Appears in *Applied Artificial Intelligence,* v19: 2005 (DRAFT)

Lessons from Emotion Psychology for the Design of Lifelike Characters

Jonathan Gratch
University of Southern California, Institute for Creative Technologies

Stacy Marsella
University of Southern California, Information Sciences Institute

Abstract

This special issue describes a number of applications that utilize lifelike characters that teach indirectly, by playing some role in a social interaction with a user. The design of such systems reflects a compromise between competing, sometimes unarticulated demands: they must realistically exhibit the behaviors and characteristics of their role, they must facilitate the desired learning, and they must work within the limitations of current technology, and there is little theoretical or empirical guidance on the impact of these compromises on learning. Our perspective on this problem is shaped by our interest in the role of emotion and emotional behaviors in such forms of learning. In recent years, there has been an explosion of interest in the role of emotion in the design of virtual humans. The techniques and motivations underlying these various efforts can seem, from an outsider's perspective, as bewildering and multifaceted as the concept of emotion itself is generally accused of being. Drawing on insights from emotion psychology, this article attempts to clarify for the designers of educational agents the various theoretical perspectives on the concept of emotion with the aim of giving guidance to designers of educational agents.

1 Introduction

The theme of this special issue is the use of educational agents that depart from the traditional role of a teacher or advisor. The bulk of these collective efforts utilize lifelike characters that teach indirectly, by playing some role in a social interaction with a user. These "virtual humans" must (more or less faithfully) exhibit the behaviors and characteristics of their role, they must (more or less directly) facilitate the desired learning, and current technology (more or less successfully) supports these demands. The design of these systems is essentially a compromise, with little theoretical or empirical guidance on the impact of these compromises on pedagogy.

Our perspective on this problem is shaped by our interest in the role of emotion and emotional behaviors in such forms of learning. In recent years, there has been an explosion of interest in the role of emotion in the design of virtual humans. Some of this work is directly motivated by the role emotion seems to play in teaching and learning, however much of it is directed more generally at making virtual characters seem more convincing, believable, and potentially more intelligent. The techniques and motivations underlying these various efforts can seem, from an outsider's perspective, as bewildering and multifaceted as the concept of emotion itself is generally accused of being. This article attempts to clarify for the designers of educational agents the various theoretical perspectives on the concept of emotion with the aim of giving guidance to designers of educational agents.

Artificial intelligence has historically taken a dim view of emotion. Following the Stoic and Enlightenment traditions, emotion has been considered, if considered at all, as a disruptive force that detracts from rational thought. Today, this view is being increasingly challenged on two fronts. On the one hand, compelling findings from neuroscience and psychology have emphasized the adaptive role emotions can play in cognition and social interaction. For example, evidence suggests that emotions are crucial for effective decision-making (Damasio, 1994; LeDoux, 1996; Mele, 2001), memory (Bower, 1991; Nasby & Yando, 1982), teaching (Lepper, 1988), coping with environmental stressors (Lazarus, 1991), communication (Brave & Nass, 2002), and social reasoning (Forgas, 1991; Frank, 1988), and such findings have motivate attempts to abstract these functions and incorporate them into general computational systems. On the other hand, advances in user interfaces have enabled increasingly sophisticated interaction between humans and computers, including life-like conversational agents (Cassell, Sullivan, Prevost, & Churchill, 2000; Cole et al., 2003; Gratch et al., 2002). There is growing evidence that people frame interactions with these systems as a social interaction and, disruptive or not, employ and are influenced by emotion behaviors. A growing list of applications include psychotherapy applications (Marsella, Johnson, & LaBore, 2000, 2003; Rothbaum et al., 1999), tutoring systems (Lester, Stone, & Stelling, 1999; Ryokai, Vaucelle, & Cassell, 2003; Shaw, Johnson, & Ganeshan, 1999), and marketing applications (André, Rist, Mulken, & Klesen, 2000). Indeed, emotion has become fashionable and the artificial intelligence community is experiencing a mini-avalanche of experimentation and innovation in “emotional” or “affective” computing.

This burgeoning interest has produced its share of growing pains. As computational metaphors go, emotion is particularly fertile, meaning different and sometimes contradictory things to different people. Even within the sciences that study human emotion, there is considerable diversity of opinion over the meaning of the term. Emotion has been variously described as (1) a fundamental set of well-specified mental primitives, (2) ad-hoc collection of unrelated processes, (3) a loose collection of communicative conventions, and (4) an epiphenomenon that distracts from fundamental underlying processes. Worse, different scientific traditions tend to adopt one of these perspectives implicitly with only occasional debate of other views. From the outside perspective of a computer science researcher looking for "the right" theory of emotion to motivate and guide computation models, these distinctions can be confusing, are easily overlooked, and certainly serve as a distraction. Not surprisingly, the result is that there is often confusion as to 'what species' of emotion is being modeled, what function it is serving, and how to evaluate its impact.

This conceptual naïveté is reflected in unsophisticated instruments used in validating emotional models. Much of the research on emotional systems attempts to improve the overall believability and/or realism of expressed behavior, but such single variable measures are simply inappropriate given the multifaceted nature of emotion. Expressive human-like interfaces have a number of potential influences over social interaction. Only a subset of these influences will likely benefit a given application. Indeed, many of the influences of expressive behavior work at cross purposes with human-computer interaction. For example, people can be more nervous in the presence of a lifelike agent (Kramer, Ti-

etz, & Bente, 2003) and tend to mask their true feelings and opinions (Nass, Moon, & Carney, 1999), properties that may complicate a teaching application. It is also clear that such effects can be differentially strengthened or mitigated, depending on how individual behaviors are realized (Cowell & Stanney, 2003; Nakanishi, Shimuzu, & Isbister, 2005). System designers must be cognizant of how these effects relate to the overall goal of their application. For example, the designers of Carmen's Bright IDEAS system utilized non-realistic behaviors to mitigate socially induced stress, as such stress conflicted with their overall goal of promoting stress reduction (Marsella, Gratch, & Rickel, 2003). Such findings call into question the utility of general measures such as "believability." Rather, to understand the role of an expressive character in any particular application, the community needs a more explicit listing and testing of individual functions of emotion and their relationship to the design goals of a given application.

This article seeks to address the general conceptual confusion surrounding research on emotional systems, focusing on their use in educational settings. We lay out a set of conceptual distinctions, drawn from the psychological literature, to help researchers clarify certain questions surrounding their work. This framework makes explicit a number of issues that are implicit and sometimes confounded in the current discourse on computational models of emotion. The following discussion is organized around the following questions: What is the function of emotion in a computational system? How can these functions be modeled? And, how can it externally manifest to a user?

2 The Function of Emotion

Computer scientists are trained to think in terms of function. When it comes to incorporating "emotion" into our computational systems, the obvious question to ask is why? Psychologists have posited a number of functions emotions serve in humans, which may, by analogy be of use to a computational entity. Emotion functions have generally been characterized from one of two very different perspectives—intra-organism functions vs. inter-organism functions—depending on if emotion is viewed as something that mediates mental processes or as something that impacts social interaction. It is important to note that, whereas this distinction is somewhat blurred in humans, computational systems can easily model one function without necessarily considering the other. For example, a webbot with no visible embodiment might usefully incorporate some mental functions of emotion, whereas an embodied agent might manifest realistic emotional behaviors via cognitively implausible mechanisms.

2.1 Cognitive (Intra-agent) function of emotion

One tradition within emotion psychology emphasizes the role emotions play in mediating cognition. Emotions impact a wide array of human cognitive processes and there is a growing consensus that this impact has adaptive value for humans and other organisms. Many emotion theorists subscribe to the psychoevolutionary view that "emotions are patterns of responses that have evolved for their ability to organize bodily systems to enable a quick and efficient response to important environmental events" (Rosenberg, 1998). Of course, it is still a matter of debate if these influences are something one would wish to incorporate into a "rational" agent architecture.

Some posited cognitive functions of emotion include:

- **Situation Awareness:** Emotion has been proposed as a mechanism that helps an organism perceive and categorize significant environmental stimuli. For example, appraisal theories of emotion argue that the mechanisms associated with emotion help an organism understand how external events and circumstances relate to internal goals and motivations (Scherer, Schorr, & Johnstone, 2001). Appraisal theories claim such events are characterized in terms of a number of criteria. Some of these criteria (e.g., the utility or likelihood of an event) are shared by traditional models of intelligence, but appraisal theories posit that additional dimensions (e.g. attributions of blame or credit) are critical for characterizing human behavior. Such appraisals feedback into the perceptual system, allowing people to rapidly interpret subsequent ambiguous stimuli (sometimes incorrectly) as consistent with these prior appraisals (Neidenthal, Halberstadt, & Setterlund, 1997). This is a very different mechanism than classical decision theory, the mechanism by which many artificial entities judge the significance of events. From a functionalist emotion perspective, decision theory provides an incomplete set of constructs for modeling human behavior, and may be incomplete from the perspective of modeling intelligent behavior in general, independent of its humanness.
- **Action selection:** A number of theorists argue that emotions are part of a mechanism that prepares an organism to act on the environment. For example, the emotion of fear leads to mental and physical arousal, focuses perceptual attention on the threat-

ening stimuli, primes rapid execution of certain responses that are heuristically adaptive (Frijda, 1987; LeDoux, 1996). There is a close association between emotion as a form of situation awareness (appraisal) and emotion as action selection in that the appraisal process is typically viewed as a sort of classifier that maps the current "person-environment relationship" into a suggested action. In this sense, emotion theory suggests that, rather than triggering responses directly from primitive features of an event, as in reactive planning systems such as (Firby, 1987), actions, at least in people and perhaps in general, should be organized around some characterization of their emotional significance.

- **Coping:** Beyond immediate responses to the environment, many emotion theorists argue that people develop persistent strategies to manage their emotional state. Some of these so-called coping strategies are shared by traditional models of intelligence (e.g. people engage in planning behaviors to deal with negative emotion), but people exhibit other "emotion-directed" coping strategies (e.g., wishful thinking, shifting blame, distancing) that are traditional characterized as irrational and large avoided by computational models of intelligence. Yet these "irrational" distortions can be adaptive, decreasing stress levels, extending life expectancy, enhancing the strength of social relationships. This adaptive nature of emotional behavior may be attributed to the fact that such coping strategies attempt to form a comprehensive response that balances the global physical and social consequences of individual beliefs and decisions.
- **Learning:** A number of studies have shown strong effects of emotion on memory and recall. Several studies by Bower and his colleagues have argued that people best

recall information when they are in the same emotional state as when they originally learned the information (Bower, 1991). There is also evidence that people selective learn and recall affect-laden concepts that are consistent with their current emotional state. For example Teasdale and Russell (1983) showed that subjects that studied positive or negative words in a normal state were better able to recall words congruent with a subsequently induced emotional state. Clearly, such effects can produce distortions and are, for example, a likely reason why depressed individuals persist in viewing the world in a negative light. However, these effects can play a useful function. For example, emotional state can be seen as a important index into retrieving memories that are relevant to the organisms current circumstances.

From the perspective of developing educational agents, modeling the intra-agent influences of emotion may provide some insights in improving the cognitive capabilities of intelligent agents in general, but the most compelling motivation is when the educational task itself demands the faithful modeling of one of the above mentioned emotional influences. For example, one might imagine a training system for clinical psychologists where they must recognize and confront the emotional distortions of virtual patients. Faithful modeling the mechanisms underlying such distortions becomes necessary whenever the interaction becomes rich and varied enough that it can no longer be scripted in advance, but requires the capabilities of an intelligent agent.

A number of computation theories and implemented systems can be viewed as encoding one or more of these functions. Carmen's Bright Ideas attempts to faithfully model the

assessment and coping functions of emotion in order to illustrate effective coping strategies to parents of pediatric cancer patients (Marsella, Johnson et al., 2003). The Mission Rehearsal Exercise (MRE) is designed to teach leadership skills in high-stress social situations, and models many of the cognitive influences of emotion to illustrate some of the cognitive biases a leader may have to recognize and manage in his subordinates (Gratch & Marsella, 2001; Rickel et al., 2002). Independent of teaching applications, many systems have focused on the potential adaptive role of emotions in the design of intelligent systems. For example, several of the early speculations on the nature of intelligent systems posited the need of emotion-like mechanisms to handle interrupts and coordinate distributed mental processes (Oatley & Johnson-Laird, 1987; Simon, 1967); Aaron Sloman and his colleagues have posited and implemented general architectural mechanisms (impacting processing characteristics, meta-reasoning, and learning) through a functional analysis of emotion's cognitive role (Beaudoin, 1995; Scheutz & Sloman, 2001); Velásquez (Velásquez, 1998) and Tyrell (Tyrrell, 1993) have proposed models of action selection inspired by emotion theory; several researchers have implemented systems or that model emotion's influence over belief (Marsella & Gratch, 2002; Thagard, 2002); and several systems capture emotion's presumed functional influence over learning (El Nasr, Yen, & Ioerger, 2000; Velásquez, 1998)

2.2 Social (Inter-agent) function of emotion

In contrast to work on cognitive influences, social psychologists and linguists emphasize the role emotional displays (and non-verbal behavior in general) play in mediating communication and other forms of social interaction. Emotional displays communicate information and are a powerful tool for shaping interpersonal behavior. Anyone that has

had a bad experience with a car salesman knows that emotions can be deliberately manipulated to achieve certain ends, but it is also increasingly accepted that emotions can benefit a social group by promoting cohesion and preventing costly misunderstandings. Indeed, many theorists argue that emotions evolved because they provided an adaptive advantage to social organisms (Darwin, 2002; de Waal, 2003; Fridlund, 1997). It is natural to ask how to utilize knowledge of such social functions for the design of computational systems.

Some posited social functions of emotion displays include:

- **Communication of mental state:** Emotional displays can communicate information about the mental state of an individual, although there it is debatable if these displays reflect true emotion or are simply communicative conventions (Manstead, Fischer, & Jakobs, 1999). From such displays, observers can form consistent interpretations of a person's beliefs (e.g., frowning at an assertion indicates disagreement), desires (e.g., joy gives information that a person values an outcome) and intentions/action tendencies (e.g. fear suggests flight). They may also provide information about the underlying dimensions along which people appraise the emotional significance of events: valence, intensity, certainty, expectedness, blameworthiness, etc. (Smith & Scott, 1997). Agents that utilize these communicative channels can potentially convey such information more efficiently and forcefully than simple text or speech messages.

- **Social manipulators:** There is evidence that in humans and other social species, emotions are part of a system of social control (Campos, Thein, & Daniela, 2003; de Waal, 2003; Fridlund, 1997). Drawing on the ethological notion of an action "releaser", certain emotional displays seem to function 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. For example, anger seems to be a mechanism for coercing actions in others and enforcing social norms; displays of guilt can elicit reconciliation after some transgression; distress can be seen as a way of recruiting social support; and displays of joy or pity are a way of signaling such support to others. Other emotion displays seem to exert control indirectly, by inducing emotional states in others and thereby influencing an observer's behavior. One example of this is *emotional contagion*, a phenomenon related to social mimicry whereby susceptible individuals can "catch" the emotions of those around them (Hatfield, Cacioppo, & Rapson, 1994). Another example is the *Pygmalion effect* (also known as "self-fulfilling prophecy") whereby our positive or negative expectations about an individual, even if expressed nonverbally, can influence them to meet these expectations (Blanck, 1993) Indeed, some education researchers have argued that such nonverbal displays can have a significant impact on student intrinsic motivation (Lepper, 1988).
- **Believability/Framing effects:** In contrast to specific communicative acts, some research has suggested that incorporating realistic emotional displays into a computer generated character can have the general effect of making it seem more believable and humanlike, and thereby cue the user to interact with the character as if they were

> interacting with another person. Designer's of educational systems can exploit, or at least need to be aware of, these various effects. In the presence of a believable agent, people are more polite, tend to make socially desirable choices and are more nervous (Kramer et al., 2003); they can exhibit greater trust of the agent's recommendations (Cowell & Stanney, 2003); and they can feel more empathy (Paiva, Aylett, & Marsella, 2004). Education agents can exploit these general framing effects to create more intrinsic motivation for learning.

If they can be captured and utilized by computational systems, such interpersonal functions could play an important role in educational applications. In that people utilize these behaviors in their everyday interpersonal interactions, modeling the function of these behaviors is essential for any application that hopes to faithfully mimic face-to-face human interaction. More importantly, however, the ability of emotional behaviors to influence a person's emotional and motivational state could potentially, if exploited effectively, guide the student towards more effective learning.

A number of applications have attempted to exploit the interpersonal function of emotion. 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 (Klesen, 2005). Nakanishi et al. (2005) and Cowell and Stanney (2003) each evaluated how certain non-verbal behaviors could communicate a character's trustworthiness for training and marketing applications, respectively. Several applica-

tions 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 (Lester, Towns, Callaway, Voerman, & FitzGerald, 2000); The VICTEC system exploits general framing effects to promote student empathy with animated characters with the goal of bullying prevention in schools; Biswas (Biswas, Schwartz, & Leelawong, 2005) also use human-like traits to promote empathy and intrinsic motivation in a learning-by-teaching system.

3 The modeling of emotion

Assuming one wishes to exploit some of the posited functions of emotion within a computational system, how should one effectively implement these functions? Existing computational approaches typically adopt one of two approaches. *Communication-driven* approaches deliberately select an emotional display purely for its communicative or manipulative effect. *Simulation-based* approaches, in contrast, attempt to simulate aspects of emotion processes: essentially giving the agent true emotions. Although this distinction is blurred in humans, computational systems can, and typically have adopted just one of these perspectives.

3.1 Communication-driven Methods

Many implemented systems treat emotional displays as a means of communication. These systems would not be viewed as "having" emotion in that there is no internal calculation of what emotion the system would naturally have given its goals and current circumstances. Rather, *communication-driven approaches* can be seen as selecting a display based on its utility, given the current state of an interaction with a user and the sys-

tem's communicative or educational goals. In most implemented systems, this utility is not explicitly calculated but, rather, is assessed by the application developer and encoded, via some scripting language, into the repertoire of agent responses. Implementations tend to be *ad hoc* and vary considerably across applications.

One promising approach to formalizing the communicative function of emotional displays is to recast the function of such displays in the same formalisms that computational linguistics has applied to understanding and modeling verbal communication. The Trindi project is representative of the linguistics approach (Larsson & Traum, 2000): the state of the interaction is represented by an explicit data structure called the *information state*; verbal and even non-verbal utterances are interpreted as a set of *speech acts* (Austin, 1962; Searle, 1969) that are formalized in terms of their impact the information state. For example, an *assertion* has the effect of establishing a commitment by the speaker that the content of the assertion holds. The same physical action can correspond to several speech acts. For example, nodding one's head in response to a request acts simultaneously to acknowledge that the request was understood, to accept the request and to give the dialogue turn back to the speaker. Along these lines, emotional displays could be formalized as "emotion acts" that have some impact on the information state, though this impact may be at a social level not typically considered by traditional dialogue systems. A few researchers have begun to formalize the communicative function of emotion along these lines. For example, Poggi and Pelachaud (1999) use facial expressions to convey the performative of a speech act, showing "potential anger" to communicate that the agent will be angry if a request is not fulfilled. Heylen (Heylen, Nijholt, & Aker, 2005)

has begun a taxonomy of emotion acts that range from traditional dialogue functions (a smile may act to communicate agreement) to more social functions (positive affect may act to increase student motivation).

Ideally, communication-driven approaches should have the means to comprehend the emotional displays of the user. That the system generates such displays will invite the user to reciprocate and such displays can contain valuable information about the current state of the interaction as well as provide feedback about the efficacy of the system's communicative moves. Even simple methods can be effective. For example, in Isbister and Nakanishi's Digital City Project, an agent watches for awkward conversational silences then asks questions to try to restart the conversation (Nakanishi et al., 2005). A number of methods have also been developed to address the notoriously difficult problem of intuiting or recognizing a user's emotional state. For example, Conati uses a Bayesian network-based appraisal model to deduce a student's emotional state based on their actions (Conati, 2002); several systems have attempted to recognize the behavioral manifestations of emotion including facial expressions (Fasel, Stewart-Bartlett, Littelwort-Ford, & Movellan, 2002; Lisetti & Schiano, 2000), physiological indicators (Haag, Goronzy, Schaich, & Williams, 2004; Picard, 1997) and vocal expression (Lee & Narayanan, 2003).

Tutoring applications usually follow a communication-driven approach, intentionally expressing emotions with the goal of motivating the students and thus increasing the learning effect. For example, in the Cosmo tutoring system, the agent's pedagogical goals

drive the selection and sequencing of emotive behaviors. In Cosmo, a congratulatory act triggers a motivational goal to express admiration that is conveyed with applause. A potential disadvantage of pure communication-driven methods in such applications, as they are not based on any internal calculation of the agent's emotional state, is that agent's displays can seem flighty or insincere – though the impact of such anecdotal observations has yet to be substantiated.

3.2 Simulation-based Methods

The second category of approaches attempt to simulate "true" emotion (as opposed to deliberately conveyed emotion). We include in this category methods that attempt to faithfully model the impact of events on human emotion as well as systems that try to capture the cognitive function of emotion at in some abstract sense. Such *simulation-based methods* can be communicative in the sense above, but rather than triggering emotion displays explicitly because of their communicative function, displays are tied to the agent's simulated emotional state. In this sense, the emotion displays reflect something about the current state of information processing of the agent and can be viewed as a window into the agent's "soul."

Most simulation-based methods trace their lineage to a psychological theory of emotion called appraisal theory. In their most general form, appraisal theories argue that emotion arises from two basic processes: appraisal and coping (see Figure 1). Appraisal is the process by which a person assesses their overall relationship with its physical and social 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.

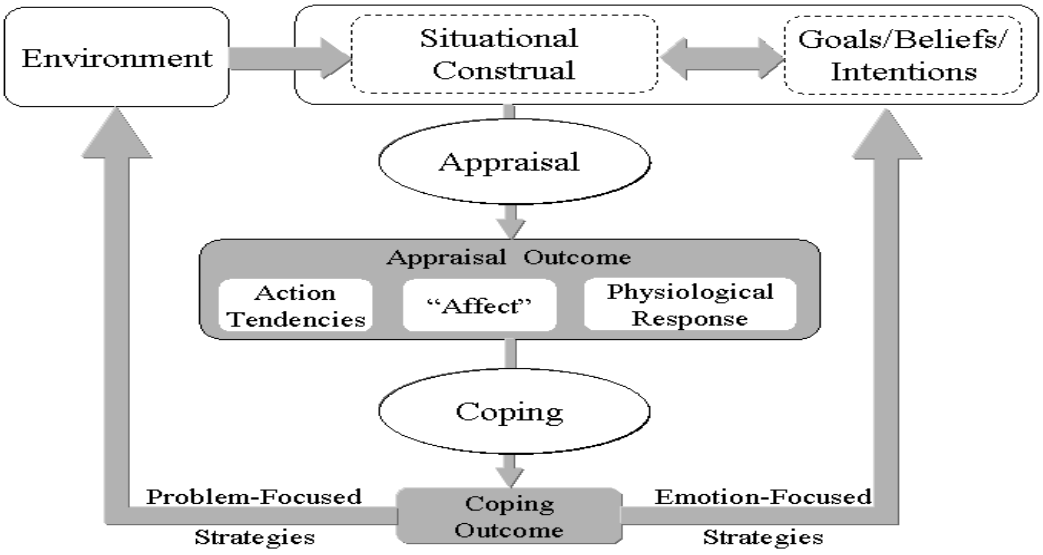

Figure 1: A process view of appraisal theory, adapted from (Smith & Lazarus, 1990)

motivated to respond to events differently depending on how they are appraised (Peacock & Wong, 1990). 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* (Lazarus, 1991) 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 (Lazarus, 1991): 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.

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 (1992) Affective Reasoner, based on the Ortony, Collins and Clore's appraisal theory (1988), 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 (2000) 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 ac-

tions 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. WILL (Moffat & Frijda, 1995) 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 (Gratch & Marsella, 2004) is one of the more comprehensive models, combing a plan-based model of appraisal with a detailed model of problem-focused and emotion-focused coping.

Some models attempt to combine both communication-driven and simulation-based methods. For example, in Carmen's Bright Ideas, the agent can make communication-driven displays and then appraise its own dialogue, thereby allowing it to regret its own statements (Marsella, Gratch et al., 2003). As another example, Prendinger uses communication-driven display rules as filters on appraised emotion (Prendinger & Ishizuka, 2001). Simulation-based models can also inform the calculation of what communication-driven displays to produce. For example, if an agent appraises a user request to be harmful, the display of alarm or disapproval is a reasonable communicative goal.

In the context of educational applications, simulation-driven methods are most appropriate when the goal is to teach users to recognize the impact of emotion on others or recognize the emotional impact of their own actions. For example, Piava et al. (2004) use a simulation-based approach based on the appraisal theory of Klaus Scherer (2001) to simulate how a school child might respond to bullying behaviors. The Mission Rehearsal

Exercise and Carmen's Bright Ideas simulates how emotions might influence the decision-making or conversational strategies of people in stressful circumstances based on the appraisal theory of Smith and Lazarus (Smith & Lazarus, 1990).

4 The Display of Emotion

Assuming a system is to display emotions to a human observer, for whatever function, the designer's must make certain decisions on how to realize this display in some medium. In studying human emotional displays and their interpretation, social psychologists make a strong distinction between behavioral *encoding* and behavioral *decoding* that is useful to consider in the design of artificial systems. Figure 2 lays out graphically this distinction. We can view the problem of conveying emotion as analogous to the problem of transmitting a message over some medium. In this case, the message is some emotional or other expressive content. This content must be encoded into some signal, in this case verbal or non-verbal behavior. An observer must then decode this signal to recover the original content. The question in designing artificial agents is whether to be faithful to how people actually encode emotional messages into their behavior, or whether to make it easy for users to decode the intended message, even if this means using unrealistic or stylized signals.

This question gains importance in that people are often bad decoders of emotional information and could misinterpret the emotional displays of an artificial agent. For example, Figure 2 uses an experimental technique known as Brunswik's lens model to demonstrate that people often focus on misleading cues when understanding the non-verbal behavior of others (Gifford, 1994). The study, which looked at indicators of personality, recorded

people with different personality types (classified based on a standard personality test) during a conversational interaction. Features of their behavior were coded and correlated with their personality type to assess how personality was encoded. A group of observers tried to assess (decode) the conversant's personality, and the cues they utilized were correlated with their assessments. Figure 2 illustrates that, as the features used for decoding are largely disjoint from those used in encoding, people in this study were particularly bad decoders of the personality trait "warmth."

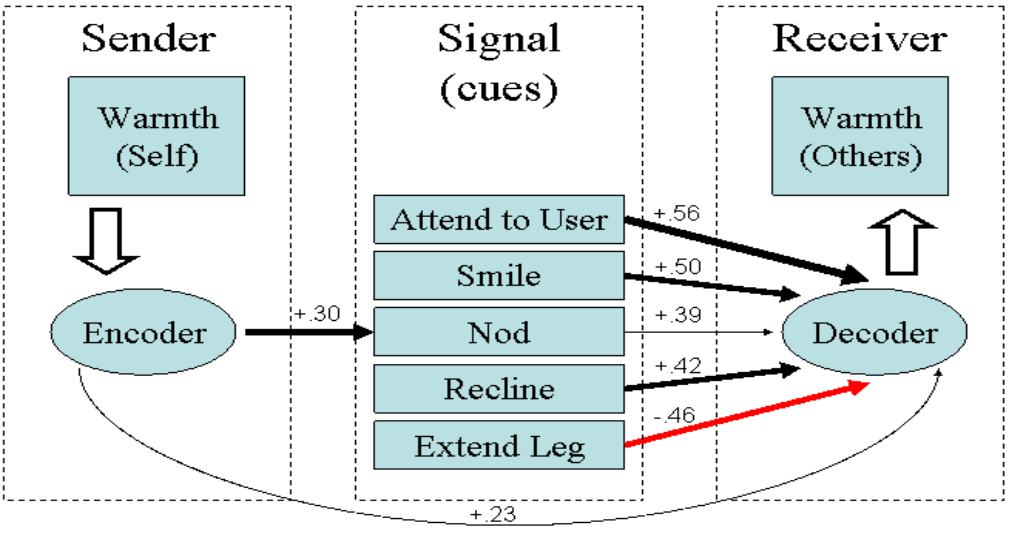

Figure 2: An application of Brunswik's lens model to a study of the nonverbal indicators of personality. In the case of the personality trait of "warmth," observers utilize a number of non-informative cues in attributing this trait to others, leading to poor recognition rates. Adapted from (Gifford, 1994)

Interestingly, some studies on the behavior of actors suggest that people are far better decoders of artificial behavior than they are of more natural human behavior. For example, Coats, Feldman and Philippot (1999) showed that actor's behavior is unnatural in several

ways: they display emotion more frequently, they use less ambiguous displays, and they rarely display emotions that don't correspond with their supposed emotional state. Animated characters tend to share these characteristics but also incorporate highly stylized or exaggerated emotional displays that people easily interpret.

Given that computational systems can use arbitrary mappings when encoding emotional content, an obvious question is how faithful should a system be to natural encodings. We may usefully distinguish between *accurate encoding models* that attempt to faithfully represent how people encode emotion for *accurate decoding models* that attempt to maximize the likelihood than an observer correctly interprets the emotional signal. Note that this distinction need not be restricted to emotion in particular, but applies to any type of human-like behavior, visual or verbal, that we might wish to model in a computational system.

4.1 Accurate encoding models

Accurate encoding models are clearly essential if the application is designed to teach observers to accurately recognize emotion. For example, an system that is designed to teach clinicians to recognize how to assess suicide risk would obviously require a fairly accurate modeling of the non-verbal signals associated with extreme distress or depression.

A few educational systems have been developed where the goal is to teach people the true indicators of emotion. For example, number of law enforcement departments use training systems to teach their officers how to read body language in order to assess if a suspect is threatening (though these are typically video-based and use actors to portray

realistic situations) and Johns Hopkins University's Applied Physics Laboratory has developed a computer-based interview training system for the FBI that incorporates a model of how people respond emotionally to certain lines of questioning (*PC-Based Human Interaction Training*). Additionally, a number of systems attempt to place people in an emotional setting they will likely face, though the goal is not explicitly to detect emotion in others. For example, Henderson (Henderson, 1998) uses a video-based system to train genetic counseling and smoking cessation practitioners how do deal with emotional patients (though again professional actors are used to portray realistic settings). We are not aware of any educational application that has attempted to validate the accuracy of their encoding model using techniques such as Brunswik's lens model.

4.2 Accurate decoding models

Accurate decoding models seek to maximize the ability of an observer to "read" the intended signal, generally through the use of exaggerated behaviors, exaggerated emotional dynamics, lack of emotional dissemblance and cinematic conventions (e.g., eyes popping out of the head to indicate surprise). These are clearly the well accepted norm for entertainment applications, and within this domain they are generally acknowledged to enhance the enjoyment, drama and interpretability of a narrative. Indeed, some researchers quite explicitly draw on acting rather psychological theory to inform the design of agent behaviors (Chi, Costa, Zhao, & Badler, 2000; Lance, Marsella, & Koizumi, 2004). Within an educational context, their use seems uncontroversial if the goal is simply to "spice up" the presentation of dry educational material; however accurate-decoding models are more controversial if the application, at least implicitly, relies on expressive behaviors to achieve specific social functions of emotion.

Specifically, the use of accurate-decoding modes raises two questions: 1) do accurate decoding models produce the same (or enhanced) social impact and 2) what implication does their use have for transfer outside of the training application. For example Paiva (this chapter) attempts to evoke empathy in children watching an animated portrayal of a bullying incident. The hope is that by empathizing with the virtual character, children will learn to empathize with real victims of bullying. The use of unnatural, stylized behaviors raises the questions: do such behaviors produce similar emotional states and social responses as would occur in a real-world interaction and will the lessons transfer to the real world where the behaviors may be more muted or obscured? Specifically, one might argue that after watching highly transparent virtual characters, people will come to underestimate the emotions of people in the real world. Following this argument, children watching the bulling virtual environment might conclude that real victims are not experiencing distress because their emotional displays are more muted than that of the virtual characters. Indeed, some evidence already suggests that watching unnatural behavioral displays can impact behavior in the real world. A study by Coats et al. (1999) showed that children that extensively watched television had difficulty recognizing that other people produced emotional displays that differed from their true emotional state.

5 Conclusion

Simulation technology has reached the state where researchers can incorporate highly expressive animated characters into educational applications, however the science of effectively exploit such characters is still in an embryonic state. In this article, we have articulated several dimensions for organizing the function expressive behavior plays in hu-

man-to-human interaction. By highlighting such distinctions, we hope to give some guidance to application designers, but more importantly, we hope to emphasize a number of fundamental scientific questions that must be addressed before we can understand the role of emotions and expressive behavior, not only in virtual characters, but in human society as well.

Acknowledgements

This work was funded by the Department of the Army under contract DAAD 19-99-D-0046. 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.

References

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

Austin, J. (1962). *How to Do Things with Words*: Harvard University Press.

Beaudoin, L. (1995). *Goal Processing in Autonomous Agents* (Ph.D Dissertation No. CSRP-95-2): University of Birmingham.

Biswas, G., Schwartz, D., & Leelawong. (2005). Learning by Teaching. A New Agent Paradigm for Educational Software. *Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors"*.

Blanck, P. D. (Ed.). (1993). *Interpersonal Expectations*. Cambridge: Cambridge University Press.

Bower, G. H. (1991). Emotional mood and memory. *American Psychologist, 31*, 129-148.

Brave, S., & Nass, C. (2002). Emotion in human-computer interaction. In J. Jacko & A. Sears (Eds.), *Handbook of human-computer interaction* (pp. 251-271). New York: Lawrence Erlbaum Associates.

Campos, J. J., Thein, S., & Daniela, O. (2003). A Darwinian legacy to understanding human infancy: Emotional expressions as behavior regulators. In P. Ekman, J. J. Campos, R. J. Davidson & F. B. M. de Waal (Eds.), *Emotions inside out: 130 years after Darwin's The Expression of the Emotions in Man and Animals*. New York: New York Academy of Sciences.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied Confersational Agents*. Cambridge, MA: MIT Press.

Chi, D., Costa, M., Zhao, L., & Badler, N. (2000). *The EMOTE Model for Effort and Shape.* Paper presented at the SIGGRAPH, New Orleans, LA.

Coats, E. J., Feldman, R. S., & Philippot, P. (1999). The influence of television on children's nonverbal behavior. In P. Philippot, R. S. Feldman & E. J. Coats (Eds.), *The social context of nonverbal behavior* (pp. 156-181). Paris: Cambridge University Press.

Cole, R., Van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., et al. (2003). Perceptive annimated interfaces: First steps toward a new paradigm for human-computer interaction. *Proceedings of the IEEE, 91*(9).

Conati, C. (2002). Probabilistic Assessment of User's Emotions in Educational Games. *Journal of Applied Artificial Intelligence, special issue on " Merging Cognition and Affect in HCI", 16*(7-8), 555-575.

Cowell, A., & Stanney, K. M. (2003). *Embodiement and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents.* Paper presented at the Intelligent Virtual Agents, Kloster Irsee, Germany.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Avon Books.

Darwin, C. (2002). *The Expression of the Emotions in Man and Animals* (3rd ed.): Oxford University Press.

de Waal, F. B. M. (2003). Darwin's Legacy and the Study of Primate Visual Communication. In P. Ekman, J. J. Campos, R. J. Davidson & F. B. M. de Waal (Eds.), *Emotions inside out: 130 years after Darwin's The Expression of the Emotions in Man and Animals*. New York: New York Academy of Sciences.

El Nasr, M. S., Yen, J., & Ioerger, T. (2000). FLAME: Fuzzy Logic Adaptive Model of Emotions. *Autonomous Agents and Multi-Agent Systems, 3*(3), 219-257.

Elliott, C. (1992). *The affective reasoner: A process model of emotions in a multi-agent system* (Ph.D Dissertation No. 32). Northwestern, IL: Northwestern University Institute for the Learning Sciences.

Fasel, I., Stewart-Bartlett, M., Littelwort-Ford, G., & Movellan, J. R. (2002). *Real time fully automatic coding of facial expressions from video.* Paper presented at the 9th Symposium on Neural Computation, California Institute of Technology.

Firby, J. (1987). *An investigation into reactive planning in complex domains.* Paper presented at the Sixth National Conference on Artificial Intelligence.

Forgas, J. P. (1991). Affect and social judgments: an introductory review. In J. P. Forgas (Ed.), *Emotion and socal judgments* (pp. 3-29). Oxford: Pergamon Press.

Frank, R. (1988). *Passions with reason: the strategic role of the emotions*. New York, NY: W. W. Norton.

Fridlund, A. J. (1997). The new ethology of human facial expressions. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The Psychology of Facial Expression* (pp. 103-129). Cambridge: Cambridge University Press.

Frijda, N. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion, 1*, 115-143.

Gifford, R. (1994). A Lens-Mapping Framework for Understanding the Encoding and Decoding of Interpersonal Dispositions in Nonverbal Behavior. *Journal of Personality and Social Psychology, 66*(2), 398-412.

Gratch, J., & Marsella, S. (2001). *Tears and Fears: Modeling Emotions and Emotional Behaviors in Synthetic Agents.* Paper presented at the Fifth International Conference on Autonomous Agents, Montreal, Canada.

Gratch, J., & Marsella, S. (2004). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research*.

Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems, July/August,* 54-61.

Haag, A., Goronzy, S., Schaich, P., & Williams, J. (2004). *Emotion recognition using bio-sensors: first steps towards an automatic system.* Paper presented at the Tutorial and Research Workshop on Affective Dialogue Systems, Kloster Irsee, Germany.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (Eds.). (1994). *Emotional Contagion*. Cambridge: Cambridge University Press.

Henderson, J. V. (1998). Comprehensive, Technology-Based clinical

Education: The "Virtual Practicum". *International Journal of Psychiatry in Medicine, 28*(1), 41-79.

Heylen, D., Nijholt, A., & Aker, o. d. (2005). Affect in Tutoring Dialogues. *Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors".*

Klesen, M. (2005). Using Theatrical Concepts for Role-Plays with Educational Agents. *Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors".*

Kramer, N. C., Tietz, B., & Bente, G. (2003). *Effects of embodied interface agents and their gestural activity.* Paper presented at the Intelligent Virtual Agents, Kloster Irsee, Germany.

Lance, B., Marsella, S., & Koizumi, D. (2004). *Towards expressive gaze manner in embodied virtual agents.* Paper presented at the AAMAS Workshop on Empathic Agents, New York.

Larsson, S., & Traum, D. (2000). Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering, 6*, 323-340.

Lazarus, R. (1991). *Emotion and Adaptation*. NY: Oxford University Press.

LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York, NY: Simon & Schuster.

Lee, C. M., & Narayanan, S. (2003). *Emotion recognition using a data-driven fuzzy interference system.* Paper presented at the Eurospeech, Geneva.

Lepper, M. R. (1988). Motivational Considerations in the Study of Instruction. *Cognition and Instruction, 5(4)*, 289-309.

Lester, J. C., Stone, B. A., & Stelling, G. D. (1999). Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments. *User Modeling and User-Adapted Instruction, 9*(1-2), 1-44.

Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L., & FitzGerald, P. J. (2000). Deictic and Emotive Communication in Animated Pedagogical Agents. In J. Cassell, S. Prevost, J. Sullivan & E. Churchill (Eds.), *Embodied Conversational Agents* (pp. 123-154). Cambridge: MIT Press.

Lisetti, C. L., & Schiano, D. (2000). Facial Expression Recognition: Where Human-Computer Interaction, Artificial Intelligence, and Cognitive Science Intersect. *Pragmatics and Cognition, 8*(1), 185-235.

Manstead, A., Fischer, A. H., & Jakobs, E. B. (1999). The Social and Emotional Functions of Facial Displays. In P. Philippot, R. S. Feldman & E. J. Coats (Eds.), *The Social Context of Nonverbal Behavior (Studies in Emotion and Social Interaction)* (pp. 287-316): Cambridge Univ Press.

Marsella, S., & Gratch, J. (2002). *A Step Toward Irrationality: Using Emotion to Change Belief.* Paper presented at the First International Joint Conference on Autonomous Agents and Multiagent Systems, Bologna, Italy.

Marsella, S., Gratch, J., & Rickel, J. (2003). Expressive Behaviors for Virtual Worlds. In H. Prendinger & M. Ishizuka (Eds.), *Life-like Characters Tools, Affective Functions and Applications*: Springer-Verlag.

Marsella, S., Johnson, W. L., & LaBore, C. (2000). *Interactive Pedagogical Drama.* Paper presented at the Fourth International Conference on Autonomous Agents, Montreal, Canada.

Marsella, S., Johnson, W. L., & LaBore, C. (2003). *Interactive pedagogical drama for health interventions.* Paper presented at the Conference on Artificial Intelligence in Education, Sydney, Australia.

Mele, A. R. (2001). *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press.

Moffat, D., & Frijda, N. (1995). *Where there's a Will there's an agent.* Paper presented at the Workshop on Agent Theories, Architectures and Languages.

Nakanishi, Shimuzu, & Isbister, K. (2005). Social Agents for Virtual Training. *Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors"*.

Nasby, W., & Yando, R. (1982). Selective encoding and retreival of affectively valent information: Two cognitive consequences of children's mood states. *Journal of Personality and Social Psychology, 43*, 1244-1253.

Nass, C., Moon, Y., & Carney, P. (1999). Are respondents polite to computers? Social desirability and direct responses to computers. *Journal of Applied Social Psychology, 29*(5), 1093-1110.

Neidenthal, P. M., Halberstadt, J. B., & Setterlund, M. B. (1997). Being happy and seeing "happy": Emotional state mediates visual word recognition. *Cognition and Emotion, 11*, 403-432.

Oatley, K., & Johnson-Laird, P. N. (1987). Cognitive Theory of Emotions. *Cognition and Emotion, 1*(1).

Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive Structure of Emotions*: Cambridge University Press.

Paiva, A., Aylett, R., & Marsella, S. (2004). *AAMAS Workshop on Empathic Agents*, from http://gaips.inesc.pt/gaips/en/aamas-ea/index.html

Paiva, A., Dias, J., Sobral, D., & Aylett, R. (2004). *Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents.* Paper presented at the Third International Joint Conference on Autonomous Agents and Multiagent Systems, New York.

PC-Based Human Interaction Training. from http://www.jhuapl.edu/education/sciencetech/factpctrain.html

Peacock, E., & Wong, P. (1990). The stress appraisal measure (SAM): A multidimensional approach to cognitive appraisal. *Stress Medicine, 6*, 227-236.

Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: MIT Press.

Prendinger, H., & Ishizuka, M. (2001). *Appraisal and filter programs for affective communication.* Paper presented at the AAAI Fall Symposium on Emotional and In-

telligent II: The Tangled Knot of Social Cognition, Technical Report FS-01-02, North Falmouth, MA.

Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., & Swartout, W. (2002). Toward a New Generation of Virtual Humans for Interactive Experiences. *IEEE Intelligent Systems, July/August,* 32-38.

Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2*(3), 247-270.

Rothbaum, B. O., Hodges, L. F., Alarcon, R., Ready, D., Shahar, F., Graap, K., et al. (1999). Virtual Environment Exposure Therapy for PTSD Vietnam Veterans: A Case Study. *Journal of Traumatic Stress*, 263-272.

Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual Peers as Partners in Storytelling and Literacy Learning. *Journal of Computer Assisted Learning, 19*(2), 195-208.

Scherer, K. R. (2001). Appraisal Considered as a Process of Multilevel Sequential Checking. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal Processes in Emotion: Theory, Methods, Research* (pp. 92-120): Oxford University Press.

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal Processes in Emotion*: Oxford University Press.

Scheutz, M., & Sloman, A. (2001). *Affect and agent control: experiments with simple affective states.* Paper presented at the IAT.

Searle, J. R. (1969). *Speech Acts*: Cambridge University Press.

Shaw, E., Johnson, W. L., & Ganeshan, R. (1999). *Pedagogical Agents on the Web.* Paper presented at the Proceedings of the Third International Conference on Autonomous Agents, Seattle, WA.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review, 74*, 29-39.

Smith, C. A., & Lazarus, R. (1990). Emotion and Adaptation. In Pervin (Ed.), *Handbook of Personality: theory & research* (pp. 609-637). NY: Guilford Press.

Smith, C. A., & Scott, H. S. (1997). A Componential Approach to the meaning of facial expressions. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The Psychology of Facial Expression* (pp. 229-254). Paris: Cambridge University Press.

Teasdale, J. D., & Russell, M. L. (1983). Differential effects of induced mood on recall of positive, negative, and neutral words. *British Journal of CLinical Psychology, 20*, 39-48.

Thagard, P. (2002). Why wasn't O. J. convicted: emotional coherence in legal inference. *Cognition and Emotion*.

Tyrrell, T. (1993). *Computational mechanisms for action selection.* Unpublished PhD, University of Edinburgh, Edinburgh.

Velásquez, J. (1998). *When robots weep: emotional memories and decision-making.* Paper presented at the Fifteenth National Conference on Artificial Intelligence, Madison, WI.