CHAPTER

8

Why and How to Build Emotion-Based Agent Architectures

Christine Lisetti *and* Eva Hudlicka

Abstract

In this chapter we explain the motivation, goals, and advantages of building artificial systems that simulate aspects of affective phenomena in humans, from architectures for rational agents to the simulation of empathic processes and affective disorders. We briefly review some of the main psychological and neuroscience theories of affect and emotion that inspire such computational modeling of affective processes. We also describe some of the diverse approaches explored to date to implement emotion-based architectures, including appraisal-based architectures, biologically inspired architectures, and hybrid architectures. Successes, challenges, and applications of emotion-based agent architectures and models are also discussed (e.g., modeling virtual patients and affective disorders with virtual humans, designing cybertherapy interventions, and building empathic virtual agents).

Key Words: computational models of emotions, affective architectures, cognitive-affective architectures, emotion-based agent architectures, virtual humans, virtual patients, affective disorder computational modeling, cybertherapy interventions, empathic virtual agents, applications of agent-based architectures

Motivation

There are many reasons for researchers and developers to be interested in creating computational models of, and agent architectures inspired from, affective phenomena. Affective phenomena include core affect, mood, emotion, as well as personality. This chapter discusses useful terminology and specific theories of affective phenomena and introduces some of the main motivations for this topic.

Building computational models of the roles of affective phenomena in human cognition is of interest to the cognitive science community. The main objective of cognitive science is to understand the human mind by developing theories of mind, creating computational models of these theories, and testing whether the input/output and timing behaviors of the resulting systems correspond to human behaviors (Thagard, 2008). Computational models of emerging cognitive and affective science theories of human emotion and affect will enable us to shed new light on the complexity of human affective phenomena.

Building emotion—or affect-based agent architectures—is also useful in subfields of computer science, such as artificial intelligence (AI), human-computer interaction (HCI), among others. AI, which focuses on developing algorithms to make rational intelligent decisions, can simulate and emulate the functional roles of affect and rational emotions in human decision making (Johnson-Laird & Oatley, 1992; Picard, 1997; Lisetti & Gmytrasiewicz, 2002). HCI on the other hand, is concerned with creating artificial agents that can adapt to users' emotion or personality to enhance adaptive human-computer interaction (Hudlicka, 2003; Picard, 1997).

The interest in building emotion-based agent architectures revolves around the notion that emotions have recently been fully acknowledged as an important part of human rational intelligence (Johnson-Laird & Oatley, 1992). Emotion research only recently emerged from its "dark ages," dated roughly from 1920 to 1960, in contrast with its classical phase, which started at the end of the nineteenth century. While psychologist William James (1984)—offering a very Darwinian view of emotion (Darwin, 1872)—restored affect as a valuable component of the evolutionary process, Cannon (1927) disagreed completely and relegated the roles of emotions to nonspecific, disruptive processes. Cannon's view contributed to the temporary demise of emotion research in the 1920s. The field of artificial intelligence, which formally emerged in 1956, founded most of its models of intelligence on previously established affectless theories of intelligence, originally rooted exclusively in logic (Russell & Norwig, 2011).

However, findings about the evidence of the universality and specificity in affective expressive behavior (Davidson and Cacioppo 1992; Ekman & Freisen, 1978; Ekman et al., 1983; among others), began the emotion research renaissance of the early 1980s. Furthermore, the 1990s benefited from neuroscience discoveries which confirmed the strong interconnections between the mechanisms mediating affective processes and those mediating cognition and reasoning (Damasio, 1994).

Since creating artificial agents that act rationally, in terms of achieving the best expected outcome, has been one of the main objectives of traditional AI (Russell & Norwig, 2011), the newly rediscovered role of emotions in rational human intelligence (de Sousa, 1990; Elster, 1999; Frank, 1988; Johnson-Laird & Oatley, 1992; Muramatsu, 2005) has begun to be modeled in architectures of rational agents in terms of their goal determination and interruption mechanisms (Frijda, 1987, 1995; Frijda & Swagerman, 1987; Jiang, 2008; Lisetti & Gmytrasiewicz, 2002; Murphy, et al., 2001; Ochs et al., 2012; Simon, 1967; Sloman, 1987; Sloman & Croucher, 1981; Sloman et al., 2001; Scheutz, 2011; Scheutz & Schermerhorn, 2009).

The more recent expressive AI endeavor (Mateas, 2011) is concerned with creating virtual agents that are socially intelligent and believable (1) in terms of their communicative expressiveness and behavior (Bates, 1994; Becker-Asano & Wachsmuth, 2009; Brave & Nass, 2002; Breazeal, 2003a, 2003b; Huang, et al., 2011; Lisetti et al., 2013; Loyall & Bates, 1997; Mateas, 2001, Pelachaud, 2009, Pütten et al., 2009) and (2) in terms of their awareness of the user's affective states (Calvo & DMello, 2010; Hudlicka & McNeese, 2002; Nasoz et al., 2010). For expressive AI, the simulation and recognition of the expressive patterns associated with emotion and personality is therefore essential.

Another reason to simulate and model some of the not-so-perfectly-rational aspects of affective human life, as well as the clearly dysfunctional ones, is emerging in domains such as entertainment, health care, medicine, and training across a variety of domains.

Creating goal-conflicted or even neurotic protagonists enhances the realism and complexity of computer games and interactive narratives in the same manner as in films and literature; complex characters engage audiences more deeply than simpler, happy, and stable characters (Campbell, 2008). Conflicted virtual characters can retain a player's interest in and engagement with the game by being unpredictable (in terms of rational behavior) and by portraying personality traits that make them unique, thereby giving the illusion of life (Bates, 1992, 1994; Johnson & Thomas, 1981; Loyall, 1997; Mateas, 2003; Ochs et al., 2012).

The design of virtual patients or mentally ill individuals has also begun to emerge to meet the recent training needs in health care, medicine, the military, and the police. These specialized personnel need to be trained in recognizing, understanding, and knowing how to deal with individuals with mental disorders (e.g., mood or personality disorders, schizophrenia, paranoia) or to help people with milder behavioral issues such as overeating, drinking, or smoking (Dunn, 2001). Emotions associated with these disorders and problematic mental states require different modeling approaches than the traditional modeling of the rationality of emotions discussed above. This modeling has begun to be addressed by the development of virtual patients (Campbell et al., 2011; Cook & Triola, 2009; Hubal et al., 2003; Rossen & Lok, 2012; Stevens et al., 2006; among others).

In the following pages we provide some background on the main psychological theories of affect and emotion and describe some of the recent progress and advances in (1) computational models of affect and emotion from a cognitive science perspective and (2) in emotion-based agent architectures from an AI and HCI perspective.

Theories of Emotion

Categorical Theories of Discrete Basic Emotions

Beginning with Darwin's evolutionary view of emotions (Darwin, 1872), Darwinian theories propose that emotions are "primary" or "basic" in the sense that they are considered to correspond to distinct and elementary forms of reactions, or action tendencies. Each discrete emotion calls into readiness a small and distinctive suite of action plans—action tendencies—that have been evolutionarily more successful than alternative kinds of reactions for survival and/or well-being and which have a large innate "hard-wired" component. Table 8.1, derived from Frijda (1986, 2008), shows a small set of the quadruples (action tendency, end state, function, emotion) that recur consistently across discrete basic emotion theories.

Although the number and choice of basic emotions vary depending on the different theories, ranging from 2 to 18 (Frijda, 1986, 1987; Izard 1971, 1992; James, 1984; Plutchik, 1980), these discrete theories share a number of features and consider emotions as (1) mental and physiological processes, (2) caused by the perception of phylogenetic categories of events,[1] (3) eliciting internal and external signals, and (4) being associated with a matching suite of innate hard-wired action plans or tendencies.

Perhaps the most well-known categorical theory of emotions to the affective computing community is Ekman's (1999), and we will show later how it has been used as a basis for modeling emotion in agent architectures. Ekman identifies seven characteristics that distinguish basic emotions from one another, and from other affective phenomena: (1) automatic appraisal, (2) distinctive universals in antecedent events, (3) presence in other primates, (4) quick onset, (5) brief duration (minutes, seconds), (6) unbidden occurrence (involuntary), and (7) distinctive physiology (e.g., autonomic nervous system, facial expressions).

According to Ekman, these seven characteristics are found in the following 17 basic emotions: amusement, anger, awe, contempt, contentment, disgust, embarrassment, excitement, fear, guilt, interest, pride in achievement, relief, sadness, satisfaction, sensory pleasure, and shame.

In addition, whereas Ekman initially thought that every basic emotion was associated with a unique facial expression (Ekman, 1984), he revised his theory in 1993 to account for emotions for which no facial signals exist (such as potentially awe, guilt, and shame) and for emotions that share the same expression (e.g., different categories of positive emotions all sharing a smile). In total, Ekman (1993) identified seven emotions with distinctive universal unique facial expressions: anger, fear, disgust, sadness, happiness, surprise, and (the one added last) contempt.

As described later, although highly popular in affective computing (Picard, 1997), the notion of basic emotions is still controversial among psychologists (Ortony, 1990; Russell & Barrett, 1999).

Dimensional Theories and Models of Core Affect and Mood

One important distinction that has been made by Russell and Barrett (1999), involves the use of the term *prototypical emotional episode* to refer to

Table 8.1 Examples of Action Tendencies

Action Tendency	End State	Function	Emotion
Approach	Access	Permission forconsummatory activity	Desire
Avoidance	Own inaccessibility	Protection	Fear
Attendance	Identification	Orientation	Interest
Rejection	Removal of object	Protection	Disgust
Antagonism	Removal of obstruction	Regaining of control	Anger
Interruption	Reorientation	Reorientation	Shock, surprise
Free activation	Action tendency's end state	Generalized readiness	Joy
Inactivity	Action tendency's end state	Recuperation	Contentment
Inhibition/preparation	Absence of response	Caution	Anxiety

Source: Adapted from Frijda, 1986.

what is typically called "emotion," and the use of the term *core affect* to refer to the most elementary affective feelings (and their neurophysiological counterparts).

According to Russell and Barrett (1999), core affect is not necessarily part of a person's consciousness, nor is it consciously directed at anything (e.g., sense of pleasure or displeasure, tension or relaxation, depression or elation). Core affect can be as free-floating as a mood but it can be directed when it becomes part of an emotional episode or emotions. Core affect is always caused, although its causes might be beyond human ability to detect (e.g., from specific events, to weather changes, to diurnal cycles). Core affect is also the underlying, always present feeling one has about whether one is in a positive or negative state, aroused or relaxed (or neutral, since core affect is always present).

Core affect elemental feeling is to be understood as included within a full-blown prototypical emotional episode, if one occurs, A prototypical emotional episode also includes behavior in relation to the object/event, attention toward and appraisal of that object, subjective experience, and physiologic responses.

In making this distinction between core affect and prototypical emotional episodes, Russell and Barrett (1999) establish that since core affect is more basic than a full-blown emotional episode it carries less information than emotions and needs to be studied and measured with fewer dimensions (although if considered as a component of an emotional episode, its low-dimensional structure is still valid).

Typically, two or three dimensions are used to represent core affect. Most frequently these are valence and arousal (Russell, 1980, 2003; Russell & Barrett, 1999; Russell & Mehrabian, 1977). Valence reflects a positive or negative evaluation, and the associated felt state of pleasure (vs. displeasure). Arousal reflects a general degree of intensity or activation of the organism. The degree of arousal reflects a general readiness to act: low arousal is associated with less energy, high arousal with more energy.

Since this two-dimensional space cannot easily differentiate among core affective states that share the same values of arousal and valence (e.g., anger and fear, both characterized by high arousal and negative valence), a third dimension is often added. This is variously termed dominance or stance (versus submissiveness). The resulting three-dimensional space is often referred to as the PAD space, for pleasure (synonymous with valence), arousal, and dominance (Mehrabian, 1995).

It is important to note that according to Russell (1980), the dimensional structure is useful only to characterize core affect (versus full-blown emotions) because full-blown emotions fall into only certain regions of the circumplex structure defined by the core affect dimensions. Qualitatively different events can appear similar or identical when only this dimensional structure is considered. For example, fear, anger, embarrassment, and disgust could share identical core affect and therefore fall in identical points or regions in the circumplex structure.

Note that the pleasure and arousal dimensions and the resulting circumplex structure represent only one component of a prototypical emotional episode, but not all of the components. These other components then differentiate among fear, anger, embarrassment, and disgust. Thus assessment devices based on the dimensional-circumplex approach can capture core affect but miss the (other) components of a prototypical emotional episode. This is an important aspect to consider when aiming to recognize emotion automatically.

Componential and Appraisal-Based Theories of Emotions

The componential perspective or appraisal-based theories emphasize distinct components of emotions (Leventhal & Scherer, 1987). The term *components* refers to both the distinct modalities of emotions (e.g., cognitive, physiologic, behavioral, subjective) but frequently also to the components of the cognitive appraisal process. In the latter case, these are referred to as appraisal dimensions or appraisal variables (Lazarus, 1991) and include novelty, valence, goal relevance, goal congruence, and coping abilities.

A stimulus, whether real or imagined, is analyzed in terms of its meaning and consequences for the agent in order to determine the affective reaction. The analysis involves assigning specific values to the appraisal variables. Once the appraisal variable values are determined by the organism's evaluative processes, the resulting vector is mapped onto a particular emotion, within the *n*-dimensional space defined by the *n* appraisal variables.

Appraisal theories of emotions have been modeled most predominantly within the affective computing community, and appraisal models are described in Gratch and Marsella's chapter in this volume. We therefore mention some of their main tenets only briefly in this chapter.

ORTONY'S OCC MODEL

The best-known theory of cognitive appraisal, and one most frequently used by the affective computing community, is a theory developed by Ortony, Collins and Clore (1988), which describes the cognitive structure of emotions. It is frequently referred to as the OCC model. Because it is covered extensively in Gratch & Marsella (this volume), we provide only a brief summary below. The OCC model describes a hierarchy that classifies 22 different types of emotions along three main branches: emotions classified in terms of (1) consequences of events (pleased, displeased), (2) actions of agents (approving or disapproving), and (3) aspects of objects (liking, disliking). Emotions are *valenced* (positive or negative) reactions to one or another of these three aspects of experience. Some subsequent branches combine to form compound emotions.

The popularity of the OCC model in the affective computing community is due in part to its relatively simple taxonomy of classes of emotions, relying on concepts such as agents and actions that are already used to conceptualize and implement agent architectures.

SCHERER'S CPT

Another influential theory of emotions in affective computing is Scherer's component process theory of emotions (CPT) (2001b). Scherer's CPT describes emotions as arising from a process of evaluation of the surrounding events with respect to their significance for the survival and well-being of the organism. The nature of this appraisal is related to a sequential evaluation of each event with regards to a set of parameters called sequential evaluation checks (SECs). SECs are chosen to represent the minimum set of dimensions necessary to differentiate among distinct emotions and are organized into four classes or in terms of four appraisal objectives. These objectives reflect answers to the following questions: How relevant is the event for me? (Relevance SECs.) What are the implications or consequences of this event? (Implications SECs.) How well can I cope with these consequences? (Coping Potential SECs.) What is the significance of this event with respect to social norms and to my self concept? (Normative significance SECs.)

One of the primary reasons for the sequential approach is to provide a mechanism whereby focusing of attention is only employed when needed and information processing (computational loading) is theoretically reduced. The SEC approach also parallels the three-layered hybrid AI architectures when it describes a three-layered emotional processing of events:

1. *Sensorimotor Level*: Checking occurs through innate feature detection and reflex systems based on specific stimulus patterns. Generally it involves genetically determined reflex behaviors and the generation of primary emotions in response to basic stimulus features.
2. *Schematic Level*: Checking is a learned automatic nondeliberative rapid response to specific stimulus patterns largely based on social learning processes.
3. *Conceptual Level*: Checking is based on conscious reflective (deliberative) processing of evaluation criteria provided through propositional memory storage mechanisms. Planning, thinking and anticipating events and reactions are typical conceptual-level actions.

Other appraisal-based theories of emotions have also been developed and, as we mention later, some of them have also influenced the affective computing community (e.g., Smith & Lazarus, 1990; Lazarus, 1991).

Challenges in Modeling Neurophysiologic Theories and Unconscious Appraisal

Neurophysiologic theories of emotions have the potential to enable the affective computing community to develop new emotion-based architectures, ones focused on how neural circuitry can generate emotions. However, these theories typically address processes that take place in the unconscious and which have not yet been widely explored in affective computing.

We briefly mention three researchers whose work is relevant for biologically inspired emotion-based agent architectures: LeDoux, Zajonc, and Damasio, although many others should also be studied.

Until recently, neuroscientists assumed that all sensory information was processed in the thalamus, then sent to the neocortex, and finally to the amygdala, where the information was translated into an emotional response. Research by LeDoux (1992) on fear conditioning and the amygdala showed that information from the thalamus can also go directly to the amygdala, bypassing the neocortex. Fear conditioning has been modeled with anatomically constrained neural networks to show how emotional information and behavior are related to anatomical and physiological observations (Armony et al., 1995).

Zajonc (1980, 1984) suggested that the processing pathways identified by LeDoux—the direct connection between the thalamus and the amygdala—are extremely important, because they indicate that emotional reactions can take place without the participation of cognitive processes. According to Zajonc, these findings would explain, for example, why individuals with phobias do not respond to logic, and the difficulty of bringing these fears or neuroses under control with psychological interventions (Zajonc, 1980, 1984b). Although Zajonc's work has not been greatly influential in affective computing, its focus on core affect may become more relevant when researchers begin to model the unconscious processes of affect (Zajonc, 1984).

It should be noted that most of LeDoux's research, which views emotions as separate from cognition, remains within the scope of one single emotion—namely fear (LeDoux, 1995). However, as LeDoux states, "[F]ear is an interesting emotion to study because many disorders of fear regulation are at the heart of many psychopathologic conditions, including anxiety, panic, phobias, and post-traumatic stress disorders." (see Riva et al., this volume to learn about how cybertherapy has been helping people with such disorders).

The somatic markers hypothesis proposed by Damasio (1994) brings another contribution to the notion that emotional guidance helps rationality. Somatic markers are those emotionally borne physical sensations "telling" those who experience them that an event is likely to lead to pleasure or pain. Somatic markers precede thought and reason. They do not replace inference or calculation, but they enhance decision making by drastically reducing the number of options for consideration.

Agent Architectures and Cognitive Models of Affective Phenomena

Identifying Theoretical Assumptions

Computational models of affect and emotion necessarily make tacit assumptions about the overall cognitive architecture of the agent, specifically, assumptions about how the agent represents the world, chooses actions, and learns over time. Cognitive theories (Ortony et al., 1988; Scherer, 2001; Smith & Kirby, 2001), which ground affect in inferences about the effects of objects, states, and events on the agent's goals, necessarily assume the presence of an inference system to make those inferences, together with a world model capable of supporting those inferences.

Neurophysiological theories by contrast, being generally grounded in human unconscious processes (Bechara et al., 1997; Damasio, 1994; Zajonc, 1980), or animal models (Gray & McNaughton, 2003; LeDoux, 1992, 1995, 2000; Rolls, 2007), are less likely to highlight the role of inference and more likely to highlight other organizations such as competition between quasi-independent behavior systems.

As discussed above, a variety of neurologic and psychologic processes are involved in producing affective phenomena: core affect, emotional episodes or full-blown emotions, moods, attitudes, and, to some extent, personality, as it influences an individual's patterns of behavior, including affective patterns.

The different theories of affect and emotions discussed above—discrete, dimensional, and componential—are applied in the context of the architectures for which they are most natural. Cognitive theories are generally applied to planner/executive architectures or reactive planners. Biological theories are generally applied to behavior-based architectures (Arbib, 2004; Arkin, 1998; Murphy, 2000). At the same time, the different theories often seek to explain different aspects of the overall phenomenon of affect.

Consequently, developing an overall theory for affect/emotion modeling would require reconciling not just the theories themselves, narrowly construed, but also their architectural assumptions. This aim, however, resembles the early dreams of strong AI, and its disillusions (Dreyfus, 1992; Dreyfus & Dreyfus, 1988) and is currently considered out of reach. Building agents and models with some limited aspects of affective phenomena, however, is feasible and desirable, as discussed earlier. It requires choosing a theory in terms of its architectural assumptions or adapting particular theoretical aspects to produce the desired functionality of the agent.

Emotion-Based Agent Architecture Overview

A typical intelligent agent architecture comprises the sensors that the agent uses to perceive its environment, a decision-making mechanism to decide what most appropriate action(s) to take at any time, and actuators that the agent activates to carry out its actions. At any time, the agent keeps track of its changing environment by using those different knowledge-representation schemas that are most relevant to the nature of its environment.

Emotion-based architectures are developed primarily for interactive intelligent agents capable of adapting to their user's affective states and manifesting affective behavior and empathy. These architectures are also developed to enhance the adaptive functioning of robots (e.g., Scheutz, 2000), and for research purposes, to explore the mechanisms of affective processes (e.g., Hudlicka, 2008). Emotion-based architectures vary in type, but they usually include (a subset of) the following components.

SENSORS

Sensors must be able to sense the user's emotional states (to some degree of accuracy appropriate for a given context) shown via one or more human emotional expressive modalities (sometimes referred to as user-centered modes). Communicative affective signals of human expression include facial expressions (which can be categorized slightly differently depending on which theory is used), gestures, vocal intonation (primarily volume, pitch), sensorimotor cues (e.g., pressure), autonomic nervous system signals associated with valence and arousal, as well natural language (which is used to communicate feelings or the subjective experience of affective states).

The agent can then capture and interpret those multimodal affective signals and translate them in terms of the most probable of the user's affective states. Depending upon the context of interaction, unimodal recognition of affect can be sufficient, whereas other types of interaction might require multimodal recognition and sensor fusion (Calvo & D'Mello, 2010; Paleari & Lisetti, 2006), as well as other nonaffective sensors.

DECISION-MAKING ALGORITHMS

Based on the agent's specific role and goals, the decision-making algorithm varies depending upon (as discussed above) which affect/emotion theory or combination of theories inspires the architecture. These decisions can be designed to have an effect not only on the agent's simulated affective state itself but also on the agent's expression of emotion via a variety of modalities (or agent-centered modes) activated by actuators.

ACTUATORS

The agent actuators can be chosen to control anthropomorphic embodiments endowed with modalities such as facial expressions, verbal, vocal intonation, or body posture. Anthropomorphic agents have the advantage that users innately understand them because they use the same social and emotional cues as those found in human-human communication. Anthropomorphic agents also elicit reciprocal social behaviors in their users (Reeves & Nass, 1996). Such actuators are most often portrayed by embodied conversational agents (Cassell et al., 2000); they can have graphical or robotic platforms (Breazeal, 2003b) or a mix of both (Lisetti et al., 2004). Other approaches to communicate affective expression have been explored in terms of nonfacial and nonverbal channels, such as appropriate social distance (see (Bethel & Murphy, 2008) for a survey), or the use of shape and color (Hook, 2004).

The majority of existing emotion-based architectures emphasize the generation of emotion via cognitive appraisal, and the effects of emotions on expressive behavior and choice of action. These are the focus of the remainder of this chapter. Less frequent are architectures that emphasize the effects of emotions on internal cognitive processing, or the cognitive consequences of emotions. A detailed discussion of these models is beyond the scope of this chapter, but examples include the MAMID architecture, which focuses on modeling affective biases on cognition (Hudlicka, 1998; 2007; 2011) and models developed by Ritter and colleagues in the context of ACT-R (Ritter & Avramides, 2000).

Basic Emotions and Agent Architectures

As mentioned earlier, categorical theories of basic emotions have had a very large influence on the affective computing community. The computational appeal of these theories lies in a clear mapping between a small set of universal antecedents to corresponding emotions along with their associated action tendencies.

Using categorical theories, an artificial agent can be designed to (1) sense a set of triggers (e.g., dangers, appeals) specific to its physicality or embodiment, (2) respond to these with action tendencies (approach, avoid, attend, reject, interrupt) implemented as a reflex-based agent (Russell and Norwig, 2011) using action-reaction rules, and (3) actuate these actions via its actuators (e.g. robot motors, two- or three-dimensional character graphics) in a manner that is psychologically valid. It should be noted however, that the reflex like nature of action tendencies is also present in noncategorical theories such as Scherer's CPT (2001), where action tendencies are activated at the lowest level of processing, namely the sensorimotor level (discussed in the previous section). For example, such an agent architecture has been implemented in two cooperative robots (Murphy et al., 2001) where states such as anger and frustration prompted robots to adjust their collaborative task strategy.

Ekman's theory of basic emotion, in particular, has had an additional appeal to the affective computing research community because (in addition to a small finite set of emotion/action tendency pairs), it provides a detailed description of the muscular activity of facial expressions. Specifically, using the widely known facial action coding system (FACS) (Ekman, 1978, 1983, 2002), Ekman's theory of basic emotions provides encoding for all of the facial movements involved in Ekman's six universal basic expressions of emotions (or EmFACS): anger, fear, disgust, sadness, happiness, and surprise (Friesen & Ekman, 1983).

Understandably, FACS, EmFACS, and the corresponding CMU-Pittsburgh AU-coded face expression image database (Kanade et al, 2000) have been very instrumental to the progress of automatic facial expression recognition and analysis, on the one hand, and of facial expression generation or synthesis on the other. Given a proper facial expression recognition sensing algorithm (Tian, 2001; Wolf et al., 2011), an agent can consistently recognize the user's state associated with the user's facial expressions. If desirable, it can also respond with synthesized facial expressions of its own (robotic head animations or a graphical virtual character's face) (Breazeal, 2003a, 2004; Pelachaud, 2009, Lisetti et al., 2013).

The quasi-exclusive focus on Ekman's six emotions, however, has limited the impact that emotion-based agents can have during human-computer interaction (HCI) in real-life scenarios. For example, users' facial expressions are often more varied than Ekman's six basic expressions (e.g., student's confusion or boredom) (Calvo & D'Mello, 2010).

Alternative approaches have studied how expressions of emotion are associated with fine-grained cognitive (thinking) processes (Scherer, 1992) (discussed earlier), or expressions that display mixed emotions (Ochs et al., 2005). Affective computing researchers need to continue to work toward including fine-grained AU-based facial expressions as a modality of agents' emotional expressions (Amini & Lisetti, 2013).

Appraisal Theories of Emotions, Agent Architectures, and Cognitive Models

COGNITIVE SCIENCE MODELS OF EMOTIONS

One of the first cognitive science modeling attempts was Newell and Simon's general problem solver (1961) which allowed a comparison of the traces of its reasoning steps with traces of human thinking processes on the same task. Other attempts followed, such as the SOAR theory of mind modeling long- and short-term memory (Laird et al., 1987; Lewis, et al., 1990) which continued to evolve (Laird & Rosenbloom, 1996). Another important cognitive science approach can be found in the adaptive control of thought-rational (ACT-R) symbolic theory of human knowledge (in terms of declarative representations of objects with schema like structures or chunks) and procedural representations of transformations in the environment (with production rules) (Anderson, 1993, 1996; Anderson and Lebiere, 1998). ACT has continued to evolve with ACT-R 5.0 (Anderson et al., 2004).

Whereas these cognitive theories of mind did not model emotions (and even considered them as noise) (Posner, 1993), recent cognitive models have begun to include the roles of emotion in cognition. EMA (Gratch, 2004; Marsella & Gratch, 2009), a rule-based domain-independent framework based on SOAR for modeling emotion, models how emotion and cognition influence each other using Lazarus' appraisal theory (1991). EMA models an agent's cognitive evaluation of a situation using a set of appraisal variables to represent the resulting emotion (possibly recalling previous situations from memory), as well as emotion-focused coping strategies that the agent can activate to reappraise the situation.

Another cognitive model of emotions is found in the SOAR-Emote model (Marinier, 2004), which is a simplified version of the basic SOAR-based cognitive appraisal model used in EMA. It uses Damasio's theory of emotions and feelings (Damasio, 1994) to also account for the influences of the body and physiology in determining affect. Furthermore following Damasio's view, the direction of causality for feelings and physiological effects in SOAR-Emote is reversed compared to EMA in which the agent first determines how it feels via cognitive appraisal and then displays appropriate body language to reflect that emotion. In subsequent work, SOAR-Emote (Marinier & Laird, 2007) comes closer to Scherer's theory of emotion generation (2001).

There have also been attempts in the ACT-R community to model emotion and motivation (Fum & Stocco, 2004).

Finally, it is interesting to note that cognitive science models of emotions and affect can also be constructed from the noncognitive nonappraisal theories of emotions (Armony et al., 1995), though much more research is called for in that domain.

APPRAISAL-BASED AGENT ARCHITECTURES

The mapping of the emotion elicitors (also referred to as emotion antecedents or emotion triggers) from the environment onto the resulting emotion (or other affective state) is the core task of the emotion generation process, implemented via cognitive appraisal. It reflects the agent's evaluation of these stimuli, in light of its goals, beliefs and behavioral capabilities and available resources.

This computational task has extensive theoretical support in the cognitive theories of emotion generation (e.g., OCC, CPT). Existing empirical data also provide a rich source of evidence regarding the nature of the trigger-to-emotion mappings (see discussion above). We know that the possibility of bodily harm triggers fear; obstruction of one's goals triggers anger; loss of love objects triggers sadness; achieving an important goal triggers happiness, and so on. When a componential model is used, a series of evaluative criteria or appraisal variables are used to represent the results of the evaluation of the triggers with respect to the agent's goals and beliefs.

As mentioned, the most commonly used set of evaluative criteria are those first identified by the OCC model, and OCC is the most frequently implemented model of emotion generation via cognitive appraisal. It uses concepts such as agents, objects, and events that are very similar to constructs used to implement virtual agents. A few of these OCC-inspired systems are Oz, EM, HAP, Affective Reasoner, FearNot!, EMA, MAMID, Greta (Adam, 2006; Andre et al., 2000; Aylett et al., 2007; Bates, 1992, 1994; De Rosis et al., 2003; Elliott, 1992; Gratch, 2004; Gratch et al., 2007b; Hudlicka, 1998; Loyall, 1997; Reilly, 1997; Marsella, 2000; Marsella & Gratch, 2009; Mateas, 2003; Predinger & Ishizuka, 2004a, Hermann et al., 2007).

The component process theory (CPT) (Scherer, 2001, 2009) has been interesting for emotion-based agents for two main reasons: (1) it considers emotions with their complex three levels (sensorimotor, schematic, and conceptual) nature and (2) it addresses human multimodal expression of emotion. CPT has been used as a guideline for developing both the generation and recognition of emotive expression and has been applied to the generation of virtual character facial expression (Paleari & Lisetti, 2006) and to sensor fusion (Paleari et al., 2007).

Few models have used appraisal variables defined by componential theorists. These include the GENESE (Scherer, 1993) and the GATE model (Wehrle & Scherer, 2001). The GATE model uses appraisal variables defined by Scherer (2001) to implement the second stage of the mapping process and maps the appraisal variable values onto the associated emotions in the multidimensional space defined by the variables.

Increasingly, models of emotion generation via cognitive appraisal are combining both the OCC evaluative criteria and appraisal variables from componential theories—for example, FLAME (El-Nasr, 2000) and EMA (Gratch & Marsella, 2004).

Furthermore, while the majority of existing symbolic agent architectures use cognitive appraisal theory as the basis for emotion generation, several models have emerged that attempt to integrate additional modalities, most often a simulation of the physiologic modality (e.g., Breazeal, 2003a; Canamero, 1997; Hiolle et al., 2012; Scheutz, 2004; Velásquez, 1997).

Dimensional Computational Models of Affective States

When the dimensional perspective is used, affective states are represented in terms of doubles or triples, representing the two dimensions of pleasure and arousal, or the three dimensions of pleasure, arousal, and dominance (see previous section).

Examples of architectures using the dimensional model for emotion generation include the social robot Kismet (Breazeal, 2003a), the WASABI architecture used for synthetic agent Max (Becker, Kopp, & Wachsmuth, 2004; Becker-Asano & Wachsmuth, 2009), and the arousal-based model of dyadic human-robot attachment interactions (Hiolle et al., 2012).

The dimensional theories of emotions have also contributed to progress in emotion recognition from physiological signals; in that respect, they are very relevant to emotion-based agents with abilities to sense the continuous nature of affect (Calvo, 2010; Gunes & Pantic, 2010; Lisetti & Nasoz, 2004; Peter & Herbon, 2006; Predinger & Ishizuka, 2004a).

Emerging Challenges: Modeling Emotional Conflicts and Affective Disorders

Modeling Affective Disorders

We have already discussed how the modeling of rationality has been one of the main motivations of AI until recently. We also want to point out that human quirks and failings can be at least as interesting to study as our intelligence.

This is not a new argument; Colby's (1975) seminal PARRY system, a model of paranoid belief structures implemented as a LISP simulation, is perhaps the earliest example of work in this vein.

Since then, a few psychologists and psychiatrists exploring the relationship between cognitive deficits and disturbances of neuroanatomy and neurophysiology (e.g., schizophrenia, paranoia, Alzheimer's disease), have built computer models of these phenomena to gather further insights into their theories (Cohen and Servan-Schreiber, 1992; Servan-Schreiber, 1986), and to make predictions of the effects of brain disturbance on cognitive function (O'Donnell, 2006). These models use a connectionist approach to represent cognitive or neural processes with artificial neural networks (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986).

One of the challenges facing affective computing researchers interested in modeling affective disorders is the identification of the mechanisms underlying psychopathology and affective disorders. Whereas the primary theories of emotions discussed above focus on adaptive affective functioning, modeling of affective disorders will require a more nuanced understanding of the mechanisms underlying psychopathology. In addition, modeling of these mechanisms will also enhance our understanding of normal affective functioning.

Work in this area is in its infancy, and more research in both psychological theories and computational approaches will be required to address these challenges. An example of this effort is a recent attempt to model alternative mechanisms underlying a range of anxiety disorders, within an agent architecture that models the effects of emotions on cognition in terms global parameters influencing multiple cognitive processes (Hudlicka, 2008).

Virtual Counseling and Virtual Humans

Some of the same psychologists at the forefront of computational models of affective disorders (Servan-Schreiber, 1986) also favored early on the notion of computerized psychotherapy.

This concept is not new either: Weizenbaum's ELIZA (1967) was the first program to simulate a psychotherapist and used simple textual pattern matching to imitate the responses of a Rogerian psychotherapist (Rogers, 1959). However, after ELIZA's unsuspected success in terms of its ability to engage users in ongoing "conversations," Weizenbaum (1976) became ambivalent about the possibility of using computers for therapy because computers would always lack essential human qualities such as compassion, empathy, and wisdom.

However, since research results established that people respond socially to computers displaying social cues (Reeves & Nass, 1996), the motivation to build socially intelligent computers as a new mode for HCI grew steadily (and, as we will see, including for therapy). The tremendous recent progress in the design of embodied conversational agents (ECAs) and intelligent virtual agents (IVAs), since their first appearance (Cassell, 2000), have changed our views of human-computer interaction. They have now become so effectively communicative in their anthropomorphic forms that they are often referred to as virtual humans (VHs) (Hill et al., 2003; Swartout et al., 2001; Swartout, 2010.

Virtual human characters now use sophisticated multimodal communication abilities such as facial expressions, gaze, and gestures (Amini & Lisetti, 2013; Bailenson, et al., 2001; De Rosis et al., 2003; Pelachaud, 2002, 2003, 2004, 2009; Poggi, 2005; Predinger and Ishizuka, 2004a, 2004b; Rutter, et al., 1984). They can establish rapport with back-channel cues such as head nods, smiles, shift of gaze or posture, or mimicry of head gestures (Gratch 2006, 2007, 2007a; Huang et al, 2011; Kang, 2008; McQuiggan, 2008; Pelachaud 2009; Prendinger & Ishizuka, 2005; Putten, et al., 2009; Wang 2010, 2009), communicate empathically (Aylett, 2007; Boukricha, 2007, 2009, 2011; McQuiggan & Lester, 2007; Nguyen, 2009; Prendinger & Ishizuka, 2005), and engage in social talk (Bickmore, 2005a, 2005; Bickmore & Giorgino, 2006; Cassell and Bickmore, 2003; Kluwer, 2011; Schulman & Bickmore, 2011).

As a result, virtual human characters open many new domains for HCI that were not feasible earlier and reopen old debates about the potential roles of computers, including the use of computers for augmenting psychotherapy (Hudlicka, 2005; Hudlicka et al., 2008, Lisetti et al., 2013). Virtual humans are making their debut as virtual counselors (Bickmore, 2010; Lisetti, 2012; Lisetti & Wagner, 2008; Rizzo et al., 2012, Lisetti et al., 2013). Robots are also being studied in a therapeutic context (Stiehl & Lieberman, 2005). Riva et al. in this volume survey some of latest progress in cybertherapy, and van den Broeck et al. (this volume) discuss the role of ECAs in health applications in general.

Virtual Patients for Mental Health

One obvious case where the simulation and modeling of human psychological problems is useful is in the treatment of such problems. Virtual patients are currently being designed to model these problems from the cognitive science approach we discussed earlier. Virtual patients are also used to train health-care and medical personnel via role playing with virtual patients exhibiting the symptoms of affective disorders before they begin to work with real patients (Cook & Triola, 2009; Cook et al., 2010; Hoffman, 2011; Hubal, 2003; Magerko, 2011; Stevens et al., 2006; Villaume, 2006).

Another potential use is to model these systems within synthetic characters with whom the patient interacts. The patient could then experiment with the character's behavior, subjecting them to different situations and observing the results, as a way of coming to better understand their own behavior. The system could also display the internal state variables of the character, such as their level of effortful control, so as to help the patient better understand the dynamics of their own behavior. Some systems have taken a similar approach (Aylett, 2007; Wilkinson et al., 2008), and their potential impact on a wide range of interventions for mental issues calls for more work in that direction.

Conflicted Protagonist Characters for Computer Games

Another case for modeling affective disorders is found in entertainment scenarios such as interactive storytelling or computer games. To build interactive games and storytelling, one needs to construct synthetic characters whose reactions are believable in the sense of making a user willing to suspend disbelief (Johnston, 1981) regardless of their overall realism. For these applications, the pauses and hesitations due to the internal inhibition of a conflicted character, or the obvious lack of inhibition of a drunken character, can be important to establishing the believability of a character.

Narrative traditionally involves characters who are presented with conflicts and challenges, often from within (Campbell, 2008). These applications provide a wonderful sandbox in which to experiment with simulations of human psychology, including humans who are not at their best.

More generally, a disproportionate amount of storytelling involves characters who are flawed or simply not at their best, particularly in applications such as interactive drama (Bates, 1992, 1994; Johnson & Thomas, 1981; Loyall, 1997; Mateas, 2003; Marsella, 2000; Hermann et al., 2007).

With the continuous rise of entertainment applications, this area of research is also very promising, where affective computing researchers can reach out to artists and vice versa.

Conclusions

In this chapter we have explained the various motivations for building emotion-based agents and provided a brief overview of the main emotion theories relevant for such architectures. We then surveyed some of the recent progress in the field, including interactive expressive virtual characters. We also briefly mentioned some less known neurophysiologic theories in the hope that they might give rise to novel approaches for biologically inspired emotion-based architectures. Finally, we discussed some of the latest application domains for emotion-based intelligent agents, such as interactive drama, mental health promotion, and personnel training. We hope to have demonstrated the importance of emotion-based agent architectures and models in current and future digital artifacts.

Note

1. In an attempt to catalogue human phylogenetic sets of affectively loaded events that consistently trigger the same emotion across human subjects, a set of emotional stimuli for experimental investigations of emotion and attention was compiled (Lang et al., 1997)—the International Affective Picture System (IAPS)—with the goal of providing researchers with a large set of standardized emotionally evocative, internationally accessible color photographs with content across a wide range of semantic categories. IAPS has been heavily used for the recognition of emotion across subjects as it attempts to provide an objective baseline for the generation of human emotion.

References

Adam, C., Gaudou, B., Herzig, A., & Longin, D. (2006). OCC's emotions: A formalization in BDI logic. In J. Euzenat & J. Domingue (Eds.), *Proceedings of the international conference on artificial intelligence: Methodology, systems, applications* (pp. 24–32). Berlin: Springer-Verlag, LNAI 4183.

Anderson, J. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355–365.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y.(2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Andre, E., Klesen, M., Gebhard, P., Allen, S., & Rist, T. (2000). Exploiting models of personality and emotions to control the behavior of animated interactive agents.... on Autonomous Agents, (October).

Amini, R., & Lisetti, C. (2013). HapFACS: An opensource API/software for AU-based generation of facial expressions for embodied conversational agents. In *Proceedings of Affective Computing and Intelligent Interactions*.

Arbib, M., & Fellous, J.-M. (2004). Emotions: From brain to robot. *Trends in Cognitive Sciences*, *8*(12), 554–561.

Arkin, R. (1998). *Behavior-based robotics*. Cambridge, MA: MIT Press.

Armony, J., Cohen, J., Servan-Schreiber, D., & Ledoux, J. (1995). An anatomically constrained neural network model of fear conditioning. *Behavioral Neuroscience*, *109*(2), 246–257.

Aylett, R., Vala, M., Sequeira, P., & Paiva, A. (2007). *Fearnot! An emergent narrative approach to virtual dramas for anti-bullying education* (Vol. LNCS 4871, pp. 202–205). Berlin: Springer-Verlag.

Bailenson, J., Blascovich, J., Beall, A., & Loomis, J. (2001). Equilibrium revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments*, *10*(6), 583–598.

Bates, J. (1992). Virtual reality, art, and entertainment. *Presence: Teleoperators and Virtual Environments*, *1*(1), 133–138.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, *37*(7).

Bechara, A., Damasio, H, Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Becker-Asano, C., & Wachsmuth, I. (2009). Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*, *20*(1), 32–49.

Bethel, C., & Murphy, R. (2008). Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems, Man, and Cybernetics*, Part C, *38*(1), 83–92.

Bickmore, T., & Giorgino, T. (2006). Methodological review: Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, *39*(5), 65–467.

Bickmore, T., & Gruber, A. (2010). Relational agents in clinical psychiatry. *Harvard Review of Psychiatry*, *18*(2), 119–130.

Bickmore, T., Gruber, A., & Picard, R. (2005). Establishing the computer patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, *59*, 21–30.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 617–638.

Boukricha, H., Becker-Asano, C., & Wachsmuth, I. (2007). Simulating empathy for the virtual human Max. In D. Reichardt & P. Levi (Eds.), *2nd Workshop on Emotion and Computing, In conjunction with 30th German Conf. on Artificial Intelligence (KI 2007)* (pp. 23–28), Osnabrück.

Boukricha, H., & Wachsmuth, I. (2011). Empathy-based emotional alignment for a virtual human: A three-step approach. *KI—Kunstliche Intelligenz*, *25*(3), 195–204.

Boukricha, H., Wachsmuth, I., Hofstatter, A., & Grammer, K. (2009). Pleasure-arousal-dominance driven facial expression simulation. In *Affective computing and intelligent interaction and workshops, 2009* (pp. 1–7). ACII 2009. 3rd International Conference on, IEEE.

Brave, S., & Nass, C. (2002). Emotion in human-computer interaction. In J. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (pp. 81–97). Mahwah, NJ: Erlbaum.

Breazeal, C. (2003a). Emotion and sociable humanoid robots. *International Journal of Human Computer Studies*, *59*(1–2), 119–155.

Breazeal, C. (2003b). Towards sociable robots. *Robotics and Autonomous Systems*, 42(3–4), 167–175.

Breazeal, C. (2004). Function meets style: Insights from emotion theory applied to HRI. *IEEE Transactions on Systems, Man and Cybernetics, Part C* (Applications and Reviews), *34*(2), 187–194.

Calvo, R., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37.

Campbell, J. (2008). *The hero with a thousand faces*, 3rd ed. New World Library.

Campbell, J. C., Hays, M. J., Core, M., Birch, M., Bosack, M., & Clark, R. E. (2011). Interpersonal and leadership skills: Using virtual humans to teach new officers. In *Interservice/Industry Training, Simulation, and Education Conference*, number 11358; 1–11.

Canamero, D. (1997). A hormonal model of emotions for behavior control. VUB. AI-Lab Memo.

Cannon, W. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *American Journal of Psychology*, *39*, 106–124.

Cassell, J. Sullivan, J., Prevost, S., & Churchill, E. (2000). Embodied conversational agents. *Social Psychology*, *40*(1), 26–36.

Cassell, J., & Bickmore, T. (2003). Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. *User Modeling and User-Adapted Interaction* 13: 89–132.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, *99*(1), 45–77.

Colby, K. (1975). *Artificial paranoia*. Pergamon Press.

Cook, D. A., Erwin, P. J., & Triola, M. M. (2010). Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Academic Medicine: Journal of the Association of American Medical Colleges*, *85*(10), 1589–602.

Cook, D. A., & Triola, M. M. (2009). Virtual patients: A critical literature review and proposed next steps. *Medical Education*, *43*(4), 303–311.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Avon.

Darwin, C. (1872). *The expression of emotions in man and animals*. London: John Murray. (Republished by University of Chicago Press, 1965.)

Davidson, R., & Cacioppo, J. (1992). New developments in the scientific study of emotion: An introduction to the special section. *Psychological Science*, *3*(1), 21–22.

De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., & de Carolis, B. (2003). From Greta's mind to her face: Modeling the dynamics of affective states in a conversational embodied agent. *International Journal of Human-Computer Studies*, *59*(1–2), 81–118.

de Sousa, R. (1990). *The rationality of emotions*. Cambridge, MA: MIT Press.

Dreyfus, H. (1992). *What computers still can't do*. Cambridge, MA: MIT Press.

Dreyfus, H., & Dreyfus, S. (Winter 1988). Making a mind versus modeling the brain: Artificial intelligence back at a branchpoint. *Dædalus*, 15–43.

Dunn, C., Deroo, L., & Rivara, F. P. (2001). The use of brief interventions adapted from motivational interviewing across behavioral domains: A systematic review. *Addiction 96*(12), 1725–1742.

Ekman, P. (1984). Expression and the nature of emotion. In K. Scherer (Ed.), *Approaches to emotion* (pp. 319–343). Hillsdale, NJ: Erlbaum.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion.* Hoboken, NJ: Wiley.

Ekman, P., & Freisen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement.* Consulting Psychologists Press.

Ekman, P., Freisen, W. V., & Hager, J. C. (2002). *Facial action coding system*, 2nd ed. (Vol. 160). Salt Lake City: Research Nexus eBook.

Ekman, P., Levenson, R. W., & Freisen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208–1210.

Elliott, C. (1992). Affective reasoner. Ph.D. thesis. The Institute for the Learning Sciences Technical Report #32. Evanston, IL: Northwestern University.

El-Nasr, M., Yen, J., & Ioerger, T. (2000). FLAME: fuzzy logic adaptive model of emotions. autonomous agents and multi-agent systems, *Autonomous Agents and Multi-Agent Systems*, 3(3), 219–257.

Elster, J. (1999). *Alchemies of the mind: Rationality and the emotions*. Cambridge, UK: Cambridge University Press.

Frank, R. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.

Friesen, W., & Ekman, P. (1983). *EMFACS-7: Emotional facial action coding system*.

Frijda, N. and Swagerman, J. (1987). Can computers feel? Theory and design of an emotional system. *Cognition and Emotion*, *1*(3), 235–257.

Frijda, N. H. (1986). *The Emotions: Studies in Emotion and Social Interaction* (Vol. 1). New York: Cambridge University Press.

Frijda, N. H. (1995). *Emotions in robots.* Cambridge, MA: MIT Press.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, *1*(2), 115–143.

Fum, D., & Stocco, A. (2004). Memory, emotion, and rationality: An ACT-R interpretation for gambling task results. In *Proceedings of the Sixth International Conference on Cognitive Modeling.*

Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, *5*(4), 269–306.

Gratch, J., Okhmatovskaia, A., & Lamothe, F. (2006). Virtual rapport. *Intelligent Virtual*.

Gratch, J., Wang, N., Gerten, J., & Fast, E. (2007a). Creating rapport with virtual agents. In *Proceedings of the international conference on intelligent virtual agents*.

Gratch, J., Wang, N., & Okhmatovskaia, A. (2007b). Can virtual humans be more engaging than real ones? In *Proceedings of the 12th international conference on human-computer interaction: Intelligent multimodal interaction environments*, HCI'07. Berlin and Heidelberg: Springer-Verlag.

Gunes, H. and Pantic, M. (2010). Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions*, *1*(1), 68–99.

Hermann, C., Melcher, H., Rank, S., & Trappl, R. (2007). Neuroticism a competitive advantage (also) for IVAs? In *International conference on intelligent virtual agents, intelligence, lecture notes in artificial intelligence* (pp. 64–71).

Hill, R. W., Gratch, J., Marsella, S., Rickel, J., Swartout, W., & Traum, D. (2003). Virtual humans in the mission rehearsal exercise system. *Kunstliche Intelligenz (KI Journal), Special issue on Embodied Conversational Agents*, *17*(4), 5–10.

Hiolle, A., Cañamero, L., Davila-Ross, M., and Bard, K. a. (2012). Eliciting caregiving behavior in dyadic human-robot attachment-like interactions. *ACM Transactions on Interactive Intelligent Systems*, *2*(1), 1–24.

Hoffman, R. E., Grasemann, U., Gueorguieva, R., Quinlan, D., Lane, D., & Miikkulainen, R. (2011). Using computational patients to evaluate illness mechanisms in schizophrenia. *Biological Psychiatry*, *69*(10), 997–1005.

Huang, L., Morency, L.-P., & Gratch, J. (2011). Virtual rapport 2.0. In *International conference on intelligent virtual agents, intelligence: Lecture notes in artificial intelligence* (pp. 68–79). Berlin and Heidelberg: Springer-Verlag.

Hubal, R. C., Frank, G. A., & Guinn, C. I. (2003). Lessons learned in modeling schizophrenic and depressed responsive virtual humans for training. In *Proceedings of the 2003 international conference on intelligent user interfaces* (IUI'03), (pp. 85–92). New York: ACM.

Hudlicka, E., & McNeese, M. (2002). User's affective & belief state: Assessment and GUI adaptation. *International Journal of User Modeling and User Adapted Interaction*, *12*(1), 1–47.

Hudlicka, E. (1998). Modeling emotion in symbolic cognitive architectures. *AAAI fall symposium on emotional and intelligent: the tangled knot of cognition. TR FS-98-03*, 92-97. Menlo Park, CA: AAAI Press.

Hudlicka, E. 2002. This time with feeling: Integrated model of trait and state effects on cognition and behavior. *Applied Artificial Intelligence*, *16*(7–8), 1–31.

Hudlicka, E. (2003). To feel or not to feel: The role of affect in human-computer interaction. *International Journal of Human-Computer Studies*, *59*(1–2).

Hudlicka, E., (2005). Computational models of emotion and personality: Applications to psychotherapy research and practice. In *Proceedings of the 10th annual cybertherapy 2005 conference: A decade of virtual reality,* Basel.

Hudlicka, E. (2007). Reasons for emotions: Modeling emotions in integrated cognitive systems. In W. E. Gray (Ed.), *Integrated models of cognitive systems* (pp. 1–37). New York: Oxford University Press.

Hudlicka, E., (2008). Modeling the mechanisms of emotion effects on cognition. In *Proceedings of the AAAI fall symposium on biologically inspired cognitive architectures.* TR FS-08-04 (pp. 82–86). Menlo Park, CA: AAAI Press.

Hudlicka, E. (2011). Guidelines for Developing Computational Models of Emotions. *International Journal of Synthetic Emotions*, 2(1), 26–79.

Hudlicka, E., Lisetti, C., Hodge, D., Paiva, A., Rizzo, A., & Wagner, E. (2008). Artificial agents for psychotherapy. In *Proceedings of the AAAI spring symposium on emotion, personality and social behavior,* TR SS-08-04, 60–64. Menlo Park, CA: AAAI.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century Crofts.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, *99*(3), 561–565.

James, W. (1884). What is an emotion? *Mind*, *9*(34), 188–205.

Gray, J., & McNaughton, N. (2003). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*, 2nd ed. Oxford, UK: Oxford University Press.

Jiang, H. (2008). *From rational to emotional agents: A way to design emotional agents*. VDM Verlag Dr. Muller.

Johnson-Laird, P. & Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition and Emotion*, *6*(3/4), 201–223.

Johnson, O., & Thomas, F. (1981). *The illusion of life: Disney animation*. Hyperion Press.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proceedings fourth IEEE international conference on automatic face and gesture recognition* (Vol. 4, pp. 46–53). Grenoble, France: IEEE Computer Society.

Kang, S.-H., Gratch, J., Wang, N., & Watt, J. (2008). Does the contingency of agents' nonverbal feedback affect users' social anxiety? In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems* (Vol. 1 pp. 120–127). International Foundation for Autonomous Agents and Multiagent Systems.

Kluwer, T. (2011). I like your shirt—Dialogue acts for enabling social talk in conversational agents. In Proceedings of Intelligent Virtual Agents 11th International Conference (IVA 2011) Reykjavik, Iceland. *Lecture Notes in Computer Science,* 6895: 14–27.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, *33*(1), 1–64.

Laird, J. E., & Rosenbloom, P. (1996). Evolution of the SOAR architecture. In D. M. Steier. &, T. M. Mitchell (Eds.), *Mind matters: A tribute to Allen Newell* (pp. 1–50). Mahwah, NJ: Erlbaum.

Lang, P. J., Bradley, M. M., & Cuthbert, B. (1997). *International affective picture system (IAPS): Technical manual and affective ratings.* Technical report. Washington, DC: NIMH Center for the Study of Emotion and Attention.

Lazarus, R. S. (1991). Cognition and motivation in emotion. *Journal of the American Psychologist*, *46*, 352–367.

Ledoux, J. (1992). Emotion and the amygdala. In *Current Opinion in Neurobiology*, (Vol. 2, pp. 339–351). Hoboken, NJ: Wiley-Liss.

Ledoux, J. (1995). Emotion: Clues from the brain. *Annual Review Psychology*, *46*, 209–235.

Ledoux, J. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, 23, 155–184.

Leventhal, H., & Scherer, K. R. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition and Emotion*, *1*, 3–28.

Lewis, R. L., Huffman, S. B., John, B. E., Laird, J. E., Lehman, J. F., Newell, A.,…Tessler, S. G. (1990). Soar as a unified theory of cognition: Spring 90. In *Twelfth annual conference of the cognitive science society* (pp. 1035–1042).

Lisetti, C. L. (2008). Embodied conversational agents for psychotherapy. In *Proceedings of the CHI 2008 conference workshop on technology in mental health* (pp. 1–12). New York: ACM.

Lisetti, C. L. & Gmytrasiewicz, P. (2002). *Can a rational agent afford to be affectless? A formal approach. in applied artificial intelligence* (Vol. 16, pp. 577–609). Taylor & Francis.

Lisetti, C. L. & Nasoz, F. (2004). Using noninvasive wearable computers to recognize human emotions from physiological signals. EURASIP. *Journal on Applied Signal Processing*, *11*, 1672–1687.

Lisetti, C. L., Amini, R., Yasavur, U., & Rishe, N. (2013). I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Transactions on Management Information Systems*, Vol. 4, No. 4, Article 19, 2013. Lisetti, C. L, Nasoz, F., Alvarez, K., & Marpaung, A. (2004). A social informatics approach to human-robot interaction with an office service robot. *IEEE Transactions on Systems, Man, and Cybernetics—Special Issue on Human Robot Interaction, 34*(2).

Loyall, B., & Bates, J. (1997). Personality-rich believable agents that use language. In *Proceedings of the first international conference on autonomous agents*.

Magerko, B., Dean, J., Idnani, A., Pantalon, M., & D'Onofrio, G. (2011). Dr. Vicky: A virtual coach for learning brief negotiated interview techniques for treating emergency room patients. In AAAI Spring Symposium.

Marinier, R., & Laird, J. (2004). Toward a comprehensive computational model of emotions and feelings. In *Proceedings of the 6th international conference on cognitive modeling*.

Marinier, R. P., & Laird, J. E. (2007). Computational modeling of mood and feeling from emotion. In *Cognitive science* (pp. 461–466).

Marsella, S. C., & Gratch, J. (2009). EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, *10*(2000), 70–90.

Marsella, S. C., Johnson, W. L., & LaBore, C. (2000). Interactive pedagogical drama. In *Proceedings of the fourth international conference on autonomous agents* (pp. 301–308).

Mateas, M. (2001). Expressive AI: A hybrid art and science practice. *Leonardo: Journal of the International Society for Arts, Sciences and Technology*, *34*(2), 147–153.

Mateas, M. (2003). Façade, an experiment in building a fully-realized interactive drama. In *Game Developer's Conference: Game Design Track*, San Jose, California, March.

McClelland, J., Rumelhart, D. & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructures of cognition: Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

McQuiggan, S., & Lester, J. (2007). Modeling and evaluating empathy in embodied companion agents. *International Journal of Human-Computer Studies*, *65*, 348–360.

McQuiggan, S., Robison, J., & Phillips, R. (2008). Modeling parallel and reactive empathy in virtual agents: An inductive approach. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, 1: 167–174.

Mehrabian, A. (1995). Framework for a comprehensive description and measurement of emotional states *Genetic, Social, and General Psychology Monographs*. *121*, 339–361.

Muramatsu, R., & Hanoch, Y. (2005). Emotions as a mechanism for boundedly rational agents: The fast and frugal way. *Journal of Economic Psychology*, *26*(2), 201–221.

Murphy, R., Lisetti, C. L., Irish, L., Tardif, R., & Gage, A. (2001). Emotion-based control of cooperating heterogeneous mobile robots. *IEEE Transactions on Robotics and Automation: Special Issue on Multi-Robots Systems*. *18*(5), 744–757.

Murphy, R. R. (2000). *Introduction to AI robotics*. Cambridge, MA: MIT Press.

Nguyen, H., & Masthoff, J. (2009). Designing empathic computers: The effect of multimodal empathic feedback using animated agent. In *Proceedings of the 4th international conference on persuasive technology* (p. 7). New York: ACM.

Ochs, M., & Niewiadomski, R. (2005). Intelligent expressions of emotions. In *Proceedings of the first affective computing and intelligent and intelligent interactions* (pp. 707–714).

Ochs, M., & Sadek, D., and Pelachaud, C. (2012). A formal model of emotions for an empathic rational dialog agent. *Autonomous Agents and Multi-Agent Systems*, *24*(3), 410–440.

O'Donnell, B. F., & Wilt, M. A. (2006). *Computational models of mental disorders. Encyclopedia of Cognitive Science*. Wiley Online Library.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge, UK: Cambridge University Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*(3), 315–331.

Paleari, M., Grizard, A., & Lisetti, C. L. (2007). Adapting psychologically grounded facial emotional expressions to different anthropomorphic embodiment platforms. In *Proceedings of the FLAIRS conference*.

Paleari, M., & Lisetti, C. L. (2006). *Toward multimodal fusion of affective cues*.

Pelachaud, C. (2009). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1535), 3539–3548.

Pelachaud, C., & Bilvi, M. (2003). Communication in multiagent systems: Background, current trends and future. In *Lecture notes in computer science* (pp. 300–317) Berlin: Springer.

Pelachaud, C., Carofiglio, V., & Poggi, I. (2002). Embodied contextual agent in information delivering application. In *Proceedings of the first international joint conference on autonomous agents & multi-agent systems*.

Pelachaud, C., Maya, V., & Lamolle, M. (2004). Representation of expressivity for embodied conversational agents. Embodied conversational agents: Balanced perception and action. In *Proceedings of the AAMAS*, *4*.

Peter, C., & Herbon, A. (2006). Emotion representation and physiology assignments in digital systems. *Interacting with Computers*, *18*(2), 139–170.

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

Plutchik, R. (1980). *Emotion theory, research and experience: Vol. 1. Theories of emotion.* New York: Academic Press.

Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., de Carolis, B., Poggi, I., . . . Carolis, B. D. (2005). *Multimodal intelligent information presentation* (Vol. 27)., New York: Springer.

Posner, M. (1993). *The foundations of cognitive science*. Cambridge, MA: MIT Press.

Predinger, H., & Ishizuka, M. (2004a). *Life-like characters*. New York: Springer.

Predinger, H., & Ishizuka, M. (2004b). What affective computing and life-like character technology can do for tele-home health care. In *Workshop on HCI and homecare: connecting families and clinicians* (pp. 1–3) in conjunction with CHI. Citeseer.

Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses user's affective states. *Applied Artificial Intelligence*, *19*, 267–285.

Pütten, A. M. V. D., Krämer, N. C., & Gratch, J. (2009). Who's there? Can a virtual agent really elicit social presence?

Reeves, B., Nass, C., & Reeves, B. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Chicago: University of Chicago Press.

Reilly, W S. N. (1997). A methodology for building believable social agents. In W. L. Johnson, and B. Hayes-Roth (Eds.), *Proceedings of the first international conference on autonomous agents agents* (pp. 114–121). New York: ACM.

Ritter, F. E., & Avraamides, M. N. (2000). *Steps towards including behavior moderators in human performance models*: College Station: Penn State University

Rizzo, A., Forbell, E., Lange, B., Galen Buckwalter, J., Williams, J., Sagae, K., & Traum, D. (2012). Simcoach: an online intelligent virtual human agent system for breaking down barriers to care for service members and veterans. In Monsour Scurfield, R., & Platoni, K. (Eds). *Healing War Trauma A Handbook of Creative Approaches.* Taylor & Francis.

Rogers, C. (1959). A theory of therapy, personality and interpersonal relation—Developed in the client-centered framework. In S. Koch (Ed.), *Psychology: The study of a science*, (Vol. 3, pp. 184–256). New York: McGraw-Hill.

Rolls, E. (2007). *Emotion explained*. Oxford Univeristy Press, Oxford, England.

Rossen, B., & Lok, B. (2012). A crowdsourcing method to develop virtual human conversational agents. *International Journal of Human-Computer Studies*, *70*(4), 301–319.

Rumelhart, D., McClelland, J., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructures of cognition: Vol. 1. Explorations in the microstructure of cognition.*

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russell, J. A. (2003). *Core affect and the psychological construction of emotion*. *Psychological Review*, *110*(1), 145–172.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819.

Russell, S., & Norwig, P. (2011). *No artificial intelligence: A modern approach*. Upper Saddle River, NJ: Prentice Hall.

Russell, J., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*, 273–294.

Rutter, D., Pennington, D., Dewey, M., & Swain, J. (1984). Eye-contact as a chance product of individual looking: Implications for the intimacy model of argyle and dean. *Journal of Nonverbal Behavior*, *8*(4), 250–258.

Scherer, K. R. (1992). What Does Facial Expression Express? In K. Strongman (Ed.), *International review of studies on emotion* (Vol. 2). Hoboken, NY: Wiley.

Scherer, K. R. (2001). *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.

Scherer, K. R. (2009). Emotions are emergent processes: They require a dynamic computational architecture. *Philosophical transactions of the Royal Society of London, Series B, Biological Sciences*, *364*(1535), 3459–3474.

Scheutz, M. (2000). Surviving in a hostile multiagent environment: How simple affective states can aid in the competition for resources. Paper presented at the Advances in Artificial Intelligence—13th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, Montreal, Quebec, Canada.

Scheutz, M. (2011). Architectural roles of affect and how to evaluate them in artificial agents. *International Journal of Synthetic Emotions*, *2*(2), 48–65.

Scheutz, M., & Schermerhorn, P. (2009). Affective goal and task selection for social robots. In J. Vallverdu &

D. Casacuberta (Eds.), *The handbook of research on synthetic emotions and sociable robotics* (pp. 74–87). Hershey, PA: IGI Publishing.

Schulman, D., Bickmore, T., & Sidner, C. (2011). In 2011 AAAI Spring Symposium Series, *An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing* (pp. 61–64).

Servan-Schreiber, D. (1986). Artificial intelligence and psychiatry. *Journal of Nervous and Mental Disease*, *174*, 191–202.

Simon, H. (1967). Motivational and emotional controls of cognition. *Psychological Review*, *1*, 29–39.

Sloman, A. (1987). Motives, mechanisms, and emotions. *Emotion and Cognition*, *1*(2), 217–234.

Sloman, A., & Croucher, M. (1981). Why robots will have emotions. In *Proceedings of the seventh IJCAI* (pp. 197–202). San Mateo, CA: Morgan-Kaufmann.

Smith, C. A., & Kirby, L. D. (2001). Toward delivering on the promise of appraisal theory. In K. Scherer, A. Schorr, and T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 121–140). New York: Oxford University Press.

Smith, C., & Lazarus, R. (1990). Emotion and adaptation. In *Handbook of personality: Theory and research* (pp. 609–637). New York: Guilford.

Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raij, A., Harrison, C.,…Lind, D. S. (2006). The use of virtual patients to teach medical students history taking and communication skills. *American Journal of Surgery*, *191*(6), 806–811.

Stiehl, W., & Lieberman, J. (2005). Design of a therapeutic robotic companion for relational, affective touch. In *ROMAN 2005 IEEE international workshop on robot and human interactive communication* (Vol. 1). AAAI Press.

Swartout, W. (2010). Lessons learned from virtual humans. *AI Magazine*, *31*(1), 9–20.

Swartout, W. Jr., Gratch, R. H., Johnson, R. H., Kyriakakis, W. L., Labore, C., Lindheim, C. M.,…Thiebaux, L. (2001). Towards the holodeck: Integrating graphics, sound, character and story. In J. P. Miller, E. Andre, S. Sen, & C. Frasson (Eds), *Proceedings of the fifth international conference on autonomous agents* (pp. 409–416). ACM.

Thagard, P. (2008). Cognitive science. In *The Stanford encyclopedia of philosophy.* Cambridge, MA: MIT Press.

Tian, Y.-I., Kanade, T., & Cohn, J. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence Analysis*, *23*(2), 97–115.

Velásquez, J. (1997). Modeling emotions and other motivations in synthetic agents. Proceedings of the National Conference on Artificial Intellligence.

Villaume, W. A., Berger, B. A., & Barker, B. N. (2006). Learning motivational interviewing: Scripting a virtual patient. *American Journal of Pharmaceutical Education*, *70*(2), 33.

Wang, N. (2009). *Rapport and facial expression*, 3rd ed. (pp. 1–6). and Workshops. ACII 2009.

Wang, N., & Gratch, J. (2010). Don't just stare at me! In *28th ACM conference on human factors in computing systems* (pp. 1241–1249). Atlanta: Association for Computing Machinery.

Wehrle, T., & Scherer, K. (2001). Toward Computational Modeling of Appraisal Theories. In Scherer, K., Schorr, A., & Johnstone, T. (Eds.). Appraisal processes in emotion: Theory, Methods, Research (pp. 350–365). New-York: Oxford University Press.

Weizenbaum, J. (1967). Contextual understanding by computers. *Communications of the ACM*, *10*(8), 474–480.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. Freeman.

Wilkinson, N., Ang, R. P., and Goh, D. H. (2008). Online video game therapy for mental health concerns: A review. *International Journal of Social Psychiatry*, *54*(4), 370–382.

Wolf, L., Hassner, T., & Taigman, Y. (2011). Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(10), 1978–1990.

Zajonc, B., & Markus, H. (1984). Affect and cognition: The hard interface. In C. Izard, J. Kagan, & R. Zajonc (Eds.), *Emotion, cognition and behavior* (pp. 73–102). Cambridge, UK: Cambridge University Press.

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151–175.

Zajonc, R. (1984). On the primacy of affect. *American Psychologist*, *39*, 117–124.