CHAPTER 5

Appraisal Models

Jonathan Gratch *and* Stacy Marsella

Abstract

This chapter discusses appraisal theory, the most influential theory of emotion in affective computing today, including how appraisal theory arose, some of its well-known variants, and why appraisal theory plays such a prominent role in computational models of emotion. The authors describe the component model framework, a useful framework for organizing and contrasting alternative computational models of emotion and outline some of the contemporary computational approaches based on appraisal theory and the practical systems they help support. Finally, the authors discuss open challenges and future directions.

Key Words: emotion, appraisal theory, computational models

Introduction

Although psychologists can afford the luxury of describing emotion in broad, abstract terms, computer scientists must get down to brass tacks. For machines to reason about a phenomenon, it must be representable in a formal language and manipulated by well-defined operations. Computer science entrants into the field of emotion are immediately confronted with the challenge of how to represent such imprecise and overlapping concepts as emotion, mood, and temperament. Emotion theory is one useful tool for confronting this imprecision. Psychological theories of emotion are by no means precise, but they posit important constraints on emotion representations and processes. Alternative theories pose quite different and potentially irreconcilable constraints and thus constitute choice points on how one approaches the problem of "implementing" affective computations. This chapter discusses appraisal theory, the most influential theory of emotion in affective computing today. We discuss how appraisal theory arose and some of its important variants and computational instantiations, but also some of the challenges this theory faces (see Reisenzein's chapter in this volume for a more general overview of emotion theories, including appraisal).

Since the very beginnings of artificial intelligence (AI), theoretical controversies about the nature of the human mind have been reflected in battles over computational techniques. The early years of AI research were dominated by controversies over whether knowledge should be represented as procedures or declarative statements (e.g., Winograd, 1975), reflecting similar debates raging in cognitive science. Within the subdomain of automated planning research, debates erupted over whether intelligent action selection was best conceptualized as processes operating on explicit plan representations or more perceptually driven reactive processes (e.g., Ginsberg, 1989; Suchman, 1987). Even within the sub-subdomain of those who favor explicit plan representations, debates rage as to whether planning is best conceptualized as a search through a space of possible world states or a search through a space of possible plans (e.g., Kambhampati & Srivastava, 1995). These choices are not simply theoretical but have clear implications for the capabilities of the resulting software systems: for example, state-based planners, in that they don't maintain explicit representations of plans, make it exceedingly difficult to compare, identify threats, and de-conflict plans of multiple agents, thus making them (arguably) an ill-suited choice for social or multiagent problem solving.

Emotion theories also have potentially profound implications for computational systems that reason about affective phenomena. As one dramatic example, consider the extremely influential theory of humours that Galen of Pergamum (AD 130–200) used to explain human mood and temperament. According to this view, mood was influenced by specific environmental and physiological processes and life events. Conceptually, mood reflected a mixture of bodily substances: phlegm, black bile, yellow bile, and blood. A predominance of yellow bile led to strong fiery emotions, was promoted by warm and dry weather, and was more common in youth or summer; conversely, phlegm produced a stolidly calm disposition that arose typically in cold and moist environments and in old age or winter. This theory suggests clear ways to represent measure and control mood. For example, the theory of humours implies that mood disorders can be treated by bleeding, blistering, sweating, or vomiting, a process that was common practice well into the eighteenth century (Duffy, 1959).

More contemporary theoretical debates center on the relationship between emotion and cognition. These debates address the following questions: Is emotional reasoning somehow distinct and qualitatively different from unemotional reasoning (Kahneman, 2003; LeDoux, 1996)? Do emotions serve adaptive functions, or do they lead to maladaptive decisions (Frank, 2004; Keltner & Haidt, 1999; Pham, 2007; H. A. Simon, 1967)? Does emotion precede or follow thought (Lazarus, 1984; Zajonc, 1984)? Appraisal theory has played a central role in shaping these debates, although, as we will see in this chapter, appraisal theorists do not always agree on how to answer these questions.

This chapter is structured as follows. We first review appraisal theories, examine why they arose, and discuss some of influential variants. We then discuss why appraisal theories play such a prominent rule in computational models and how some of their properties are particularly well-suited for computational realization. We outline some of the contemporary computational approaches based on appraisal theory and the practical systems they help support. Finally, we discuss open challenges and future directions.

Appraisal Theory

Appraisal theory is currently a predominant force among psychological perspectives on emotion and is arguably the most fruitful source for those interested in the design of AI systems because it emphasizes and explains the connection between emotion and the symbolic reasoning processes that AI favors. Indeed, the large majority of computational models of emotion stem from this tradition. In appraisal theory, emotion arises 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 are cognitive in nature but not necessarily conscious or controlled. They characterize personally significant events in terms of a fixed set of specific criteria, sometimes referred to as *appraisal variables* or *appraisal dimensions* and include considerations such as whether events are congruent with the individual's goals, expected, or controllable. (Table 5.1 illustrates appraisal variables proposed by some prominent appraisal theorists.) According to appraisal theory, specific emotions are associated with specific patterns of appraisal. For example, a surprising and uncontrollable event might provoke fear. In several versions of appraisal theory, appraisals also trigger cognitive responses, often referred to as *coping strategies*—e.g., planning, procrastination, or resignation—that feed back into a continual cycle of appraisal and reappraisal (Lazarus, 1991, p. 127).

The assumption underlying appraisal theory (i.e., that emotions arise from subjective evaluations) has reoccurred many times in history and can be found in the writings of Aristotle and Hume. The recent usage of the term "appraisal" commences with the writings of Magda Arnold (1960) and was subsequently reinforced by the work of Richard Lazarus (1966). The development of appraisal theory was motivated, in part, by the observation that different individuals might respond quite differently to the same event and

Table 5.1 Appraisal variables proposed by several appraisal theorists.

Scherer	Frijda	Roseman	Smith/Ellsworth
Novelty	Change		Attentional activity
• Suddenness			
• Familiarity	Familiarity		
• Predictability			
Intrinsic pleasantness	Valence		Pleasantness
Goal significance		Appetitive/Aversive	
• Concern relevance	Focality	Motives	importance
Outcome probability	Certainty	Certainty	certainty
• Expectation	Presence		
• Conduciveness	Open/Closed	Motive consistency	Perceived obstacle/Anticipated effort
• Urgency	Urgency		
Coping potential			
• Cause: agent	Intent/Self–other	Agency	Human agency
• Cause: motive			
• Control	Modifiability	Control potential	Situational control
• Power	Controllability		
• Adjustment			
Compatibility Standards			
• External	Value relevance		Legitimacy
• Internal			

Terms that line up horizontally refer to comparable processes, despite the fact that the respective authors use different labels (adapted from Scherer, 2005).

the desire to posit specific mechanisms that could explain these differences. Appraisal theories have been studied for many years, and there is substantial experimental evidence that supports the basic claims underlying this theory (for a more detailed introduction into the different variants of appraisal theory, see Scherer, Schorr, & Johnstone, 2001).

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 (see Parkinson, 1997, for one critical perspective on this view). 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 & Scherer, 2003, for a discussion). Much of the research has focused on the structural relationship between appraisal variables and specific discrete emotions—that is, which pattern of appraisal variables would elicit hope (see Ortony, Clore, & Collins, 1988), or on the structural relationship between appraisal variables and specific behavioral and cognitive responses—that is, which pattern of appraisal variables would elicit certain facial expressions (Scherer & Ellgring, 2007; Smith & Scott, 1997) or coping tendencies (Lazarus, 1991). Appraisal theorists allow that the same situation may elicit multiple appraisals and, in some cases, that these appraisals can occur at multiple levels of reasoning (Scherer, 2001), but most 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 (corresponding to discrete emotion categories) or more dimensional representations (such as valence and arousal).

Today, most emotion researchers accept that appraisal plays a role in emotion, although they may differ on the centrality of this process. Active research on appraisal theory has moved away from demonstrating the existence of appraisal and has turned to more specific questions about how it impacts individual and social behavior. Some work examines the processing constraints underlying appraisal—to what extent is it parallel or sequential (Moors, De Houwer, Hermans, & Eelen, 2005; Scherer, 2001)? Does it occur at multiple levels (Scherer, 2001; Smith & Kirby, 2000)? Some work seeks to create a better understanding of the cognitive, situational, and dispositional factors that influence appraisal judgments (Kuppens & Van Mechelen, 2007; Smith & Kirby, 2009). Other work focuses more on the consequences of emotion on subsequent appraisal and decision making (Han, Lerner, & Keltner, 2007; Horberg, Oveis, & Keltner, 2011). Finally, a very active area of interest concerns the implications of appraisal theory on social cognition (Gratch & Marsella, 2014; Hareli & Hess, 2009; Manstead & Fischer, 2001).

Although there are many appraisal theorists, work in affective computing has been most influenced by a small set of appraisal theories. The most influential among these has been the so-called *OCC model* (Figure 5.1A) of Ortony, Clore, and Collins (1988)—the name reflects the first initial of each author. OCC is most naturally seen as a structural model (in the sense of structural equation modeling) in that it posits a small set of criteria (appraisal variables) that distinguish between different emotion terms. Thus, it can be seen as an easily implemented decision tree for classifying emotion-evoking situations, which perhaps explains its seduction for computer scientists.

At the top level, the OCC divides emotions into three broad classes. First, objects or events might lead to emotion in that they are intrinsically pleasing/displeasing for a given individual (e.g., "I love chocolate but hate rock concerts"). Second, objects or events might evoke emotion based how they relate to an individual's goals (e.g., "I'm afraid this traffic will make me late for my date"). Finally, from a social perspective, emotions may arise due to how an object (typically a person) or event impact social norms (e.g., "I disapprove of his stealing"). Within these broad categories, emotions are further distinguished by the extent to which they are positive/negative, impact self or other, and so forth. The OCC also posits a large number of criteria that can impact the intensity of emotional reactions.

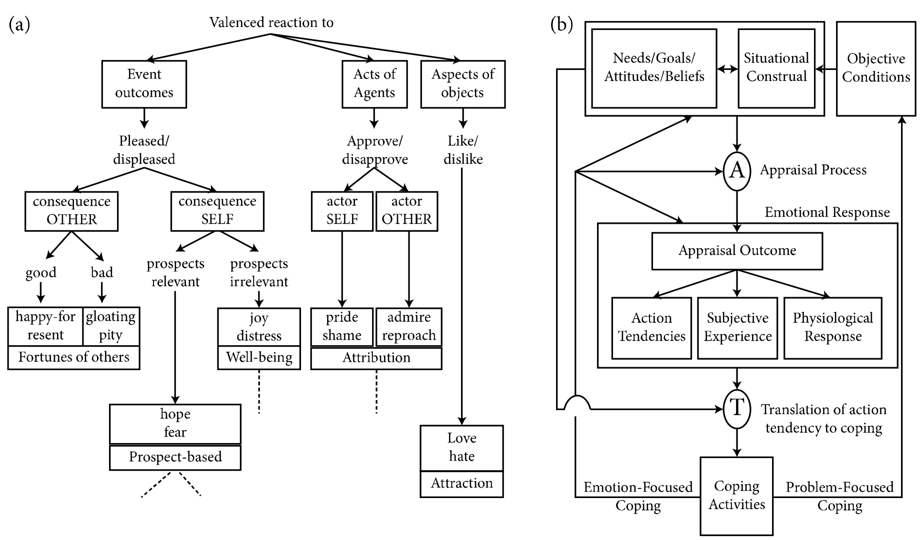

Fig. 5.1 ~~Gratch Appraisal Models.~~ (A) The figure on the left illustrates the OCC model of emotion (adapted from Ortony, Clore, & Collins, 1988). (B) The figure on the right is one visual representation of Lazarus's appraisal theory proposed by Smith and Lazarus (adapted from Smith & Lazarus, 1990).

Whereas the OCC emphasizes the structure of emotion-eliciting events, the work of Richard Lazarus (1991) takes a broader and more process-oriented view that emphasizes both the antecedents and consequences of emotion. Lazarus's work is rich and nuanced, but affective computing researchers have been most influenced by the description of this theory outlined in a joint paper with Craig Smith (Smith & Lazarus, 1990), which recasts the theory in more computational terms (as illustrated in Figure 5.1B). Lazarus's theory follows a similar approach to OCC with regard to emotion antecedents (i.e., emotions arise from patterns of judgments on how objects or event impact beliefs, attitudes, and goals). But inspired by his work with clinical populations, this theory further emphasizes that appraisals shape broader patterns of behavior, which he calls coping strategies, and thus influence subsequent appraisals in a dynamic, cyclical process of appraisal and reappraisal. Coping strategies are roughly grouped into problem-focused strategies (e.g., planning and seeking instrumental social support) that act on the world and emotion-focused strategies (e.g., distancing or avoidance) that act on the self. In either case, coping strategies serve to modify the person–environment relationship to maintain emotional well-being.

Klaus Scherer's *sequential checking theory* (SCT) is the most recent and certainly the most elaborate appraisal theory to significantly impact affective computing researchers (Figure 5.2). As with the OCC and Lazarus's cognitive mediational theory, the SCT posits a set of appraisal dimensions that relate to the assessment of the person–environment relationship. As with Lazarus, the SCT adopts the view of appraisal as a process that unfolds over time, with later stages feeding back and modifying initial appraisals in a cyclical process of appraisal and reappraisal. The SCT goes further than Lazarus in positing a fixed sequential structure to appraisals. First, novel events are appraised as if they are self-relevant. If relevant, they are judged for their implication for the individual's goals. Next, coping potential is assessed, and finally events are appraised with regard to their compatibility with social norms.

Although the SCT is the most elaborate appraisal theory, this doesn't necessarily make it the most suitable starting point for an affective computing researcher. There are many other variants of appraisal theory, and affective computing researchers would benefit by considering multiple theoretical sources and understanding the different processing and representational commitments they might entail. For example, Rainer Reisenzein's belief-desire theory of emotion (Reisenzein, 2009) is a simple and elegant

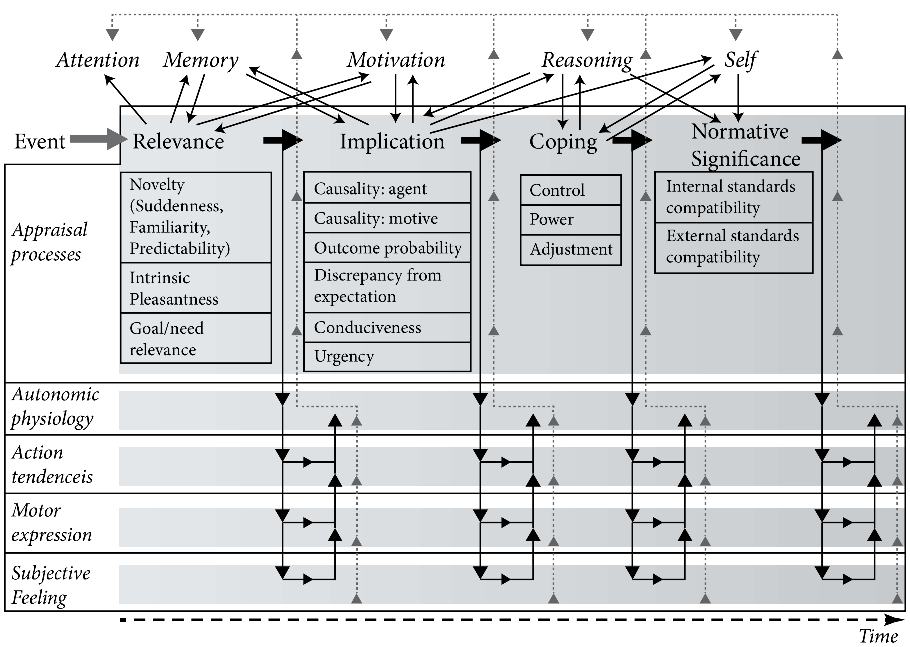

Fig. 5.2 ~~Gratch Appraisal Model.~~ Sequential checking theory (originally appearing in Sander, Grandjean, & Scherer, 2005).

formalization that especially appeals to those interested in formal computational models, and the work of Frijda (1988) emphasizes the connection between emotion and action tendencies and has influenced several computational models (Moffat & Frijda, 1995; Frijda & Swagerman, 1987).

Each of these appraisal theories shares much in common. Each emphasizes that emotions are a relational construct: they represent how an individual is doing vis-à-vis his or her environment. They further share that this relationship is assessed in terms of specific appraisal variables that characterize this relationship. They differ in detail, and this may have implications for affective computing researchers who aim to exploit these theories. For example, appraisals of control and coping potential play a prominent role in SCT and Lazarus's theory (see also Roseman, 2001) but not in the OCC. These differences impact the conceptualization of emotion (e.g., control is a key component of anger for Lazarus but not OCC) as well as the relationship between emotion and behavior (e.g., for Lazarus, appraisals of control dictate the coping strategy an individual will adopt).

Computational Appraisal Theory

Artificial intelligence grew out of cognitive and symbolic approaches to modeling human decision making (e.g., Simon, 1969) so it is hardly surprising that a cognitive theory like appraisal theory should have such affinity for computational scientists of emotion. Unlike some alternative perspectives on emotion (e.g., Russell & Barrett, 1999), appraisal theory aspires to provide a detailed information processing description of the mechanisms underlying emotion production (although perhaps less detailed descriptions of other aspects of emotion processing, such as its bidirectional associations with bodily processes). Further, well-known appraisal theorists (e.g., Andrew Ortony and Craig Smith) were trained in computational methods and tend to describe their theories in ways that resonate with affective computing researchers.

Within affective computing research, models derived from appraisal theories of emotion 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 & Marsella, 2004; Marsella & Gratch, 2009), reactive plans (Neal Reilly, 2006; Rank & Petta, 2005; Staller & Petta, 2001), Markov decision processes (El Nasr, Yen, & Ioerger, 2000; Si, Marsella, & Pynadath, 2008), or detailed cognitive models (Marinier, Laird, & Lewis, 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 labels 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 the later discussion of core affect) that summarizes the effect of recent emotion elicitations (Gebhard, 2005; Gratch & Marsella, 2004; Neal Reilly, 1996). Some more recent models attempt to capture the impact of momentary emotion and mood on the appraisal process (Gebhard, 2005; Gratch & Marsella, 2004; Marsella & Gratch, 2009; Paiva, Dias, & Aylett, 2005).

In most computational models, appraisal is not an end of itself, but a means to influence behavior (such as an agent's emotional expressions or decision making). Models make different choices on how behavior is related to appraisal, emotion, and mood. Some systems encode a direct connection between appraisals and behavior. For example, in Gratch and Marsella's (2004) EMA model, coping behaviors are triggered directly from appraisals (although the choice of which appraisal to focus on is moderated by a mood state). Other systems associate behaviors with emotion (Elliott, 1992) or mood (Gebhard, 2005), essentially encoding the theoretical claim that affect mediates behavior. In the former case, the emotional state as a label is not so critical if there are clear rules that determine behavior (including expression) as a function of appraisal patterns.

There are now several computational models based on appraisal theory (for recent overviews, see Hudlicka, 2008; Marsella, Gratch, & Petta, 2010), but modelers often fail to build on each other's work, tending rather to start anew from original psychological sources. This is beginning to change, and there is now a family tree of sorts, illustrated in Figure 5.3. Yet, even when models build on each other, this tends to be at an abstract, conceptual level. Methods may adopt a similar general approach (e.g., cast appraisal as inference over some plan-like data structures) but rarely share identical algorithm or representational choices, as in other fields of computer science.

Steunebrink et al. (2012) used KARO to formalize the cognitive-motivational preconditions

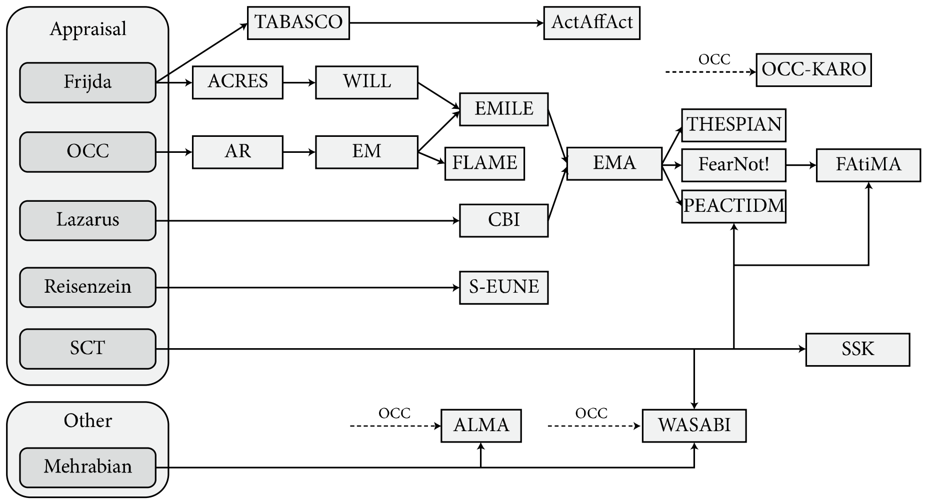

Fig. 5.3 ~~Gratch Appraisal Model~~. A family history of appraisal models. Blocks on the left correspond to original psychological sources. Blocks on the right correspond to models, and arrows correspond to conceptual links. Models mentioned include ACRES (Frijda & Swagerman, 1987), AR (Elliott, 1992), TABASCO (Staller & Petta, 2001), WILL (Moffat & Frijda, 1995), EM (Neal Reilly, 1996), FLAME (El Nasr et al., 2000), EMILE (Gratch, 2000), CBI (Marsella, Johnson, & LaBore, 2000), S-EUNE (Macedo, Reisenzein, & Cardoso, 2004), ALMA (Gebhard, 2005), ActAffAct (Rank, 2009), EMA (Gratch & Marsella, 2004), THESPIAN (Si et al., 2008), FearNot (Dias & Paiva, 2005), PEACTIDM (Marinier et al., 2009), WASABI (Becker-Asano, 2008), OCC-KARO (Steunebrink, Dastani, & Meyer, 2008), FAtiMA (Dias, Mascarenhas, & Paiva, 2011), and SSK (Broekens, DeGroot, & Kosters, 2008).

of the 22 emotions considered in the OCC theory (Ortony et al., 1988).

When approaching a computational approach to appraisal theory, we encourage affective computing researchers to avoid the temptation to simply "implement" a specific psychological theory. In his 1969 book, *The Sciences of the Artificial*, Herb Simon outlined several ways in which computational scientists bring a unique and complementary perspective to the challenge of understanding human intelligence. First, in contrast to the natural sciences, which seek to describe intelligence as it is found in nature, the "artificial sciences" seek to describe intelligence as it ought to be. This normative emphasis often leads to serviceable abstractions that crisply capture the essence of phenomena while avoiding the messy details of how these functions are implemented in biological organisms. Second, computational scientists approach the problem of achieving function with a mindset emphasizing process, and, from this perspective, apparently complex behavior can often be reduced to simple goal-directed processes interacting over time with a complex environment. Finally, computational scientists produce working artifacts, and these allow theories to be tested in novel and important ways. For example, we might model an artificial ant in terms of a minimal number of theoretically posited functions, simulate the interaction of this model with complex environments, and thereby empirically work out the implicit consequences of our assumptions. Indeed, Simon's original argument was that psychological theories were sorely in need of rational reinterpretation using the computational tools of function and process. Thus, by faithfully "implementing" a psychological theory, computational scientists are doing a disservice both to computational science and the original theory.

New tools often transform science, opening up new approaches to research and allowing previously unaddressed questions to be explored, as well as revealing new questions. Computational appraisal theory, although still in its infancy, has begun to have an impact in several distinct areas of research, including the design of artificially intelligent entities and research on human–computer interactions. It is even beginning to flow back and shape the psychological research from which it sprang. In terms of AI and robotics, appraisal theory suggests ways to generalize and extend traditional rational models of intelligence (Antos & Pfeffer, 2011; Gmytrasiewicz & Lisetti, 2000). In terms of human–computer interaction, appraisal theory posits that emotions reflect the personal significance of events, thus computers that either generate or recognize emotion may foster a better shared understanding of the beliefs, desires, and intentions of the human–machine system (Conati, 2002; Gratch & Marsella, 2005*b*). Finally, the exercise of translating psychological appraisal theory into a working artifact allows theory to be tested in novel and important ways.

A Component Model View

Elsewhere, we have argued that research into computational models of emotions could be considerably advanced by a more incremental and compositional approach toward model construction (Marsella et al., 2010), and we summarize these arguments here (see also Hudlicka, 2011; Reisenzein, 2001). This perspective emphasizes that appraisal involves an ensemble of information processing, and, as a consequence, an emotional model is often assembled from individual "submodels" and these smaller components could be (and in some cases, already are) shared. 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, providing insights to all models that share this subapproach. Thus, this chapter summarizes this componential perspective when reviewing computational approaches to appraisal theory.

Figure 5.4 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, although 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; Parkinson, 2009; Scherer, 2001): 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 encode these relationships explicitly but must support their derivation. Examples of this include the decision-theoretical planning representations in EMA (Gratch & Marsella, 2004), which combines decision-theoretic planning representation with belief-desire-intention formalisms or the partially observable Markov decision representations in THESPIAN (Si et al., 2008).

• *Appraisal-derivation model.* An appraisal-derivation model transforms some representation of the person–environment relationship into a set of appraisal variables. For example, if an agent's

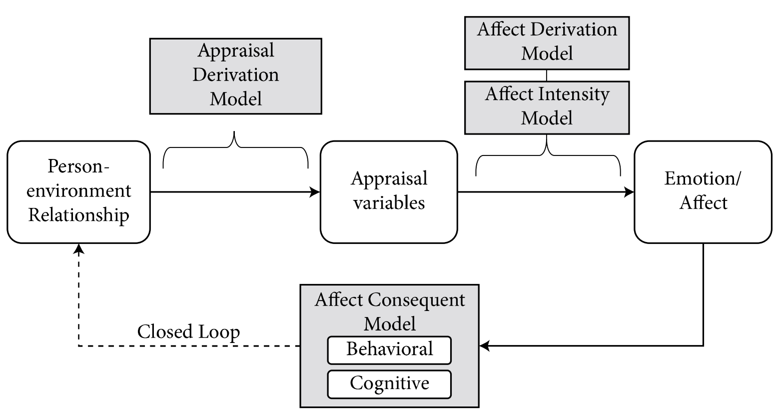

Fig. 5.4 ~~Gratch Appraisal Model~~. A component model view of computational appraisal models.

goal is potentially thwarted by some external action, an appraisal-derivation model should be able to automatically infer that this circumstance is undesirable, assess its likelihood, and calculate the agent's ability to cope. Several computational appraisal models do not provide an appraisal-derivation model or treat its specification as something lying outside of the system (e.g., Gebhard, 2005), whereas others treat it as a central contribution of their approach (Gratch & Marsella, 2004). Models also differ in the processing constraints that this component should satisfy. For example, models influenced by Scherer's SCT incorporate assumptions about the order in which specific appraisal variables should be derived (Marinier, 2008).

• *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 models adopt different sets of appraisal variables or dimensions depending on their favorite appraisal theorist. For example, many approaches utilize the set of variables proposed by the work of Ortony, Collins and Clore (1999), known as the "OCC model," including AR (Elliott, 1992), EM (Neal Reilly, 1996), FLAME (El Nasr et al., 2000), and ALMA (Gebhard, 2005). Others favor the variables proposed by SCT including WASABI (Becker-Asano & Wachsmuth, 2008) and PEACTIDM (Marinier 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. 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, a core affect, or even some combination of these factors. For example, Elliott's AR (1992) maps appraisal variables into discrete emotion labels, Becker-Asano's WASABI (Becker-Asano & Wachsmuth, 2008) maps appraisals into a dimensional (e.g., PAD) representation of emotion, and Gebhard's (2005) ALMA does both simultaneously.

• *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; however, it is useful to view these separately because they can be independently varied—indeed, computational systems with the same affect-derivation model often have quite different intensity equations (Gratch, Marsella, & Petta, 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 & Siegle, 1993).

• *Emotion/Affect.* Affect, in the present context, is a representation of the agent's current emotional state. This could be a discrete emotion label, a set of discrete emotions, a 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 whether the circumstances that provoked the emotion are explicitly represented. Emotions are often viewed as being about something (e.g., I am angry *at Valarie*), and behavioral or coping responses are typically directed at that target. 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 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 & Wachsmuth, 2008) breaks this link.

• *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 open- or closed-loop). With regard to the inner- versus outer-directed dimension, *behavior consequent models* summarize how affect alters an agent's observable physical behavior (e.g., facial expressions), and *cognitive consequent models* determine how affect alters the nature or content of cognitive processes (e.g., coping strategies).

Most embodied computational systems model the former mapping: for example, WASABI maps regions of core affect into facial expressions (Becker-Asano, 2008, p. 85). Examples of the later include EMA's (Gratch & Marsella, 2004) implementation of *emotion-focused coping strategies* like wishful thinking. The second dimension distinguishes 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 (i.e., an open-loop strategy) does not address the underlying causes of the fear, whereas one that translates this fear into an action tendency to seek power (i.e., a closed-loop strategy) is attempting to address its underlying cause. Open-loop models may be appropriate in multiagent setting where the display is presumed to recruit resources from other agents (e.g., building a robot that expresses fear makes sense if there is a human around who can recognize this display and plug it in). Closed-loop models attempt to realize a direct impact to regulate emotion and suggest ways to enhance the autonomy of intelligent agents and naturally implement a view of emotion as a continuous cycle of appraisal, response, and reappraisal.

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. The FAtiMA modular appraisal framework is a recent important tool that helps these sort of comparisons (Dias, Mascarenhas, & Paiva, 2011). It essentially implements the conceptual framework outlined in Figure 5.4 and allows affective computing researchers to explore the interaction of different modules. 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.

Challenges and Future Directions

Although influential, appraisal theory is by no means the final word on emotion. Several aspects of the theory are under intense scrutiny from both with the community of appraisal proponents (suggesting novel mechanisms and novel domains within which to explore the implications of the theory) and without (raising sustained criticism of core assumptions and calling the whole enterprise into question). These issues raise interesting challenges and opportunities in the field of affective computing.

Many of the criticisms of appraisal theory attack the core assumption, held by many but not all appraisal theorists, that cognitive processes precede emotional responses. In general, appraisal theory has been criticized for being overly cognitive, and it seemingly fails to capture the apparent reflexive and uncontrolled nature of emotional responses. This conflict was explored in great detail through a series of debates between Richard Lazarus and Robert Zajonc (Lazarus, 1984; Zajonc, 1984), with Zajonc taking the position that affective processes are primary and serve to motivate and recruit subsequent cognitive responses. Appraisal theorists respond with evidence that different emotions arise depending on the content of mental states (e.g., beliefs and goals).

From the perspective of computer science, many of these arguments seem to fall flat and seem to devolve into arguments over definitions: such as, what is cognition, deliberation, and consciousness? For example, AI has helped to highlight how much complexity underlies presumably reactive reasoning, making distinctions between levels less obvious. Nonetheless, this debate extends to more fundamental issues about the sequencing and linkage between cognition and emotion. For example, a strong cognitive perspective on appraisal theory would suggest that my anger is preceded by judgments that another has intentionally caused me harm. However, quite a bit of evidence suggests that the process could act in reverse (at least in some circumstances). According to this view, an initial feeling of anger would motivate cognitive judgments that "rationalize" the feeling: I'm mad, therefore I blame you! (see Parkinson, 1997). Such theorists don't rule out cognitions as a potential source of affective reactions, but take a broader view, arguing that many factors may contribute to a feeling of emotion including symbolic intentional judgments (e.g., appraisal) but also subsymbolic factors such as hormones. Most importantly, this broader perspective argues that the link between any preceding intentional meaning and emotion is not explicitly represented and must be recovered after the fact, sometimes incorrectly (Clore & Plamer, 2009; Clore, Schwarz, & Conway, 1994; Russell, 2003). For example, Russell argues for the following sequence of emotional components: some external event occurs (e.g., a person walks out on a shaky suspension bridge), the event results in a dramatic change in affect (e.g., arousal); this change is attributed to some "object" (which could be the correct object—the potential for falling—or some irrelevant object, such as a person of the opposite sex standing on the bridge); and only then is the object cognitively appraised in terms of its goal relevance, causal antecedents, and future prospects (e.g., get off the bridge following a correct attribution or ask person out for a date for the incorrect one).

The sequencing of emotion and appraisal becomes less relevant if one adopts the view of Lazarus: that is, that emotion involves a continuous cycle of appraisal, response, and reappraisal. In a cycle, deciding if the egg or chicken came first becomes more academic. Nonetheless, misattribution effects—irrelevant factors, such as how the weather may impact decisions about whether to invest in the stock market (Hirshleifer & Shumway, 2003)—present challenges for the current crop of computational models of appraisal processes.

Another challenge to appraisal theory argues that it is not a theory of true emotion, but rather a theory of how people think about emotion (e.g., see Johnson-Laird & Oatley, 1992, for one discussion of this debate). This challenge emphasizes that much of the evidence in support is introspective. For example, people are presented an imaginary situation and decide how they might feel (e.g., see Gratch & Marsella, 2005*a*). Indeed, the OCC model grew, in part, out of a series of meetings at the University of Illinois where Ortony, Clore, and others tried to sort emotion terms into different categories. Although there is evidence on both sides of this debate, an interesting question for the affective computing researcher is whether it matters. If the goal is to create a robot that effectively communicates emotion or to capture what third-party observers infer from emotional displays, a "folk-theory" of emotion may produce more accurate conclusions.

Even if one accepts the basic tenets of appraisal theory, the approach has been criticized as being too limited in scope for many of the domains of interest in affective computing. One major concern is its relevance for social interaction. Much of appraisal theory has taken the individual as the unit of analysis: how do emotions arise in the individual, how do they impact individual behavior (e.g., physiology and expressions), and how do they impact subsequent cognitions? As a consequence, "social emotions" such as guilt, shame, and embarrassment are underdeveloped in many appraisal theories, and the impact of emotion displays on the behavior and judgments of others has been explored even less, at from the appraisal perspective (but see de Melo, Gratch, & Carnevale, 2011; Hareli & Hess, 2009; Manstead & Fischer, 2001). This situation is changing, and many of the most exciting developments in appraisal theory deal with its extension to social phenomena (e.g., see Kappas, 2013). For a review of recent developments in the area of social appraisals, see Gratch and Marsella (2014).

Conclusion

Appraisal theory is an influential theory of emotion and an especially useful framework for computational scientists interested in building working models of how emotion influences cognitive and behavioral processes. By postulating that emotions arise from patterns of judgments/information processing, appraisal theory draws fruitful connections with other areas of automated reasoning. Although rarely specified in enough detail to directly inform computational systems (with the consequence that many quite different computer models might be consistent with the same theory), it is specified in sufficient detail to posit clear and falsifiable constraints on information process. Recent research on the social antecedents and consequences of emotion are especially interesting and emphasize the relevance of appraisal theory to those interested in building social systems.

References

Antos, D., & Pfeffer, A. (2011). *Using emotions to enhance decision-making.* Paper presented at the Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume One.

Arnold, M. (1960). *Emotion and personality.* New York: Columbia University Press.

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

Becker-Asano, C., & Wachsmuth, I. (2008). *Affect simulation with primary and secondary emotions.* Paper presented at the 8th International Conference on Intelligent Virtual Agents, Tokyo.

Broekens, J., DeGroot, D., & Kosters, W. A. (2008). Formal models of appraisal: Theory, specification, and computational model. *Cognitive Systems Research*, *9*(3), 173–197.

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

Clore, G., Schwarz, N., & Conway, M. (1994). Affect as information. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 121–144). Mahwah, NJ: Lawrence Erlbaum.

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.

de Melo, C., Gratch, J., & Carnevale, P. J. (2011). *Reverse appraisal: Inferring from emotion displays who is the cooperator and the competitor in a social dilemma*. Paper presented at the Cognitive Science Conference, Boston.

Dias, J., Mascarenhas, S., & Paiva, A. (2011). *Fatima modular: Towards an agent architecture with a generic appraisal framework.* Paper presented at the Proceedings of the International Workshop on Standards for Emotion Modeling, Leiden, The Netherlands.

Dias, J., & Paiva, A. (2005). *Feeling and Reasoning: A computational model for emotional agents*. Paper presented at the Proceedings of 12th Portuguese Conference on Artificial Intelligence, EPIA 2005, Covilhã, Portugal.

Duffy, J. (1959). Medical practice in the ante bellum South. *The Journal of Southern History*, *25*(1), 53–72. doi: 10.2307/2954479

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

Elliott, C. (1992). *The affective reasoner: A process model of emotions in a multi-agent system.* Northwestern, IL: Northwestern University Institute for the Learning Sciences.

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

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

Frank, R. H. (2004). Introducing moral emotions into models of rational choice. In A. Manstead, N. Frijda, & A. Fischer (Eds.), *Feelings and emotions* (pp. 422–440). Cambridge, UK: Cambridge University Press.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist*, *43*, 349–358.

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

Gebhard, P. (2005). *ALMA—A Layered Model of Affect.* Paper presented at the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, Utrecht.

Ginsberg, M. L. (1989). Universal planning: An (almost) universally bad idea. *AI Magazine*, *10*(4), 40.

Gmytrasiewicz, P., & Lisetti, C. (2000). *Using decision theory to formalize emotions for multi-agent systems.* Paper presented at the Second ICMAS-2000 Workshop on Game Theoretic and Decision Theoretic Agents, Boston.

Gratch, J. (2000). *Émile: Marshalling passions in training and education.* Paper presented at the Fourth International Conference on Intelligent Agents, Barcelona, Spain.

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

Gratch, J., & Marsella, S. (2005a). Evaluating a computational model of emotion. *Journal of Autonomous Agents and Multiagent Systems*, *11*(1), 23–43.

Gratch, J., & Marsella, S. (2005b). Lessons from emotion psychology for the design of lifelike characters. *Applied Artificial Intelligence*, *19*(3–4), 215–233.

Gratch, J., & Marsella, S. (Eds.). (2014). *Social emotions in nature and artifact*. Cambridge, MA: Oxford University Press.

Gratch, J., Marsella, S., & Petta, P. (2009). Modeling the antecedents and consequences of emotion. *Journal of Cognitive Systems Research*, *10*(1), 1–5.

Han, S., Lerner, J. S., & Keltner, D. (2007). Feelings and consumer decision making: The appraisal-tendency framework. *Journal of Consumer Psychology*, *17*(3), 158–168.

Hareli, S., & Hess, U. (2009). What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition and Emotion*, *24*(1), 128–140.

Hirshleifer, D., & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *Journal of Finance*, *58*, 1009–1032.

Horberg, E. J., Oveis, C., & Keltner, D. (2011). Emotions as moral amplifiers: An appraisal tendency approach to the influences of distinct emotions upon moral judgment. *Emotion Review*, *3*(3), 237–244.

Hudlicka, E. (2008). Review of cognitive-affective architectures. In G. Zacharias, J. McMillan, & S. Van Hemel (Eds.), *Organizational modeling: From individuals to societies*. Washington, DC: National Academies Press.

Hudlicka, E. (2011). Guidelines for designing computational models of emotions. *International Journal of Synthetic Emotions*, *2*(1),pp. 26–79.

Johnson-laird, P. N., & Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition & Emotion*, *6*(3–4), 201–223. doi: 10.1080/02699939208411069

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720.

Kambhampati, S., & Srivastava, B. (1995). *Universal classical planner: An algorithm for unifying state-space and plan-space planning*. Paper presented at the New Directions in AI Planning, EWSP, Assisi, Italy.

Kappas, A. (2013). Social regulation of emotion: messy layers. *Frontiers in Psychology*, *4*, 1–11

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*(5), 505–521.

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

Lazarus, R. S. (1966). *Psychological stress and the coping process*. New York: McGraw-Hill.

Lazarus, R. S. (1984). On the primacy of cognition. *American Psychologist*, *39*(2), 124–129. doi: 10.1037/0003–066X.39.2.124

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

LeDoux, J. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon & Schuster.

Macedo, L., Reisenzein, R., & Cardoso, A. (2004). *Modeling forms of surprise in artificial agents: Empirical and theoretical study of surprise functions*. Paper presented at the 26th Annual Conference of the Cognitive Science Society, Chicago.

Manstead, A. S. R., & Fischer, A. H. (2001). Social appraisal: The social world as object of and influence on appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal* *processes in emotion: Theory, methods, research* (pp. 221–232)). New York: Oxford University Press.

Marinier, R. P. (2008). *A computational unification of cognitive control, emotion, and learning.* (PhD), University of Michigan, Ann Arbor, MI.

Marinier, R. P., Laird, J. E., & Lewis, R. L. (2009). A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*, *10*(1), 48–69.

Marsella, S., & Gratch, J. (2009). EMA: A process model of appraisal dynamics. *Journal of Cognitive Systems Research*, *10*(1), 70–90.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. In K. R. Scherer, T. Bänziger & E. Roesch (Eds.), *A blueprint for affective computing: A sourcebook and manual* (pp. 21–46). New York: Oxford University Press.

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

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

Moors, A., De Houwer, J., Hermans, D., & Eelen, P. (2005). Unintentional processing of motivational valence. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1043–1063.

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*. Paper presented at the Eighteenth European Meeting on Cybernetics and Systems Research, Vienna, Austria.

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

Paiva, A., Dias, J., & Aylett, R. (2005). Learning by feeling: Evoking empathy with synthetic characters. *Applied Artificial Intelligence* (*special issue on "Educational Agents—Beyond Virtual Tutors"*), *19*(3–4), 235–266.

Parkinson, B. (1997). Untangling the appraisal-emotion connection. *Personality and Social Psychology Review*, *1*(1), 62–79.

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

Pham, M. T. (2007). Emotion and rationality: A critical review and interpretation of empirical evidence. *Review of General Psychology*, *11*(2), 155–178.

Rank, S. (2009). *Behaviour coordination for models of affective behavior.* Ph.D. dissertation, Vienna University of Technology, Vienna, Austria.

Rank, S., & Petta, P. (2005). *Appraisal for a character-based story-world.* Paper presented at the 5th International Working Conference on Intelligent Virtual Agents, Kos, Greece.

Reisenzein, R. (2001). Appraisal processes conceptualized from a schema-theoretic perspective. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 187–201). New York: Oxford University Press.

Reisenzein, R. (2009). Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion. *Journal of Cognitive Systems Research*, *10*(1), 6–20

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research*. New York: Oxford University Press.

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

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

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, *18*, 317–352.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.

Scherer, K. R. (2005). Appraisal theory. *Handbook of cognition and emotion*, T. Dagleish and M. J. Power (Eds.), John Wiley and Sons, West Sussex, England. 637–663.

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, *7*(1), 113–130. doi: 10.1037/1528–3542.7.1.113

Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.). (2001). *Appraisal processes in emotion.* New York: Oxford University Press.

Si, M., Marsella, S. C., & Pynadath, D. V. (2008). *Modeling appraisal in theory of mind reasoning.* Paper presented at the 8th International Conference on Intelligent Virtual Agents, Tokyo, Japan.

Simon, H. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

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

Smith, C. A., & Kirby, L. (2000). Consequences require antecedents: Toward a process model of emotion elicitation. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 83–106). New York: Cambridge University Press.

Smith, C. A., & Kirby, L. D. (2009). Putting appraisal in context: Toward a relational model of appraisal and emotion. *Cognition and Emotion*, *23*(7), 1352–1372.

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

Smith, C. A., & Scott, H. S. (1997). A componential approach to the meaning of facial expressions. In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 229–254). Paris: Cambridge University Press.

Staller, A., & Petta, P. (2001). Introducing emotions into the computational study of social norms: A first evaluation. *Journal of Artificial Societies and Social Simulation*, *4*(1), 27–60.

Steunebrink, B. R., Dastani, M. M., & Meyer, J. -J. C. (2012). A formal model of emotion triggers: an approach for BDI agents. *Synthese* 185.1, 83–129.

Steunebrink, B. R., Dastani, M. M., & Meyer, J. -J. C. (2008). *A formal model of emotions: Integrating qualitative and quantitative aspects.* Paper presented at the 18th European Conference on Artificial Intelligence, Patras, Greece.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. New York: Cambridge University Press.

Winograd, T. (1975). Frame representations and the declarative/procedural controversy. *Representation and understanding: Studies in cognitive science,* D. G. Bobrow and A. Collins (Eds.). Academic Press, Inc. Orlando, FL.185–210.

Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist*, *39*(2), 117–123. doi: 10.1037/0003–066X.39.2.117