DRAFT MANUSCRIPT – PLEASE DO NOT CITE VERBATIM

Computational Models of Emotion

Stacy Marsella, Jonathan Gratch, Paolo Petta

Chapter to be published in Scherer, K.R., Bänziger, T., & Roesch, E. (Eds.) A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing. Oxford: Oxford University Press, in press

Introduction

Recent years have seen a significant expansion in research on computational models of human emotional processes, driven both by their potential for basic research on emotion and cognition as well as their promise for an ever increasing range of applications. This has led to a truly interdisciplinary, mutually beneficial partnership between emotion research in psychology and computational science, of which this volume is an exemplar. To understand this partnership and its potential for transforming existing practices in emotion research across disciplines and for disclosing important novel areas of research, we explore in this chapter the history of work in computational models of emotion including the various uses to which they have been put, the theoretical traditions that have shaped their development, and how these uses and traditions are reflected in their underlying architectures.

For an outsider to the field, the last fifteen years have seen the development of a seemingly bewildering array of competing and complementary computational models. Figure 1 lists a “family tree” of a few of the significant models and the theoretical traditions from which they stem. Although there has been a proliferation of work, the field is far from mature: the goals that a model is designed to achieve are not always clearly articulated; research is rarely incremental, more often returning to motivating theories than extending prior computational approaches; and rarely are models contrasted with each other in terms of their ability to achieve their set goals. Contributing to potential confusion is the reality that computational models are complex systems embodying a number of, sometimes unarticulated, design decisions and assumptions inherited from the psychological and computational traditions from which they emerged, a circumstance made worse by the lack of a commonly accepted lexicon for even designating these distinctions.

In this chapter, we lay out the work on computational models of emotion in an attempt to reveal the common uses to which they may be put and the underlying techniques and assumptions from which the models are built. Our aim is to present conceptual distinctions and common terminology that can aid in discussion and comparison of competing models. Our hope is this will not only facilitate an understanding of the field for outside researchers but work towards a lexicon that can help foster the maturity of the field towards more incremental research.

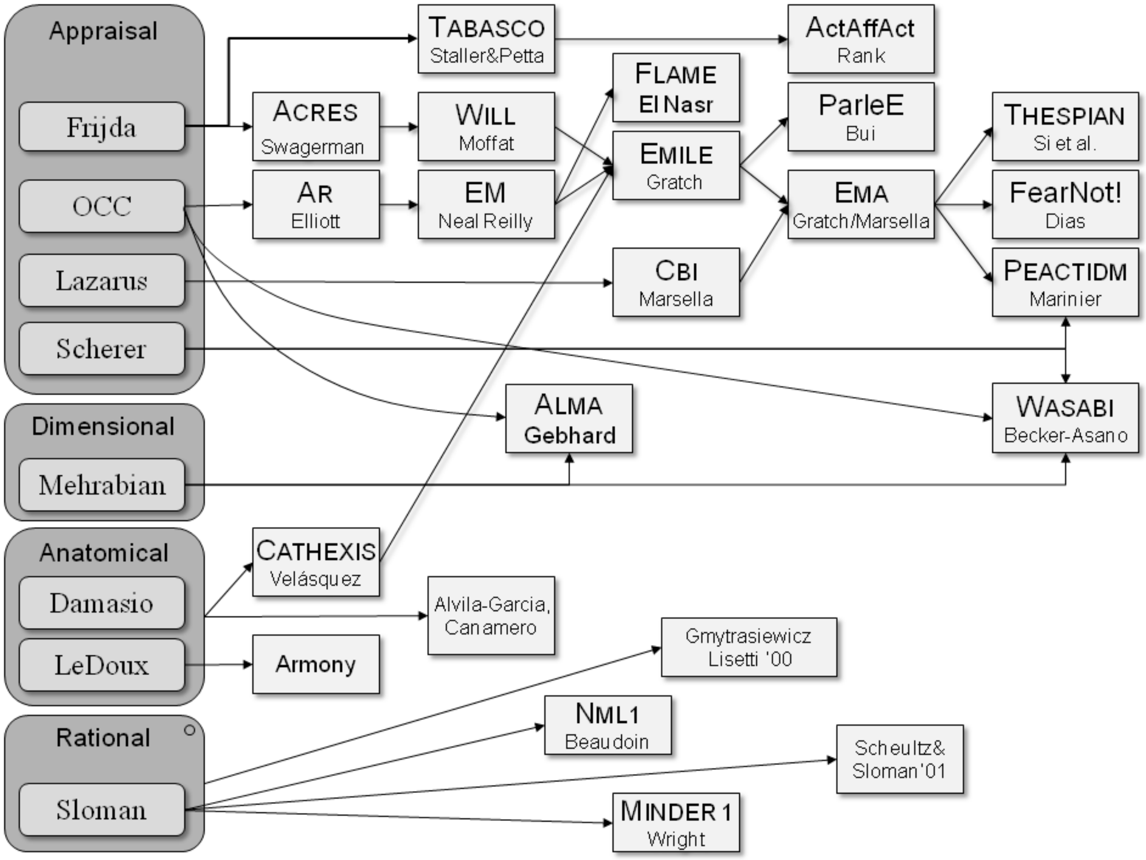

Figure 1: A history of computational models of emotion

In characterizing different computational models of emotion, we begin by describing interdisciplinary uses to which computational models may be put, including their uses in improving human-computer interaction, in enhancing general models of intelligence, and as methodological tools for furthering our understanding of human behavior. We next discuss how models have been built, including the underlying theoretical traditions that have shaped their development. These differing theoretical perspectives often conceptualize emotion in quite different ways, emphasizing different scenarios and proposed functions, different component processes and different linkages between these components. It should then come as no surprise that such differences are also reflected in the underlying design of the computational models. We next narrow our focus to cognitive appraisal theory, undeniably one of the most influential theoretical perspectives within computational research. To help organize and dissect research on computational appraisal models, we introduce a generic appraisal architecture, a *component model view of appraisal models*, that conceptualizes emotion as a set of component models and relations between these components. We discuss how different computational systems address some, but typically not all, of these component models and describe differing processing choices that system developers have used in realizing their component model variants. Finally, we illustrate how this component model view can help guide work in evaluating and contrasting alternative computational models of emotion.

The uses of computational models: an interdisciplinary partnership

New tools often transform science, opening up new approaches to research, allowing previously unaddressed questions to be explored, as well as revealing new questions. To appreciate the transformative role that computational models of emotion can have on research, we consider three aspects in this section: the impact on emotion research in psychology, the impact on artificial intelligence (AI) and finally the impact on work in human-computer interaction.

Impact on psychological research on emotion

Work in computational models of emotion impacts research in human emotion by transforming how theories are formulated and evaluated. One way this occurs is through a process of concretizing concepts in the theory. Psychological theories of emotion have typically been cast at an abstract level and through informal (natural language) descriptions. Concepts in the theory are usually not defined formally, and how processes work may not be laid out in systematic detail. The formulation of a computational model enforces more detail. The structures and processes of the theory must be explicitly and formally defined in order to implement them in a computational model, thus making a computer model a particularly concrete realization of the theory. The process of realizing the model can reveal implicit assumptions and hidden complexities, thereby forcing them to be addressed explicitly in some documented fashion. For example, appraisal theories often argue that a key variable in appraisal is an attribution of blameworthiness for an event deemed motivationally incongruent (e.g., Lazarus 1991). But the process by which a person makes such an attribution and therefore whether a particular situation would be deemed blameworthy, and the related required resources and capacities are typically not carefully laid out. And yet this attribution process may in itself be quite involved (e.g., Shaver 1985; Weiner 1995).

As computational modeling exposes hidden assumptions in the theory, addressing those assumptions can extend the scope of the theorizing. Seen in this way, computational models become not only a way to concretize theories, but also a framework for theory construction. In so doing, computational modeling also extends the language of emotion theorizing by incorporating concepts, processes, and metaphors drawn from computation, much as concepts such as *information processing* and *symbol systems* impacted psychology in general. For example, several computational models have recast the appraisal theory in terms of concepts drawn from AI, including knowledge representation (e.g., Gratch and Marsella 2004), planning (e.g., Gratch 2000; Dias and Paiva 2005), neural networks (Sander, Grandjean et al. 2005) and decision-making (Lisetti and Gmytrasiewicz 2002; Ito, Pynadath et al. 2008). Incorporation of the models into larger simulations can also expose hidden questions behind traditional conceptualizations and extend the scope of theorizing. For example, several computational models of emotion have been incorporated into larger simulation systems that seek to model emotion's role in human mental processes and behavior (Marsella and Gratch 2001; Dias and Paiva 2005; Becker-Asano 2008; Rank 2009). This has led researchers to address fundamental architectural questions about the relation of appraisal processes to other cognitive processes, perception, and behavior (Marsella and Gratch 2009; Rank 2009). Of course, a central challenge here is to ensure that increases in the scope of the theorizing do not endanger the parsimony often critical to a model's explanatory power.

Coupled to this transformation of the theory formation process through modeling and simulation runs of the model, the computational realization of a theory can also increase the capacity to draw predictions from theory. In particular, computational models provide a new empirical framework for studying emotion processes that goes beyond what is feasible in more traditional laboratory settings. Computer simulations of the model *behave*: they provide a means to explore systematically the temporal dynamics of emotion processes and form predictions about the time course of those processes. Manipulations of experimental conditions may be explored more extensively first with a computational model, such as ablating certain functionalities or testing responses under adverse conditions that may be costly, risky or raise ethical concerns in vivo (e.g., Armony, Servan-Schreiber et al. 1997). Simulations can reveal unexpected model properties that suggest further exploration. Additionally, models of emotion and affective expression have been incorporated into virtual humans, software artifacts that look and act like humans, capable of perception and action in a virtual world that they can co-habit with people. These systems essentially allow for the study of emotion in a virtual ecology, a form of synthetic in vivo experimentation.

Finally, the computational modeling of emotion and emotional expression has led to new ways to create stimuli for human subject experimentation. Virtual humans are in some ways the experimenter's ultimate confederate. A virtual human can be manipulated systematically to elicit behavior from human subjects. For example, virtual humans have been used to show that subtle changes in physical appearance or behavior can profoundly impact social interaction, including changes to people's willingness to cooperate (Krumhuber, Manstead et al. 2007), the fluidity of their conversation (Gratch, Wang et al. 2007), learning outcomes (Baylor and Kim 2008) and even their level of social aggression (McCall, Blascovich et al. 2009).

Impact on Artificial Intelligence & Robotics

Modern research in the psychology, cognitive science and neuroscience of emotion has led to a revolution in our thinking about the relation of emotion to cognition and social behavior, and as a consequence is also transforming the science of computation. Findings on the functional, often adaptive role that emotions play in human behavior have motivated AI and robotics research to explore whether computer analogues of human emotion can lead to more intelligent, flexible and capable systems. Early work by Simon (Simon 1967) argued that emotions serve the crucial function of interrupting normal cognition when unattended goals require servicing. Viewing emotion as serving this critical interrupt capacity provides a means for an organism to balance competing goals as well as incorporate reactive behaviors into more deliberative processing. A range of studies point to emotions as the means by which the individual establishes values for alternative decisions and decision outcomes. Busemeyer et al. (Busemeyer, Dimperio et al. 2007) argue that emotional state influences the subjective utility of alternative choice. Studies performed by Damásio and colleagues suggest that damage to ventromedial prefrontal cortex prevents emotional signals from guiding decision making in an advantageous direction (Bechara, Damasio et al. 1999).

Other authors have emphasized how social emotions such as anger and guilt may reflect a mechanism that improves group utility by minimizing social conflicts, and thereby explains peoples "irrational" choices to cooperate in social games such as prison's dilemma (Frank 1988). Similarly, "emotional biases" such as wishful thinking may reflect a rational mechanism that is more accurately accounting for certain social costs, such as the cost of betrayal when a parent defends a child despite strong evidence of their guilt in a crime (Mele 2001).

Collectively, these findings suggest that emotional influences have important social and cognitive functions that would be required by *any* intelligent system. This view is not new to Artificial Intelligence (Simon 1967; Sloman and Croucher 1981; Minsky 1986) but was in large measure ignored in AI research of the late 20th century which largely treated emotion as antithetical to rationality and intelligence. However, in the spirit of Hume's famous dictum: "reason is, and ought only to be the slave of the passions…" (Hume 2000, 2.3.3.4), the question of emotion has again come to the fore in AI as models have begun to catch up to theoretical findings. This has been spurred, in part, by an explosion of interest in integrated computational models that incorporate a variety of cognitive functions (Bates, Loyall et al. 1991; Anderson 1993; Rickel, Marsella et al. 2002). Indeed, until the rise of broad integrative models of cognition, the problems emotion was purported to solve, for example, juggling multiple goals, were largely hypothetical. More recent cognitive systems embody a variety of mental functions and face very real challenges how to allocate resources. A re-occurring theme in emotion research in AI is the role of emotion in addressing such control choices by directing cognitive resources towards problems of adaptive significance for the organism (Scheutz and Sloman 2001; Staller and Petta 2001; Blanchard and Cañamero 2006; Scheutz and Schermerhorn 2009).

Impact on Human Computer Interaction

Finally, research has revealed the powerful role that emotion and emotion expression play in shaping human social interaction, and this in turn has suggested that computer interaction can exploit (and indeed must address) this function. Emotional displays convey considerable information about the mental state of an individual. Although there is a lively debate whether these displays reflect true emotion or are simply communicative conventions (Manstead, Fischer et al. 1999), pragmatically there is truth in both perspectives. From emotional displays, observers can form interpretations of a person's beliefs (e.g., frowning at an assertion may indicate 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 and Scott 1997). With such a powerful signal, it is not surprising that emotions can be a means of social control (Fridlund 1997; Campos, Thein et al. 2003; de Waal 2003). Emotional displays seem to function to elicit particular social responses from other individuals ("social imperatives", Frijda 1987) and arguably, such responses can be difficult to suppress. 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. Specific examples of this are *emotional contagion,* that can lead individuals to "catch" the emotions of those around them (Hatfield, Cacioppo et al. 1994) and 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). Given this wide array of functions in social interactions, many have argued that emotions evolved because they provide an adaptive advantage to social organisms (Darwin 2002; de Waal 2003).

To the extent that these functions can be realized in artificial systems, they could play a powerful role in facilitating interactions between computer systems and human users. This has inspired several trends in human-computer interaction. 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 a user's emotion including facial expressions (Lisetti and Schiano 2000; Fasel, Stewart-Bartlett et al. 2002; Haag, Goronzy et al. 2004), physiological indicators (Picard 1997; Haag, Goronzy et al. 2004) and vocal expression (Lee and Narayanan 2003).

A related trend in HCI work is the use of emotions and emotional displays in virtual characters that interact with the user. As animated films (Thomas and Johnston 1995) so poignantly demonstrate, emotional displays in an artificially generated character can have the general effect of making it seem human or lifelike, and thereby cue the user to respond to, and interact with, the character as if it were another person. A growing body of research substantiates this view. In the presence of a lifelike agent, people are more polite, tend to make socially desirable choices and are more nervous (Kramer, Tietz et al. 2003); they can exhibit greater trust of the agent's recommendations (Cowell and Stanney 2003); and they can feel more empathy (Paiva, Aylett et al. 2004). 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 a user towards more effective interactions. For example, education researchers have argued that nonverbal displays can have a significant impact on student intrinsic motivation (Lepper 1988).

A number of applications have attempted to exploit this interpersonal function of emotional expression. 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) et al. 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 applications have also tried to manipulate a student's motivations through emotional behaviors of a virtual character: Lester utilized praising and sympathetic emotional displays to provide feedback and increase student motivation in a tutoring application (Lester, Towns et al. 2000). Researchers have also looked at emotion and emotional expression in characters as a means to engender empathy and bonding between between learners and virtual characters (Marsella, Gratch et al. 2003; Paiva, Dias et al. 2005); Biswas (Biswas, Leelawong et al. 2005) also use human-like traits to promote empathy and intrinsic motivation in a learning-by-teaching system.

In summary, computational models of emotion serve differing roles in research and applications. Further, the evaluation of these models is in large measure dependent on those roles. In the case of the psychological research that uses computational models, the emphasis will largely be on fidelity with respect to human emotion processes. In the case of work in AI and Robotics, evaluation often emphasizes how the modeling of emotion impacts reasoning processes or leads in some way to improved performances such as an agent or robot that achieves a better fit with its environment. In HCI work, the key evaluation is whether the model improves human-computer interaction such as making it more effective, efficient or pleasant.

Overall, the various roles for computational models of emotion have led to a number of impressive models being proposed and developed. To put this body of work into perspective, it is critical for the field to support a deeper understanding of the relationship between these models. To assist in that endeavor, we now turn to presenting some common terms and distinctions that can aid in the comparison of competing models.

A Component Perspective on the Design of Computational Models

Each of the computational models listed in Figure 1 is a very different entity, with incompatible inputs and outputs, different behaviors, embodying irreconcilable processing assumptions and directed towards quite different scientific objectives. What we argue here, however, is that much of this variability is illusory. These models are complex systems that integrate a number of component "sub-models." Sometimes these components are not clearly delineated, but if one disassembles models along the proper joints, then a great many apparent differences collapse into a small number of design choices. To facilitate this decomposition, this section describes the component processes underlying emotion, with a particular emphasis on components posited in connection with appraisal theory. These components are not new—indeed they are central theoretical constructs in many theories of emotion—but some of the terminology is new as we strive to simplify terms and de-conflict them with other terminology more commonly used in computer science. We begin by describing the various theoretical traditions that have influenced computational research and the components these theories propose.

Theoretical traditions

A challenge in developing a coherent framework for describing computational models of emotion is that the term "emotion" itself is fraught with ambiguities and contrasting definitions. Emotions are a central aspect of everyday life and people have strong intuitions about them. As a consequence, the terms used in emotion research (appraisal, emotion, mood, affect, feeling) have commonsense interpretations that can differ considerably from their technical definition within the context of a particular emotion theory or computational model (Russell 2003). This ambiguity is confounded by the fact that there are fundamental disputes within psychological and neuroscience research on emotion over the meaning and centrality of these core concepts. Theories differ in which components are intrinsic to an emotion (e.g., cognitions, somatic processes, behavioral tendencies and responses), the relationships between components (e.g. do cognitions precede or follow somatic processes), and representational distinctions (e.g. is anger a linguistic fiction or a natural kind) – see Chapter 1 for an overview of different theoretical perspectives on emotion.

Understanding these alternative theoretical perspectives on emotion is essential for anyone that aspires to develop computational models, but this does not imply that a modeler must be strictly bound by any specific theoretical tradition. Certainly, modelers should strive for a consistent and well-founded semantics for their underlying emotional constructs and picking and integrating fundamentally irreconcilable theoretical perspectives into a single system can be problematic at best. If the goal of the computational model is to faithfully model human emotional processes, or more ambitiously, to contribute to theoretical discourse on emotion, such inconsistencies can be fatal. However, some "fundamentally irreconcilable" differences are illusory and evaporate when seen from a new perspective. For example, disputes on if emotion precedes or follows cognition dissipate if one adopts a dynamic systems perspective (i.e., a circle has no beginning). Nonetheless, theoretical models provide important insights in deriving a coherent computational model of emotion and deviations from specific theoretical constraints, ideally, will be motivated by concrete challenges in realizing a theory within a specific representational system or in applying the resulting model to concrete applications. Here we review some of the theoretical perspectives that have most influenced computational modeling research.

Appraisal theory

Appraisal theory, discussed in detail in Chapter 1, is currently a predominant force among psychological perspectives on emotion and arguably the most fruitful source for those interested in the design of symbolic AI systems, as it emphasizes and explains the connection between emotion and cognition. Indeed, the large majority of computational models of emotion stem from this tradition. In appraisal theory, emotion is argued to arise from patterns of individual judgment concerning the relationship between events and an individual's beliefs, desires and intentions, sometimes referred to as the *person-environment relationship* (Lazarus 1991). These judgments, formalized through reference to devices such as *situational meaning structures* or *appraisal variables* (Frijda 1987), characterize aspects of the personal significance of events. Patterns of appraisal are associated with specific physiological and behavioral reactions. In several versions of appraisal theory, appraisals also trigger cognitive responses, often referred to as *coping strategies*—e.g., planning, procrastination or resignation—feeding back into a continual cycle of appraisal and re-appraisal (Lazarus 1991 p. 127).

In terms of underlying components of emotion, appraisal theory foregrounds appraisal as a central process. Appraisal theorists typically view appraisal as the cause of emotion, or at least of the physiological, behavioral and cognitive changes associated with emotion. Some appraisal theorists emphasize "emotion" as a discrete component within their theories, whereas others treat the term emotion more broadly to refer to some configuration of appraisals, bodily responses and subjective experience (see Ellsworth and Scherer 2003 for a discussion). Much of the work has focused on the structural relationship between appraisal variables and specific emotion labels – i.e., which pattern of appraisal variables would elicit hope (see Ortony, Clore et al. 1988) – or the structural relationship between appraisal variables and specific behavioral and cognitive responses – i.e., which pattern of appraisal variables would elicit certain facial expressions (Smith and Scott 1997; Scherer and Ellgring

2007) or coping tendencies (Lazarus 1991). Indeed, although appraisal theorists allow that the same situation may elicit multiple appraisals, theorists are relatively silent on how these individual appraisals would combine into an overall emotional state or if this state is best represented by discrete motor programs or more dimensional representations. More recent work has begun to examine the processing constraints underlying appraisal – to what extent is it parallel or sequential (Scherer 2001; Moors, De Houwer et al. 2005)? does it occur at multiple levels (Smith and Kirby 2000; Scherer 2001)? – and creating a better understanding of the cognitive, situational and dispositional factors that influence appraisal judgments (Kuppens and Van Mechelen 2007; Smith and Kirby 2009).

Models derived from appraisal theories of emotion, not surprisingly, emphasize appraisal as the central process to be modeled. Computational appraisal models often encode elaborate mechanisms for deriving appraisal variables such as decision-theoretic plans (Gratch and Marsella 2004; Marsella and Gratch 2009), reactive plans (Staller and Petta 2001; Rank and Petta 2005; Neal Reilly 2006), Markov-decision processes (El Nasr, Yen et al. 2000; Si, Marsella et al. 2008), or detailed cognitive models (Marinier, Laird et al. 2009). Emotion itself is often less elaborately modeled. It is sometimes treated simply as a label (sometimes with an intensity) to which behavior can be attached (Elliott 1992). Appraisal is typically modeled as the cause of emotion with specific emotion label being derived via *if-then rules* on a set of appraisal variables. Some approaches make a distinction between a specific emotion instance (allowing multiple instances to be derived from the same event) and a more generalized "affective state" or "mood" (see discussion of *core affect,* below*)* that summarizes the effect of recent emotion elicitations (Neal Reilly 1996; Gratch and Marsella 2004; Gebhard 2005). Some more recent models attempt to model the impact of momentary emotion and mood on the appraisal process (Gratch and Marsella 2004; Gebhard 2005; Paiva, Dias et al. 2005; Marsella and Gratch 2009).

Computational appraisal models have been applied to a variety of uses including contributions to psychology, AI and HCI. For example, Marsella and Gratch have used EMA to generate specific predictions about how human subjects will appraise and cope with emotional situations and argue that empirical tests of these predictions have implications for psychological appraisal theory(Gratch, Marsella et al. 2009; Marsella, Gratch et al. 2009). Several authors have argued that appraisal processes would be required by any intelligent agent that must operate in real-time, ill-structured, multi-agent environments (e.g., Staller and Petta 2001). The bulk of application of these techniques, however, has been for HCI applications, primarily for the creation of real-time interactive characters that exhibit emotions in order to make these characters more compelling (e.g., Neal Reilly 1996), more realistic (e.g., Traum, Rickel et al. 2003; Mao and Gratch 2006), more able to understand human motivational state (e.g., Conati and MacLaren 2004) or more able to induce desirable social effects in human users (e.g., Paiva, Dias et al. 2005).

Dimensional Theories

Dimensional theories of emotion argue that emotion and other affective phenomena should be conceptualized, not as discrete entities but as points in a continuous (typically two or three) dimensional space (Mehrabian and Russell 1974; Watson and Tellegen 1985; Russell 2003; Barrett 2006). Indeed, many dimensional theories argue that discrete emotion categories (e.g., hope, fear and anger) are folkpsychological concepts that have unduly influenced scientific discourse on emotion and have no "reality" in that there are no specific brain regions or circuits that correspond to specific emotion categories (Barrett 2006). Not surprisingly, dimensional theories de-emphasize the term emotion or relegate it to a cognitive label attributed, retrospectively, to some perceived body state. Rather they emphasize concepts such as mood, affect or more recently *core affect* (Russell 2003). We adopt this later term in subsequent discussion. A person is said to be in exactly one affective state at any moment (Russell 2003, p. 154) and the space of possible core affective states is characterized in terms of broad, continuous dimensions. Many computational dimensional models build on the three-dimensional "PAD" model of Mehrabian and Russell (1974) where these dimensions correspond to *pleasure* (a measure of valence), *arousal* (indicating the level of affective activation) and *dominance* (a measure of power or control).

It is worth noting that there is a relationship between the dimensions of core affect and appraisal dimensions – the pleasure dimension roughly maps onto appraisal dimensions that characterize the valence of an appraisal-eliciting event (e.g., intrinsic pleasantness or goal congruence), dominance roughly map onto the appraisal dimension of coping potential, and arousal a measure of intensity. However, they have quite different meaning: appraisal is a relational construct characterizing the relationship between some specific object/event and the individual's beliefs desires and intentions and several appraisals may be simultaneously active; core affect is a non-relational construct summarizing a unique overall state of the individual.

Dimensional theories emphasize different components of emotion than appraisal theories and link these components quite differently. Dimensional theories foreground the structural and temporal dynamics of core affect and often do not address affect's antecedents in detail. Most significantly, dimensional theorists question the tight causal linkage between appraisal and emotion that is central to appraisal accounts. Dimensional theorists conceive of core affect as a "non-intentional" state, meaning the affect is not about some object (as in "I am angry at *him*). In such theories, many factors may contribute to a change in core affect including symbolic intentional judgments (e.g., appraisal) but also sub-symbolic factors such as hormones and drugs (Schachter and Singer 1962), but most importantly, the link between any preceding intentional meaning and emotion is broken (as it is not represented within core affect) and must be recovered after the fact, sometimes incorrectly (Clore, Schwarz et al. 1994; Clore and Plamer 2009). For example, Russell argues for the following sequence of emotional components: some external event occurs (e.g., a bear walks out of the forest), it is perceived in terms of its affective quality; this perception results in a dramatic change in core affect; this change is attributed to some "object" (e.g., the bear); and only then is the object cognitively appraised in terms of its goal relevance, causal antecedents and future prospects (see also, Zajonc 1980).

Models influenced by dimensional theories, not surprisingly, emphasize processes associated with core affect and other components (e.g., appraisal) tend to be less elaborately developed. Core affect is typically represented as a continuous time-varying process that is represented at a given period of time by a point in 3-space that is "pushed around" by eliciting events. Computational dimensional models often have detailed mechanisms for how this point changes over time – e.g., decay to some resting state

– and incorporating the impact of dispositional tendencies such as personality or temperament (Gebhard 2005).

Computational dimensional models are most often used for animated character behavior generation, perhaps because it translates emotion into a small number of continuous dimensions that can be readily mapped to continuous features of behavior such as the spatial extent of a gesture. For example, PAD models describe all behavior in terms of only three dimensions whereas modelers using appraisal models must either associate behaviors with a larger number of appraisal dimensions (see Smith and Scott 1997; Scherer and Ellgring 2007) or map appraisals into a small number of discrete, though perhaps intensity-varying, expressions (Elliott 1992). For a similar reason, dimensional models also frequently used as a good representational framework for systems that attempt to recognize human emotional behavior and there is some evidence that they may better discriminate user affective states than approaches that rely on discrete labels (Barrett 2006).

The relationship between core affect and cognition is generally less explored in dimensional approaches. Typically the connection between emotion-eliciting events and current core-affective state is not maintained, consistent with Russell's view of emotion as a non-intentional state (e.g., Becker-Asano and Wachsmuth 2008). Interestingly, we are not aware of any computational models that follow the suggestion from Zajonc and Russell that appraisal is a *post hoc* explanation of core affect. Rather, many computational models of emotion that incorporate core affect have viewed appraisal as the mechanism that initiates changes to core affect. For example Gebhard's (2005) ALMA model includes Ortony, Clore and Collins (1988) inspired appraisal rules and WASABI (Becker-Asano and Wachsmuth 2008) incorporates appraisal processes inspired by Scherer's sequential-checking theory into a PAD-based model of core affect. Some computational models explore how core affect can influence cognitive processes. For example, HOTCO 2 (Thagard 2003) allow explanations to be biased by dimensional affect (in this case, a one-dimensional model encoding valence) but this is more naturally seen as the consequence of emotion on cognition (e.g., the modeling of an emotion-focused coping strategy in the sense of Lazarus 1991).

Other approaches

Anatomic approaches: Anatomic theories stem from an attempt to reconstruct the neural links and processes that underlie organisms' emotional reactions (LeDoux 1996; Panskepp 1998; Öhman and Wiens 2004). Unlike appraisal theories, such models tend to emphasize sub-symbolic processes. Unlike dimensional theories, anatomic approaches tend to view emotions as different, discrete neural circuits and emphasize processes or systems associated with these circuits. Thus, anatomically-inspired models tend to foreground certain process assumptions and tend to be less comprehensive than either appraisal or dimensional theories, with researchers focusing on a specific emotion such as fear. For example, LeDoux, emphasizes a "high-road" vs. "low-road" distinction in the fear circuit with the later reflecting automatic/reflexive responses to situations whereas the former is mediated by cognition and deliberation. Computational models inspired by the anatomic tradition often focus on low-level perceptual-motor tasks and encode a two-process view of emotion that argues for a fast, automatic, undifferentiated emotional response and a slower, more differentiated response that relies on higher-level reasoning processes (e.g., Armony, Servan-Schreiber et al. 1997).

Rational approaches: Rational approaches start from the question of what adaptive function does emotion serve and then attempt to abstract this function away from its "implementation details" in humans and incorporate these functions into a (typically normative) model of intelligence (Simon 1967; Sloman and Croucher 1981; Frank 1988; Scheutz and Sloman 2001; Anderson and Lebiere 2003; Doyle 2006). Researchers in this tradition typically reside in the field of artificial intelligence and view emotion as window through which one can gain insight into adaptive behavior, albeit it a very different window than has motivated much of artificial intelligence research. Within this tradition, cognition is conceived as a collection of symbolic processes that serve specific cognitive functions and are subject to certain architectural constraints on how they interoperate. Emotion, within this view, is simply another, albeit often overlooked, set of processes and constraints that have adaptive value. Models of this sort are most naturally directed towards the goal of improving theories of machine intelligence.

Communicative approaches: Communicative theories of emotion argue that emotion processes function as a communicative system; both as a mechanism for informing other individuals of one's mental state – and thereby facilitate social coordination – and as a mechanism for requesting/demanding changes in the behavior of others – as in threat displays (Keltner and Haidt 1999; Parkinson 2009). Communicative theories emphasize the social-communicative function of displays and sometimes argue for a disassociation between internal emotional processes and emotion displays which need not be selected on the basis of an internal emotional state (e.g., see Fridlund 1997; Gratch 2008). Computational models inspired by communicative theories often embrace this disassociation and dispense with the need for an internal emotional model and focusing on machinery that decides when an emotional display will have a desirable effect on a human user. For example, in the Cosmo tutoring system (Lester, Towns et al. 2000), 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. Not surprisingly, computational models based on communicative theories are most often directed towards the goal of achieving social influence.

Dissecting Computational Appraisal Theory

Appraisal theory, by far, dominates the work on computational models of emotion so here we spend some time laying out some terminology that is specific to this class of models (although some of this terminology could apply to other approaches). As we discussed earlier, our aim is to promote incremental research on computational models of emotion by presenting a compositional view of model building, emphasizing that an emotional model is often assembled from individual "sub-models" and these smaller components are often shared and can be mixed, matched, or excluded from any given implementation. More importantly, these components can be seen as embodying certain content and process assumptions that can be potentially assessed and subsequently abandoned or improved as a result of these assessments. In presenting this, we attempt to build as much as possible on terminology already introduced within the emotion literature.

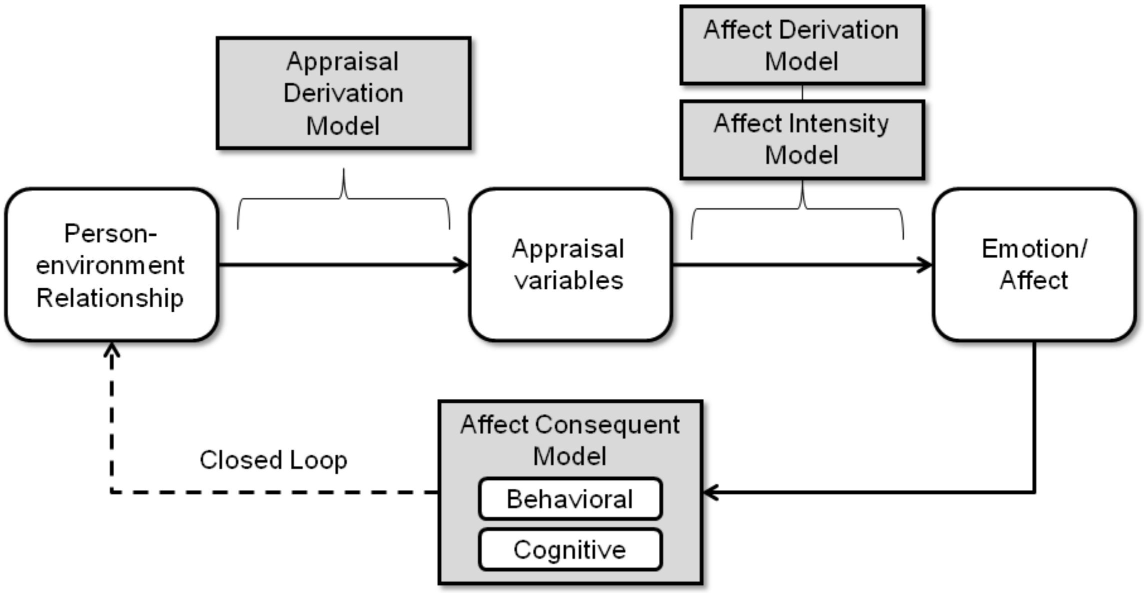

Figure 2: A component model view of computational appraisal models

Component Models

Figure 2 presents an idealized computational appraisal architecture consisting of a set of linked component models. This figure presents what we see as natural joints at which to decompose appraisal systems into coherent and often shared modules, though any given system may fail to implement some of these components or allow different information paths between components. In this architecture, information flows in a cycle as argued by several appraisal theorists (Lazarus 1991; Scherer 2001; Parkinson 2009): some representation of the person-environment relationship is appraised; this leads to an affective response of some intensity; the response triggers behavioral and cognitive consequences; these consequences alter the person-environment; this change is appraised; and so on. Each of these stages can be represented by a model that represents or transforms state information relevant to emotion-processing. Here we introduce terminology associated with each of these:

Person-environment relationship: Lazarus (1991) introduced this term to refer to some representation of the agent's relationship with its environment. This representation should allow an agent, in principle, to derive the relationship between external events (real or hypothetical) and the beliefs, desires and intentions of the agent or other significant entities in the (real or hypothetical) social environment. This representation need not represent these relationships explicitly but must support their derivation. Examples of this include the decision-theoretical planning representations in EMA (Gratch and Marsella 2004) which combines decision-theoretic planning representation with belief-desire-intention formalisms or the partially-observable Markov-decision representations in THESPIAN (Si, Marsella et al. 2008)

Appraisal-derivation model: An appraisal-derivation model transforms some representation of the person-environment relationship into a set of appraisal variables.[1] For example, if an agent's goal is potentially thwarted by some external action, an appraisal-derivation model should be able to automatically derive appraisals that this circumstance is undesirable, assess its likelihood, and calculate the agent's ability to cope, i.e., by identifying alternative ways to achieve this goal. Several computational appraisal models don't provide an appraisal-derivation model or treat its specification as something that is outside of the system. For example, ALMA (Gebhard 2005) allows domain developers to author a the relational model by specifying how certain states or actions should be appraised (e.g., if Sven attacks Valerie, she should appraise this as undesirable). Other researchers treat the appraisal-derivation as a central contribution of their approach. For example, EMA provides a series of domain-independent inference rules that derive appraisal variables from syntactic features of the person-environment relationship (e.g., if the effect of an action threatens a plan to achieve a desired state, this is undesirable). Models also differ in the processing constraints that this model should satisfy. For example, models influenced by Scherer's sequential checking theory incorporate assumptions about the order in which specific appraisal variables should be derived (Marinier 2008). Appraisal-derivation models are often triggered by some eliciting event, though this is not always the case (e.g., EMA simultaneously appraises every goal in an agent's working memory and updates these appraisals continuously as new information about these goals is obtained.

Appraisal variables: Appraisal variables correspond to the set of specific judgments that the agent can use to produce different emotional responses and are generated as a result of an appraisal-derivation model. Different computational appraisal models adopt different sets of appraisal variables, depending on their favorite appraisal theorist. For example, many approaches utilize the set of variables proposed by Ortony, Clore and Collins (1988) including AR (Elliott 1992), EM (Neal Reilly 1996), FLAME (El Nasr, Yen et al. 2000) and ALMA (Gebhard 2005). Others favor the variables proposed by Scherer (Scherer 2001) including WASABI (Becker-Asano and Wachsmuth 2008) and PEACTIDM (Marinier, Laird et al. 2009).

Affect-derivation Model: An affect-derivation model maps between appraisal variables and an affective state, and specifies how an individual will react emotionally once a pattern of appraisals has been determined.[2] As noted in the discussion of different theoretical perspectives above, there is some diversity in how models define "emotion" and here we consider any mapping from appraisal variables to affective state, where this state could be either a discrete emotion label, a set of discrete emotions, core

[1] Smith and Kirby Smith, C. A. and L. D. Kirby (2009). "Putting appraisal in context: Toward a relational model of appraisal and emotion." Cognition and Emotion **00**(00). propose the term *relational model* to refer to this mapping, building on Lazarus' idea that appraisal is a relational construct relating the person and the environment. They introduced the term to draw attention to the fact that many appraisal theories emphasize the mapping from appraisal variable to emotion but neglect the situational and dispositional antecedents of appraisal. As "relation" and "relational" often has a very different meaning within computer science, we prefer a different term.

[2] Smith and Kirby Ibid. use the term *structural model* to refer to this mapping, drawing analogy to structural equation modeling Kline, R. B. (2005). Principles and Practice of Structural Equation Modeling, The Guilford Press., the statistical technique for estimating the causal relationships between variables that appraisal theorists often use to derive these mappings. As the term "structural model" is often used to contrast with "process models" (a distinction we ourselves use later), we prefer the different term.

affect, or even some combination of these factors. For example, Elliott's AR (Elliott 1992) maps appraisal variables into discrete emotion labels, Becker-Asano's WASABI (Becker-Asano and Wachsmuth 2008) maps appraisals into a dimensional (e.g. PAD) representation of emotion, and Gebhard's ALMA (Gebhard 2005) does both simultaneously. Many computational systems adopt the affect-derivation model proposed by Ortony, Clore and Collins (1988) whereby twenty-two emotion labels are defined as conjunctions of appraisal variables – this will be henceforth referred to as the OCC model. Others have implemented models based on the work of Lazarus (e.g., Gratch and Marsella's EMA) and Scherer (e.g., Becker-Asano's WASABI and Marinier's PEACTIDM). Much of the empirical work in psychological appraisal theory has focused on identifying the affect-derivation model that best conforms to human behavior but the results of these studies are far from definitive and can be interpreted to support multiple proposed models.

Affect-Intensity model: An affect-intensity model specifies the strength of the emotional response resulting from a specific appraisal. There is a close association between the affect-derivation model and intensity model (the intensity computation is often implemented as part of the affect-derivation model), however it is useful to conceptualize these separately as they can be independently varied – indeed computational systems with the same affect-derivation model often have quite different intensity equations (Gratch, Marsella et al. 2009). Intensity models usually utilize a subset of appraisal variables (e.g., most intensity equations involve some notion of desirability and likelihood), however they may involve several variables unrelated to appraisal (e.g., Elliott and Siegle 1993). Although less studied than appraisal-derivation models, some research has investigated which intensity model best conforms to human behavior (Mellers, Schwartz et al. 1997; Gratch, Marsella et al. 2009; Reisenzein 2009).

Emotion/Affect: Affect is a representation of the agent's current emotional state. This could be a discrete emotion label, a set of discrete emotions, core affect (i.e., a continuous dimensional space), or even some combination of these factors. An important consideration in representing affect, particularly for systems that model the consequences of emotions, is if this data structure preserves the link between appraisal factors and emotional state. As noted above in the discussion of appraisal and dimensional theories, emotions are often viewed as being about something (e.g., I am angry *at Valarie*). Agents that model affect as some aggregate dimensional space must either preserve the connection between affect and domain objects that initiated changes to the dimensional space, or they must provide some attribution process that *post hoc* recovers a (possibly incorrect) domain object to apply the emotional response to. For example, EM (Neal Reilly 1996) has a a dimensional representation of core affect (valence and arousal) but also maintains a hierarchal data structure that preserves the linkages through each step of the appraisal process to the multiple instances of discrete emotion that underlie its dimensional calculus. In contrast, WASABI (Becker-Asano and Wachsmuth 2008) breaks this link. Some models propose some hybrid. For example, EMA maintains discrete appraisal frames that represent specific discrete emotion instances but then allow a general dimensional "mood" to moderate which discrete emotion raises to the level of awareness.

Affect-consequent model: An affect-consequent model maps affect (or its antecedents) into some behavioral or cognitive change. Consequent models can be usefully described in terms of two dimensions, one distinguishing if the consequence is inner or outer directed (cognitive vs. behavioral), and the other describing whether or not the consequence feeds into a cycle (i.e., is closed-loop).

Emotion can be directed outward into the environment or inward, shaping a person's thoughts. Reflecting this, *behavior consequent models* summarize how affect alters an agent's observable physical behavior and *cognitive consequent models* that determine how affect alters the nature or content of cognitive processes. Most embodied computational systems model the mapping between affect and physical display, such as facial expressions. For example, WASABI maps regions of core affect into one of seven possible facial expressions (Becker-Asano 2008, p. 85) and ParleE (Bui 2004) maps from an emotion state vector (the intensity of six discrete emotion labels) to a facial muscle contraction vector (specifying the motion of 19 facial action units). Emotions can also trigger physical actions. For example, in a process called *problem-focused coping,* EMA (Gratch and Marsella 2004; Marsella and Gratch 2009) attempts to mitigate negative emotions by changing features in the environment that led to the initial undesirable appraisal. In contrast, cognitive consequent models change some aspect of cognition as a result of affective state. This can involve changes in how cognition processes information – for example, Gmytrasiewicz and Lisetti (2000) propose a model that changes the depth of forward projection as a function of emotional state in order to model some of the claimed effects of emotion on human decision making. Cognitive consequent models can also change the content of cognitive processes – for example, EMA implements a set of *emotion-focused coping strategies* like wishful thinking, distancing and resignation that alter an agent's beliefs, desires and intentions, respectively.

We can further distinguish consequences by whether or not they form a cycle by altering the circumstances that triggered the original affective response. For example, a robot that merely expresses fear when its battery is expiring does not address the underlying causes of the fear, whereas one that translates this fear into an action tendency (e.g., seek power) is attempting to address its underlying cause. In this sense, both behavioral and cognitive consequences can be classified as *closed-loop* if they directly act on the emotion eliciting circumstances or *open-loop* if they do not.

Open-loop models are best seen as making an indirect or mediated impact on the agent's emotional state. For example, open-looped behavioral consequences such as emotional displays make sense in multi-agent setting where the display is presumed to recruit resources from other agents. For example, building a robot that expresses fear makes sense if there is a human around that can recognize this display and plug it in. Gmytrasiewicz and Lisetti's (2000) work on changing the depth of decision making can similarly be seen as having an indirect on emotion: by altering the nature of processing to one best suited to a certain emotional state, the cognitive architecture is presumably in a better position to solve problems that tend to arise when in that state.

Closed-loop models attempt to realize a direct impact to regulate emotion and suggest ways to enhance the autonomy of intelligent agents. Closed-loop models require reasoning about the cognitive and environmental antecedents of an emotion so that these factors can ultimately be altered. For example, EMA implements problem-focused coping as a closed-loop behavioral strategy that selects actions that address threats to goal achievement, and implements emotion-focused coping as a closed-loop cognitive strategy that alters mental state (e.g., abandons a goal) in response to similar threats. Closed-loop models naturally implement a view of emotion as a continuous cycle of appraisal, response and re-appraisal: In EMA, an agent might perceive another agent's actions to be a threat to its goals, resulting in anger, which triggers a coping strategy that results in the goal being abandoned, which in turn lowers the agent's appraised sense of control, resulting in sadness (see Marsella and Gratch 2009).

The component model in Figure 2 is, of course, only one of many possible ways to dissect and link emotion components but we have found it pragmatically useful in our own understanding of different computation approaches, as we illustrate below. Additionally, many of the components we identify have previously been highlighted as important distinctions with the literature on emotion research. For example Smith and Kirby (2009) highlight appraisal-derivation as an important but understudied aspect of appraisal theory. In their work they propose the term *relational model* to refer to this component, building on Lazarus' idea that appraisal is a relational construct relating the person and the environment. As "relation" and "relational" often has a very different meaning within computer science, we prefer a different term appraisal-derivation model. Appraisal-derivation models are frequently identified within the appraisal theory under the term *structural model*, drawing analogy to structural equation modeling (Kline 2005), the statistical technique for estimating the causal relationships between variables that appraisal theorists often use to derive these mappings. As the term "structural model" is not emotion-specific, and is often used to contrast with "process models" (a distinction we ourselves use later), we prefer the term appraisal-derivation model. The idea of closed-loop models has been proposed by a variety of appraisal theorists. Most recently, Brian Parkinson has used the term *transactional model* to highlight the incremental unfolding nature of emotional reactions (Parkinson 2009), although we prefer the term closed-loop, again drawing on computer metaphors.

Processing Assumptions

Computational appraisal models can vary, not only by which sub-components they choose to implement, but how these individual components are realized and interact. Some computational systems make strong commitments to *how* information is processed (e.g., is in parallel or sequential?). Others make strong commitments concerning *what* information is processed (e.g., states, goals, plans). Here we introduce terminology that characterizes these different processing choices.

Process Specificity: Computational models vary considerably in term of the claims they make about how information is process. At the most abstract level, a *structural model* specifies a mapping between inputs and outputs but makes no commitment to how this mapping is realized – the term structural comes from structural equation modeling (Kline 2005), a statistical method whereby the relationship between input and output can be inferred. In contrast, a process model posits specific constraints on how inputs are transformed into outputs. For example, Ortony, Clore and Collins present a structural affect-derivation model that maps from appraisal values to an emotion label, whereas Scherer's Sequential Checking Theory is a process appraisal-derivation model that not only specifies the structure of appraisal but proposes a set of temporal processing constraints on how appraisal variables should be derived (e.g., goal relevance should be derived before normative significance). The distinction between structural and process is not clear-cut and is best seen as a continuum. Psychological process theories only specify processes to some level of detail and different theories vary considerably in terms of their specificity. In contrast, a computational model must be specified in sufficient detail for it to be realized as working software, however many of these process details are pragmatic and do not correspond to strong theoretical claims about how such processes should be realized. For example, Elliott's Affective Reasoner implements affect-derivation via a set of *ad hoc* rules, but this should not be seen as a claim that appraisal should be implemented in this manner, but rather as a short cut necessary to create a working system.

Processing constraints can be embedded within an individual appraisal component or can emerge from the interactions of individual components. For example, Scherer's sequential checking theory posits temporal ordering constraints with its model of appraisal derivation. In contrast, Gratch and Marsella's EMA model posits that emotion arises from a continuous cycle of appraisal, coping and re-appraisal and that such temporal properties arise from the incremental evolution of the person-environment relationship (see Marsella and Gratch 2009 for an in-depth discussion of this point).

Processing constraints can be asserted for a variety of reasons. In psychology, process models are typically used to assert theoretical claims about the nature of human mental processes, such as if appraisal is a sequential or parallel process. Within computational systems, the story is more complex. For computational systems that model human psychological processes, the aim is the same: faithfully reflect these theoretical claims into computational algorithms. For example, Marinier (2008) translates Scherer's processing assumptions into architectural constraints on how information is processed in his PEACTIDM model. However, processing constraints can be introduced for a variety of other reasons having nothing to do with fidelity to human psychological findings. These include, for example, formalizing abstract mappings into precise language (Meyer 2006; Lorini and Castelfranchi 2007), proving that a mapping is computable, illustrating efficient or robust algorithms to achieve a mapping, etc.

Representational Specificity: Regardless of how component models process information, computational systems vary considerably in the level-of-detail of the information they process. Some models emotional processes as abstract black boxes (exploring, for example, the implications of different relationships between components) whereas others get down the nitty-gritty of realizing these processes in the context of concrete application domains. This variance is perhaps easiest to see when it comes to appraisal derivation. For example, all appraisal models decompose the appraisal process into a set of individual appraisal checks. However, some models stop at this level, treating each check as a representational primitive (e.g., Sander, Grandjean et al. 2005), whereas others further decompose appraisal checks into the representational details (e.g., domain propositions, actions, and the causal relationships between them) that are necessary for an agent to appraise its relationship to the environment (e.g., Neal Reilly 1996; El Nasr, Yen et al. 2000; Gratch and Marsella 2004; Dias and Paiva 2005; Mao and Gratch 2006; Becker-Asano 2008; Si, Marsella et al. 2008).

Process specificity can vary independently from representational specificity. For example, Sander and colleagues (2005) provide a detailed neural network model of how appraisals are derived from the person-environment relationship, but the person-environment relationship itself is only abstractly represented. Process and representational specificity also vary across component models within the same system. For example, WASABI (Becker-Asano 2008) incorporates detailed representational and process commitments for its model of affect-derivation but uses less detail for its model of appraisal derivation. Such differences often result from the fact that specific systems are directed at addressing a subset of the components involved in emotion processes but that the authors often require a complete working system to assess the impact of their proposed contribution and these other components may be rudimentary or *ad hoc*.

Domain specific vs. Domain independent: In addition to their processing constraints and representational specificity, algorithms can be characterized by the generality of their implementation. A domain-independent algorithm enforces a strict separation between details of a specific domain, typically encoded as a *domain theory* and the remaining code which is written in such a way that it can be used without modification. For example, planning algorithms operate on a domain theory consisting of a set of states and actions that describe a domain and provide general algorithms that operate syntactically on those representations to generate plans. Computational appraisal models differ in terms of how domain-specific knowledge is encoded and which components require domain-specific input. Most systems incorporate domain-independent affect-derivation models (Neal Reilly 1996; Bui 2004; Gratch and Marsella 2004; Gebhard 2005; Becker-Asano 2008; Marinier 2008). Fewer systems provide domain-independent algorithms for appraisal-derivation (e.g., Neal Reilly 1996; El Nasr, Yen et al. 2000; Gratch and Marsella 2004; Si, Marsella et al. 2008).

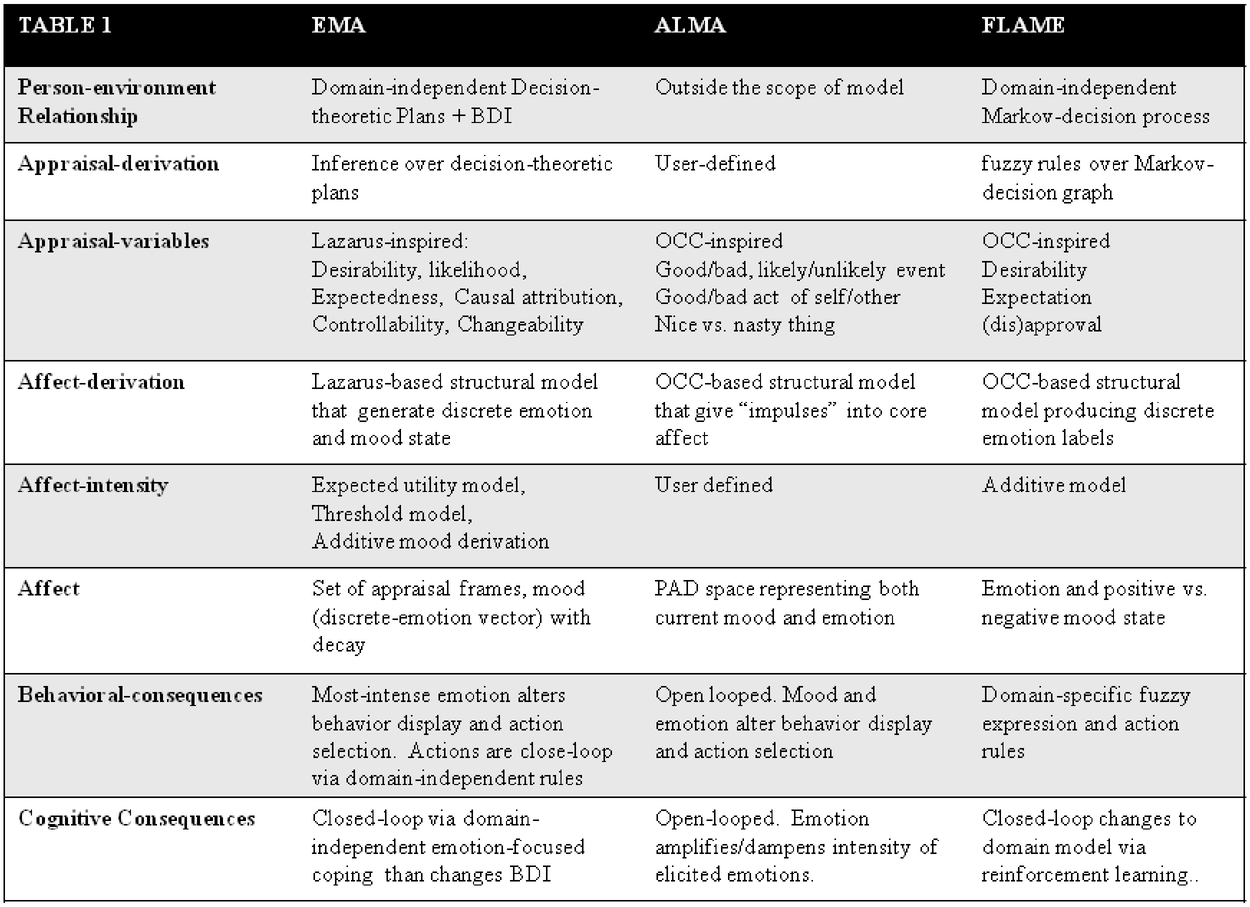

TABLE 1	EMA	ALMA	FLAME
Person-environment Relationship	Domain-independent Decision-theoretic Plans + BDI	Outside the scope of model	Domain-independent Markov-decision process
Appraisal-derivation	Inference over decision-theoretic plans	User-defined	fuzzy rules over Markov-decision graph
Appraisal-variables	Lazarus-inspired: Desirability, likelihood, Expectedness, Causal attribution, Controllability, Changeability	OCC-inspired Good/bad, likely/unlikely event Good/bad act of self/other Nice vs. nasty thing	OCC-inspired Desirability Expectation (dis)approval
Affect-derivation	Lazarus-based structural model that generate discrete emotion and mood state	OCC-based structural model that give "impulses" into core affect	OCC-based structural model producing discrete emotion labels
Affect-intensity	Expected utility model, Threshold model, Additive mood derivation	User defined	Additive model
Affect	Set of appraisal frames, mood (discrete-emotion vector) with decay	PAD space representing both current mood and emotion	Emotion and positive vs. negative mood state
Behavioral-consequences	Most-intense emotion alters behavior display and action selection. Actions are close-loop via domain-independent rules	Open looped. Mood and emotion alter behavior display and action selection	Domain-specific fuzzy expression and action rules
Cognitive Consequences	Closed-loop via domain-independent emotion-focused coping than changes BDI	Open-looped. Emotion amplifies/dampens intensity of elicited emotions.	Closed-loop changes to domain model via reinforcement learning..

Example applications of this framework

Viewing a computational model of emotion as a model of models allows more meaningful comparisons between systems. Systems that appear quite different on the surface can be seen as adopting similar choices along some dimensions and differing in others. Adopting a component model framework can help highlight these similarities and differences, and facilitate empirical comparisons that assess the capabilities or validity of alternative algorithms for realizing component models.

Table 1 illustrates how the component model framework can highlight conceptual similarities and differences between emotion models. This table characterizes three quite different systems: EMA is the authors' own work on developing a general computational model of appraisal and coping motivated by the appraisal theory of Richard Lazarus (Lazarus 1991) and has been applied to driving the behavior of embodied conversational agents (Swartout, Gratch et al. 2006); FLAME is an OCC-inspired appraisal model that drives the behavior of characters in interactive narrative environments (El Nasr, Yen et al. 2000); and ALMA is intended as a general programming tool to allow application developers to more easily construct computational models of emotion for a variety of applications (Gebhard 2005). Some observations that can be made from this table include:

- EMA and FLAME both focus on appraisal derivation. They provide domain-independent techniques for representing the person-environment relationship and derive appraisal variables via domain-independent inferences rules, although the approaches adopt somewhat different representational and inferential choices. In contrast, ALMA does not address appraisal derivation;
- All systems in Table 1 use rules to derive affect from a set of appraisal variables. ALMA and FLAME both adopt OCC-style appraisal variables and affect-derivation models whereas EMA uses a model inspired by Lazarus;
- Each system adopts a different choice for how the intensity of an emotion is calculated.
- All systems incorporate some notion of core affect, though they adopt different representations. EMA has a mood state that summarizes the intensity of all active emotional appraisals and this mood biases the selection of a single emotional appraisal that can impact behavior. ALMA represents both a current emotion and a more general mood in a three-dimensional (PAD) space (either of which can impact behavior). FLAME has a one-dimensional (positive vs. negative) mood state that can influence behavior.
- EMA and FLAME propose closed-loop consequence models that allow emotion to feed back into changes in the mental representation of a situation, although they adopt quite different algorithmic choices for how to realize this function.

Besides allowing such conceptual comparisons, the key benefit of decomposing systems into component models is that it allows individual design decisions to be empirically assessed independent of other aspects of the system. For example, in Table 1, FLAME and EMA adopt different models for deriving the intensity of an emotional response: both systems calculate intensity as a function of the utility and probability of goal attainment but FLAME adds these variables whereas EMA multiples them for prospective emotions (e.g., hope and fear) and uses a threshold model for retrospective emotions (e.g., joy and sadness). An advantage of the component model view is these alternative choices can be

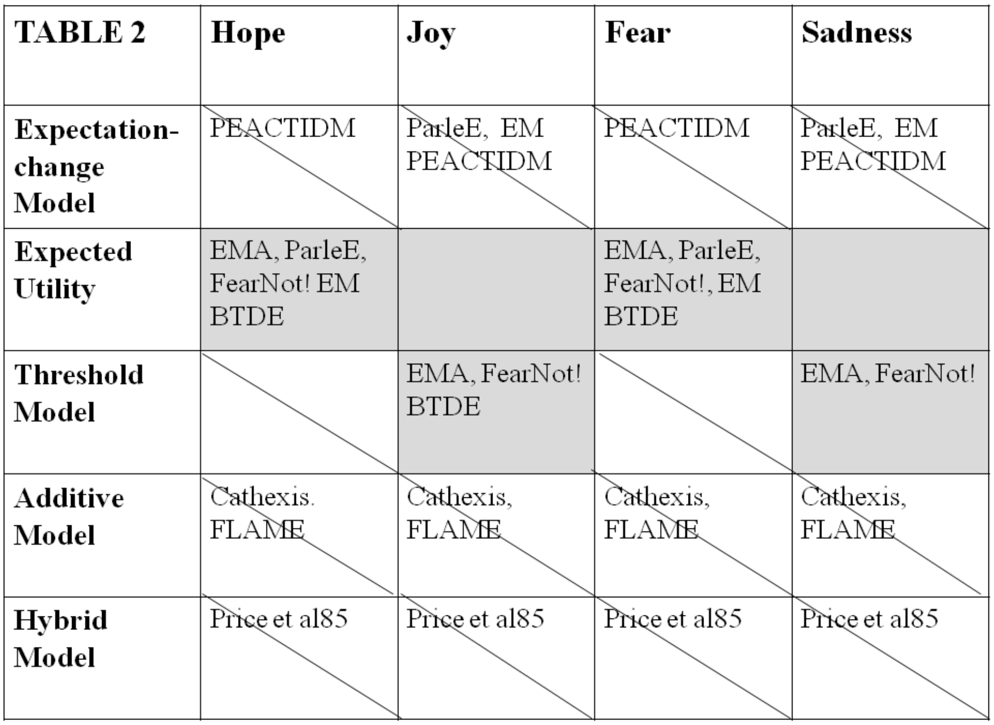

TABLE 2	Hope	Joy	Fear	Sadness
Expectation-change Model	~~PEACTIDM~~	~~ParleE, EM PEACTIDM~~	~~PEACTIDM~~	~~ParleE, EM PEACTIDM~~
Expected Utility	EMA, ParleE, FearNot! EM BTDE		EMA, ParleE, FearNot!, EM BTDE	
Threshold Model		EMA, FearNot! BTDE		EMA, FearNot!
Additive Model	~~Cathexis. FLAME~~	~~Cathexis, FLAME~~	~~Cathexis, FLAME~~	~~Cathexis, FLAME~~
Hybrid Model	~~Price et al85~~	~~Price et al85~~	~~Price et al85~~	~~Price et al85~~

directly compared and evaluated, independently of the other choices adopted in the systems from which they stem.

Gratch and Marsella recently applied this component-model perspective to an empirical comparison of different affect-derivation models (Gratch, Marsella et al. 2009). Besides the two approaches proposed by EMA and FLAME, researchers have proposed a wide range of techniques to calculate the intensity of an affective response. In their study, Gratch and Marsella analyzed several competing approaches for calculating the intensity of a specific emotional response to a situation and classified these approaches into a small number of general approaches (this includes approaches used in a variety of systems including: Price, Barrell et al. 1985; Neal Reilly 1996; Velásquez 1998; El Nasr, Yen et al. 2000; Bui 2004; Dias and Paiva 2005; Marinier, Laird et al. 2009; Reisenzein 2009). They then devised a study to empirically test these competing appraisal intensity models, assessing their consistency with the behavior of a large number of human subjects in naturalistic emotion-eliciting situations. In the study they had subjects play a board game (Battleship™ by Milton Bradley™) and assessed subjects self-reported emotional reactions as the game unfolded and as a consequence of if they were winning or losing (which was manipulated experimentally).

Table 2 summarizes the results of this study that compared the behavior of EMA to several other systems proposed in the literature. These include ParleE (Bui 2004), a system that uses appraisal models to drive facial animation; BTDE (Reisenzein 2009), an appraisal theory that attempts to reduce appraisal-derivation, affect-derivation and affect-intensity to operations over beliefs and desires; FLAME, described above, Cathexis, an anatomical approach that views emotions as arising from drives; FearNot! (Dias and Paiva 2005), an appraisal model based on EMA; EM (Neal Reilly 1996) an OCC-inspired model that drives the behavior of interactive game characters; and a model proposed by Price and colleagues (Price, Barrell et al. 1985) that inspired the design of FLAME. Although these models vary in many ways, when it comes to affect-intensity, they can be described in terms of four basic methods have been proposed in the literature for deriving the intensity of an emotional response.

As noted in the table, different systems used different intensity models depending on the emotion type. The intensity models, listed in the first column, include expected utility (i.e., the intensity of emotional response is proportional to the utility of a goal times its probability of attainment), expectation-change (i.e., the intensity is proportional to the change in probability caused by some event), and additive (i.e., the intensity is proportional to the sum of probability and utility). The cells in the table indicate the intensity model that a particular system applies to calculate the intensity of a given emotion. The table also summarizes the results of how well these different models explain the data elicited from the study. A slash through the box indicates the model cannot explain the results of the experiment. This analysis lends support for the expected utility model for all emotions, with a particularly strong fit for the prospective emotions (i.e., hope and fear), though allows that some modified version of a threshold model might explain the results of retrospective motions like joy and sadness. If the goal of an emotion model is to realistically model human emotional responses, expected utility is probably a good choice for that appraisal intensity component model.

Of course, the behavior of a specific component is not necessarily independent of other design choices so such a strong independence assumption should be treated as a first approximation for assessing how alternative design choices will function in a specific system. However, unless there is a compelling reason to believe choices are correlated, such an analysis should be encouraged. Indeed, a key advantage of the compositional approach is that it forces researchers to explicitly articulate what these dependencies might be, should they wish to argue for a component that is repudiated by an empirical test that adopts a strong assumption of independence.

Dividing computational emotion models into components enables a range of such empirical studies that can assess the impact of these design choices on the possible uses of emotion models that were outlined at the start of this chapter – i.e., their impact on psychological theories of emotion; their impact on artificial intelligence and robotics; and their impact on human-computer interaction. Here we touched on some studies that more naturally apply to the first goal and several examples of this now exist including evaluations of the psychological validity of cognitive consequent models (Marsella, Gratch et al. 2009) and appraisal-derivation models (e.g., Mao and Gratch 2006; Tomai and Forbus 2007). However, the same approach can be equally applied to these other overall objectives. For example, de Melo and colleagues present evidence that the appraised expression of emotion can influence human-computer interaction in the context of social games such as iterated prisoner's dilemma (de Melo, Zheng et al. 2009) and it would be interesting to consider how different appraisal-derivation and intensity models might impact the power of this effect. Other researchers have explored how emotions might improve the decision-making capabilities of general models of intelligence (Scheutz and Sloman 2001; Ito, Pynadath et al. 2008) and a component model analysis can give greater insight into which aspects of these models contribute to enhanced performance.

Conclusion

In this chapter, we provided an overview of research into computational models of emotion that details the common uses of the models and the underlying techniques and assumptions from which the models are built. Our goals were two-fold. For researchers outside the field of computational models on emotion, we want to facilitate an understanding of the field. For research in the field, our goal is to provide a framework that can help foster incremental research, with researchers relying on careful comparisons, evaluations and leveraging to build on prior work, as a key to forward progress.

To achieve those goals, we presented several conceptual distinctions that can aid in evaluation of competing models. We identified several roles for models, in psychological research, in human-computer interaction and in AI. Evaluation, of course, must be sensitive to these roles. If, for example, the model is being used as a methodological tool for research in human emotions or in human-computer interaction research as a means to infer user emotional state, then fidelity of the model with respect to human behavior will be critical. If the model is to be used to create virtual characters that facilitate engagement with, or influence of, humans then fidelity may be less important, even undesirable, while effectiveness in the application becomes more important.

Our assumption is that regardless of the specific details of the evaluation, research progress in computational models of emotion critically hinges not only in evaluations of specific models but also in the comparison across models. Due to complexity of some of these models, and their emphasis on different aspects of the overall emotion process, it may not be reasonable or desirable to undertake comparison and evaluation *in toto*. Rather component-by-component analyses, based on a common lexicon, will be both more revealing and often easier. Our hope is that the application of the component analyses exemplified above may serve as a means to facilitate this component-by-component evaluation and lead to additional work in this direction.

References

Anderson, J. R. (1993). Rules of the Mind. Hillsdale, NJ, Lawrence Erlbaum.

Anderson, J. R. and C. Lebiere (2003). "The Newell Test for a theory of cognition." Behavioral and Brain Sciences **26**: 587-640.

Armony, J. L., D. Servan-Schreiber, et al. (1997). "Computational modeling of emotion: Explorations through the anatomy and physiology of fear conditioning." Trends in Cognitive Science **1**: 28-34.

Barrett, L. F. (2006). "Emotions as natural kinds?" Perspectives on Psychological Science **1**: 28-58.

Bates, J., B. Loyall, et al. (1991). "Broad Agents." Sigart Bulletin **2**(4): 38-40.

Baylor, A. L. and S. Kim (2008). The Effects of Agent Nonverbal Communication on Procedural and Attitudinal Learning Outcomes. International Conference on Intelligent Virtual Agents. Tokyo, Springer**:** 208-214.

Bechara, A., H. Damasio, et al. (1999). "Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-Making." Journal of Neuroscience **19**(13): 5473-5481.

Becker-Asano, C. (2008). WASABI: Affect simulation for agents with believable interactivity. Bielefeld, University of Bielefeld. **PhD**.

Becker-Asano, C. and I. Wachsmuth (2008). Affect Simulation with Primary and Secondary Emotions. 8th International Conference on Intelligent Virtual Agents, Tokyo, Springer.

Biswas, G., K. Leelawong, et al. (2005). "Learning by Teaching. A New Agent Paradigm for Educational Software." Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors" **19**(3-4): 393-412.

Blanchard, A. and L. Cañamero (2006). Developing Affect-modulated behaviors: stability, exploration, exploitation, or imitation? 6th International Workshop on Epigenetic Robotics, Paris.

Blanck, P. D., Ed. (1993). Interpersonal Expectations. Studies in Emotion and Social Interaction. Cambridge, Cambridge University Press.

Bui, T. D. (2004). Creating emotions and facial expressions for embodied agents. Department of Electrical Engineering, Mathematics and Computer Science. Enschede, University of Twente. **PhD**.

Busemeyer, J. R., E. Dimperio, et al. (2007). Integrating emotional processing into decision-making models. Integrated models of cognitive systems. W. D. Grey. Oxford, Oxford University Press.

Campos, J. J., S. Thein, et al. (2003). A Darwinian legacy to understanding human infancy: Emotional expressions as behavior regulators. Emotions inside out: 130 years after Darwin's *The Expression of the Emotions in Man and Animals*. P. Ekman, J. J. Campos, R. J. Davidson and F. B. M. de Waal. New York, New York Academy of Sciences**:** 110-134.

Clore, G. and J. Plamer (2009). "Affective guidance of intelligent agents: How emotion controls cognition." Cognitive Systems Research **10**(1): 21-30.

Clore, G., N. Schwarz, et al. (1994). Affect as information. Handbook of affect and social cognition. J. P. Forgas. Mahwah, NJ, Lawrence Erlbaum**:** 121-144.

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.

Conati, C. and H. MacLaren (2004). Evaluating a probabilistic model of student affect. 7th International Conference on Intelligent Tutoring Systems, Maceio, Brazil.

Cowell, A. and K. M. Stanney (2003). Embodiement and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents. Intelligent Virtual Agents, Kloster Irsee, Germany, Springer-Verlag.

Darwin, C. (2002). The Expression of the Emotions in Man and Animals, Oxford University Press.

de Melo, C., L. Zheng, et al. (2009). Expression of Moral Emotions in Cooperating Agents. 9th International Conference on Intelligent Virtual Agents. Amsterdam, Springer.

de Waal, F. B. M. (2003). Darwin's Legacy and the Study of Primate Visual Communication. Emotions inside out: 130 years after Darwin's *The Expression of the Emotions in Man and Animals*. P. Ekman, J. J. Campos, R. J. Davidson and F. B. M. de Waal. New York, New York Academy of Sciences**:** 7-31.

Dias, J. and A. Paiva (2005). Feeling and Reasoning: a Computational Model for Emotional Agents. Proceedings of 12th Portuguese Conference on Artificial Intelligence, EPIA 2005, Springer.

Doyle, J. (2006). Extending Mechanics to Minds: The Mechanical Foundations of Psychology and Economics. London, UK, Cambridge University Press.

El Nasr, M. S., J. Yen, et al. (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. Northwestern, IL, Northwestern University Institute for the Learning Sciences.

Elliott, C. and G. Siegle (1993). Variables influencing the intensity of simulated affective states. AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications. Palo Alto, CA, AAAI**:** 58-67.

Ellsworth, P. C. and K. R. Scherer (2003). Appraisal processes in emotion. Handbook of the affective sciences. R. J. Davidson, H. H. Goldsmith and K. R. Scherer. New York, Oxford University Press**:** 572-595.

Fasel, I., M. Stewart-Bartlett, et al. (2002). Real time fully automatic coding of facial expressions from video. 9th Symposium on Neural Computation, California Institute of Technology.

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. The Psychology of Facial Expression. J. A. Russell and J. M. Fernández-Dols. Cambridge, Cambridge University Press**:** 103-129.

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

Gebhard, P. (2005). ALMA - A Layered Model of Affect. Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, Utrecht.

Gmytrasiewicz, P. and C. Lisetti (2000). Using Decision Theory to Formalize Emotions for Multi-Agent Systems. Second ICMAS-2000 Workshop on Game Theoretic and Decision Theoretic Agents, Boston.

Gratch, J. (2000). Émile: marshalling passions in training and education. Fourth International Conference on Intelligent Agents, Barcelona, Spain.

Gratch, J. (2008). True emotion vs. Social Intentions in Nonverbal Communication: Towards a Synthesis for Embodied Conversational Agents. Modeling Communication with Robots and Virtual Humans. I. Wachmuth and G. Knoblich. Berlin, Springer. **LNAI 4930**.

Gratch, J. and S. Marsella (2004). "A domain independent framework for modeling emotion." Journal of Cognitive Systems Research **5**(4): 269-306.

Gratch, J. and S. Marsella (2004). Evaluating the modeling and use of emotion in virtual humans. 3rd International Joint Conference on Autonomous Agents and Multiagent Systems, New York.

Gratch, J., S. Marsella, et al. (2009). "Modeling the Antecedents and Consequences of Emotion." Journal of Cognitive Systems Research **10**(1): 1-5.

Gratch, J., S. Marsella, et al. (2009). Assessing the validity of appraisal-based models of emotion. International Conference on Affective Computing and Intelligent Interaction. Amsterdam, IEEE.

Gratch, J., N. Wang, et al. (2007). Creating Rapport with Virtual Agents. 7th International Conference on Intelligent Virtual Agents. Paris, France.

Haag, A., S. Goronzy, et al. (2004). Emotion recognition using bio-sensors: first steps towards an automatic system. Tutorial and Research Workshop on Affective Dialogue Systems, Kloster Irsee, Germany, Springer.

Hatfield, E., J. T. Cacioppo, et al., Eds. (1994). Emotional Contagion. Studies in Emotion and Social Interaction. Cambridge, Cambridge University Press.

Hume, D. (2000). A treatise of human nature. Oxford, Oxford University Press.

Ito, J., D. Pynadath, et al. (2008). Modeling Self-Deception within a Decision-Theoretic Framework 8th International Conference on Intelligent Virtual Agents, Tokyo, Springer.

Keltner, D. and J. Haidt (1999). "Social Functions of Emotions at Four Levels of Anysis." Cognition and Emotion **13**(5): 505-521.

Klesen, M. (2005). "Using Theatrical Concepts for Role-Plays with Educational Agents." Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors" **19**(3-4): 413-431.

Kline, R. B. (2005). Principles and Practice of Structural Equation Modeling, The Guilford Press.

Kramer, N. C., B. Tietz, et al. (2003). Effects of embodied interface agents and their gestural activity. Intelligent Virtual Agents, Kloster Irsee, Germany, Springer.

Krumhuber, E., A. Manstead, et al. (2007). "Facial dynamics as indicators of trustworthiness and cooperative behavior." Emotion **7**(4): 730-735.

Kuppens, P. and I. Van Mechelen (2007). "Interactional appraisal models for the anger appraisals of threatened self-esteem, other-blame, and frustration." Cognition and Emotion **21**: 56-77.

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. and S. Narayanan (2003). Emotion recognition using a data-driven fuzzy interference system. Eurospeech, Geneva.

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

Lester, J. C., S. G. Towns, et al. (2000). Deictic and Emotive Communication in Animated Pedagogical Agents. Embodied Conversational Agents. J. Cassell, S. Prevost, J. Sullivan and E. Churchill. Cambridge, MIT Press**:** 123-154.

Lisetti, C. and P. Gmytrasiewicz (2002). "Can a rational agent afford to be affectless? A formal approach." Applied Artificial Intelligence **16**: 577-609.

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

Lorini, E. and C. Castelfranchi (2007). "The cognitive structure of Surprise: looking for basic principles." Topoi: An International Review of Philosophy **26**(1): 133-149.

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

Mao, W. and J. Gratch (2006). Evaluating a computational model of social causality and responsibility. 5th International Joint Conference on Autonomous Agents and Multiagent Systems, Hakodate, Japan.

Marinier, R. P. (2008). A Computational Unification of Cognitive Control, Emotion, and Learning. Computer Science. Ann Arbor, MI, University of Michigan. **PhD**.

Marinier, R. P., J. E. Laird, et al. (2009). "A computational unification of cognitive behavior and emotion." Cognitive Systems Research **10**(1): 48-69.

Marsella, S. and J. Gratch (2001). Modeling the interplay of plans and emotions in multi-agent simulations. Cognitive Science Society, Edinburgh, Scotland.

Marsella, S. and J. Gratch (2009). "EMA: A Model of Emotional Dynamics." Journal of Cognitive Systems Research **10**(1): 70-90.

Marsella, S., J. Gratch, et al. (2003). Expressive Behaviors for Virtual Worlds. Life-like Characters Tools, Affective Functions and Applications. H. Prendinger and M. Ishizuka. Berlin, Springer-Verlag: 317-360.

Marsella, S., J. Gratch, et al. (2009). Assessing the validity of a computational model of emotional coping. International Conference on Affective Computing and Intelligent Interaction. Amsterdam, IEEE.

McCall, C., J. Blascovich, et al. (2009). "Proxemic behaviors as predictors of aggression towards Black (but not White) males in an immersive virtual environment." Social Influence **4**(2): 138 - 154.

Mehrabian, A. and J. A. Russell (1974). An Approach to Environmental Psychology. Cambridge, Mass, The MIT Press.

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

Mellers, B. A., A. Schwartz, et al. (1997). "Decision affect theory: Emotional reactions to the outcomes of risky options." Psychological Science **8**(6): 423-429.

Meyer, J.-J. C. (2006). "Reasoning about Emotional Agents." International journal of intelligent systems **21**(6): 601-619.

Minsky, M. (1986). The Society of Mind. New York, Simon and Schuster.

Moors, A., J. De Houwer, et al. (2005). "Unintentional processing of motivational valence." The Quarterly Journal of Experimental Psychology.

Nakanishi, H., S. Shimizu, et al. (2005). "Sensitizing Social Agents for Virtual Training." Applied Artificial Intelligence special Issue "Educational Agents - Beyond Virtual Tutors" **19**(3-4): 341-362.

Neal Reilly, W. S. (1996). Believable Social and Emotional Agents. Pittsburgh, PA, Carnegie Mellon University.

Neal Reilly, W. S. (2006). Modeling what happens between emotional antecedents and emotional consequents. Eighteenth European Meeting on Cybernetics and Systems Research. Vienna, Austria, Austrian Society for Cybernetic Studies: 607-612.

Öhman, A. and S. Wiens (2004). The concept of an evolved fear module and cognitive theories of anxiety. Feelings and Emotions. A. Manstead, N. Frijda and A. H. Fischer. Cambridge, Cambridge University Press: 58-80.

Ortony, A., G. Clore, et al. (1988). The Cognitive Structure of Emotions, Cambridge University Press.

Paiva, A., R. Aylett, et al. (2004). "AAMAS Workshop on Empathic Agents." from http://gaips.inesc.pt/gaips/en/aamas-ea/index.html.

Paiva, A., J. Dias, et al. (2005). "Learning by Fealing: Evoking Empathy with Synthetic Characters." Applied Artificial Intelligence special issue on "Educational Agents - Beyond Virtual Tutors" **19**(3-4): 235-266.

Panskepp, J. (1998). Affective Neuroscience: The Foundations of Human and Animal Emotions. New York, Oxford University Press.

Parkinson, B. (2009). "What holds emotions together? Meaning and response coordination." Cognitive Systems Research **10**: 31-47.

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

Price, D. D., J. E. Barrell, et al. (1985). "A quantitative-experiential analysis of human emotions." Motivation and Emotion **9**(1).

Rank, S. (2009). Behaviour Coordination for Models of Affective Behavior. Vienna, Austria, Vienna University of Technology. **Ph.D.**

Rank, S. and P. Petta (2005). Appraisal for a Character-based Story-World. 5th International Working Conference on Intelligent Virtual Agents, Kos, Greece, Springer.

Reisenzein, R. (2009). "Emotions as Metarepresentational States of Mind: Naturalizing the Belief-Desire Theory of Emotion." Journal of Cognitive Systems Research **10**(1).

Rickel, J., S. Marsella, et al. (2002). Toward a New Generation of Virtual Humans for Interactive Experiences. IEEE Intelligent Systems. **July/August:** 32-38.

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

Sander, D., D. Grandjean, et al. (2005). "A systems approach to appraisal mechanisms in emotion." Neural Networks **18**: 317-352.

Schachter, S. and J. E. Singer (1962). "Cognitive, social, and physiological determinants of emotional state." Psychological Review **69**: 3790399.

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

Scherer, K. R. and H. Ellgring (2007). "Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal?" Emotion.

Scheutz, M. and P. Schermerhorn (2009). Affective Goal and Task Selection for Social Robots. The Handbook of Research on Synthetic Emotions and Sociable Robotics. J. Vallverdú and D. Casacuberta.

Scheutz, M. and A. Sloman (2001). Affect and agent control: experiments with simple affective states. IAT, World Scientific Publisher.

Shaver, K. G. (1985). The attribution of blame: Causality, responsibility, and blameworthiness. NY, Springer-Verlag.

Si, M., S. C. Marsella, et al. (2008). Modeling Appraisal in Theory of Mind Reasoning. 8th International Conference on Intelligent Virtual Agents, Tokyo, Japan, Springer.

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

Sloman, A. and M. Croucher (1981). Why robots will have emotions. International Joint Conference on Artificial Intelligence, Vancouver, Canada.

Smith, C. A. and L. Kirby (2000). Consequences require antecedents: Toward a process model of emotion elicitation. Feeling and Thinking: The role of affect in social cognition. J. P. Forgas, Cambridge University Press**:** 83-106.

Smith, C. A. and L. D. Kirby (2009). "Putting appraisal in context: Toward a relational model of appraisal and emotion." Cognition and Emotion **00**(00).

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

Staller, A. and P. Petta (2001). "Introducing Emotions into the Computational Study of Social Norms: A First Evaluation." Journal of Artificial Societies and Social Simulation **4**(1).

Swartout, W., J. Gratch, et al. (2006). "Toward Virtual Humans." AI Magazine **27**(1).

Thagard, P. (2003). "Why wasn't O. J. convicted: emotional coherence in legal inference." Cognition and Emotion **17**: 361-383.

Thomas, F. and O. Johnston (1995). The Illusion of Life: Disney Animation. New York, Hyperion.

Tomai, E. and K. Forbus (2007). Plenty of Blame to Go Around: A Qualitative Approach to Attribution of Moral Responsibility. Proceedings of Qualitative Reasoning Workshop. Aberystwyth, U.K.

Traum, D., J. Rickel, et al. (2003). Negotiation over tasks in hybrid human-agent teams for simulation-based training. International Conference on Autonomous Agents and Multiagent Systems, Melbourne, Australia.

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

Watson, D. and A. Tellegen (1985). "Toward a consensual structure of mood." Psychological Bulletin **98**: 219-235.

Weiner, B. (1995). The Judgment of Responsibility, Guilford Press.

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