Advocating a Componential Appraisal Model to Guide Emotion Recognition

Mortillaro, Marcello; Meuleman, Ben; Scherer, Klaus R.

How to cite

MORTILLARO, Marcello, MEULEMAN, Ben, SCHERER, Klaus R. Advocating a Componential Appraisal Model to Guide Emotion Recognition. In: International Journal of Synthetic Emotions, 2012, vol. 3, n° 1, p. 18–32. doi: 10.4018/jse.2012010102

This publication URL: https://archive-ouverte.unige.ch/unige:97219

Publication DOI: 10.4018/jse.2012010102

Advocating a Componential Appraisal Model to Guide Emotion Recognition

Marcello Mortillaro, Swiss Center for Affective Sciences–University of Geneva, Switzerland

Ben Meuleman, Swiss Center for Affective Sciences–University of Geneva, Switzerland

Klaus R. Scherer, Swiss Center for Affective Sciences–University of Geneva, Switzerland

ABSTRACT

Most models of automatic emotion recognition use a discrete perspective and a black-box approach, i.e., they output an emotion label chosen from a limited pool of candidate terms, on the basis of purely statistical methods. Although these models are successful in emotion classification, a number of practical and theoretical drawbacks limit the range of possible applications. In this paper, the authors suggest the adoption of an appraisal perspective in modeling emotion recognition. The authors propose to use appraisals as an intermediate layer between expressive features (input) and emotion labeling (output). The model would then be made of two parts: first, expressive features would be used to estimate appraisals; second, resulting appraisals would be used to predict an emotion label. While the second part of the model has already been the object of several studies, the first is unexplored. The authors argue that this model should be built on the basis of both theoretical predictions and empirical results about the link between specific appraisals and expressive features. For this purpose, the authors suggest to use the component process model of emotion, which includes detailed predictions of efferent effects of appraisals on facial expression, voice, and body movements.

Keywords: *Appraisal, Automatic Recognition System, Computational Model of Emotion, Emotion Production, Emotion Recognition*

INTRODUCTION

Research on the facial expression of emotion dramatically improved in the quality of experimental design and the cumulativeness of findings when the Facial Action Coding System (Ekman & Friesen, 1978), a comprehensive coding scheme that describes facial expressions on the basis of the activity of single facial muscles (Action Units), started to be widely used. The clear and objective definition of Action Units solved the problem of comparing results between different studies and allowed research to focus on the theoretical predictions generated by underlying models of expression.

A similar evolution is happening in the field of affective computing. Several groups are now working on the development of systems that automatically recognize action units (Valstar,

DOI: 10.4018/jse.2012010102

Mehu, Pantic, & Scherer, in press). Current results are promising and we can expect that in the near future these systems will become fully reliable and perform in a satisfactory way. As the detection problem is getting solved, attention should now focus on what is the best model to attribute an emotional meaning[1]. Indeed, emotion recognition systems can be conceived as made of two parts, a detection component and an inference component. The detection component performs the analysis of the facial movements; the inference component outputs the attribution of an emotional meaning to the movements detected by the first component. While for the detection component there is one recognized standard (FACS), for the inference component we have to turn to emotion psychology where multiple theoretical models currently co-exist. Most researchers in affective computing choose a pragmatic approach and avoid theoretical controversies, but every system necessarily implies theoretical assumptions (Calvo & D'Mello, 2010). In the next paragraphs we will first present different theoretical models of emotion and discuss their use for automatic emotion recognition. The goal is not to provide an exhaustive review of the available systems, but rather a brief description of the pros and cons of each choice. We will then introduce a specific componential appraisal model of emotion that has already been used in computational models of emotion production and discuss the possibility of using this model for emotion recognition.

THREE THEORETICAL PERSPECTIVES ON EMOTION

Most research on emotion expression implicitly or explicitly used a discrete emotion perspective (Scherer, Clark-Polner, & Mortillaro, 2011) and the same is true for automatic emotion recognition systems. Discrete emotion theory has been formulated on the basis of findings concerning few intense emotions –called basic emotions – that are expected to have prototypical facial expressions and emotion-specific physiological signatures (Ekman, 1992, 1999; Ekman, Levenson, & Friesen, 1983; Ekman, Sorenson, & Friesen, 1969). This theory dominated the field for decades and it is still the most widely used. There is robust evidence about the existence of some facial configurations that are cross-culturally labeled with the same emotion terms. However, several studies show that people frequently report the experience of emotional states that are not part of this set of basic emotions (Scherer & Ceschi, 2000; Scherer, Wranik, Sangsue, Tran, & Scherer, 2004), and, more importantly, that for spontaneous and enacted emotional expression, these complete prototypical expressions rarely occur (Naab & Russell, 2007; Russell & Fernandez-Dols, 1997; Scherer & Ellgring, 2007a). Discrete emotion theorists tried to solve these problems by suggesting the concept of emotion families (Ekman, 1992). An emotion family includes several lexically marked variations of basic emotions labels; all terms within a family share a common theme (characteristics unique to the family) and variations due to individual, cultural, and contextual factors. Nevertheless, it is not clear how these variations should be modeled and whether they can be identified from vocal or facial expressions.

Automatic recognition systems that are based on the detection of prototypical expressions (by means of the detection of either single AUs or global facial configuration) are extremely successful with intense and posed expressions – i.e., expressions carefully built on the basis of Ekman's predictions, e.g., those included in the Cohn-Kanade database (Kanade, Cohn, & Tian, 2000) or in the JACFEE (Biehl et al., 1997). Most advanced methods are also extremely accurate at distinguishing other types of emotion expressions, including expressions that are not prototypical (Valstar et al., in press). These technologically impressive advancements still cannot solve a number of theoretical and methodological issues. First, strictly speaking, systems based on the detection of global configurations of AUs are capable of recognizing expressions, not the emotions. The "emotional validity" of these expressions relies completely on basic emotion theory. Second, although these prototypes have a certain amount of heuristic utility, spontaneous emotional expressions are rather different from the prototypes, and expressions are much more subtle and less differentiated (Gunes & Pantic, 2010; Zeng et al., 2007). Third, systems that use these predictions to infer an emotion state from a series of detected AUs can account only for very few, generally six to seven, basic emotions. For emotions outside of this restricted class, and for mixed or blended emotions, it is likely that multiple expression configurations are involved, making the search for discrete expressive signatures a rather unsuccessful exercise.

Dimensional models of emotion provide an alternative theoretical framework. These models define emotions as states that can be represented on a common multidimensional space. Dimensional models, first demonstrated in the form of emotional connotations of words (Osgood, 1962, 1964) and subsequently in the structural organization of self-reported affective states (Russell & Mehrabian, 1974, 1977), have also been suggested as a framework for emotion expression (Russell, Bachorowsky, & Fernandez-Dols, 2003). The original models included three dimensions: pleasure, arousal, and dominance. Pleasure (or valence) refers to the hedonic quality of the emotion – positive or negative; arousal refers to the physical activation of the organism; dominance (or power) refers to the degree of control that the person has in the situation. More recent versions propose a bi-dimensional space organized along the axes of valence and arousal and suggest that the subjective feeling of an emotion is the result of an interaction between *core affect* (i.e., the position in the valence per arousal space) and a cognitive component such as interpretation or attribution (Russell, 2003). On one hand, this approach has the merit of reducing the complexity of the emotion recognition task, making it easier to attribute a global affective meaning (positive or negative) to subtle expressions. On the other hand, two or three dimensions seem too limited to capture the complexity of the emotion space and to explain the wide variety of individuals' subjective feelings: for example, by using only two dimensions, it is difficult to distinguish between emotions like anger and fear, since both are characterized by negative valence and high arousal. Furthermore, it is not clear how to set the boundaries between emotion labels in the multidimensional continuous space. So far, there has been little empirical research on emotion expression using this approach (Russell & Fernandez-Dols, 1997; Russell et al., 2003; Russell & Bullock, 1985).

Systems for the automatic recognition of emotion based on dimensional models of emotion started to be used only recently. The most challenging problem in this framework is the mapping of emotion labels onto the continuous multidimensional space. Most systems simplified this problem to two separate binary decisions (high vs. low arousal; positive vs. negative), but this choice cannot be considered satisfactory as it ultimately returns back to the discrete approach while at the same time losing the better descriptive quality that is afforded by a direct use of emotion terms. It is evident that this approach is still in its pioneering state, but it is hard to see how the problem of labeling can be satisfactorily solved in this fashion. There is need for more theoretical reasoning and basic psychological research to fill in the gap between dimensional representation of emotion expressions and emotion labeling. This is not to deny that dimensional models may be very useful in applications that require a simple evaluation of valence and arousal and do not require precise specification of the underlying emotion through classification.

Appraisal models are a third alternative perspective on emotion: They combine elements of dimensional models – emotions as emergent results of underlying dimensions – with elements of discrete theories – emotions have different subjective qualities – and add a definition of the cognitive mechanisms at the basis of the emotion. Starting from the original suggestion of Arnold (1960) - who defined appraisal as a direct, immediate and intuitive evaluation able to distinguish qualitatively among different emotions - appraisal theorists argue that the experience of an emotion is determined by a series of cognitive evaluations on different levels of processing (for a review, see Ellsworth & Scherer, 2003; Scherer, Schorr, & Johnstone, 2001) and it is the specific set of responses in each individual case that determines which emotion will be experienced (Scherer, 2009a). Appraisal theories offer a more flexible framework than discrete and dimensional models, being able to account for individual differences and variations of responses to the same stimulus by the same individual at two different moments in time (Roseman & Smith, 2001), as well as for some cultural differences (for example in the form of appraisal biases; Scherer & Brosch, 2009). Appraisal theories do not generally assume the existence of a one-to-one relationship between a situation and a response or between a single appraisal and a specific emotion (Nezlek, Vansteelandt, van Mechelen, & Kuppens, 2008). It is the pattern of appraisals that determines the emotion experienced (Frijda, 1986; Lazarus, 1991; Ortony, Clore, & Collins, 1988; Scherer, 1984, 2001, 2009a). Appraisal theories attempt to explain the differentiation of emotion states by different configurations of the underlying appraisal dimensions and try to map emotion labels on this multidimensional space. Some authors advanced predictions about the typical appraisal structure of the most frequently occurring emotions. So, for example, the event that causes fear would be typically appraised as unexpected, unpleasant, obstructive to personal goals, and hard to cope with (Scherer, 1994). Since the foundation of affective computing, emotion modeling has preferentially looked at appraisal models of emotion (with the OCC model being the most frequently used, Ortony et al., 1988). The same cannot be said for research oriented towards the automatic recognition of emotion from expressive signals. In the next paragraph we will introduce the Component Process Model, an appraisal model that is currently used for modeling emotion. Here we advocate using this model as a theoretical framework for automatic emotion recognition systems.

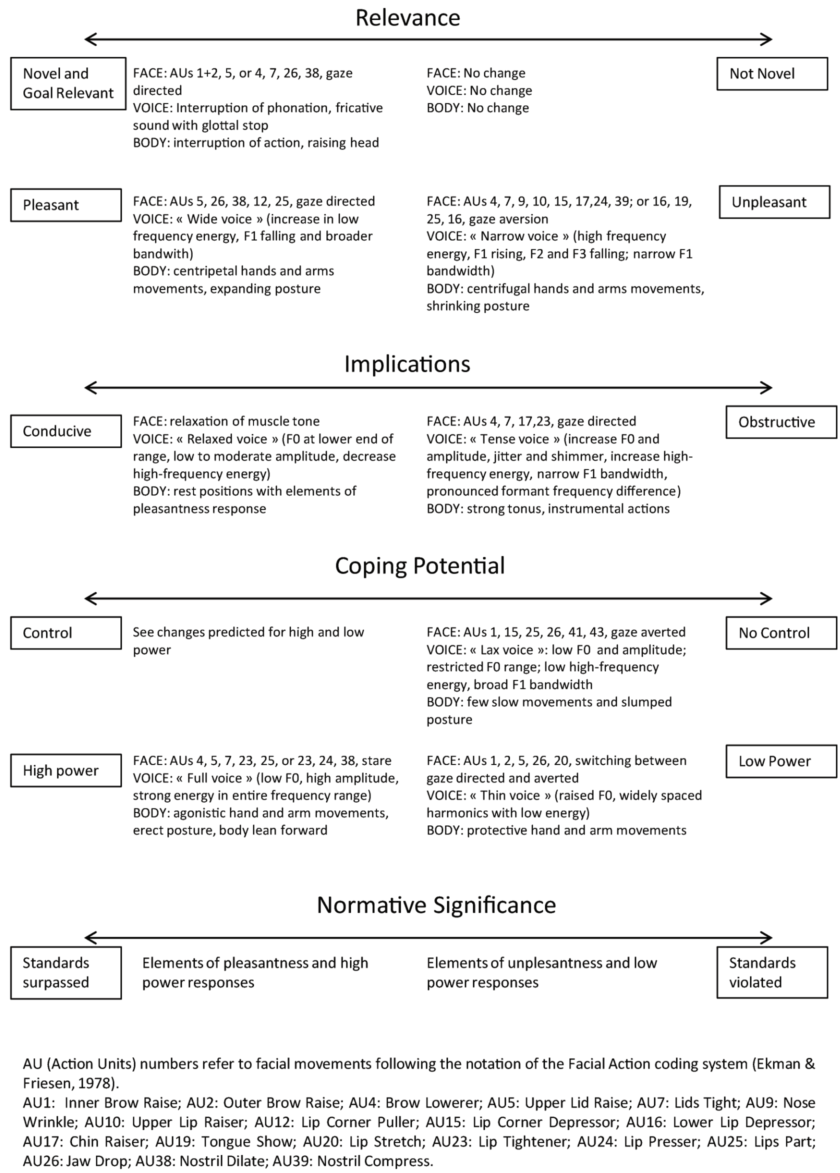
The Component Process Model

According to the Component Process Model (CPM) an emotion is a process that involves five functional components: cognitive, peripheral efference, motivational, motor expression and subjective feeling (Scherer, 2001, 2009a). Emotion is defined as "an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" (Scherer, 2001, p. 93). Emotions are mostly determined by the cognitive component, i.e., the appraisal that consists of a fixed sequence of evaluation checks. This proposed sequence comprises four main steps: a) *relevance* check, an evaluation of the novelty of the stimulus, its relevance for the goals of the person, and its intrinsic pleasantness – a relevant stimulus requires attention, further information processing and potential action; b) *implication* check, an evaluation of the implications and consequences of the stimulus (conducive or obstructive to the goals of the person); c) *coping potential* check, an evaluation of the ability to cope with the situation (agency, control, power, and adjustment); d) *normative significance* check, an evaluation of the overall compatibility of the stimulus event with personal and social norms and values (Scherer, 2001).

The event-stimulus is evaluated along these dimensions and the appraisals guide the adaptive response of the organism and have effects on all the other components. The CPM includes predictions of appraisal efferent effects on face, body and voice. These predictions are based on empirical evidence and physiological considerations. A detailed list of predictions of appraisal-related expressive features can be found in Figure 1.

Although the empirical investigation of the efferent effects of appraisals is extremely complex, some preliminary evidence is available for facial expression (Kaiser & Wehrle, 2001), and vocal expression (Johnstone, van Reekum, & Scherer, 2001). A recent study by Scherer and Ellgring (2007a) analyzed the frequency

Figure 1. Predictions of the Component Process Model for facial expressions, voice, and body movements (adapted from Scherer, 2001, 2010).



of facial action units in a corpus of enacted expressions of emotions. Results indicated that, overall, individual facial action units were used in a way that could be plausibly explained by the appraisals that underlie the emotions. Other studies tested more specific links between single appraisals and specific facial movements. EMG studies found that: a) the activity of the *corrugator supercilii* (the muscle whose activity produces frowning, Action Unit 4 using the notation of FACS; Ekman & Friesen, 1978) was correlated with appraisals of perceived obstacle and motivational incongruence (Pope and Smith, 1994; Smith, 1989), as well as with goal obstructive events and unpleasant stimuli (Aue, Flykt, & Scherer, 2007; Aue & Scherer, 2008; Delplanque et al., 2009); b) the activity of the *zygomaticus major* (the muscle whose activity pulls the corners of the lips up, i.e., smile, Action Unit 12 using FACS notation) was associated with the appraisal of subjective pleasantness and with goal conducive events (Aue et al., 2007; Aue & Scherer, 2008); and c) the activity of the *frontalis* (the muscle whose activity pulls the eyebrows up – Action Units 1 and 2 using FACS notation) was higher for novel stimuli than for familiar stimuli. In a recent study Mortillaro, Mehu, and Scherer (2011) tested the plausibility of an appraisal approach as a way to distinguish facial expressions of different positive emotions. In this study, the authors used the Facial Action Coding System to code dynamic expressions of four positive emotions and found evidence for some of the predictions advanced by the CPM: appraisal of novelty was reflected in the degree of eyes opening, while the appraisal of intrinsic pleasantness was associated with cheek raise.

Banse and Scherer (1996) analyzed the acoustic features of vocal expressions of emotions and found that they could be plausibly explained by the hypothesized appraisals and indirectly confirmed a sizeable amount of CPM predictions. Johnstone, van Reekum, Hird, Kirsner, and Scherer (2005) used a computer game to investigate emotional speech following goal conducive (or obstructive) events and pleasant or unpleasant stimuli. Results showed that speech following goal obstructive events was higher in intensity and fundamental frequency than speech following goal conducive events; for the pleasantness dimension there was a greater proportion of energy in the higher frequencies after unpleasant stimuli than after pleasant stimuli. Patel, Scherer, Björkner, and Sundberg (2011) analyzed the acoustic features of 100 vocal affect bursts - representing 5 emotions - and found acoustic features that seem to reflect the power and the control parts of the appraisal of coping potential; these results were in line with predictions of the CPM.

The CPM also makes predictions for the effects that different appraisals would have on body movements and postures. Scherer and Ellgring (2007b) used an appraisal perspective to investigate emotion expression integrating face, voice, and body movements; Coulson (2009) used the CPM to model emotion postures and postural shifts. Dael, Mortillaro, and Scherer (2011) analyzed body movements shown in 120 emotion expressions taken from the Geneva Multimodal Emotion corpus (Bänziger & Scherer, 2010; Bänziger, Mortillaro, & Scherer, 2011). Using a newly developed coding system for body action and postures (Dael, Mortillaro, & Scherer, in press), the authors found that behavior could be plausibly explained by the appraisals that underlie the emotions under investigation.

Although all these studies showed the existence of a link between appraisals and expressive features, the possibility of detecting appraisals from expressive features still needs research. In the following paragraphs we will suggest one framework for testing this possibility.

An Appraisal Framework for Emotion Recognition

Appraisal models define a series of cognitive evaluations and predict which emotion would be experienced based on the resulting appraisals. Similarly, appraisals are seen as the causal mechanisms at the basis of physiological modifications and nonverbal expressions. Some researchers have suggested the possibility that this mechanism could be reversed for emotion recognition: inferring specific appraisals from specific expression configurations. Mortillaro, Mehu, and Scherer (2012) listed "inference of underlying appraisals" as one of the mechanisms that human observers could use to infer an emotion state from someone's expressions. The possibility that observers could directly infer appraisals from expressions and then use the appraisals to attribute an emotion label was already advanced in the first componential models of emotions. Indeed componential views of emotion expression postulate that while a configuration of cues may be immediately perceived as an expression of one specific emotion, individual elements of the expression (a facial movement or an acoustic feature) may be still meaningful because they would carry information about the underlying appraisals (Smith & Scott, 1997).

Two recent studies investigated the possibility that observers would be able to infer appraisal information from nonverbal expressions. Scherer and Grandjean (2008) asked participants to judge facial expressions of seven emotions using a) emotion labels; b) social messages; c) appraisals; or d) action tendencies. Judges were more accurate and more confident in their judgment when they used emotion labels and appraisals. Laukka et al. (2010) developed a corpus of emotion vocal expressions and had these expressions evaluated on several appraisal dimensions; results showed that a) listeners could reliably infer appraisals from vocal expressions; b) inferred appraisals were in line with predictions of appraisal theory; and c) appraisals were significantly correlated with a number of acoustic features.

These studies suggest that the appraisal framework lends itself for fruitful application in the context of emotion recognition. Importantly, such an approach would address three concerns that are currently relevant in the field of (automatic) emotion recognition, including:

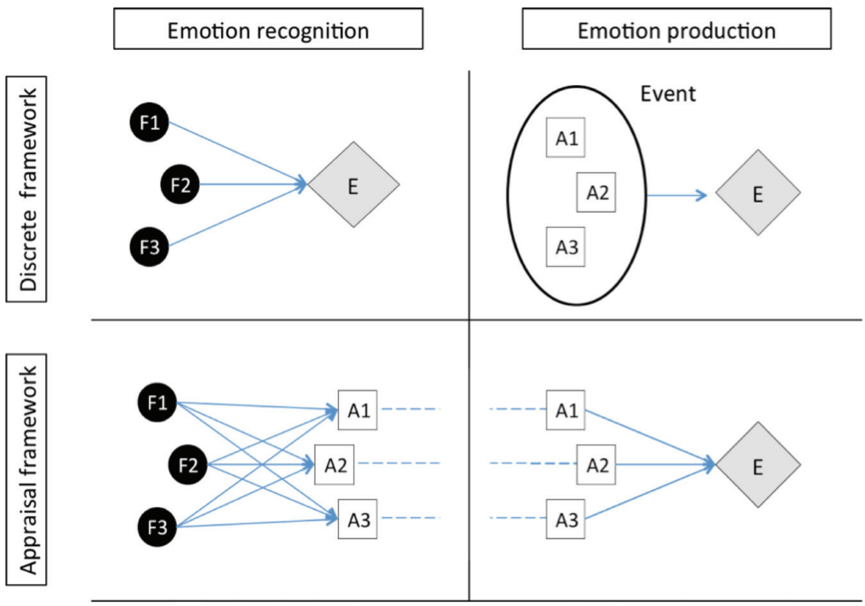
1. How to establish a link between models of emotion recognition and emotion production
2. How to add contextual information to systems of emotion recognition
3. How to increase the sensitivity with which weak, subtle, or complex emotion states can be detected

The first and most important issue concerns the link between emotion production and recognition. In the discrete view on (automatic) emotion recognition, such a link has either not been established or is treated as secondary to the problem of accurately inferring emotional states (Castellano, Caridakis, Camurri, Karpouzis, Volpe, & Kollias, 2010). Within the appraisal framework, by contrast, the link between emotion recognition and emotion production is rendered explicit. In emotion production, it is assumed that humans make appraisals and then experience emotional states based on these appraisals. In emotion recognition, it seems plausible that observers detect appraisal results and their consequences on autonomic symptoms (blushing) and motor behavior (facial muscle movements) and then infer emotional states based on these appraisals.

This congruence allows direct communication between models of recognition and production, as depicted in Figure 2 (lower panel). In this diagram, black circles denote expression variables (e.g., individual facial action units), white squares denote appraisal variables (e.g., goal obstruction, pleasantness), and grey diamonds denote emotion labels (e.g., anger, fear). In the classical framework, the recognition model outputs an emotion label (upper left panel) without the detailed specification of a production model. In terms of production, an undetermined "affect program" is assumed without separately considering the underlying appraisals (upper right panel).

In the appraisal framework, on the other hand, using appraisals as responses in the recognition model naturally leads to a feedforward compatibility with appraisals in the production model (lower panel, dashed lines). In particular, data or predictions at the response side of the recognition model can be inputted directly into the production model to estimate a plausible emotion label. In the discrete recognition framework, one could in theory still attempt to make a reverse prediction of a subject's appraisals, by submitting the output of a discrete recognition model (upper left panel) to an appraisal production model (lower right panel), but this is unattractive due to (a) the reversal of causality this implies, (b) the different pro-

Figure 2. Discrete vs. appraisal framework for emotion recognition and production. Black circles denote expression variables (e.g., individual facial action units), white squares denote appraisal variables (e.g., goal obstructiveness, pleasantness), and grey diamonds denote emotion labels (e.g., anger, fear).



cesses that each model represents, and (c) the assumption that appraisals can only be inferred when first an emotion label is obtained. Importantly, an emotion recognition system that outputs appraisals would be extremely useful in all those cases in which it would be difficult to attribute a single emotion label (e.g., when multiple emotion labels have the same likelihood estimate, or none is applicable).

The second issue concerns how to integrate contextual information into systems of emotion recognition. In the classical (discrete) framework, emotions are identified on the basis of a relatively isolated set of facial and vocal expressions associated with a particular emotion (e.g., joy, anger, etc.). No information about the environment (e.g., work, home), the subject (e.g., personality traits), or the current situation (e.g., performing a task, relaxing) is taken into account, yet these factors are known to strongly affect the observed emotion. At present, nearly all automatic systems of emotion recognition are context-insensitive (Zeng et al., 2009). In the appraisal framework, on the other hand, there is an implicit relation between emotion and environment. This is because appraisals themselves represent abstractions of contextual information. For instance, when we observe a person frowning and infer an appraisal of goal obstructiveness, we not only increase the likelihood of attributing an anger label to the person, we also gain information about the causes of the frown, which is goal obstruction occurring in the person's environment. In other words, by inferring appraisals from behavior we infer more than just markers of emotion, we infer information about the causes of that emotion (Castellano et al., 2010).

The third issue concerns the sensitivity with which systems of emotion recognition can detect weak, subtle, or mixed emotion states. By inferring appraisals rather than prototypical emotion categories we broaden the scope of emotion recognition both quantitatively and qualitatively. Qualitative broadening means that we increase the information content of the inferred states by establishing links with contextual determinants, as described in the previous paragraph. Quantitative broadening means that we can describe both emotional and non-emotional states with an expanded set of continuous appraisal dimensions. This could result in a much richer and complex space than either the discrete emotion framework or the valence-arousal dimensional framework can account for. Notably, a large part of this space spans mental states which are not necessarily emotional, or may contain just traces of emotional states. For instance, a faintly raised eyebrow will not signal a full-blown emotion but may be recognized as a signal of unexpectedness. Since prototypical expressions of emotion tend to be rare in everyday life, it should be expected that subtlety and complexity are the rule rather than the exception. Expressions of appraisal are predicted at a much finer level of resolution than emotion categories, and thus allow dealing with much greater complexity. In recent reviews, authors have called for recognition systems to incorporate non-basic affective states (Zeng et al., 2009; Castellano et al., 2010), and some progress has already been made with the detection of mental states such as agreeing or disagreeing (el Kaliouby & Robinson, 2004), pain (Littlewort et al., 2007) and interest (Yeasin et al., 2006).

In this section we have argued that the appraisal framework offers numerous advantages to studies of emotion recognition. By replacing emotion categories as the target state of inference with continuous appraisal dimensions, it becomes possible to link models of emotion recognition and production, add contextual information to the recognition process, and extend the scope of mental states to be inferred. In the next section, we take a look at how this approach can be implemented in practice and how studies on both emotion recognition and emotion production can benefit from adopting a common methodology.

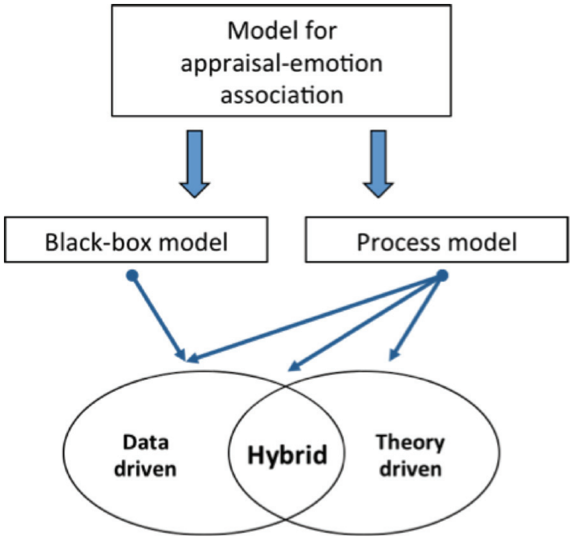
A Common Methodology for Modeling Recognition and Production

How could the above strategy be accomplished in practice? A full implementation of the appraisal framework requires both a computational model of emotion recognition and one of emotion production. Unfortunately, research in these two fields currently shares little overlap with regard to the methodology that is being employed, each having more or less developed its practices independently from the other.

In Figure 3 we consider the general class of models that wish to model an association between appraisal and other emotion components (whether in the field of recognition or production). A first important distinction can be made between black-box models and process models (Wehrle & Scherer, 2001). In a black-box model, one attempts to model the relation between appraisal variables and other emotion components as accurately as possible without taking into account the interpretability of the final model. This is the strategy that has generally been adopted in the field of automatic emotion recognition: a model is trained on empirically gathered data and its success measured by its predictive accuracy (e.g., Castellano et al., 2010; Devillers, Vidrascu, & Layachi, 2010). Typically these models are advanced statistical models originating from the field of machine learning (e.g., support vector machines, *k*-nearest neighbor classifiers). The advantage of this approach is the flexibility with regard to modeling that can be attempted, for instance by contrasting linear to nonlinear models. The major drawback of this approach is often the lack of interpretability, which is either inherent to the model being used or simply outside the interest of the researchers.

In a process model, one attempts to describe the association between appraisal and other emotion components (possibly over time) in a meaningful way. Here, it is useful to make a further distinction between theory-driven and data-driven models as outlined in Figure 3. In a theory-driven model, the relations between

Figure 3. Taxonomy of the models for the association appraisal-emotion



input and output are completely specified by the user, whereas, in a data-driven model, the model, based on empirical data, estimates these relations. Evidently, black-box models are data-driven by their very nature. With regard to computational models of emotion production, most of the current models in existence are heavily theory driven. In contrast to the field of emotion recognition, however, emotion production turned to appraisal theory as a venue for modeling at an early stage, to the extent that many of the current models of emotion production are based on appraisal (Marsella, Gratch, & Petta, 2010). The majority of these models rely on the OCC theory by Ortony, Clore, and Collins (1988), while others have attempted to employ structural or conceptual features of the CPM proposed by Scherer, such as WASABI (Becker-Asano, 2008), PEACTDIM (Marinier, 2008), and GATE (Wehrle, 1995; Wehrle & Scherer, 2001). The major drawback of the current approach to modeling appraisal is the lack of flexibility offered by statistically founded, data-driven approaches such as those used in emotion recognition. This is particularly problematic in the face of increasing consensus that emotions are dynamic, nonlinear phenomena (Lewis, 2005; Scherer 2000, 2009b), casting doubt on the ability of theory-driven models to capture such complex patterns.

Ideally, however, a process model should be able to incorporate both theory-driven and data-driven properties. This naturally leads us to consider so-called hybrid models, where one part of the model is estimated from empirical data while the other is specified by theoretical constraints (Figure 3, diagram intersection). Examples of hybrid models are neural networks. These models are particularly attractive for computational modeling of emotion (Sander, Grandjean, & Scherer, 2005; Scherer, 2009b), because they allow flexibility with regard to the architectural choices the user can make such as the number of layers, the number of connections, and the possibility to include feedback loops. The hybrid approach to computational modeling may be extended to models that are capable of adapting to new data as they are processed (online learning) and that allow emergent phenomena through dynamic input and feedback between components (Scherer, 2009b).

CONCLUSION

Individuals' appraisals are key elements for emotion production (e.g., Ellsworth & Smith, 1988; Roseman & Evdokas, 2004; Siemer, Mauss, & Gross, 2007; Smith & Kirby, 2004), but their role in emotion recognition is still largely unexplored. A handful of recent studies, however, already suggest that the appraisal framework may be a fruitful venue for studying emotion recognition. We believe that this approach may offer specific advantages over the discrete and dimensional frameworks of emotion. First, by adopting the appraisal framework, we can establish formal links between models of emotion production and recognition. Second, by linking appraisals to expressive features, we implicitly introduce contextual information to models of emotion recognition. Third, by expanding the dimensional basis of emotion categories, we increase the sensitivity with which subtle or complex emotion states may be detected.

We propose that emotion recognition and emotion production should be studied together within the appraisal framework, and that research on emotion systems should adopt a common strategy to modeling that achieves an optimal trade-off between data- and theory-driven methods. With regard to emotion recognition, the black-box methodology that currently dominates the field severely limits the interpretability of the fitted models. It is extremely important that, when developing emotion recognition systems, researchers take an explicit informed decision on the theoretical model they use to infer the emotion, and use this theory to guide the modeling process. With regard to emotion production, current models of appraisal suffer from being overdetermined by theory. Here, studies on emotion production could benefit by adopting more data-driven methods from the field of emotion recognition.

Despite theoretical controversies, all theoretical models can be useful for emotion recognition system under different conditions. When the focus is on basic emotions and intense expressions, a discrete classification approach can provide satisfactory results. When the goal is to detect broad elements like valence or intensity, a dimensional model can be the preferred solution. However, as we have argued, when the goal is to address a potentially large number of affective states, not necessarily characterized by extreme prototypical expressions, we suggest using fine-grained appraisal dimensions, and use them to estimate the likelihood of different emotion labels.

REFERENCES

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

Aue, T., Flykt, A., & Scherer, K. R. (2007). First evidence for differential and sequential efferent effects of goal relevance and goal conduciveness appraisal. *Biological Psychology*, *74*, 347–357. doi:10.1016/j.biopsycho.2006.09.001

Aue, T., & Scherer, K. R. (2008). Appraisal-driven somatovisceral response patterning: Effects of intrinsic pleasantness and goal conduciveness. *Biological Psychology*, *79*, 158–164. doi:10.1016/j.biopsycho.2008.04.004

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636. doi:10.1037/0022-3514.70.3.614

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2011). Introducing the Geneva Multimodal Expression corpus for experimental research on emotion perception. *Emotion*.

Bänziger, T., & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus. In Scherer, K. R., Bänziger, T., & Roesch, E. B. (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 271–294). New York, NY: Oxford University Press.

Becker-Asano, C. (2008). *WASABI: Affect simulation for agents with believable interactivity* (Doctoral dissertation, University of Bielefeld). Amsterdam, The Netherlands: IOS Press.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*, *21*, 3–21. doi:10.1023/A:1024902500935

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*, 18–37. doi:10.1109/T-AFFC.2010.1

Castellano, G., Caridakis, G., Camurri, A., Karpouzis, K., Volpe, G., & Kollias, S. (2010). Body gesture and facial expression analysis for automatic affect recognition. In Scherer, K. R., Bänziger, T., & Roesch, E. B. (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 245–255). New York, NY: Oxford University Press.

Coulson, M. (2009). Expressing emotion through body movement: A component process approach. In Canamero, L., & Aylett, R. (Eds.), *Animating expressive characters for social interaction (Advances in Consciousness Research Series)* (Vol. 74). Amsterdam, The Netherlands: Benjamins.

Dael, N., Mortillaro, M., & Scherer, K. R. (2011). Emotion expression in body action and posture. *Emotion*. doi: 10.137/a0025737

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). The body action and posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior*. doi:10.1007/s10919-012-0130-0

Delplanque, S., Grandjean, D., Chrea, C., Coppin, G., Aymard, L., & Cayeux, I. (2009). Sequential unfolding of novelty and pleasantness appraisals of odors: evidence from facial electromyography and autonomic reactions. *Emotion (Washington, D.C.)*, *9*(3), 316–328. doi:10.1037/a0015369

Devillers, L., Vidrascu, L., & Layachi, O. (2010). Automatic detection of emotion from vocal expression. In Scherer, K. R., Bänziger, T., & Roesch, E. B. (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 232–244). New York, NY: Oxford University Press.

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological Science*, *3*, 34–38. doi:10.1111/j.1467-9280.1992.tb00253.x

Ekman, P. (1999). Facial expressions. In Dalgleish, T., & Power, M. J. (Eds.), *Handbook of cognition and emotion* (pp. 301–320). New York, NY: John Wiley & Sons.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*, 1208–1210. doi:10.1126/science.6612338

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*, 86–88. doi:10.1126/science.164.3875.86

El Kaliouby, R., & Robinson, P. (2004). Real time inference of complex mental states from facial expressions and head gestures. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Workshop on Real Time Computer Vision for Human Computer Interaction* (Vol. 10, p. 154).

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

Ellsworth, P. C., & Smith, C. A. (1988). Shades of joy: Patterns of appraisal differentiating pleasant emotions. *Cognition and Emotion*, *2*, 301–331. doi:10.1080/02699938808412702

Frijda, N. H. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.

Gunes, H., & Pantic, M. (2010). Automatic, dimensional and continuous emotion recognition. *International Journal of Synthetic Emotions*, *1*, 68–99. doi:10.4018/jse.2010101605

Johnstone, T. van Reekum, & Scherer, K. R. (2001). Vocal expression correlates of appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion* (pp. 271-284). New York, NY: Oxford University Press.

Johnstone, T., van Reekum, C. M., Hird, K., Kirsner, K., & Scherer, K. R. (2005). Affective speech elicited with a computer game. *Emotion (Washington, D.C.)*, *5*(4), 513–518. doi:10.1037/1528-3542.5.4.513

Kaiser, S., & Wehrle, T. (2001). Facial expressions as indicators of appraisal processes. In Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.), *Appraisal processes in emotion* (pp. 285–300). New York, NY: Oxford University Press.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proceedings of the International Conference on Face and Gesture Recognition* (pp. 46-53).

Laukka, P., Elfenbein, H. A., Chui, W., Thingujam, N. S., Iraki, F. K., Rockstuhl, T., & Althoff, J. (2010). Presenting the VENEC corpus: Development of a cross-cultural corpus of vocal emotion expressions and a novel method of annotating emotion appraisals. In *Proceedings of the LREC Workshop on Corpora for Research on Emotion and Affect* (pp. 53-57).

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

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *The Behavioral and Brain Sciences*, *28*, 169–245. doi:10.1017/S0140525X0500004X

Littlewort, G. C., Bartlett, M. S., & Lee, K. (2007). Faces of pain: automated measurement of spontaneous facial expressions of genuine and posed pain. In *Proceedings of the 9th International Conference on Multimodal Interfaces* (pp. 15-21).

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

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. In Scherer, K. R., Bänziger, T., & Roesch, E. B. (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 21–41). New York, NY: Oxford University Press.

Mortillaro, M., Mehu, M., & Scherer, K. R. (2011). Subtly different positive emotions can be distinguished by their facial expressions. *Social Psychological and Personality Science*, *2*, 262–271. doi:10.1177/1948550610389080

Mortillaro, M., Mehu, M., & Scherer, K. R. (in press). The evolutionary origin of multimodal synchronization in emotional expression. In Altenmüller, E., Schmidt, S., & Zimmerman, E. (Eds.), *Evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man*. New York, NY: Oxford University Press.

Naab, P. J., & Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion (Washington, D.C.)*, *7*, 736–744. doi:10.1037/1528-3542.7.4.736

Nezlek, J. B., Vansteelandt, K., Van Mechelen, I., & Kuppens, P. (2008). Appraisal-emotion relationships in daily life. *Emotion (Washington, D.C.)*, *8*, 145–150. doi:10.1037/1528-3542.8.1.145

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511571299

Osgood, C. E. (1962). Studies of the generality of affective meaning systems. *The American Psychologist*, *17*, 10–28. doi:10.1037/h0045146

Osgood, C. E. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist*, *66*, 171–200. doi:10.1525/aa.1964.66.3.02a00880

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, *87*, 93–98. doi:10.1016/j.biopsycho.2011.02.010

Pope, L. K., & Smith, C. A. (1994). On the distinct meanings of smiles and frowns. *Cognition and Emotion*, *8*, 65–72. doi:10.1080/02699939408408929

Roseman, I., & Evdokas, A. (2004). Appraisals cause experienced emotions: Experimental evidence. *Cognition and Emotion*, *18*, 1–28. doi:10.1080/02699930244000390

Roseman, I. J., & Smith, C. A. (2001). Appraisal theory: Overview,assumptions, varieties, controversies. In Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 3–19). New York, NY: Oxford University Press.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172. doi:10.1037/0033-295X.110.1.145

Russell, J. A., Bachorowski, J.-A., & Fernandez-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, *54*, 329–349. doi:10.1146/annurev.psych.54.101601.145102

Russell, J. A., & Bullock, M. (1986). On the dimensions preschoolers use to interpret facial expressions of emotion. *Developmental Psychology*, *22*, 97–102. doi:10.1037/0012-1649.22.1.97

Russell, J. A., & Fernandez-Dols, J. M. (1997). *The psychology of facial expression*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511659911

Russell, J. A., & Mehrabian, A. (1974). Distinguishing anger and anxiety in terms of emotional response factors. *Journal of Consulting and Clinical Psychology*, *42*, 79–83. doi:10.1037/h0035915

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*, 273–294. doi:10.1016/0092-6566(77)90037-X

Sander, D., Grandjean, D., & Scherer, K. R. (2005). A systems approach to appraisal mechanisms in emotion. *Neural Networks*, *18*, 317–352. doi:10.1016/j.neunet.2005.03.001

Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. In P. Shaver (Ed.), *Review of personality and social psychology: Vol. 5. Emotions, relationships and health* (pp. 37-63). Thousand Oaks, CA: Sage.

Scherer, K. R. (1994). Toward a concept of "modal" emotions. In Ekman, P., & Davidson, R. J. (Eds.), *The nature of emotion: Fundamental questions* (pp. 25–31). New York, NY: Oxford University Press.

Scherer, K. R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In Lewis, M. D., & Granic, I. (Eds.), *Emotion, development, and self- organization: Dynamic systems approaches to emotional development* (pp. 70–99). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511527883.005

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

Scherer, K. R. (2009a). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*, 1307–1351. doi:10.1080/02699930902928969

Scherer, K. R. (2009b). Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society B*, *364*, 3459–3474. doi:10.1098/rstb.2009.0141

Scherer, K. R. (2010). The component process model: Architecture for a comprehensive computational model of emergent emotion. In Scherer, K. R., Bänziger, T., & Roesch, E. B. (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 47–70). New York, NY: Oxford University Press.

Scherer, K. R., Clark- Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, *46*, 401–435. doi:10.1080/00207594.2011.626049

Scherer, K. R., & Brosch, T. (2009). Culture-specific appraisal biases contribute to emotion dispositions. *European Journal of Personality*, *288*, 265–288. doi:10.1002/per.714

Scherer, K. R., & Ceschi, G. (2000). Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Personality and Social Psychology Bulletin*, *26*, 327–339. doi:10.1177/0146167200265006

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

Scherer, K. R., & Ellgring, H. (2007b). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion (Washington, D.C.)*, *7*, 158–171. doi:10.1037/1528-3542.7.1.158

Scherer, K. R., & Grandjean, D. (2008). Facial expressions allow inference of both emotions and their components. *Cognition and Emotion*, *22*, 789–801. doi:10.1080/02699930701516791

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

Scherer, K. R., Wranik, T., Sangsue, J., Tran, V., & Scherer, U. (2004). Emotions in everyday life: Probability of occurrence, risk factors, appraisal and reaction pattern. *Social Sciences Information. Information Sur les Sciences Sociales*, *43*, 499–570. doi:10.1177/0539018404047701

Siemer, M., Mauss, I., & Gross, J. J. (2007). Same situation--different emotions: how appraisals shape our emotions. *Emotion (Washington, D.C.)*, *7*, 592–600. doi:10.1037/1528-3542.7.3.592

Smith, C. A. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology*, *56*, 339–353. doi:10.1037/0022-3514.56.3.339

Smith, C. A., & Kirby, L. D. (2004). Appraisal as a pervasive determinant of anger. *Emotion (Washington, D.C.)*, *4*, 133–138. doi:10.1037/1528-3542.4.2.133

Smith, C. A., & Scott, H. (1997). A componential approach to the meaning of facial expressions. In Russell, J., & Fernandez-Dols, J. (Eds.), *The psychology of facial expression* (pp. 229–254). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511659911.012

Valstar, M., Mehu, M., Pantic, M., & Scherer, K. R. (in press). Meta-analysis of the first facial expression recognition and analysis challenge. *IEEE Transactions on Systems, Man, and Cybernetics*.

Wehrle, T. (1995). *The Geneva Appraisal Theory Environment (GATE). Unpublished computer software*. Geneva, Switzerland: University of Geneva.

Wehrle, T., & Scherer, K. R. (2001). Toward computational modelling of appraisal theories. In Scherer, K. R., Schorr, A., & Johnstone, T. (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York, NY: Oxford University Press.

Yeasin, M., Bullot, B., & Sharma, R. (2006). Recognition of facial expressions and measurement of levels of interest from video. *IEEE Transactions on Multimedia*, *8*(3), 500–507. doi:10.1109/TMM.2006.870737

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(1), 39–58. doi:10.1109/TPAMI.2008.52

ENDNOTE

1 Although here we focus on facial expressions, the model that we suggest is valid for any expressive modalities.