Section 1

Theoretical approaches to the study of emotion in humans and machines

Throughout history android automata have fascinated scientists and the general public, an interest that reached an apex in the eighteenth century (Riskin 2007). It is probably no accident that La Mettrie wrote his influential *L'homme machine* [Machine man] in a period when Jacques de Vaucanson had an immense success with an automatic flute player and a defecating mechanical duck. One of the most remarkable androids from that period is probably the lady musician built by Swiss watchmakers Pierre and Henri-Louis Jacquet-Droz (Voskuhl 2007). This android played several pieces on the harmonium, showing all the skills of an emotionally engaged performer, moving her head, eyes, and lips—and even breathing—in synchrony with musical affect (see Figure 0.1.1). The harmonium player, a sensation in the courts and capitals of Spain and France at the time, might well be considered a first rudimentary effort at *affective computing* in the sense of an embodied agent.

The term *affective computing*, coined by Rosalind Picard (1997), has gradually been accepted as the label for 'computing that relates to, arises from, or deliberately influences emotion or other affective phenomena', as deployed in emotional robots or affectively competent autonomous, and possibly conversational, agents, and which informs emotion theory with the genuine contributions of computational operationalizations and models. In principle, this is a modern update of La Mettrie's machine man endowed with advanced emotional competences—being able to recognize human emotions, to convincingly express emotional signals, and possibly even *have* emotions (or at least an underlying architecture that simulates human-like emotion processes). This is a tall order indeed and one that, as outlined in the preface, can only be attacked through massively interdisciplinary approaches.

One can reasonably argue that any attempt in the direction of creating believable emotionally competent agents requires a guideline, or an architectural blueprint, in terms of a comprehensive, conceptually sound, model of emotion. Unfortunately, currently there is little consensus on such a model, the history of theorizing about emotion in several disciplines having been controversial (see the entries 'emotion definitions' and 'emotion theories and concepts' in Sander and Scherer 2009). In consequence, in this first section, an overview of available approaches is provided.

In Chapter 1.1, 'Emotion and emotional competence: conceptual and theoretical issues for modelling agents', Scherer provides a description of some of the elements theoretically needed to construct a virtual agent with the ability to display human-like emotions and to respond

Fig. 1 'The harmonium player', the first emotionally competent automaton produced by Pierre and Henri-Louis Jacquet-Droz (see Voskuhl 2007; image downloaded from http://drnorth.wordpress.com/category/georges-melies/).

appropriately to human emotion. It includes distinctive definitions of affective concepts and a definition of emotional competence in this context. He presents a comparative overview of major psychological models/theories and their relative advantages or disadvantages with respect to the computational modelling of the proposed mechanisms. As different theories cover different aspects of emotion, the chapter discusses how their contributions can be integrated. The chapter concludes with a survey of desirable features for emotion theories that make them ideal blueprints for agent models.

Then, in Chapter 1.2, 'Computational models of emotion', Marsella, Gratch, and Petta demonstrate how computer simulations of emotion models provide a means to question traditional conceptualizations and expose hidden assumptions. This approach also allows the systematic comparison of the temporal dynamics of emotion processes and makes it possible to formulate predictions about the time course of these processes. Marsella and colleagues point out that findings on the functional, often adaptive role that emotions play in human behaviour have motivated artificial intelligence and robotics research to explore whether modelling emotion processes can contribute to making progress in elusive areas such as perception and sociability, leading to more intelligent, flexible, and capable systems. Most importantly, they provide a detailed discussion of the potential roles for computational models of emotion: a methodological tool for emotion theories; a new approach to artificially intelligent systems; and a means to enhance human–computer interaction. In addition to providing a detailed overview of major computational models described in the literature, Marsella and colleagues discuss their relative advantages and disadvantages in current applications. They conclude by identifying future directions for this research and outlining its potential impact on emotion research.

Chapter 1.1

Emotion and emotional competence: conceptual and theoretical issues for modelling agents

Klaus R. Scherer

Summary

After a brief definition of the term affective competence, it is argued that this notion must be squarely based on a viable architectural model of emotion that reflects the evidence established by centuries of conceptualization and empirical research about the nature and function of emotion. There seems to be massive convergence in the literature on the notion that emotions are mechanisms that facilitate an individual's adaptation to constantly and complexly changing environmental contingencies. In order for this mechanism to be successful it must be based on an evaluation or appraisal of these contingencies using criteria that are specific to the unique motivational structure, both dispositional and transient, of that individual and it must activate action tendencies that are appropriate to the respective situational demands and affordances. As environmental contingencies and also transient motivational states can change rapidly and as information processing may constantly alter situational appraisal, emotion must consist of a recursive process of synchronized changes in several components.

Given these design features of the emotion mechanism, competence to manage the continuously evolved emotion mechanism, which integrates both psychobiological and sociocultural determinants, needs to be defined in terms of function or adaptation. I suggest that highly emotionally competent individuals are characterized by an optimal functioning of the emotion mechanism with respect to two major domains—emotion production and emotion perception—each of which is constituted by different facets of competence. Emotion production refers to the total pattern of bodily and behavioural changes that characterizes the adaptive function of emotion, allowing the organism to cope with events of major relevance for well-being. These changes are outwardly visible and constitute important social signals for interaction partners, informing them about the individual's reaction and probable behavioural intention. Given this important role of emotion signalling in social intercourse, individuals need to be able to accurately perceive and interpret the emotional state of others. This is what will be referred to as emotion perception competence.

This conceptualization provides a description of the elements theoretically needed to construct a virtual agent with the ability to display human-like emotions and to respond appropriately to human emotion. It is shown that the computational model of emotion used for the architecture of an emotionally intelligent virtual agent must correspond to the design features of emotion and emotional competence in human individuals—in other words, it must be dynamic, recursive, and emergent.

The main contribution of this chapter is a comparative overview of psychological models/theories and their relative advantages or disadvantages with respect to the computational modelling

of the proposed mechanisms. As different theories cover different aspects of emotion, the chapter discusses how their contributions can be integrated. The chapter concludes with a survey of desirable features for emotion theories that would make them ideal blueprints for agent models.

Defining affective, or emotional, competence

While the term *affective computing* is now widely used to refer to computational modelling of emotion and implementations of autonomous agents capable of affective processing, the notion of *affect* remains ill defined. I have suggested (Scherer 2005*b*) that this term is used to cover emotions, preferences, attitudes, affective dispositions, and interpersonal stances that differ widely with respect to event focus, intrinsic and transactional appraisal, degree of synchronization, rapidity of change, behavioural impact, intensity, and duration. In consequence, the term affective competence used with respect to autonomous agents that can recognize and express affective states covers a very wide range of phenomena. However, in virtually all cases the term is mainly used to refer to different types of *emotions*. In consequence, given the absence of an appropriate body of literature on other types of affect, I will focus on the more established notion of *emotional competence* (EC). This term is similar to but not synonymous with *emotional intelligence* (EI). Emotional intelligence has been conceived of in parallel to cognitive intelligence (Mayer and Salovey 1993), defined as a bundle of abilities or skills that have a common underlying factor (small *g*). In contrast, emotional competence is based squarely on the nature and function of the emotion mechanism and describes the abilities and skills needed to use the latter to the best advantage of the individual and/or his reference group. In the following paragraphs, the definition of emotional competence is summarized, based an earlier overview (Scherer 2007).

The central assumption is that differences in the competence to manage the continuously evolved emotion mechanism, which integrates both psychobiological and sociocultural determinants, need to be defined in terms of *function or adaptation*. Thus, highly emotionally competent individuals are characterized by an optimal functioning of the emotion mechanism with respect to two major domains—emotion production and emotion perception—each of which involves different facets of competence. *Emotion production competence* refers to the total pattern of bodily and behavioural changes that characterize the adaptive function of emotion, allowing the organism to cope with events of major relevance for well-being. In many cases the 'raw' emotion generated on the basis of immediate appraisal is modified, either due to reappraisal of the situation, strategic intentions, or social rules. The capacity to achieve such modification in an efficient fashion will be called emotion regulation competence. The outwardly visible manifestations of the regulated emotion constitute important social signals for interaction partners, informing about the individual's reaction and probable behavioural intention. This function requires two skills that will be subsumed under the notion of emotion communication competence. Given the important role of emotion signalling in social intercourse, individuals need to be able to: (1) generate appropriate and convincing cues of their true emotion (or the emotion they want to advertise to observers) and (2) accurately perceive and interpret the emotional state of other individuals.

We can distinguish three components of production competence: (1) producing the most appropriate emotional reaction to different types of events based on adequate appraisal of internal goal states, coping potential, and the probable consequences of events; (2) being able to adaptively regulate one's emotional states, both with respect to internal set points and according to the sociocultural and situational context; (3) efficiently communicating in social interaction through appropriate expression of one's own emotional state. The discussion of these three aspects is based in part on a specific model of emotion, the component process model (CPM) of emotion (see Scherer 2001, 2009*a*; Chapter 2.1, this volume).

Appraisal competence

In order to produce an appropriate response, the significance of the triggering events must be correctly appraised (Ellsworth and Scherer 2003; Scherer 2001; see also Chapter 2.1, this volume). One can distinguish two facets of appraisal competence: (1) appropriate emotion elicitation and (2) appropriate emotion differentiation. Appropriate emotion elicitation refers to the ability to rapidly detect significant objects and events that require an emotional response. Our relevance detection ability is of paramount importance, especially as it often relies on unconscious processes. It is one thing to react emotionally when it is required, another to react with the appropriate emotion. The criterion for an emotion being appropriate for a given context is difficult to define and relies in large part on circumstantial evidence. One approach is to define appropriateness negatively, by identifying emotional disturbances such as anhedonia, euphoria, dysphoria, depression, panic attacks, and the like. There is widespread social consensus that such enduring emotional response dispositions are signals of ill health and require therapy, indicating that the emotional reactions of the respective individuals are considered as *inappropriate* or pathological by society at large.

Another approach to defining the appropriateness of an emotional reaction is constituted by the notion of *valid appraisal*. Appropriate emotion differentiation requires evaluating the implications of an event in a realistic fashion and correctly estimating one's coping potential. In addition, emotions such as pride, shame, guilt, and anger require an accurate representation of social expectations, norms, and moral standards. One essential prerequisite for accurate appraisal is to evaluate each event on its merits and to avoid being influenced by evaluative biases or stereotypical judgements. Examples for such biases are causal attribution biases (e.g. tendencies toward other-blaming or self-blaming; exaggerated optimism or pessimism) or an over- or underestimation of one's power or coping potential. Scherer and Brosch (2009) have suggested that cultural goal, belief, and value systems may encourage certain types of appraisal bias and may thus provide an explanation for vestiges of culture-specific emotion dispositions.

If appraisal competence is high and pertinent events are evaluated realistically, appropriate response preparation should normally follow automatically. It seems reasonable to assume that the results of the individual evaluation checks drive changes in the other emotion components, in other words, autonomic physiology, motor expression, and action tendencies (see Scherer 2001, 2009*a*; Chapter 2.1, this volume). Ideally, synchronized response patterning, appropriately shaped by appraisal, should result in the preparation of adaptive action tendencies. However, it is possible that the translation of appraisal results into response patterning and action tendency preparation will malfunction because of 'hardware' problems (e.g. lesions in mediating brain circuits) or biases produced by specific learning histories (e.g. a preponderance of a specific kind of response due to strong reinforcement in the past). Thus, when an individual responds in a seemingly incompetent fashion to emotionally arousing events, it may be necessary to examine the appropriateness of the appraisal mechanism and the way in which appraisal results trigger response patterns separately.

Regulation competence

Virtually all theories of EI or EC assign a central role to emotion regulation ability. One important function of emotion regulation is to correct inappropriate emotional responses that might have been produced by unrealistic appraisals. Often, our social environment will alert us to the fact that an emotional reaction is inappropriate in kind or intensity. Given a certain sluggishness of the response system, especially physiological arousal, emotions cannot be turned on or off like an electric light, and control and management strategies are required. One might think that emotion regulation skills are not needed if one commands exceptional appraisal competence—in that case the emotions triggered by the appraisal results should always be appropriate. However, this is rarely the case. First, fine-tuning is required, as appraisal changes rapidly and abrupt reappraisal as a result of new information requires strong regulation skills. In addition, emotional reactions are subject to strong normative control in most societies. Thus, even though a strong anger reaction to a veiled insult may be an appropriate behaviour preparation in an evolutionary sense, rules of politeness might prohibit such reactions. Many authors have described the existence and operation of display and even feeling rules in different societies (Ekman 1972; Ekman and Friesen 1969; Hochschild 1979; Matsumoto 1990). Although the importance of emotion regulation is often underlined, relatively little is known about the details of the underlying mechanisms (but see the contributions in Philippot and Feldman, 2004). Three components of regulation competence can be distinguished (Scherer 2007).

1 *Monitoring competence,* which consists of: (a) appropriate reflection and integration of all emotion components; (b) balanced conscious and unconscious processing; and (c) accurate proprioceptive feedback of peripheral responses to a central monitoring system and their appropriate interpretation. Basically, the idea is that the processes of cross-modality and temporal integration (of appraisal results and the corresponding response patterns; see Scherer 2004), as well as the interaction between unconscious and conscious processing (see Scherer 2005*a*), can operate in a more or less optimal manner (see Scherer 2007 for examples).

2 *Automatic unconscious regulation*, involving the automatic allocation of attentional resources, is of major importance. Upon detection of potential relevance, the executive space needs to be largely allocated to the further processing of the respective stimulus or event. Individuals differ in the rapidity of reactions, task switching capacity, and parallel processing ability. There are major differences in cognitive ability, specifically with respect to executive processes, which could account for differential competence in automatic regulation. Indeed, there is some evidence in the literature that the automatic regulation of emotion may depend on available executive processing resources (Baumeister 2002; Derryberry and Reed 2002; Van der Linden 2004).

3 *Controlled conscious regulation*. Almost all of the research conducted on emotional regulation to date deals with conscious monitoring and control attempts. In particular, the pioneering work of Gross and his associates (Gross and John 2003; John and Gross 2004) has demonstrated the effects of reaction suppression and cognitive reappraisal. Reaction suppression refers to the attempt to reduce emotional intensity by controlling or suppressing physiological reactivity and overt motor expression, with the effect presumably being due to the diminution of proprioceptive feedback. This explanation is consistent with the claim that subjective feeling is an integration of the complete representation of all component changes, including proprioceptive feedback from the periphery (Scherer, 2004; Chapter 2.1, this volume). If less autonomic arousal and motor expression activity is integrated into the total reflection of component changes, subjective feeling will change qualitatively and quantitatively (in terms of intensity).

The effect of reappraisal was early posited and empirically demonstrated by Lazarus and his collaborators (Lazarus 1968). If the results of appraisal determine the nature of the ensuing emotion, a reappraisal of a central criterion will obviously change the nature of the emotion and consequently of the subjective feeling. Modern componential theories, and particularly the CPM, conceptualize appraisal as a recursive process. In consequence, rather than focusing on single acts of reappraisal, these theorists envisage a constant effort to refine appraisal results and bring them into line with reality. This is achieved by continuous processing of incoming information and continuous search for the most appropriate schemata or criteria for the comparison of currently experienced events and their features to internally stored experiences. The result is a constant change of the qualitative nature and intensity of the resulting emotion (and its subjective experience), something that the notion of emotional states in the sense of a few basic emotions hardly does justice to.

Communication competence

Emotions are often socially shared by expressing them or talking about them (Rimé 2009). We can distinguish two major subcompetences, related to: (1) the *sending or encoding* of information about one's emotional state via appropriate verbal and/or nonverbal expression for optimal impact on a receiver (e.g. an interaction partner) and (2) the ability to *receive or decode* the verbal and/or nonverbal expressions of others in the sense of correctly interpreting these signals and being able to correctly infer the underlying emotional state. The second subcompetence comprises the 'emotion perception competence' component of emotional competence.

Sending (encoding) communication competence

Emotional expression, as an integral part of emotion production, informs interaction partners about the way in which an individual has appraised an action or event, the consequent reaction, and, most importantly, the probability of different behavioural consequences (Scherer 1984, 2001). Obviously, then, it is part of EC to produce emotional expressions that are optimally suited to that purpose. At this point production competence and regulation competence converge. Thus, in many cases regulation competence requires that the automatically produced expressions, e.g. as effects of the ongoing physiological changes, be modified. Thus, it is suboptimal to send inappropriate or ambiguous signals about reactions and action tendencies, as this will encourage misunderstanding and seriously complicate interaction processes. If one produces signs toward a partner that he or she can interpret as an anger reaction even though one is worried about the future of the relationship, one is likely to produce unwanted effects of spiralling anger escalation. In some sense, emotion expression always has a strategic aspect that can be more or less pronounced. Scherer and his collaborators have distinguished *push effects*, which represent the automatic motor consequences of the internal processes, from *pull effects,* which reflect cultural templates of socially desirable or strategically useful expressions (see Chapter 3.2, this volume; Scherer 1985).

The appropriate control of emotional expression has been intensively discussed in the literature. Ekman and Friesen (1969) coined the term 'display rules' to refer to cultural norms that govern the licence to express different emotions in social situations (see also Ekman 1972, 2003*c*; Matsumoto 1990). Clearly, this is a competence that needs to be acquired in the socialization process via which a child becomes a well-functioning member of a particular society. Issues concerning both the understanding and the execution of expression control have consequently been an important part of the literature on the socialization of affect (Ceschi and Scherer 2003; Saarni 1979).

But expression control goes much beyond suppression or modification. Clearly, one needs to add *fabrication*, in other words, showing an emotion one does not feel at all. The strategic use of emotional expression is a central element of emotional skills. A nice example is provided by what may be one of the first formal statements on EC in the history of philosophy and psychology: In his *Nicomachean ethics*, Aristotle (1941) exhorts us to react to an insult with the appropriate amount of anger, at the appropriate time, directed at the appropriate person, in order to avoid being seen as a social fool. The secret is the measured response, avoiding overreacting (to avoid being seen as hysterical or stressed out) or underreacting (being seen as a 'social fool').

The anthropologist Erving Goffman has brilliantly expanded on this important idea by showing the powerful human tendency for positive self-presentation (Goffman 1959). Importantly, emotional expression, even if strategically regulated or manipulated, needs to be credible and convey the impression of being authentic. One of the essential skills in this respect is to produce congruent expression in different modalities, something that requires important skills, given the difficulty of monitoring and manipulating many different cues at the same time In consequence, one would assume that *sending or encoding communication competence*, as an important part of EC, involves being able to produce a skilful blend of push effects, to appear authentic, and pull effects, to conform to norms and pursue one's interactional aims.

Receiving (decoding) communication competence (perception or recognition competence)

In addition to efficient signal production, communication competence requires accurate signal perception and recognition or interpretation (receiving ability). This implies a high ability to recognize emotional states of others in different modalities such as the face, voice, or body, as well as in verbal content, even though the pertinent cues may be controlled or concealed. For example, accurate emotion recognition is important in negotiations to understand when someone gets edgy or irritated to the point that negotiations may break off.

Clearly, individuals differ greatly in this capacity. Not surprisingly, the issue of 'social intelligence', which includes emotion recognition competence, has been early appreciated by the pioneers of intelligence testing, and attempts were made to produce valid tests of these abilities (O'Sullivan and Guilford 1975; Ruisel 1992). Later, the field of nonverbal communication produced a large amount of work on nonverbal sensitivity, which also concerns emotion recognition ability (Hall and Bernieri 2001). It is surprising that current efforts to develop tests of EI either completely ignore this important competence (the personality trait/adjustment approach) or deal with it exclusively from the point of view of socially convergent interpretation (see Scherer 2007). Our group has recently validated a new multimodal performance test of nonverbal recognition ability (MERT; see Bänziger *et al.* 2009).

Emotional competence in virtual agents

Which of these aspects of EC are important for an emotionally competent virtual agent and how can these be implemented? Clearly, this depends on what kind of virtual agent one intends to build. Much of the current effort in affective computing seems to be directed at the development of what I will call a 'service robot', be it in the form of a virtual agent on the screen or a real, physical robot. The EC of such service robots required by the respective applications seems to be limited to the component of EC called *communication competence* above. One widely desired competence is related to the receiving or decoding ability. The robot has to be able to correctly recognize the emotion of the human it is supposed to serve and to adapt its service in consequence, e.g. adopting a soothing attitude upon detecting sadness. Increasingly, sending or encoding competence is equally required. The robot should show the context-appropriate emotion in delivering a specific message, e.g. regret upon having to deny a request, or during a specific service activity, e.g. enjoyment in being able to help. The conceptualization and implementation of such skills are being pursued in many laboratories all over the world, with massive support from industry, as it is hoped that such communication competence will greatly augment the commercial viability of such robots. Obviously, neither the production nor the regulation competence is required as the robot is not supposed to have any emotion. The decoding or receiving competence is limited to successful pattern recognition based on the analysis of spoken utterances or facial or vocal expression configurations. The encoding or sending competence is limited to the operation of pull effects as described above, as the robot has no underlying emotion processes that push out expression. And even these pull effects seem to be limited to the realization of socioculturally desirable or prescribed expression templates, as the robot is unlikely to have his/her own strategic aims that require specific expressions (except possibly those built in by the manufacturer).

While the implementation of such low-level abilities seems rather straightforward, it is fraught with difficulties. In terms of receiving or recognition ability, most current approaches use automatic learning algorithms to acquire prototypical emotion expression patterns (based on the input of stimuli for which the 'ground truth' is provided) in order to detect similar patterns in testing stimuli and classify the underlying emotion accordingly. These approaches consist of *holistic matching* methods as they seem to be based on a match between a particular emotion, conceived of as an invariant unity, and the associated prototypical expression pattern, which is also conceived of as an invariant unity. While such methods enjoy some success in laboratory settings with constrained stimulus material, they do not seem to fare particularly well in more realistic settings. This is not very surprising as there is mounting evidence that neither emotions nor their expressions are invariant unities and that, in consequence, there are no unambiguous prototypical expressions, even for so-called basic emotions, let alone the myriad of so-called complex or of mixed emotions (see Mortillaro, *et al.* in preparation; Scherer 1992, 2001; Scherer and Ellgring 2007*a*). To give a single, but powerful example—one would expect that infants show clear prototypical patterns of very basic emotions like surprise; however, a wealth of empirical evidence shows that this is not the case (Scherer *et al.* 2004*b*). Thus, if there are no prototypical emotion expressions, there are no ground truths that could be learned by automatic algorithms (quite apart from the problems of the choice of a representative corpus). The exception might be expression patterns from the extreme end of the pull effect continuum—emoticons. But even actor portrayals of emotion, which can also be considered to be driven mostly by pull effects (Bänziger and Scherer, submitted), are extremely variable and only rarely show the predicted prototypes, for example, in terms of complete facial action unit configurations (Scherer and Ellgring 2007*a*).

It may well be that holistic matching methods lead us into a blind alley. They are unlikely to produce recognition results in realistic contexts that can be reliably used as a guide to behaviour. One could go even further and claim that there is no ground truth in emotion expression, given the variability and the rapid changeability of emotion processes as well as the tremendous amount of individual differences.

What is the alternative? Clearly, the most appropriate approach would be to teach service robots the same perception mechanism that is used by humans. Humans do not use holistic matching but integrative analytic inference. Specifically, I have suggested that human emotion perception can be best described by the Brunswikian lens model (adapted from Bänziger and Scherer, submitted; see also Scherer 2003). The sender expresses (consciously or unconsciously) underlying emotional states by a vector of distal (objectively measurable) cues in facial, vocal, or bodily behaviour. These cues are transmitted via the appropriate communication channels to a receiver. They may be distorted or weakened in the transmission process. The sensorium and brain association areas of the receiver will represent these cues in a proximal fashion, i.e. as subjective impressions, which may be more or less faithful representations of the distal quality of the cues (due to perception, attention, or short-term memory processes). The receiver then infers or attributes an emotional state to the sender on the basis of stored probabilistic relationships. Both encoding and decoding mechanisms are different for push and pull effects.

How can this model help in our quest to improve decoding competence? One important issue is that the model does not assume correct recognition of the true underlying state. This would require that the distal cues completely map the state, that the transmission is flawless, that the proximal cues are equivalent to the distal ones, and that the inference rules exactly mirror the encoding rules. It is unlikely that this happens very frequently. In addition, emotion is a process, not a state and, in consequence, states, and corresponding emotions, change extremely rapidly. Therefore we might best operationally define receiving competence as the ability to infer and attribute emotion processes in others as well or better than the upper quartile of the population (using a quartile is arbitrary of course; it could also be any top percentile of the distribution). What is essential is the definition of a criterion in terms of reaching a convergent attribution as made by a group with known competence in emotion inference from expressive cues, rather than in terms of recognizing the 'true' emotion of the sender (which the latter may also be partially unaware of; see Scherer 2005*a*).

One might well be able to teach the requisite skills to both humans and virtual agents *if* we had the appropriate empirical data. We can use structural statistical models to represent the process described by the Brunswikian model and estimate its *ecological validity* and the patterns of *cue utilization* by skilled receivers. Such data do not currently exist and unfortunately there is very little research activity in this domain. In addition, the inference and attribution process needs to be broken down further as it is likely that in many cases emotions are indirectly inferred from expressive features that are driven by appraisal and action tendencies. One can hypothesize that receivers first infer these direct causes of the expression and make an emotion attribution on this basis (Scherer 1992, 2003).

What about sending or encoding communication competence? As mentioned above, in the case of service robots, this should be entirely driven by pull effects. If there are no socially desirable or prescribed (or strategic interest) target expressions, no emotional expression should be shown (except possibly noncommittal baseline friendliness). If there are putative sociocultural templates, they should be expressed in a fashion that is as authentic as possible. Again, in order to teach these skills to humans and our service robots we would need empirical data (obtained across cultures and subcultures) concerning the putative templates for different situations and contexts. Obviously, it would have to be certain that the distal cues used in the template are indeed interpreted in the desired way by skilled receivers (using structural modelling based on a Brunswikian approach; see Bänziger and Scherer, submitted). Again, pertinent research hardly exists although the methodology required does not represent any major difficulties.

Differential utility of emotion theories for dynamic modelling

Let us now turn towards more lofty aims—the implementation of autonomous virtual agents that are actually capable of having emotions, albeit only virtual or artificial ones. Clearly, if these agents are to be emotionally competent, the complete list of competencies outlined above would be required. These skills would have to be implemented as part of a complete emotional architecture of the agent. In consequence, it is essential to decide on the definition of emotion one wants to adopt (as there are unfortunately many different ways of defining an emotion; Scherer 2005*a*), specifying its nature and function, and choosing a theory of emotion that allows computer implementation. In the following section we will examine the different contenders in terms of their utility for the purpose at hand.

Psychological theories of emotion differ with respect to their assumptions on how the emotion components—cognitive processes, peripheral physiological responses, motivational changes, ED: Pls check and confirm S ok here?

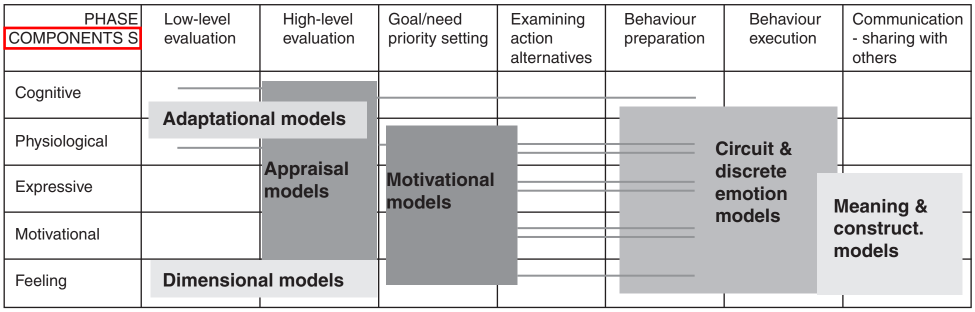

Fig. 1.1.1 Mapping competing emotion theories in a space defined by phases of the emotion process and type of emotion component. (Reproduced from Scherer and Peper 2001.)

motor expression, and subjective feeling—are integrated and, particularly, how qualitatively different emotional states are to be differentiated with respect to their patterning. They also differ in their focus on specific stages of the process. Therefore, psychological models of emotion are difficult to assess in terms of their merits for research on, and development of, emotionally competent agents. This certainly makes it difficult for computational modellers to decide which model is the most appropriate for a particular research question or a particular application.

We will now review the major 'families' of emotion theories that are proposed in the literature (for further details, see Moors 2009; Scherer 2000*a*, 2009*c*) and that can be considered as competitors for the modeller's choice. The purpose is to unravel some of the different strands pursued by different theorists and to place the different models on a map that can serve for navigation. One can use a two-dimensional coordinate system to plot different psychological theories of emotion (see Scherer and Peper 2001), as illustrated in Figure 1.1.1. One dimension consists of the different components of the emotion, as described above: cognitive processes, peripheral physiological responses, motivational changes, motor expression, and subjective feeling. The second dimension consists of different phases of the emotion process and its consequences: low-level evaluation, high-level evaluation, goal/need priority setting, examining action alternatives, behaviour preparation, behaviour execution, and communication/sharing. Each of the different families of theories described below is marked by a box situated in the region that represents the major preoccupation of each theory with respect to phases and components. Lines emanating horizontally from some theories indicate that the theory also treats other phases in connection with its focus. In what follows, a brief synthesis of the major families of theories is provided (following Scherer and Peper 2001).

Adaptation theories

Under this heading, one can group theorists who emphasize that the emotion system has an important adaptive function, and is primed by evolution to detect stimuli that are vitally significant for the organism's well-being. Öhman (1987) suggests that organisms are evolutionarily 'prepared' for the evaluation of certain contingencies, allowing the detection of threat stimuli in a pre-attentive mode and preparing appropriate physiological orienting or defence reactions. LeDoux (1996) also highlights pre-attentive emotion elicitation and postulates direct projections within the brain from the sensorium and thalamus to the amygdala, which in turn trigger rudimentary viscero-motor and behavioural responses. Both theorists acknowledge that there is a second phase, characterized by higher-level attention-driven evaluation. Because of the emphasis on biologically prepared, pre-attentive processes, both theorists focus on fear-inducing stimuli, such as electric shocks, spiders, or snakes, and the resulting emotion of fear, and were able to confirm their hypotheses in several empirical studies.

Dimensional theories

Emotions can be easily differentiated by their position on a pleasantness–unpleasantness (or valence) and an arousal (or activation) dimension (varying on a continuum from active to passive). This allows one to distinguish between negative and positive emotions of different degrees of intensity and arousal. Dimensional theories have been very popular because of the economical fashion in which they capture valence and arousal differences between emotions. Like Wundt's (1905) pioneering work (who, in addition to pleasantness–unpleasantness and rest–activation, suggested a third dimension, relaxation–attention), most dimensional models focus on the 'subjective feeling' component of emotion. In consequence, much of the research in this tradition uses verbal labels (Davitz 1969; Russell 1980, 1983), in ways that are quite comparable to the earlier three-dimensional model of verbal meaning (Osgood, *et al.* 1957). A large number of factor-analytic studies have supported the fact that verbal labels can be very reliably mapped into a valence by arousal space (Barrett and Russell 2009).

Appraisal theories

Appraisal theories posit that most (but not all) emotions are elicited by a cognitive (but not necessarily conscious or controlled) evaluation of antecedent situations and events (see Ellsworth and Scherer 2003; Scherer 1999*a*; Scherer, *et al.* 2001 for overviews) and that the patterning of the reactions in the different response domains is driven by the results of this evaluation process. Arnold (1960) and Lazarus (1968) pioneered the explicit assumption that subjective appraisal, specifically the evaluation of the significance of an event for the organism and its ability to cope with the event, determines the nature of the respective emotion. Appraisal theorists following this tradition have refined the conceptualization of appraisal (see Ellsworth and Scherer 2003; Roseman and Smith 2001; Scherer 1999*a*). Thus, at one extreme, Lazarus (1991) has postulated a theme-based appraisal reminiscent of discrete emotion theories. The component process model proposed by Scherer (1984, 2001, 2009*c*), which assumes that there are as many different emotional states as there are differential patterns of appraisal results, is located at the other extreme of appraisal approaches. Intermediate positions are represented by appraisal theorists such as Ellsworth (1991; Smith and Ellsworth 1985), Smith (1989; Smith and Kirby 2001), Roseman (1984; Roseman, *et al.* 1994), Frijda (1986, 2007), and Weiner (1985). These theorists take a more eclectic view of the issue concerning the number and the 'basicness' of emotions. However, they all propose that a specific set of cognitive appraisal or evaluation dimensions or criteria allows us to predict which type of emotion will be experienced by an individual on the basis of the results of the appraisal process.

Contrary to other models, appraisal theories render the link between elicitation of emotion and response patterning more explicit. While dimensional emotion theorists doubt the existence of differential emotion patterning (see Barrett 2006) or reduce response specificity to neurophysiological circuits or programs (as in the discrete emotion and circuit models; see Panksepp 1998*a*), componential appraisal theorists make detailed predictions as to specific physiological, expressive, and motivational changes expected to be driven by appraisal results (Smith 1989; Smith and Scott 1997; Scherer 1984, 1986, 2001, 2009*c*). This is justified by the assumption that the evaluation or appraisal of an event will lead to specific requirements for further information processing, or to specific response or action tendencies, which are in turn determined by the motivational tendency suggested as an adaptive response by the appraisal outcome. In consequence, appraisal theories include all components and phases of the emotion process shown in Figure 1.1.1, suggesting a central determining role for the results of the evaluation on lower and higher cognitive levels. As will be argued below, this aspect makes appraisal theories ideal candidates for computational modelling.

Over the last 20 years numerous empirical studies based on appraisal theories have been conducted and substantial experimental evidence for many of the predictions has been published. For an overview, the reader is referred to reviews in a volume on appraisal theories (Scherer, Schorr, and Johnstone 2001, in particular, chapters by Johnstone, van Reekum, and Scherer; Kaiser and Wehrle; Pecchinenda, Roseman, and Smith; Smith and Kirby; Scherer). In addition to using self-report, much of the work has made extensive use of objectively measured indicators of appraisal processes such as physiological parameters and expressive behaviour. Thus, several studies have demonstrated the efferent effects of appraisal checks on somatovisceral changes and motor expression as markers of appraisal results (Aue and Scherer 2008; Aue, *et al.* 2007; Banse and Scherer 1996; Johnstone, *et al.* 2005; Johnstone, *et al.* 2007; Scherer and Ellgring, 2007*a*; Van Reekum, *et al.* 2004) and evidence for the sequential processing of appraisal check predicted by the CPM (Aue, *et al.* 2007; Lanctôt and Hess 2007; Delplanque, *et al.* 2009; Flykt, *et al.* 2009; Grandjean and Scherer 2008). Preliminary evidence on the synchronization predicted for felt emotional experiences has also been reported(Dan Glauser and Scherer 2008).

Motivational theories

The close relationship between emotional and motivational phenomena is often neglected (Lazarus, *et al.* 1982). Some theorists, however, base their models centrally on this relationship. One of the oldest theories in this group is that of Plutchik (1980), who has argued that the major types of emotions can be derived from evolutionary continuous motivational primitives. Thus love is seen as a correlate of parental care, fear as a signal of danger inducing a flight response, and anger as an antagonistic emotion that prepares the organism to fight. Many other emotion theories postulate similar motivational underpinnings (see also Frijda 1986). But Plutchik's theory is a special case in that it bases emotion classification directly on fundamental kinds of psychobiological motivation. Another motivation-based account has been provided by Buck (1985), who views emotions as 'read-outs' of motivational tendencies. There has been relatively little experimental work in this tradition.

Circuit theories

Psychological theories of emotion have continually been influenced by evolving neuroscientific knowledge of pathways or circuits in the brain. Such approaches attempt to use evidence from functional neuroanatomy in order to understand emotion elicitation and differentiation in a comparative perspective. Emotion networks had already been described by the pioneers of affective neuroscience. Prominent advocates of this tradition have been Gray (1990) and Panksepp (1998*a*). These models are all based on the assumption that the differentiation and the number of fundamental emotions are determined by genetically coded neural circuits. They have stimulated a considerable amount of neuropsychological and psychophysiological research with human subjects.

Discrete (or basic) emotion theories

The most popular theoretical accounts of emotion are based on Darwin's (1872/1998) influential book *The expression of emotion in man and the animals*. These theories claim the existence of a limited number of basic or fundamental emotions such as anger, fear, joy, sadness, and disgust. These models can be located close to the end of the emotion process shown in Figure 1.1.1, the differentiation being mostly explained by patterning of effector mechanisms in behaviour preparation or execution (similar to dimensional and circuit models). Following Darwin, theorists in this tradition suggest that, during the course of evolution, a limited number, generally between 7 and 14, of basic or fundamental emotions have evolved. Each of these basic emotions has its own eliciting conditions and its own physiological, expressive, and behavioural reaction patterns.

Discrete emotion theory was pioneered by Tomkins (1962) who argued that a number of basic or fundamental emotions could be conceived of as phylogenetically stable neuromotor programs. These programs automatically trigger a pattern of reactions ranging from peripheral physiological responses to muscular innervation, particularly in the face. This tradition has been most strongly developed by Ekman and Izard, who extended Tomkins' theory. They described the discrete patterning of universal, prototypical facial expressions for a number of basic emotions (Ekman 1972, 1992, 2003*c*; Ekman and Rosenberg 2005; Izard 1977, 1992; Levenson, *et al.* 1992). A patterning of autonomic–endocrine reactions has also been suggested (Levenson, *et al.* 1990).

Lexical theories

The richness of emotion terms in most languages has given rise to a number of psychological and philosophical models of emotion. One of the basic assumptions of these approaches seems to be that semantic structure will point the theoretician to the underlying organization and determinants of the emotion domain. Thus, Oatley and Johnson-Laird (1996) focus on goal structures implied by major emotion terms. Ortony, *et al.* (1988) provide a theoretical analysis of the semantic implicational structure underlying major emotion words. Shaver, *et al.* (1987) use cluster analysis to show the hierarchical meaning structure of the emotion lexicon. Only a small number of experimental studies have examined these theories.

Social constructivist theories

An approach to defining emotion that is favoured by sociologists and anthropologists suggests that the meaning of emotion is mostly constituted by socioculturally determined behaviour and value patterns (Averill 1980; Shweder 1993). Advocates of this approach consider the psychobiological reaction components of emotion as secondary to the meaning of the emotion in a specific sociocultural context. Often, theorists in this tradition also consider the emotion labels available in a language as indicative of the emotional meaning structures in the respective culture (Lutz and White 1986). Both the lexical and the constructivist theories focus on the final phase of an emotion process: the communication or the sharing of the emotional experience with the social environment. This places heavy, if not exclusive, emphasis on the subjective feeling component. Much of the evidence is based on field work.

Overlap

As is easily seen from the preceding discussion, there is quite a bit of overlap between these traditions. Adaptation and motivation theories are quite compatible with appraisal theories. Similarly, circuit and discrete (basic) emotion theories share many assumptions. These two groups differ mostly with respect to the focus on different phases in the emotion process. Whereas circuit and discrete (basic) emotion theories focus on the response end, assuming specific patterning (elicited by typical situations), appraisal theories (and adaptation and motivation theories) focus on the elicitation and evaluation phase at the beginning of the process, assuming that responses are driven directly by the results. Lexical, dimensional, and constructivist theories focus on a still later stage, that of categorization and labelling in the service of communication. Based on this comparative evaluation, Scherer (2009*c*) has proposed that three large families of theories be distinguished:

- appraisal theories in the widest sense, based on the writings of many philosophers and psychologists, and comprising theories focusing on adaptation and motivation;
- basic (or discrete) emotion theories, based on Darwin, including circuit theories;
- constructivist theories, loosely based on James (1890/1898) and Schachter and Singer (1962) on the one hand and cultural relativism on the other, comprising dimensional, social constructivist, and lexical theories.

Choosing a theory as the basis for computational modelling of emotion

The preceding review of the competing models suggests that there is reasonable convergence on the view that emotion is considered by most theorists as a *bounded episode* in the life of an organism that is characterized as an *emergent pattern* of *component synchronization* preparing adaptive *action tendencies* to relevant events as defined by their *behavioural meaning* and seeking *control precedence* over behaviour (see also Frijda and Scherer 2009).

Each of the theoretical models described above captures and explains important facets of the emotion phenomenon thus defined. As illustrated by the structural decomposition in Figure 1.1.2, it is essential to determine exactly which of the many aspects of the emotion process are highlighted by the various theories, and to what extent they can be mapped on to each other. As shown above, many of the models can be integrated, if one assumes that different models describe different components and phases of the emotion phenomenon (see also Scherer 2000*a*). For example, one can argue that the valence and arousal dimensions represent a higher-order

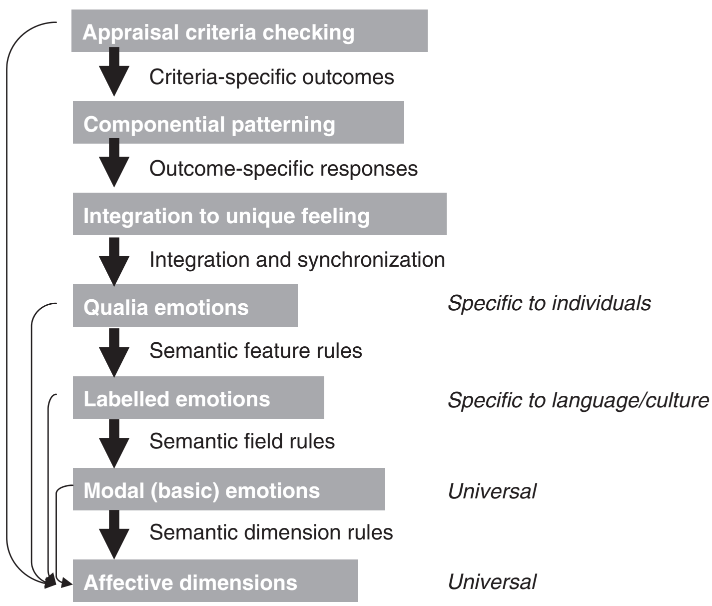

Fig. 1.1.2 The hierarchy of levels of description for emotion processes and their mapping into lower-dimensional space.

factor space into which the so-called discrete or basic emotions can be plotted. Similarly, the basic emotion 'families' (see Ekman 1992) can be seen as higher-order factors with respect to the highly variable outcomes of appraisal processes. Scherer has proposed the concept of *modal emotions* (1984, 1994*b*) to account for the existence of a limited number of such 'families', referring to frequently occurring patterns of appraisal of events that are universally encountered by individuals, such as sadness in the case of loss, or anger in the case of blocked goals. These common elements account for the fact that languages group these states together, using a single label. One may need to go to an even lower level to identify individual emotion family members that share common appraisal profiles (characterized by brief expressions, such as 'righteous anger'). The lowest level might consist of the continuous adaptive changes that—according to component-process theory—are produced by single appraisals. Examples are the startle as well as defence and orienting responses (which may be a part of a higher-order emotion such as surprise or fear). Figure 1.1.2 shows these different levels as well as the mechanisms that seem to underlie the grouping of lower-order units on a higher level. The figure also shows predictions as to which mechanisms are likely to be specific to individuals, to language or culture, or universally shared.

It seems reasonable to start with the most comprehensive and most detailed theoretical approach, and examine how this high-dimensional set of information can be mapped into a lower-dimensional space. It can be argued that the componential appraisal approaches, briefly reviewed above, represent the most comprehensive and detailed attempt to model emotion. This is due to the fact that most of them provide a detailed account of the elicitation mechanisms that produce differentiated emotions (i.e. appraisal criteria checking), predict concrete response patterns based on these appraisal profiles, and consider the construction of subjective feeling as based on these processes. In contrast, most other theories focus on higher levels of aggregation.

What are the elements that would be needed to construct a virtual agent with the ability to respond with something approaching human emotion? If one knows exactly what level of resolution one needs in an emotionally competent agent it is clearly appropriate to choose the model that provides the most economical solution. It is rarely useful to gather a lot of information that is never used thereafter. It is much easier to represent emotion in a two-dimensional space representing pleasantness and arousal, or to use one of a finite number of basic emotions, than to use detailed information on appraisal and response patterning. However, in choosing such an economical model, developers of affective agents need to be aware of their needs, given the design specification of the agents to be built. In what follows, a non-exhaustive and non-systematic list of the requirements for constructing a process model of human emotions is reviewed, and the usefulness of the major theoretical models to fulfil the respective requirements is evaluated.

- *Number of components*. As shown above, despite the consensus that emotions have several components, which interact with one other, many theories emphasize particular components and neglect others (see Figure 1.1.2). If computational modelling or agent implementation requires the participation and interaction of several components, care should be taken to choose a guiding emotion model that provides specifications for this essential feature.
- *Relevant events*. Generally, events or situations are seen as elicitors of emotion episodes. Appraisal theorists assume that it is not the event itself, but the appraisal by the individual, that is decisive and that may change over time, in the course of reappraisal. Constructivist theories do not clearly specify how events affect continuous core affect. According to Russell (2003) individuals may attribute a certain core affect to an event.
- *Behavioural meaning*. Appraisal theorists assume that the transactional evaluation of the event constitutes the behavioural meaning for the individual, insisting on the fact that it is only through the specific behavioural meaning of an event for an individual that the action

preparation following the appraisal process can have adaptive value. This is not a meaningful feature for basic or constructivist theories. The former take the type of event as the discriminating factor; the latter see categorization and conceptualization of core affect as independent from event evaluation. In consequence, it would seem difficult to model the assumptions of basic or constructivist theories into an agent model as the former would require that specific affect programs for a large number of events would need to be built into the model (which would make the model circular and trivial) and the latter implies that the behavioural meaning of an event can only be understood in an ideographic fashion, which would require the modelling of all possible individual differences and situational effects.

- *Adaptive responses*. Most emotion theories assume some degree of functionality of emotion (Nesse 2009). For basic emotion theories, the affect program is pre-programmed to deal with the eliciting event. In contrast, appraisal theories define the adaptive functions in terms of the efferent results of individual appraisal checks that add up cumulatively to prepare appropriate action tendencies (Ellsworth and Scherer 2003; Scherer 2001). Constructivist theories generally endorse the adaptive value of emotion but there is no justification for this claim in terms of the postulated architecture and the criteria for functionality are not defined (see Barrett 2006). However, the modelling of adaptive function seems essential for the attempt to create an emotion architecture for an autonomous virtual agent as adaptation to a specific environment and current goal states is of central importance.
- *Component synchronization*. While the componential architecture of emotions is generally admitted, only some appraisal theories, in particular the CPM (Scherer 2004, 2005*a*, *b*; see Chapter 2.1, this volume), strongly insist on a process of synchronization and desynchronization of components within the bounded episode, to the point of making the degree of coherence a central criterion for the existence of an emotion (Scherer 2005*b*; Dan Glauser and Scherer 2008). The synchronization assumption can be considered a major advantage for computational implementation as it allows the building of some degree of coherence between different response modalities into the model, which should add to stability.
- *Process modelling and dynamic change*. Despite the general acceptance of the notion that emotion is a dynamic process, most emotion theories deal with discrete, unchanging emotional states and do not specify mechanisms that allow analysing or modelling dynamic change over time. While it is simpler and more economical to restrict modelling to highly circumscribed discrete states, a realistic computational model of emotion requires true process modelling of dynamic change over time with varying inputs. Only componential appraisal theories provide the tools for such an endeavour. While modern constructivist theories (Russell 2003) also propose continuously varying core affect, this is restricted to valence by arousal and no determining input factors are specified.
- *Emergent properties*. This refers to unexpected features of the process that may occur and influence the overall process at different levels. The emotion process is intrinsically dynamic and probably nonlinear, and the linear perspective often taken by theorists and modellers may be insufficient to describe its complexity (Scherer 2000*b*, 2009*b*; Sander, *et al.* 2005). For instance, eliciting events (input) are not limited to prototypical categories or predetermined locations in a valence × arousal space, and the system should not be equated to simple lookup functions with predetermined databases. Similarly, on the output side, care should taken to model response patterning in detail, instead of postulating a small number of prototypical affect programs with a single output. The basic emotion model is deterministic on a macro level—a given stimulus or event will determine the occurrence of one of the basic emotions (through a process of largely automatic appraisal). In contrast, appraisal theories are

deterministic on a micro level—specific appraisal results or combinations thereof are expected to determine, in a more molecular fashion, specific action tendencies and the corresponding physiological and motor responses. Thus, appraisal theorists espouse *emergentism* assuming that the combination of appraisal elements in a recursive process is unfolding over time and that the ensuing reactions will form emergent emotions that are more than the sum of their constituents and more than instantiations of rigid categories, namely, unique emotional experiences in the form of qualia (see Scherer 2004, 2009*a*).

- *Bounded episode*. Consistent with popular assumptions, both basic and appraisal theories consider emotions as bounded episodes in time, having a clear onset and a somewhat fuzzy offset. In contrast, constructivist theories assume that the stream of continuously varying core affect is segmented only by mental categorization and conceptualization. The latter are seen to depend entirely on the individual's construction and will thus vary widely over individuals, which makes it impossible to build a nomothetic model that can be applied, in a lawful manner, to different individuals.
- *Number and type of emotions*. On one extreme, we find the notion of a limited number of evolutionarily continuous adaptive emotion systems (held by many basic emotion theorists) and, on the other, that of fuzzy, unpredictable state changes that achieve coherence only by their place in a valence/arousal space and by conceptual classification, espoused by some constructivists. In this debate, appraisal theorists are somewhere in the middle—they neither accept the idea of a limited repertoire of basic, homogeneous emotions with highly prototypical characteristics nor that of emotions being individually labelled points in two-dimensional affect space. Rather, while assuming that there are widely varying types of emotions, they postulate the existence of modal emotion families (Scherer 1994) with frequently occurring appraisal profiles that have adaptive functions in dealing with quintessential contingencies in animal and human life. The insistence of constructivist theorists on individual and situational differences and the absence of predictions do not predestine these theories as guides for model building. Basic emotion theories have the disadvantage that they describe only a few emotions in detail, some of which may not be too useful for modelling in agents (e.g. disgust). In contrast, appraisal theorists allow much more flexibility, including predictions for emotional states that have no standard linguistic labels but can be distinguished on the basis of appraisal-driven response patterns. This may permit the modelling of emotions that are quite specific to agents in a specific environment.
- *Variations of intensity and duration*. Regrettably, there has been little interest in the study of emotion intensity and duration (but see Edwards 1998; Frijda, *et al.* 1991; Sonnemans and Frijda 1994). In general, emotion researchers working with experimental subjects have recorded self-reported intensity, and sometimes duration, but have rarely used this information in their theorizing. However, in animating an affective agent, the issue of intensity and duration of emotional expressions is crucial. In consequence, care should be taken to choose a guiding model that makes predictions concerning differences in intensity or duration of emotion processes. The guiding theory should at least take account of major differences between members of emotion families such as hot versus cold anger; anxiety versus fear, sadness versus despair, happiness versus elation, etc. (see Banse and Scherer 1996). Currently, all models of emotion are underspecified in this respect, although some, like appraisal models, are more suited to deal with such variations because of their more fine-grained structures.
- *Qualia differences*. Specifications of discrimination ability vary for models and agents. In some cases, only the most rudimentary distinctions need to be made, for example, absence or presence of a particular emotion (anger in a client's voice, or the presence of general negative arousal).

In others, much finer discrimination is required. Clearly, the choice of model is based on this fundamental requirement. One of the issues of concern should be the capacity of the model to evolve. In many cases, one tends to start with a simple undifferentiated model but quickly sees the need for further development and differentiation. As shown above, a forward mapping from complex, high-information, lower-order models to simpler, low-dimensional, higher-order models is always possible, while the reverse is not true. Thus starting with a simple model allowing little differentiation may impede further development.

- *Response patterning*. Virtual agents are becoming increasingly sophisticated, and it is predictable that ever more emphasis will be placed on affective agents that have a highly differentiated repertoire of response patterns in several modalities. Currently, much effort is being made to improve facial and to some extent vocal expression. Interest is also developing in more natural animation of gestures and body movements. It is to be expected that interest in modelling brain and peripheral physiological responses will increase in the future. In consequence, the choice of guiding models should be informed by the capacity of different models to predict complex dynamic processes in response patterning. It would seem to be in the interest of flexibility and realism that this should occur in a molecular and emergent fashion rather than in a holistic and deterministic one.
- *Moving back and forth between levels*. As shown above, different emotion theories focus on different levels, dealing with phenomena of lower or higher order. As we have suggested, mapping between theories is possible in one direction, but once information about a lower level is lost, it cannot be retrieved. In consequence, it is important to determine to what extent computational modelling or implementation of virtual agents requires moving back and forth between levels, or whether it is sufficient to remain on a single level of analysis and synthesis. The emotion models presented here vary with respect to affording this possibility.

Conclusion

Stepping back to match these requirements against the many different emotion theories described above, it is easy to see that some theories are better suited to address certain of these requirements than others. Basic emotion theories have been (and still are to some extent) the models of choice in computer sciences and engineering. However, as shown above, if one accepts the central features of emotion outlined above, they do not fare so well both from a point of mapping theory to underlying processes and with respect to the specification of mechanisms that allow the modelling of the essentially emergent nature of dynamic emotion processes. In addition, the notion of fairly rigid affect programs for a small number of basic emotions seriously limits the construction of open, emergent architectures. If the aim is to compute emotion, in terms of a multicomponential process over time, rather than constructively assigned labels of emotion, likely to vary greatly from one individual to another in a rather unpredictable way, constructivist theories need to be discarded (especially as the determining factors are underspecified and precise predictions of mechanisms are absent).

The review of theories above suggests that appraisal theories of emotion constitute the most comprehensive way to represent the complexity of the emotion process, spanning the whole gamut from low-level appraisals of the eliciting event to high-level influence over behaviour. In addition, they present specific hypotheses for the underlying mechanisms that have received consistent support in experimental research. They may thus be the theories of choice for designing adaptive and evolving systems in complex environments as well as for experimental exploration of the emotion mechanism in virtual reality. First attempts to construct partial models based on appraisal theory have yielded promising results (Gratch and Marsella 2004*a*,*b*; Marsella and Gratch 2009; Scherer 1993; Wehrle and Scherer 2001).

To assure a sufficient degree of ecological validity and flexibility, it is essential to base computational models on the most recent insights concerning the architecture of the emotion process and the essential role of dynamic change. In consequence, modellers should define clearly the aims of their simulations against the requirements described above before choosing a specific emotion theory to guide their work.