated by the scale. Consistent with how the SCPQ is used, we allow the model to propose coping strategies, but these proposals do not influence subsequent phases (the model proposes strategies but their effects are not actually implemented). The evolving phases in each episode are encoded by changing the perceived likelihood of future outcomes at each phase in the episode. The SCPQ provides the basic causal structure of the scenarios but we must set two parameters to complete each model, specifically the subjective probability of future actions in each phase and the utility of action outcomes.

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% percent, 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 [18], 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.

Results

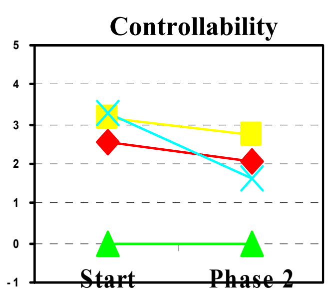
Trends 1.1 and 1.3 are supported by the model: the aversive condition is appraised as more controllable and changeable and negative valence increases across phases in both conditions. Trend 1.2 is fully supported for the aversive condition but only partially supported in the loss condition: EMA correctly deduces that the situation is less likely to change across phases, but it decides that the agent has no control over the loss, even in phase 1. 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. 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.

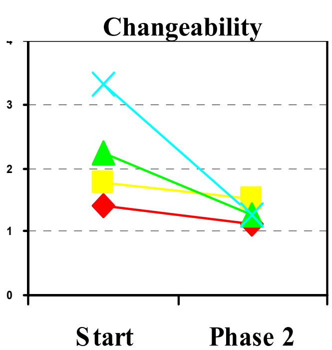
Trends 2.1 and 2.2 are both supported. In the aversive condition, the model forms an intention to restore the loss only when its probability of success is high (phase 1). In the loss condition, no known action can influence the pending loss so control is low and no problem-directed strategies are selected.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

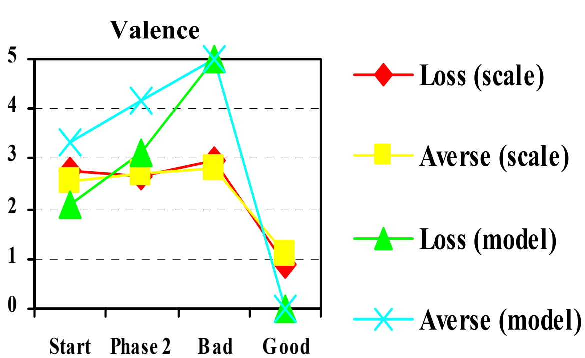
Trends 3.1 and 3.2 are fully supported. When the model finds the situation likely to improve on its own (high changeability), it proposes monitoring the truth-value of the state predicate that has high probability of changing. As changeability drops, the model proposes strategies that suppress the monitoring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

Trend 4 is supported. 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 information) 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).

Table 1	Aversive	Loss
Phase 1	Seek information Take action	Suppress information Procrastinate Seek instrument. support
Phase 2	Mental disengagement Suppress information	Mental disengagement Suppress information Resignation Wishful thinking
Good	Accept responsibility	
Bad	Mental disengagement Suppress information	Mental disengagement Suppress information







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 indicates that the OCC model's definition of "distress" 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., [12]). This alternative definition could be straightforwardly added to our model.

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 an assumption and could be relaxed by adding some other action to the domain model executable by the agent that could influence the likelihood of the loss.

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 mode, potentially biasing our results.

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.

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 paper, we presented one methodology 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. [22, 25]). This paper contributes to this evolution.

Acknowledgements

This work was funded by the Department of the Army under contract DAAD 19-99-D-0046. Lilia Moshkina produced the stimuli used the social referencing experiment. Any opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of the Department of the Army.

[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 AAMAS03 , Melbourne, Australia, 2003.

[3] 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*, July/August, 2002, pp. 32-38.

[4] S. Marsella, W. L. Johnson, and C. LaBore, "Interactive pedagogical drama for health interventions," presented at Artificial Intelligence in Education, Sydney, Australia, 2003.

[5] 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,.

[6] M. Cavazza, F. Charles, and S. J. Mead, "Agents' Interaction in Virtual Storytelling," presented at 3rd Int. Workshop on Intelligent Virtual Agents, 2001.

[7] 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.

[8] R. W. Picard, *Affective Computing*. Cambridge, MA: MIT Press, 1997.

[9] S. van Mulken, E. André, and J. Muller, " The Persona Effect: How Substantial Is It," presented at HCI Conference, Berlin, 1998.

[10] 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.

[11] S. Marsella, W. L. Johnson, and C. LaBore, "Interactive Pedagogical Drama," presented at 4th Int. Conference on Autonomous Agents, Montreal, 2000.

[12] R. Lazarus, *Emotion and Adaptation*. NY: Oxford University Press, 1991.

[13] J. Velásquez, "When robots weep: emotional memories and decision-making.," presented at AAAI98, Madison, WI, 1998.

[14] E. Peacock and P. Wong, "The stress appraisal measure (SAM): A multidimensional approach to cognitive appraisal," *Stress Medicine*, vol. 6, pp. 227-236, 1990.

[15] J. Blythe, "Decision Theoretic Planning," in *AI Magazine*, vol. 20(2), 1999, pp. 37-54.

[16] B. Grosz and S. Kraus, "Collaborative Plans for Complex Group Action," *AIJ*, vol. 86, 1996.

[17] K. G. Shaver, *The attribution of blame: Causality, responsibility, and blameworthiness*. NY: Springer-Verlag, 1985.

[18] A. Ortony, G. Clore, and A. Collins, *The Cognitive Structure of Emotions*: Cambridge Press., 1988.

[19] 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.

[20] 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.

[21] D. Moffat and N. Frijda, "Where there's a Will there's an agent," presented at Workshop on Agent Theories, Architectures and Languages, 1995.

[22] H. Prendinger, S. Mayer, J. Mori, and M. Ishizuka, "Persona Effect Revisited," presented at Intelligent Virtual Agents, Kloster Irsee, Germany, 2003.

[23] 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.

[24] M. Perrez and M. Reicherts, *Stress, Coping, and Health*. Seattle, WA: Hogrefe and Huber Pub., 1992.

[25] 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.